Operations Research Proceedings

Stefan Helber · Michael Breitner
Daniel Rösch · Cornelia Schön
Johann-Matthias Graf von der Schulenburg
Philipp Sibbertsen · Marc Steinbach
Stefan Weber · Anja Wolter *Editors*

Operations Research Proceedings 2012

Selected Papers of the International Annual Conference of the German Operations Research Society (GOR), Leibniz Universität Hannover, Germany, September 5–7, 2012

GOR
Gesellschaft für Operations Research e.V.

Springer

Operations Research Proceedings

GOR (Gesellschaft für Operations Research e.V.)

For further volumes:
http://www.springer.com/series/722

Stefan Helber · Michael Breitner
Daniel Rösch · Cornelia Schön
Johann-Matthias Graf von der Schulenburg
Philipp Sibbertsen · Marc Steinbach
Stefan Weber · Anja Wolter
Editors

Operations Research Proceedings 2012

Selected Papers of the International Annual Conference of the German Operations Research Society (GOR), Leibniz Universität Hannover, Germany, September 5–7, 2012

Springer

Editors
Stefan Helber
Cornelia Schön
Anja Wolter
Institute of Production Management
Leibniz Universität Hannover
Hannover
Germany

Michael Breitner
Institute for Information Systems Research
Leibniz Universität Hannover
Hannover
Germany

Daniel Rösch
Institute for Banking and Finance
Leibniz Universität Hannover
Hannover
Germany

Johann-Matthias Graf von der Schulenburg
Institute for Risk and Insurance
Leibniz Universität Hannover
Hannover
Germany

Philipp Sibbertsen
Institute of Statistics
Leibniz Universität Hannover
Hannover
Germany

Marc Steinbach
Institute of Applied Mathematics
Leibniz Universität Hannover
Hannover
Germany

Stefan Weber
Institute of Probability and Statistics
Leibniz Universität Hannover
Hannover
Germany

ISSN 0721-5924
ISBN 978-3-319-00794-6 ISBN 978-3-319-00795-3 (eBook)
DOI 10.1007/978-3-319-00795-3
Springer Cham Heidelberg New York Dordrecht London

Preface

This volume contains a selection of short refereed papers related to the presentations given at OR 2012, the International Annual Conference of the German Operations Research Society. The conference was held from September 4–7, 2012, at Leibniz Universität Hannover. More than 500 participants from about 40 countries attended more than 300 contributed talks in 18 different streams, ranging from "Applied Probability" to "Traffic and Transportation".

Special attention was given to those OR-problems that are related to the numerous aspects and interactions of "Energy, Markets and Mobility''. The choice of this main topic reflected not only current challenges of society at large, but also important strengths of the hosting institutions, Leibniz Universität Hannover, as well as its business environment in the German state of Lower Saxony. A large number of presentations, both invited and contributed, addressed this field. However, the conference also provided ample opportunity to present and learn about the newest developments in operations research in general.

As in previous years, the presentations of the prize winners were one of the highlights of the conference. The excellent works submitted mainly by junior researchers again confirmed how attractive and vivid operations research is as a field of both theory development and application.

The editors of this book served as the local organizing committee. We are deeply indebted to the many institutions, firms, and individuals who worked hard and often invisibly or donated generously to make the conference a success. To all of them this volume is dedicated.

Stefan Helber
Michael Breitner
Daniel Rösch
Cornelia Schön
Johann-Matthias Graf von der Schulenburg
Philipp Sibbertsen
Marc Steinbach
Stefan Weber
Anja Wolter

Contents

Part I
Award Winners

Product Line Design with Pricing Kits

Pascal Lutter

1 Introduction

Product design and pricing are important business decisions. On the one hand, product characteristics induced by product design should match different customer preferences. On the other hand, product pricing should utilize customers' willingness-to-pay (WTP) as much as possible to enhance profit. The resulting conflicts should be anticipated as well as possible to find a compromise between both clients' and corporate interests. Such approaches are being pursued in the literature and there are many specific mathematical programming models. All models support the design and pricing of only complete products. Following the increasing use of modularization and mass customization, it is possible to offer customized products at lower costs. Besides solely offering a variety of different product variants, it is also possible to allow customers to configure the product according to their own preferences. Pricing kits are innovative modular pricing systems in the form of list prices for modular product components and help to determine a corresponding pricing system for individual product components. Applications can be found in the configuration of personal computers or in the automotive sector. Previous descriptions in literature left the crucial aspect of pricing unanswered.

With this contribution, a mathematical optimization model is developed in order to determine a profit maximizing pricing kit. After a brief presentation of current methods for measuring customer preferences, product configuration and pricing, the new concept of pricing kits will be defined and a mathematical formulation will be provided. Then different forms of integration are discussed and compared with pure product line design models in a simulation study. The paper concludes with a brief outlook.

P. Lutter (✉)
Ruhr University Bochum, Institute of Management, Universitätsstr. 150, 44801 Bochum, Germany
e-mail: pascal.lutter@rub.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_1,

2 Product Line Design and Pricing

Hereinafter a product will be defined as a pool of characteristics. The characteristics of products are determined by their attributes m and corresponding attribute levels a. One attribute illustrates one component that is observable or at least conceptually ascertainable. For example, the internal memory of a PC is an attribute whose levels are given by its memory size. The combination of particular levels of a given set of attributes to a complete product is called product design, e.g. a PC is described by the attributes processor, operating system, memory, hard drive, optical drive, video card and monitor. Choosing all attributes at their lowest levels one gets a simple office PC. Another product variant emerges by varying attribute levels.

2.1 Modeling Customer Choice

The anticipation of customer choice requires knowledge of the corresponding preferences in terms of utility for each attribute and level u_{kma}. In practice, Conjoint Analysis is a common method for investigating customer utility. With these observed utility values, a corresponding WTP can be calculated in order to predict customer decisions for any specific combination of product and price. To anticipate decisions of customers, a deterministic approach, the so-called First-Choice (FC) Rule, also known as the Maximum Utility Rule [6], is applied. This rule assumes perfectly rational customer choices and the exact knowledge of customer utility as well as all other factors that might affect customer choice. Even though these assumptions seem to be very restrictive, the resulting predictions have the same quality as other rules of choice [4].

2.2 Pure Product Line Design and Pricing

Since heterogeneity of demand is increasing and globalization of markets is rising, the pressure on businesses to shape their products in line with the market conditions is growing. In a wider sense product line design contains both determination of quantity of product configurations and price setting. Existing mathematical models can be subdivided into single-stage and two-stage approaches. Single-stage models, also known as product line design (PLD) models, generate a predetermined quantity of different product variants as well as their corresponding prices [5]. It is important to mention that these models assume that product attributes cannot be customized afterwards. Hence, it is not possible to change some product attribute levels due to customer requests. If components are exchanged ex post, all product prices have to be recalculated in order to prevent cannibalization effects. The basis for PLD approaches forms the attribute level, so that product configuration and price setting takes place simultaneously, whereas two-stage models, also known as product line selection (PLS) models, address the level of complete products. In the first step the overall set of possible products is reduced to a smaller amount, so that in a second step the

final choice of products and their prices can be found in a subset of all possibilities of products [3]. It remains unclear how product preselection of the first stage has to take place, since PLS approaches only provide models for the second stage. In a comparison of these two approaches, the PLS approaches can only approximate PLD solutions [6].

3 Pricing Kits

A pricing kit is a pricing system in terms of list prices for standardized, combinable, complementary as well as substitutive and individually alienable components that can be provided as a complete product at the market. In contrast to product line design where customers are offered fixed products, customers can customize their own product. Precondition for pricing kits are the so-called product kits. A product kit consists of standardized and combinable performance based components [2]. Therefore, it shows the close link to the principle of modularization [1, 7]. Essential for this concept is the transition from the idea of impersonal mass production to a new approach of an individual product design. This concept still differs from complete made-to-measure production because of the use of standardized components. Pricing kits provide the opportunity to set individual prices for all attribute levels. Thus, prices for individual products can easily be calculated. A popular example for the successful usage of this pricing system is the well-known computer company Dell.

3.1 Pricing Kit Model

In the following it is assumed that all relevant customers are allocated in K homogenous customer segments, each consisting of S_k customers. A product consists of M attributes with each A_m levels per attribute m. For each customer segment k the WTP for all attribute levels R_{kma} is known. Furthermore, the variable costs C_{ma} for all attribute levels are given. The binary variables x_{kma} represent the customer choice with $x_{kma} = 1$ if the chosen product consists of attribute m with level a and otherwise $x_{kma} = 0$. The realized net benefit per segment k for attribute m is represented by s_{km}. The price for all attributes and corresponding levels is given by p_{ma}. It is assumed that a product necessarily consists of all attributes. In cases of optional attributes a further attribute level can be implemented with $C_{ma} = 0$. Finally, it is assumed that the company acts as a monopolist and all customers purchase one product at most. To determine a pricing system in terms of a pricing kit maximizing the contribution margin, the following model is proposed:

$$\max \quad \sum_{k=1}^{K} S_k \sum_{m=1}^{M} \sum_{a=1}^{A_m} x_{kma} \, (p_{ma} - C_{ma}) \tag{1}$$

$$\text{s.d.} \quad s_{km} = \sum_{a=1}^{A_m} x_{kma} \left(R_{kma} - p_{ma}\right) \qquad \forall k, m \qquad (2)$$

$$s_{km} \geq R_{kma} \sum_{a'=1}^{A_m} x_{kma'} - p_{ma} \qquad \forall k, m, a \qquad (3)$$

$$\sum_{m=1}^{M} s_{km} \geq 0 \qquad \forall k \qquad (4)$$

$$\sum_{a=1}^{A_m} x_{kma} = \sum_{a=1}^{A_{m'}} x_{km'a} \qquad \forall k, m, m' \qquad (5)$$

$$\sum_{a=1}^{A_m} x_{kma} \leq 1 \qquad \forall k, m \qquad (6)$$

$$x_{kma} \in \{0, 1\} \qquad \forall k, m, a \qquad (7)$$

$$s_{km}, p_{ma} \in \mathbb{R} \qquad \forall k, m, a \qquad (8)$$

The objective function (1) maximizes the contribution margin, which is the sum of products of the segment-related contribution margins with the corresponding segment size. Segment-specific contribution margins can be calculated as the difference between the costs and the price of the individual product. Customers' behavior is forecasted using the First Choice Rule: For each attribute customers choose the level which generates the highest net benefit. If the sum of all net benefits is non-negative the product is bought. This behavior is implemented in (2)–(4). (2) calculates the net benefit for all attributes, (3) assures rationality of customer choice and (4) guarantees non-negativity of net benefit of purchased products. The complete specification from the acquired product is ensured with (5), whereas (6) guarantees that each attribute contains not more than one attribute level. Altogether, these restrictions assure that all customized products consist of exactly M attributes each with one level. Finally, the domain of the decision variables is given in (7) and (8). The model can be linearized introducing customer individual prices p_{kma} which are equal to p_{ma} if customer k buys attribute m in level a and 0 otherwise. The resulting formulation can be solved using standard solvers like Xpress.

3.2 Integration of Pricing Kits

Offering pure pricing kits (PK) is a straightforward way of using pricing kits. In such cases, the entire product line only consists of a single pricing kit. For many products, this seems to be an inadequate consideration of consumers behavior. One cannot assume that each customer segment consists of willingness and necessary knowledge to choose the right product components. Although there may exist many

consumers who are familiar with the product in general and take pleasure in a full customization of the product, there are also customers showing lower interest in product customization. In particular, those customers could be deterred by a pure PK. A combination of pricing kits and pre-configured products could overcome this drawback. Promising extensions are combined pricing kits (CPK) and upgrade pricing kits (UPK). CPK consist of one or more product variants as well as a pricing kit, while UPK consist of basic product variants that can be customized exchanging certain components. This causes several advantages compared to PK. First, customers with low product involvement can easily afford a fully functional basic version. Second, customers with high product involvement are able to customize the product according to their preferences. The main advantage from a corporate perspective is the use of several basic types and different pricing kits which lead to a better price discrimination.

4 Computational Results

In the evaluation of the proposed pricing kit as well as with its extensions a broad simulation study was conducted, which examines the decision situation under certainty as well as under uncertainty. The performance of proposed models is tested against pure PLD strategies using random instances with 4–6 different customer segments of the same size.

All models are solved in their linear formulation using Xpress. The instances differ with respect to variable costs and the structure of willingness-to-pay. The observed products have 6–8 attributes with 2–4 levels. In contrast to the new models it turns

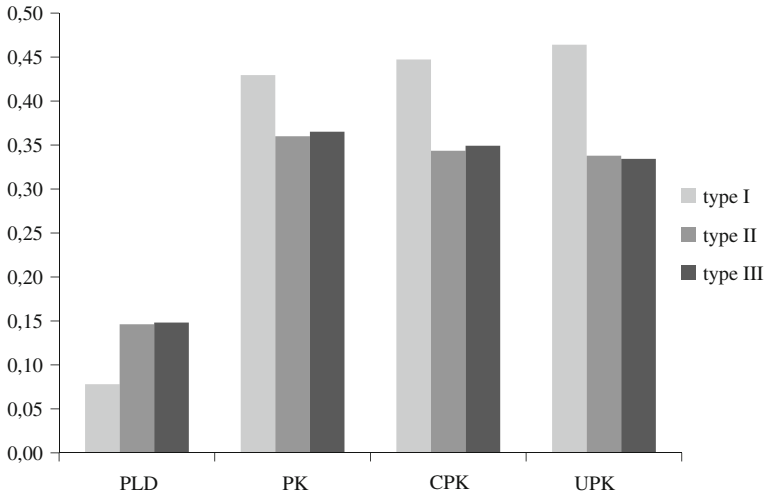


Fig. 1 Purchase probability of test segments

out that competitive margins are only achieved offering every customer segment its own product variant. Almost equal margins are generated using the new models with two or less pre-configured product variants.

To take account of uncertainties, the influence of another customer segment that has not been considered in the optimization is analyzed. For this purpose, all models were optimized using 4 basic segments. Then the purchase probability of a different test segment is estimated. Three different types of WTP structure are considered. The results are illustrated in Fig. 1. One can draw the conclusion that the pure PLD strategy performs significantly worse than the new pricing strategies. This results from the fact that pricing kits and its extensions provide much more flexibility in the product configuration. Hence, the possibility to change some attribute levels leads to significantly higher purchase probabilities.

5 Outlook

In this paper a mathematical model to determine a pricing kit that maximizes contribution margin was presented and compared with traditional strategies of pure product line design. Further enhancements of pricing kits were discussed and analyzed. It turns out that the proposed pricing strategies are not inferior to the traditional strategies in deterministic situations and show a clear advantage in cases of uncertainty.

References

1. Ben-Arieh, D., Easton, T., Choubey, A.M.: Solving the multiple platforms configurations problem. Int. J. Prod. Res. **47**, 1969–1988 (2009)
2. Diller, H.: Preisbaukästen als preispolitische Option. Wist **6**, 270–275 (1993)
3. Green, P.E., Krieger, A.M.: Models and Heuristics for Product Line Selection. Mar. Sci. **4**, 1–19 (1985)
4. Heidbrink, M.: Relibilität und Vailidität von Verfahren zur Präferenzmessung: Ein meta-analytischer Vergleich verschiedener Verfahren der Conjoint-Analyse. VDM Verlag Dr. Müller e. K, Saarbrücken (2008)
5. Kohli, R., Sukumar, R.: Heuristics for Product-Line Design Using Conjoint Analysis. Man. Sci. **36**, 1464–1478 (1990)
6. Schoen, C.: On the Optimal Product Line Selection Problem with Price Discrimination. Man. Sci. **56**, 896–902 (2010)
7. Werners, B., Pietschmann, U.: Optimale dynamische und flexibilitätssteigernde Modulstrategien. In: Bertram, F., Czymmek, F. (eds.) Modulstrategie in der Beschaffung, pp. 53–68. Wolfsburg (2011)

Sparsity of Lift-and-Project Cutting Planes

Matthias Walter

1 Introduction

This work is an extended abstract of the author's diploma thesis and contains the most important concepts and results. It is about a numerical property of a certain cutting plane method in mixed-integer (linear) programming. Many practical problems can be modeled as MIPs. Examples are settings where decisions are modeled with binary variables which are then connected via linear constraints.

Solving MIPs is NP-hard in general, nevertheless large problems can be solved using the combination of many different techniques which evolved during the past decades. Typically, a so-called *branch & cut* method is used which utilizes the linear relaxation to bound the quality of subproblems. The relaxation is created by omitting the integrality constraints and is a *linear program* (LP). The set of feasible solutions of such an LP is a *polyhedron*, the intersection of finitely many halfspaces. Without loss of generality we assume that it is a rational polyhedron of the following form:

$$P = \{x \in \mathbb{Q}^n \colon Ax \leq b\}$$

where $A \in \mathbb{Q}^{m \times n}$ and $b \in \mathbb{Q}^m$ associated to a linear objective function $c^\intercal x$ which is to be minimized.

A solution $x \in \mathbb{Q}^n$ is feasible for the MIP if $x_i \in \mathbb{Z}$ for all $i \in I$ where I is a specified subset of the variables. One of the most important observations in mixed-integer programming is that the convex hull

$$P_I := \operatorname{conv}\{x \in P : x_i \in \mathbb{Z}\ \forall i \in I\}$$

of all feasible points is again a polyhedron.

M. Walter (✉)
Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany
e-mail: matthias.walter@ovgu.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_2,

Applying a bounding procedure based on the LP-relaxation improves the running time dramatically compared to pure enumeration techniques when solving a mixed-integer program. This effect is stronger if the relaxation P approximates P_I more tightly. Hence a fundamental way of improving solver software is by adding more valid inequalities, so-called *cutting planes* (see [1, 6]).

One method to generate valid cutting planes is *lift-and-project* and was introduced in [3]. Here the polyhedron P is split into two polyhedra P_1 and P_2 via a disjunction $\pi^\mathsf{T} x \in (-\infty, \pi_0] \cup [\pi_0 + 1, +\infty)$ where $\pi \in \mathbb{Z}^n$ and $\pi_0 \in \mathbb{Z}$. Then the convex hull of the union $P_1 \cup P_2$ is a stronger relaxation. All inequalities valid for this relaxation which cut off a given point $\hat{x} \in P$ can be found efficiently via the so-called *cut generating LP* (CGLP). The point $\hat{x}$ shall be separated via the cut $\alpha^\mathsf{T} x \leq \beta$.

$$\begin{aligned} \min\ & \beta - \alpha^\mathsf{T} \hat{x} && \text{(CGLP)} \\ \text{s.t. } & \alpha^\mathsf{T} = w^\mathsf{T} A + w_0 \pi^\mathsf{T} \\ & \alpha^\mathsf{T} = v^\mathsf{T} A - v_0 \pi^\mathsf{T} \\ & \beta \geq w^\mathsf{T} b + w_0 \pi_0 \\ & \beta \geq v^\mathsf{T} b - v_0 (\pi_0 + 1) \\ & w, v \in \mathbb{Q}^m_+ \\ & w_0, v_0 \in \mathbb{Q}_+ \end{aligned}$$

We call a cutting plane *sparse* if only a few of its coefficients α_i are non-zero. This property is especially relevant from a practical point of view since modern LP-solvers only work with the non-zeros in memory and hence their number influences the running time. In experiments we mostly measure *density* which is the converse concept.

2 Normalization Constraints

Since (CGLP) is a polyhedral cone it must be truncated in order to be able to optimize over it. This is done by so-called *normalization constraints*. The choice of a good normalization is very important because there is a large number of possible lift-and-project cutting planes. We now shortly present the most important ones that are studied in literature and also suggest another one that attempts to generate sparser cuts.

There are three "primal" normalizations which try to bound the α_i or β.

$$\sum_{i=1}^{n} |\alpha_i| = 1 \qquad (\alpha_1\text{-NC})$$

$$|\alpha_i| \leq 1 \quad (\forall i \in [n]) \qquad (\alpha_\infty\text{-NC})$$

$$|\beta| = 1 \qquad (\beta\text{-NC})$$

The other category of normalization constraints is based on the idea of bounding the multipliers w, w_0, v, and v_0. All of the following truncations have a very attractive property: There is a correspondence between the bases of (CGLP) and the bases of the original LP (see [4]).

$$w_0 + v_0 = 1 \tag{TNC}$$

$$\sum_{i=1}^{m} w_i + w_0 + \sum_{i=1}^{m} v_i + v_0 = 1 \tag{SNC}$$

The second normalization highly depends on the representation of a certain inequality $A_{i,*}x \leq b_i$ in that a scaled version of a row (scaled by some $\lambda > 1$) is preferred over the original row. This happens because it needs a smaller multiplier to get the same result.

If we interpret the values of the multipliers as a resource with capacity equal to 1 we will (on average) use approximately half of it for w and half of it for v. This means that the resulting cut is almost a convex combination of some inequalities of $Ax \leq b$ scaled by $1/2$. The scaling factor in turn means that incorporating a generated lift-and-project cut into another lift-and-project cut is penalized. This fact is considered as a reason that with the SNC the rank of the lift-and-project inequalities remains small even after several rounds of cut generation. Because typical MIPs usually have sparse rows rank 1 cuts with a small dual support (a small number of positive multipliers) are sparse as well. A scaled variant was developed by Fischetti, Lodi and Tramontani in [5].

$$\sum_{i=1}^{m} ||A_{i,*}||_2(w_i + v_i) + ||\pi||_2(w_0 + v_0) = 1 \tag{ENC}$$

It makes the choice of all constraints fair by rescaling the weights. Now we want to additionally penalize dense constraints.

$$\sum_{i=1}^{m} |\text{supp}\left(A_{i,*}\right)| \cdot ||A_{i,*}||_2(w_i + v_i) + ||\pi||_2(w_0 + v_0) = 1 \tag{DNC}$$

Note that by $|\text{supp}(x)|$ we denote the number of non-zeros of vector x.

The interpretation in terms of resources is simple. The CGLP is allowed to incorporate two sparser inequalities instead of a single dense inequality with the same average of multipliers.

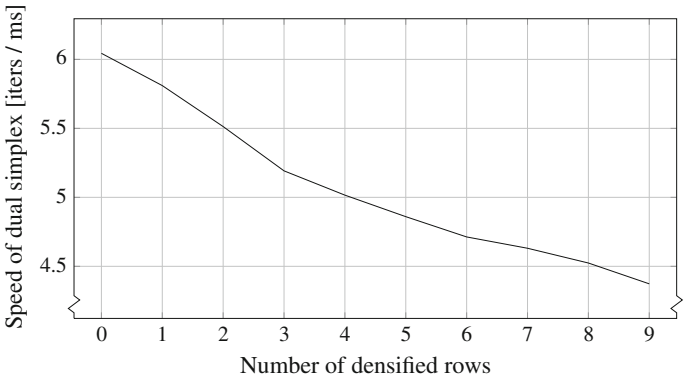


Fig. 1 Simplex speed for densified rows

3 Results

This section is divided into three parts. In the first, the impact of sparsity on the running time of the simplex method is measured. The second summarizes the measurements for the lift-and-project cutting planes and the third describes the main result of the thesis.

3.1 Effects of Sparsity

Before testing the normalization constraints in practice we devised the following experiment in order to quantify the effects of using dense rows. For each instance of the MIPLIB 2003 [2] we created a valid very dense equation $\alpha^\intercal x = \beta$ as a linear combination of other equations. We then ran CPLEX in order to solve the instance, except that we hooked Algorithm 1 into the branching decision callback of CPLEX.

Algorithm 1 Pseudocode for Densification Experiment

1. Get the current optimal basis B from the node LP.
2. For $d = 0, \ldots, 9$, carry out Steps 3.1,. . .,3.2.
3. Copy LP to LP$'$ and apply the CPLEX branching steps (bound tightening) to LP$'$.
4. Add $\alpha^\intercal x = \beta$ to the first d rows of LP$'$.
5. Feed B as a warm-start basis into LP$'$.
6. Solve LP$'$ with the dual simplex method. Measure the number of simplex iterations (pivot steps) and the solving time.

Table 1 Relative densities of cuts for the MIPLIB 2003 instances in % (geometric mean).

Type	Rank	α_∞-NC	α_1-NC	β-NC	TNC	SNC	ENC	DNC
Primal density	$= 1$	82.2	1.9	6.6	8.6	5.4	5.5	5.2
Primal density	≤ 10	86.9	3.5	11.6	23.3	11.5	14.0	10.7
Dual density	$= 1$	23.2	1.8	1.8	2.8	1.2	1.2	1.2
Dual density	≤ 10	23.3	3.3	4.6	5.7	1.9	2.2	1.8

This procedure does not change the remaining behavior of CPLEX and hence the LPs to be solved in Step 3.1 are essentially the same except that d inequalities were "densified" with help of $\alpha^\intercal x = \beta$. So from the polyhedral point of view the LPs were the same, but nevertheless, the solving times were different as Fig. 1 shows.

3.2 Actual and Possible Sparsity

The goal of this part of the work was to compare the different normalizations with respect to sparsity. First we observed that (α_∞-NC) leads to horribly dense cuts and approve former results stating that (β-NC) is instable. Interestingly, the cuts obtained from (α_1-NC) are the best when we are interested only in sparsity but, as shown in [3], they are provably not as strong as other cuts.

Via a big-M-based MIP model we were able to force cut coefficients to zero. Since the (CGLP) has the cut violation as an objective value we measured the cut violation that can be obtained for a given maximal number of non-zero coefficients. The result is that if we strive for a sparsest cut for it helps to allow cut violations of 20 % below. On the other hand, allowing weaker cut violations does not help much for sparsity.

Another known result is that cuts of higher *cut rank* are typically denser than rank-1 cuts. This effect is present especially for the multiplier-based normalizations. Table 1 provides some details although many more can be found in the thesis.

3.3 Dual Sparsity

We also measured the *dual density* which is defined as the average number of non-zero coefficients in the multiplier vectors w and v. Our experiments indicate a strong correlation between the primal and dual density (see Fig. 2). From that we derived our suggested improvement of the Euclidean normalization constraint introduced in [5].

Among the four related multiplier-based normalizations this normalization (DNC) performs best. It is ongoing work to evaluate the normalization in a realistic branch & cut solving process.

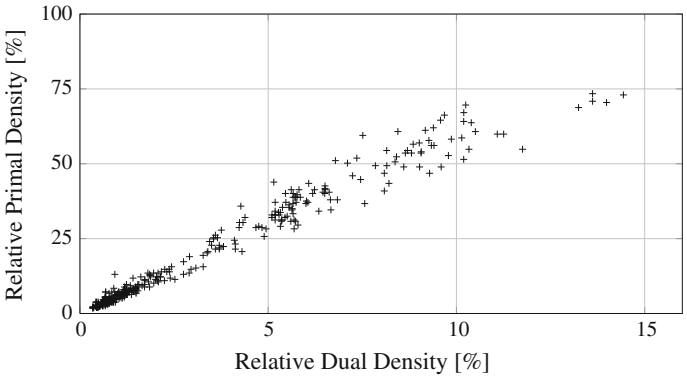


Fig. 2 Correlation between primal and dual density for instance `pp08aCUTS`

Acknowledgments I owe many thanks to my internship advisor Laci Ladanyi in the CPLEX group at IBM as well as my university advisor Volker Kaibel. Additionally, I want to thank Andrea Lodi, Tobias Achterberg and Roland Wunderling for several stimulating discussions.

References

1. Achterberg, T.: SCIP: solving constraint integer programs. Math. Program. Comput. **1**(1), 141 (2009)
2. Achterberg, T., Koch, T., Martin, A.: MIPLIB 2003. Operations Res. Lett. **34**, 361–372 (2006)
3. Balas, E., Ceria, S., Cornuéjols, G.: A lift-and-project cutting plane algorithm for mixed 0–1 programs. Math. Program. **58**, 295324 (1993)
4. Balas, E., Perregaard, M.: A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed integer gomory cuts for 0–1 programming. Math. Program. **94**, 221–245 (2003)
5. Fischetti, M., Lodi, A., Tramontani, A.: On the separation of disjunctive cuts. Math. Program. **128**, 205–230 (June 2011)
6. Wolter, K.: Implementation of cutting plane separators for mixed integer programs. Masters thesis (2006)

Railway Track Allocation

Thomas Schlechte

1 Micro-Macro Transformation of Railway Networks

A major challenge is modeling railway systems to allow for resource and capacity analysis. Railway capacity has basically two dimensions, a space dimension which are the physical infrastructure elements as well as a time dimension that refers to the train movements, i.e., occupation or blocking times, on the physical infrastructure. Railway safety systems operate on the same principle all over the world. A train has to reserve infrastructure blocks for some time to pass through. Two trains reserving the same block of the infrastructure within the same point in time is called *block conflict*. Therefore, models for railway capacity involve the definition and calculation of reasonable running and associated reservation and blocking times to allow for a conflict free allocation.

There are *microscopic* models that describe the railway system extremely detailed and thorough. Microscopic models have the advantage that the calculation of the running times and the energy consumption of the trains is very accurate. A major strength of microscopic models is that almost all technical details and local peculiarities are adjustable and are taken into account. Railway system on a microscopic scale covers the behavior of trains and the safety system completely and correctly. Those models of the railway infrastructure are already very large even for very small parts of the network. The reason is that all signals, incline changes, and switches around a railway station have to be modeled to allow for precise running time calculations of trains. In general microscopic models are used in simulation tools which are nowadays present at almost all railway companies all over the world. The most important field of application is to validate a single timetable and to decide whether a timetable is operable and realizable in practice. However, microscopic models are inappropriate for mathematical optimization because of the size and the high level of detail. Hence,

T. Schlechte (✉)
Zuse Institute Berlin, Takustr.7, 14195 Berlin, Germany
e-mail: schlechte@zib.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_3,

most optimization approaches consider simplified, so called *macroscopic*, models. The challenging part is to construct a reliable macroscopic model for the associated microscopic model and to facilitate the transition between both models of different scale.

In order to allocate railway capacity significant parts of the microscopic model can be transformed into aggregated resource consumption in space and time. We develop a general macroscopic representation of railway systems which is based on minimal headway times for entering tracks of train routes and which is able to cope with all relevant railway safety systems. We introduce a novel bottom-up approach to generate a macroscopic model by an automatic aggregation of simulation data produced by any microscopic model. The transformation aggregates and shrinks the infrastructure network to a smaller representation, i.e., it conserves all resource and capacity aspects of the results of the microscopic simulation by conservative rounding of all times. The main advantage of our approach is that we can guarantee that our macroscopic results, i.e., train routes, are feasible after re-transformation for the original microscopic model. Because of the conservative rounding macroscopic models tend to underestimate the capacity. Furthermore, we can control the accuracy of our macroscopic model by changing the used fixed time discretization. Finally, we provide a priori error estimations of our transformation algorithm, i.e., in terms of exceeding of running and headway times. We implemented our new transformation algorithm in a tool called `NETCAST`. The technical details can be found in [9].

2 Optimal Railway Track Allocation

The main application of railway track allocation is to determine the best operational implementable realization of a requested timetable, which is the main focus of our work. But, we want to mention that in a segregated railway system the track allocation process directly gives information about the infrastructure capacity. Imaging the case that two trains of a certain type, i.e., two train slots, are only in conflict in one station. A potential upgrade of the capacity of that station allows for allocating both trains. This kind of feedback to the department concerning network design is very important. Even more long-term infrastructure decisions could be evaluated by applying automatically the track allocation process, i.e., without full details only on a coarse macroscopic level but with different demand expectations. Hence, suitable extensions or simplifications of our models could support infrastructure decisions in a quantifiable way. For example major upgrades of the German railway system like the high-speed route from Erfurt to Nürnberg or the extension of the main station of Stuttgart can be evaluated from a reliable resource perspective. The billions of euros for such large projects can then be justified or ranked by reasonable quantifications of the real capacity benefit with respect to the given expected demand.

The optimal track allocation problem for macroscopic railway models can be formulated with a graph-theoretic model. In that model optimal track allocations correspond to conflict-free paths in special time-expanded graphs.

More specifically, we studied four types of integer programming model formulations for the track allocation problem: two standard formulations that model resource or block conflicts in terms of packing constraints, and two novel coupling or "configuration" formulations. In both cases we considered variants with either arc variables or with path variables. The classical way of formulating the train pathing problem is by using single flow formulations for the trains which are coupled by additional packing constraints to exclude conflicts between those trains, see [2] or [7]. In [1] we introduced an alternative formulation for the track allocation problem. The key idea of the new formulation is to use additional "configuration" variables that are appropriately coupled with the standard "train" flow variables to ensure feasibility. We show that these models are a so called "extended" formulations of the standard packing models. This arc coupling formulation (ACP) formulation is based on the concept of feasible arc *configurations*, i.e., sets of time-expanded arcs on a track without occupation conflicts. Given is a standard time expanded scheduling graph D, in which time expanded paths represent train routes and departure and arrival times for the individual train requests $i \in I$ in the subgraph $D_i \subseteq D$. Operational feasiblilty with respect to headway times between two consecutive trains is handled by artificial digraphs D_j for each $j \in J$. Each two train moves on track j which respect the minimal headway time are connected by artificial arcs in D_j.

In case of a transitive headway matrix paths in D_j represent valid consecutive successions of discrete train moves on track j. Note that each arc representing a train move is contained in exactly one D_j and one D_i and thus induces two variables x_a and y_a. More details on the construction of D_i and D_j can be found in [8].

In the ACP formulation variables x_a, $a \in A_i$, $i \in I$ control the use of arc a in D_i and y_b, $b \in A_j$, $j \in J$ in D_J, respectively. Note that only arcs which represent track usage, i.e., $A_S = A_I \cap A_J$, have to be coupled in that formulation by two variables x_a and y_b. The objective, denoted in ACP (1), is to maximize the weight of the track allocation, which can be expressed as the sum of the arc weights w_a. Equalities (2) and (4) are well-known *flow conservation constraints* at intermediate nodes for all trains flows $i \in I$ and for all flows on tracks $j \in J$, (3) and (5) state that at most one flow, i.e., train and track, unit is realized. Equalities (6) link arcs used by train routes and track configurations to ensure a conflict-free allocation on each track individually. Packing constraints (7) ensures that no station capacity is violated in its most general form. Finally, (8) states that all variables are binary.

The success of an integer programming approach usually depends on the strength of the linear programming (LP) relaxation. Hence, we analyze the LP relaxations of our model formulations.

We show, that in case of block conflicts, the packing constraints in the standard formulation stem from cliques of an interval graph and can therefore be separated in polynomial time. It follows that the LP relaxation of a strong version of this model, including all clique inequalities from block conflicts, can be solved in polynomial time. We prove that the LP relaxation of the extended formulation for which the number of variables can be exponential, can also be solved in polynomial time, and that it produces the same LP bound. The path variant of the coupling model provides a strong LP bound, is amenable to standard column generation techniques,

and therefore suited for large-scale computation.

(ACP)

$$\max \sum_{a \in A} w_a x_a \tag{1}$$

$$s.t. \sum_{a \in \delta^i_-(v)} x_a - \sum_{a \in \delta^i_+(v)} x_a = 0, \quad \forall i \in I, v \in V_i \backslash \{s_i, t_i\} \tag{2}$$

$$\sum_{a \in \delta^i_-(s_i)} x_a \leq 1, \quad \forall i \in I \tag{3}$$

$$\sum_{a \in \delta^i_-(v)} y_a - \sum_{a \in \delta^i_+(v)} y_a = 0, \quad \forall j \in J, v \in V_j \backslash \{s_j, t_j\} \tag{4}$$

$$\sum_{a \in \delta^i_-(s_j)} y_a \leq 1, \quad \forall j \in J \tag{5}$$

$$x_a - y_a = 0, \quad \forall a \in A_S \tag{6}$$

$$\sum_{a \in A_c} x_a \leq \kappa_c, \quad \forall c \in C \tag{7}$$

$$x_a, y_a \in \{0, 1\}, \quad \forall a \in A_I, a \in A_J. \tag{8}$$

Furthermore, we present a sophisticated solution approach that is able to compute high-quality integer solutions for large-scale railway track allocation problems in practice. Our algorithm is a further development of the rapid branching method introduced in [3] (see also the thesis [10]) for integrated vehicle and duty scheduling in public transport. The method solves a Lagrangean relaxation of the track allocation problem as a basis for a branch-and-generate procedure that is guided by approximate LP solutions computed by the bundle method. This successful second application in public transportation provides evidence that the rapid branching heuristic guided by the bundle method is a general heuristic method for large-scale path packing and covering problems. All models and algorithms are implemented in a software module `TS-OPT`, see [5] and [4].

3 Computational Studies

Finally, we go back to practice and present several case studies using the tools `NETCAST` and `TS-OPT`. In [8] we provide a computational comparison of our new models and standard packing models used in the literature. Our computational experience indicates that our approach, i.e., “configuration models”, outperforms other models. Moreover, the rapid branching heuristic and the bundle method enables

us to produce high quality solutions for very large scale instances, which has not been possible before.

The highlights are results for the Simplon corridor in Switzerland, see [6]. We optimized the train traffic through this tunnel using our models and software tools. To the best knowledge of the author and confirmed by several railway practitioners this was the first time that fully automatically produced track allocations on a macroscopic scale fulfill the requirements of the originating microscopic model, withstand the evaluation in the microscopic simulation tool `OpenTrack`, and exploit the infrastructure capacity. This documents the success of our approach in practice and the usefulness and applicability of mathematical optimization to railway track allocation.

Acknowledgments Finally, I thank Martin Grötschel, Gottfried Ilgmann, and Klemens Polatschek for their important support in organizing and realizing this project, and in particular the Simplon case-study. Most of all I want to thank my supervisor Dr. habil. Ralf Borndörfer and special thanks goes also to Dr. Steffen Weider. Applied research is really applied only if it is done and evaluated in close collaboration with an industrial and operating partner. Therefore, I am very thankful for all discussions with external experts in particular from Swiss Federal Railways (SBB). Special thanks go to Thomas Graffagnino, Martin Balser, and Elmar Swarat for explaining various technical details of railway systems and discussing several results. In addition, I want to thank Daniel Hürlimann for his support for the simulation tool `OpenTrack`.

References

1. Borndörfer, R., Schlechte, T.: Models for railway track allocation, in ATMOS 2007–7th workshop on algorithmic approaches for transportation modeling, optimization, and systems. In: Liebchen, C., Ahuja, R.K., Mesa, J.A. (eds.) Vol. 07001 of Dagstuhl Seminar Proceedings, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany (2007).
2. Brännlund, U., Lindberg, P., Nou, A., Nilsson, J.-E.: Railway timetabling using langangian relaxation. Transp. Sci. **32**, 358–369 (1998)
3. Borndörfer, R., Löbel, A., Weider, S.: A bundle method for integrated multi-depot vehicle and duty scheduling in public transit. Computer-aided systems in public transport (CASPT 2004). In: Hickman, M., Mirchandani, P., Voß, S. (eds.) Vol. 600 of Lecture Notes in Economics and Mathematical Systems, pp. 3–24. Springer (2008).
4. Borndörfer, R., Löbel, A., Reuther, M., Thomas, S., Weider, S.: Rapid branching, ZIB report 12–10, ZIB, Takustr. 7, 14195. Submitted 17.02.2011 to Proceedings of Conference on Advanced Systems for Public, Transport 2012 (CASPT12), Berlin (2010).
5. Borndörfer, R., Schlechte, T., Weider, S.: Railway track allocation by rapid branching. In: Erlebach, T., Lübbecke, M. (eds.) Proceedings of the 10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. OpenAccess Series in Informatics (OASIcs), vol. 14, pp. 13–23. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2010).
6. Borndörfer, R., Erol, B., Graffagnino, T., Schlechte, T., Swarat, E.: Optimizing the simplon railway corridor. Ann. Oper. Res. 1–14 (2012). <error l="84" c="Undefined command " />.
7. Caprara, A., Monaci, M., Toth, P., Guida, P.L.: A lagrangian heuristic algorithm for a real-world train timetabling problem. Discrete Appl. Math. **154**, 738–753 (2006)

8. Schlechte, T.: Railway Track Allocation-Models and Algorithms. PhD thesis, Technische Universität Berlin (2012).
9. Schlechte, T., Borndörfer, R., Erol, B., Graffagnino, T., Swarat, E.: Micro-macro transformation of railway networks. J. Rail Transp. Plann. Manag. **1**, 38–48 (2011)
10. Weider, S.: Integration of Vehicle and Duty Scheduling in Public Transport. PhD thesis, TU Berlin (2007).

Integrating Routing Decisions in Network Problems

Marie Schmidt

1 Introduction

To model and solve optimization problems arising in public transportation, data about the passengers is necessary and must be included in the models in any phase of the planning process. Many solution approaches (see, e.g., [7, 9, 10, 12]) assume a two-step procedure: firstly, the data about the passengers is distributed over the public transportation network using traffic-assignment procedures. Secondly, the actual planning of lines, timetables, disposition timetables, etc. is done. This approach ignores that for most passengers there are many possible ways to reach their destinations, thus the actual connections the passengers take depend strongly on the decisions made during the planning phase.

Hence, in order to find better and more realistic solutions, the traffic assignment (or *routing*) procedure should be integrated in the optimization process. Some attempts to integrate the routing step and the optimization step have been made in [1, 2] for line planning and in [6, 8, 13] for timetabling and a hardness result for delay management with dynamic path choices is given in [4].

However, to the extend of our knowledge, [11] is the first comprehensive approach to a conceptual integration of routing decisions in different public transportation problems.

In this article, we focus on the problem delay management: Given an operating public transportation system and a set of delays occurring in the daily operations, the task of delay management is to decide whether trains should wait for delayed feeder trains or depart on time and to update the existing timetable to a *disposition timetable*.

Our goal in the *delay management with routing* (*DMwR*) problem is to determine a disposition timetable which is optimal in the sense that it allows a routing with

M. Schmidt (✉)
Institute for Numerical and Applied Mathematics, Lotzestraße 16-18, 37083 Göttingen, Germany
e-mail: m.schmidt@math.uni-goettingen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_4,

travel time shorter than or equal to the travel time of routings that are possible for other disposition timetables.

In Sect. 2 we formalize the definition of DMwR before in Sects. 3 and 4 we give an overview on the algorithms and complexity results for DMwR obtained in [11].

2 Delay Management with Routing

In delay management problems we are given a *public transportation network* $\mathscr{G} = (\mathscr{S}, \mathscr{T})$ which represents the underlying infrastructure consisting of *stations* $\mathscr{S}$ and *tracks* $\mathscr{T}$. There is a set of trains operating on this network according to a *timetable* π which assigns times to every *arrival* and every *departure* of a train in the network. This information is often coded in an *event-activity network* (*EAN*) (see Fig. 1), as introduced in [9] for timetabling problems and used for classical delay management problems in [12]. The nodes of this network correspond to the arrival and departures of trains and are called *events*. The events are connected by directed arcs which represent the driving of trains (*driving activities* $\mathscr{A}_{\text{drive}}$), the waiting of trains at stations (*waiting activities* $\mathscr{A}_{\text{wait}}$) and the transfers of passengers from one train to another train at a station (*changing activities* $\mathscr{A}_{\text{change}}$). The activities are assigned lower bounds representing minimal time durations of the driving of trains, train waiting times at stations to let passengers get on and off, and the time passengers need to transfer from one train to another, respectively.

Furthermore, there are *source delays* d_i assigned to departures and arrivals i. The 'classical' task of delay management is to decide which of the connections, modeled by the changing activities in the EAN, should be maintained (*wait-depart decisions*) and to adapt the timetable to a feasible *disposition timetable* x accordingly. In terms of the EAN, the feasibility of a disposition timetable x can be formulated as

$$x_i \geq \pi_i + d_i \ \forall i \in \mathscr{E} \text{ and } x_j \geq x_i + l_{(i,j)} \ \forall (i,j) \in \mathscr{A}_{\text{drive}} \cup \mathscr{A}_{\text{wait}} \cup \mathscr{A}_{\text{fix}}$$

where $\mathscr{A}_{\text{fix}} \subset \mathscr{A}_{\text{change}}$ denotes the set of changing activities corresponding to the maintained connections. The first condition says that no event can happen earlier than scheduled in the original timetable (plus the amount of source delay in the event). The second condition ensures that in the disposition timetable the lower bounds on the driving, waiting and changing times are respected.

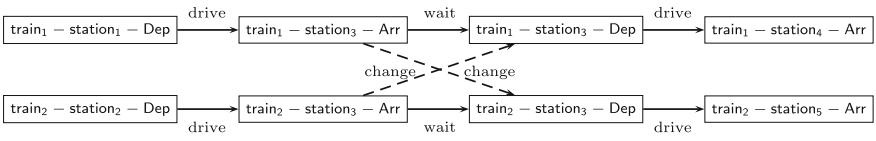


Fig. 1 Cutout of an EAN

In [12] it has been shown that given the wait-depart decisions, i.e., a set $\mathscr{A}_{\text{fix}} \subset \mathscr{A}_{\text{change}}$, a disposition timetable $x(\mathscr{A}_{\text{fix}})$ which minimizes the arrival times at all arrival events simultaneously can be found in polynomial time by the *critical path method*.

In DMwR, passenger demand is given in form of OD-pairs (u, v, r) specifying origin station u, destination station v and planned starting time r of a passenger or a group of passengers. We denote the set of OD-pairs as $\mathscr{OD}$. To allow that passengers are freely routed, *origin events* and *destination events* are added to the EAN and, in case of origin events, connected to all departure events at the origin station that take place after the planned starting time r by *origin activities* $\mathscr{A}_{\text{org}}$, or, in case of destination events to all arrival events at the destination station that take place after time r by *destination activities* $\mathscr{A}_{\text{dest}}$.

Now every path from origin event to destination event in the EAN defines a possible route for the corresponding passenger(s). We call a set of routes $\mathscr{R} = \{P_{(u,v,r)} : (u, v, r) \in \mathscr{OD}\}$, where $P_{(u,v,r)}$ is a path for OD-pair (u, v, r), a *routing*. Certainly, given a routing $\mathscr{R}$, every changing activity contained in a path $P \in \mathscr{R}$ must be contained in $\mathscr{A}_{\text{fix}}$, i.e., passengers can only transfer if connections are maintained. Or, the other way round: given the wait-depart decisions, a routing has to be found in the EAN $\mathscr{N}(\mathscr{A}_{\text{fix}}) := (\mathscr{E}, \mathscr{A}_{\text{drive}} \cup \mathscr{A}_{\text{wait}} \cup \mathscr{A}_{\text{fix}} \cup \mathscr{A}_{\text{org}} \cup \mathscr{A}_{\text{dest}})$ where all changing activities which correspond to dropped connections are removed.

Given a set of maintained connections $\mathscr{A}_{\text{fix}}$, the corresponding disposition timetable $x := x(\mathscr{A}_{\text{fix}})$, and a routing $\mathscr{R} = \{P_{(u,v,r)} : (u, v, r) \in \mathscr{OD}\}$ in $\mathscr{N}(\mathscr{A}_{\text{fix}})$, the *travel time* for an OD-pair (u, v, r) can be calculated as the time span $l(\mathscr{A}_{\text{fix}}, P_{(u,v,r)}) := x_j - r$ between planned departure time r and the arrival time x_j at the last arrival event (the one at station v) on the corresponding path $P_{(u,v,r)}$ in the EAN.

Note that this corresponds to the length of path $P_{(u,v,r)}$, when we assign arc lengths

$$L_{(i,j)} := \begin{cases} x_j - x_i & \text{if } (i, j) \in \lambda\ \mathscr{A}_{\text{drive}} \cup \mathscr{A}_{\text{wait}} \cup \mathscr{A}_{\text{fix}}), \\ x_j - r & \text{if } i \text{ is an origin event}, \\ 0 & \text{if } j \text{ is a destination event}. \end{cases} \tag{1}$$

to the arcs of the $\mathscr{N}(\mathscr{A}_{\text{fix}})$.

In DMwR, our objective is to find a solution, consisting of a set of maintained connections and a routing such that the passengers' overall travel time

$$\min \sum_{(u,v,r)\in\mathscr{OD}} w_{(u,v,r)} l(\mathscr{A}_{\text{fix}}, P_{(u,v,r)})$$

is minimized where $w_{(u,v,r)}$ denotes the number of passengers in OD-pair (u, v, r).

Note that this is equivalent to finding a solution with minimal overall delay or minimal overall arrival time.

3 Exact Solution Approaches to DMwR

Even for *traditional delay management problems*, i.e., when passengers are assumed to stay on their initially planned routes and are not allowed to adapt their itineraries to changes in the timetable, polynomial-time solution algorithms are known only for very restricted cases (see [4, 12]) and NP-hardness is shown under quite hard restrictions [5].

Intuitively, the hardness of the delay management problem stems from the fact that the wait-depart-decisions are ambiguous: If a train tr_1 arrives with a delay at a station and passengers want to transfer to another train tr_2, we could either

1. make tr_2 depart on time: then the transferring passengers have to wait until the next period or to switch to another, possibly much longer connection to their destination,
2. or make train tr_2 wait for the passengers from train tr_1: in this case, tr_2 will also be delayed, resulting in delayed arrivals of the passengers that are already in this train and wait-depart decisions for the connecting trains at the next stations visited by tr_2.

Hence, one would expect that if there are no such conflicts, e.g., because there is only one passenger, delay management should be fairly easy to solve. Surprisingly, this is not the case when passengers are allowed to adapt their routes to timetable changes: In [11] we show by reduction from *Hamiltonian Path* that without further restrictions on the problem instances, DMwR is strongly NP-hard even if there is only one passenger.

However, using a modified Dijkstra-type shortest path algorithm in the corresponding EAN which keeps track of the delays a passenger 'accumulates' on his route, good solution to DMwR with one OD-pair can be found. More precisely, the Dijkstra-type algorithm finds an optimal solution if the EAN and the OD-pair are structured in such a way that the passenger cannot find a (reasonably short) route where he/she would enter a train more than once. Otherwise, the algorithm provides a feasible solution that is a 2-approximation (with respect to the travel time) as well as a lower bound on the optimal objective value. The latter result can be generalized to obtain a lower bound on the objective value of instances of DMwR with more than one OD-pairs, by calculating and summing up the lower bounds for every single OD-pair. This property has been successfully used in [3] to improve running times when solving an integer programming formulation of DMwR. The Dijkstra-type algorithm is detailed in [11], a previous version can be found in [3].

Dijkstra's algorithm for finding a shortest path in a graph $G = (V, E)$ from an origin u to a destination v can be easily generalized to find a set of shortest paths from u to all other nodes $v \in V \setminus \{u\}$. It is hence a natural question, whether such a generalization is also possible for our Dijkstra-type algorithm for one OD-pair. In [3, 11] we answer this question, showing that DMwR is strongly NP-hard, even if all OD-pairs have the same origin and no passenger can find a route where he/she would enter a train more than once.

Since DMwR is NP-hard already in very restricted cases, we propose to solve the problem using integer programming techniques. A flow-based integer programming formulation has been given in [3, 11]. This formulation has been successfully tested on real-world instances in [3], using the lower bound provided by the polynomial-time algorithm for one OD-pair to improve running times.

4 Heuristic Solution Approaches to DMwR

Although in principle DMwR can be solved to optimality by integer programming, solution times grows big for large problem instances. Hence it seems suitable to search for heuristic approaches for solving DMwR.

Two types of decisions are considered in the problem DMwR:

1. *Wait-depart decisions*: For every connection between two trains it has to be decided whether it is maintained or not.
2. *Routing decisions*: For every OD-pair a route from origin to destination has to be chosen.

Given the wait-depart decisions, for every OD-pair an optimal routing can be found using a shortest-path algorithm in $\mathcal{N}(\mathcal{A}_{\text{fix}})$ with arc lengths as defined in (1). On the other hand, given a routing and assuming that all connections on these routes have to be maintained, an optimal disposition timetable for the given routing can be found in polynomial time using the critical path method [12].

Hence a reasonable and fast approach to find good solutions to DMwR is to iterate wait-depart decision and routing decisions. If all individual steps of this heuristic approach are solved optimally, starting with an arbitrary feasible solution, in every step of the heuristic algorithm we find a feasible solution and the objective values cannot deteriorate during the procedure. See [11] for a detailed description and analysis for this approach which is applicable not only for DMwR but for a whole class of network problems with integrated routing.

Unfortunately, there is no guaranteed approximation ratio for solving DMwR using the heuristic approach described above. In [11] we hence investigated the existence of polynomial-time approximation algorithms for DMwR. When considering approximation algorithms, the choice of the objective function, i.e., whether we consider the delay of the passengers or their travel times is crucial. It follows directly from the aforementioned NP-hardness proof that even considering only instances with one passenger there is no polynomial-time constant-factor approximation algorithm for DMwR with *delay objective* unless $P = NP$. In contrast, considering the *travel time objective*, the aforementioned Dijkstra-type algorithm finds a 2-approximation for DMwR with one OD-pair. However, this result cannot be generalized to more than one OD-pair: Even for two OD-pairs having the same origin there is no polynomial-time constant-factor approximation algorithm for DMwR unless $P = NP$.

5 Conclusions

Integrating the passengers' routing decisions into the process of making optimal wait-depart decisions leads to a more realistic representation of passengers' routes and hence to solutions with better objective values than traditional delay management approaches. Unfortunately, DMwR is strongly NP-hard and hard to approximate even in very restricted cases. However, good lower bounds can be found using a Dijkstra-type algorithm, the problem can be solved to optimality by integer programming, and heuristic solutions can be determined using an iterative approach.

References

1. Borndörfer, R., Grötschel, M., Pfetsch, M.E.: Models for line planning in public transport. In: Hickman, M., Mirchandani, P., Voß, S. (eds.) Computer-aided Systems in Public Transport. Lecture Notes in Economics and Mathematical Systems, pp. 363–378. Springer, Berlin (2008)
2. Borndörfer, R., Grötschel, M., Pfetsch, M.E.: A column-generation approach to line planning in public transport. Transp. Sci. **41**(1), 123–132 (2007)
3. Dollevoet, T., Huisman, D., Schmidt, M., Schöbel, A.: Delay management with re-routing of passengers. Transp. Sci. **46**(1), 74–89 (2011)
4. Gatto, M., Glaus, B., Jacob, R., Peeters, L., Widmayer, P.: Railway delay management: exploring its algorithmic complexity. In: Proceedings of 9th Scandinavian Workshop on Algorithm Theory (SWAT). Lecture Notes in Computer Science, vol. 3111, pp. 199–211 (2004)
5. Gatto, M., Jacob, R., Peeters, L., Schöbel, A.: The computational complexity of delay management. In: Kratsch, D. (ed.) Graph-Theoretic Concepts in Computer Science: 31st International Workshop (WG 2005). Lecture Notes in Computer Science, vol. 3787 (2005)
6. Kinder, M.: Models for periodic timetabling. Master's thesis, TU Berlin (2008)
7. Liebchen, C.: Periodic timetable optimization in public transport. PhD thesis, TU Berlin (2006) (published by dissertation.de)
8. Lübbe, J.: Passagierrouting und Taktfahrplanoptimierung. Master's thesis, TU Berlin (2009)
9. Nachtigall, K.: Periodic Network Optimization and Fixed Interval Timetables. Deutsches Zentrum für Luft- und Raumfahrt, Institut für Flugführung, Braunschweig (1998). Habilitationsschrift
10. Peeters, L.: Cyclic railway timetable optimization. PhD thesis, Erasmus University Rotterdam (2003)
11. Schmidt, M.: Integrating routing decisions in network problems. PhD thesis. Universität Göttingen (2012)
12. Schöbel, A.: Optimization in public transportation. Stop location, delay management and tariff planning from a customer-oriented point of view. Optimization and Its Applications. Springer, New York (2006).
13. Siebert, M.: Integration of routing and timetabling in public transportation. Master's thesis, Georg-August-Universität Göttingen (2011)

Supply Chain Coordination in Case of Asymmetric Information: Information Sharing and Contracting in a Just-in-Time Environment

Guido Voigt

1 Introduction

Supply chain interactions are frequently analyzed in idealized and normative model settings, in which the system behavior is described for fully rational and profit maximizing decision makers. Optimal decisions and strategies in these normative frameworks are typically derived by applying Operations Research methods in order to make predictions regarding the supply chain performance even for complex settings involving, e.g., uncertainties or information asymmetries. However, recent work in experimental economics and behavioral operations management is showing that such a normative approach cannot describe supply chain outcomes that are characterized by bounded rational decisions or objectives different from (expected) profit maximization (e.g., decision makers apply heuristics or care for fairness in the business relationship). The underlying work, therefore, combines methods of mathematical model analysis and experimental economics in order to identify critical assumptions that should be relaxed in existing normative modeling approaches in order to test the external validity of the proposed models.

Supply chain coordination is a prominent research field applying normative modeling approaches. A major challenge of supply chain coordination is to align the incentives of the supply chain parties such that decentralized decision making leads to an overall supply chain efficient outcome. Yet, if the incentives are not sufficiently aligned, then the supply chain members might deliberately exploit the supply chain counterpart to enhance own financial ratios. In this situation, the implementation of specific concepts like, e.g., just-in-time delivery is likely to fail unless there are mechanisms avoiding the pitfalls of opportunistic behavior in advance. Very prominent tools in the supply chain management literature for avoiding opportunistic behavior are contracts that legally stipulate the business relation between the supply chain par-

G. Voigt (✉)
Faculty of Management and Economics, Operations Management,
Otto-von-Guericke University Magdeburg, Magdeburg, Germany
e-mail: guido.voigt@ovgu.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_5,

ties. These contracts are typically derived by applying advanced Operations Research methods ranging from stochastic programming in the newsvendor context to non-linear optimization in settings under asymmetric information and strive for linking the actions of the company to individual performance measures, e.g., costs or revenues to the overall supply chain performance. In other words, contracts try to align the incentives such that optimizing the individual performance is in line with optimizing the supply chain performance.

In this context, the information availability within the supply chain becomes particularly important, since opportunistic behavior might translate to deception and mistrust and, thus, to an overall deterioration of supply chain performance. If all parties are acting fully rational and opportunistic, it can be shown that the supply chain optimal outcome cannot be achieved. However, in this context non-linear contract schemes (so-called "Screening Contracts" or "Menu-of-Contracts") can be derived by applying Operations Research methods of non-linear optimization (e.g., lagrangian method or optimal control theory) which coordinate the supply chain to the second-best outcome, i.e., the best outcome achievable if decentralized decision makers are fully rational and opportunistic. These non-linear "Menu-of-Contracts" incentivize the superior informed party to (ex-post) reveal the private information by the contract choice. The underlying work (see [1]) provides a comprehensive review over the contributions that are applying screening theory in the Supply Chain and Operation Management context. Moreover, the work contributes to the existing literature on game-theory under asymmetric information by applying Operations Research methods of non-linear optimization in a supply chain coordination context in order to gain deeper insights into the impact of communication in supply chain interactions, since communication is regarded as a main facilitator for establishing cooperative supply chain behavior. In the following, the main contributions presented in [1] are outlined.

2 A Strategic Lot Sizing Framework

The screening theory is applied to a strategic economic lot sizing framework in which the supply chain parties (i.e., buyer and supplier) negotiate the means of delivery in a supply chain context (see also [2]). In this model framework, the buyer possesses private information regarding his advantages of low order sizes (namely, the buyer possesses superior information regarding his holding costs), while the supplier only possesses an a-priori distribution over possible holding cost realizations. Moreover, the model depicts the supplier's option to invest in process improvements in order to facilitate lower order sizes. The model under investigation is characterized by a non-linear objective function, whose properties (convexity/concavity) specifically depend on the investment cost function, i.e., the cost occurring for process improvements. In particular, the optimization problem becomes more difficult compared to classical approaches in the screening literature, since the objective function loses its convexity property for linear and concave investment functions. The analysis,

therefore, concentrates on the marginal revenues and marginal costs that are related to the investment decision and it can be shown that asymmetric information lead to inefficient outcomes regardless of the specific form of the investment cost function, because the supplier tends to overinvest in process improvements. Since all typical investment cost functions are incorporated into the analysis, the model allows for identifying industrial settings in which overinvestments in process improvements are a major threat for the overall supply chain efficiency.

3 Laboratory Testing of Screening Theory

One main contribution is the presentation of an empirical test of the screening theory obtained by several controlled laboratory experiments (see also [3]). The problems that arise under asymmetric information heavily rely on the assumptions of the supply chain members acting opportunistically, and having therefore an incentive to misrepresent their private information. Moreover, it is implicitly assumed that the actors are fully rational, i.e., they easily identify how to misrepresent their information or how to exploit truthfully shared information. Yet, there is a growing body of research in the field of Behavioral Economics and Operation Management showing that these assumptions need to be handled carefully. In the context of asymmetric information, the assumptions of rationality and opportunistic behavior imply that supply chain parties will prefer inefficient complicated screening-contracts (non-linear optimization problems being solved with methods like, e.g., optimal control theory or lagrangian method) under asymmetric information to efficient and less complicated contract structures under full information. In other words, theory predicts that supply chain members are more likely to use complicated inefficient contracts instead of establishing truthful communication and trust while using simpler contract schemes. However, so far no empirical assessment of different contract structures under asymmetric information has been performed, which is not surprising, since there are substantial methodological difficulties (e.g., full access to critical information is typically not granted, benchmark is suffering from statistical errors) to assess the amount of truth-telling and the extent of trust in supply chain interactions. For this reason, the present work employs laboratory experiments as a technique to address the above questions in a controlled environment, in which communication between the supply chain parties is perfectly observable. The laboratory experiments allow to directly control for the buyer's actual private information while observing if communication is truthful or deceptive and if communication is trusted or mistrusted.

The first experimental design measures the adjustment of the supplier's statistical forecast of the private information as a reaction on the buyer's communicated report. This adjusted forecast is the input the supplier is giving into a decision support system. This decision support system allows for measuring the supplier's tendency to trust or mistrust the reported information while abstracting from the complexities arising from non-linear optimization. This measurement of trust and trustworthiness allows for analyzing the interaction of behavioral types (trusting and mistrusting suppliers,

deceptive and truthful buyers) under asymmetric information. The results show that there are two critical assumptions within the screening theory which are not fully supported by the conducted experiment. First of all, it is shown that communication under asymmetric information is not fully ignored by the supply chain parties, but instead factored in when building an expectation regarding the information that is asymmetrically distributed. On average, the results indicate that those supply chains perform best in which trust (i.e., believing in reported information) and trustworthiness (i.e., information are true and consistent) is prevalent. Second, the critical assumption stating that even very small-pay-off differences suffice to coordinate the supply chain parties to the screening benchmark, proves to be a major source of inefficiencies within the supply chain. This is, because the superior informed party faces very little cost when deviating from the equilibrium path while the impact on the less informed party is huge. Interestingly, this critical assumption leads to the outcome that even the second-best outcome cannot be achieved, even if communication between the supply chain parties is facilitated.

In another experimental design, it is analyzed how the supply chain performance can be improved by introducing theoretical irrelevant but behavioral highly relevant decision variables into the supply chain interaction. For this purpose, a second series of experiments was performed in which the focus of the analysis was rather on the complexity resulting from non-linear contract types than on the subtle measurement of statistical forecast adjustments. For this purpose, the supplier's contract offer options were limited to the screening contract based on the a-priori probabilities and the simple wholesale-price contract. Thus, the supplier's decision was either to trust and to offer the simple wholesale price contract, or to mistrust and to offer the screening contract. The results replicate the deficiencies identified in the experiments in Chap. 4 and show that there are two crucial behavioral factors that may enhance supply chain performance towards the screening benchmark. First, if the screening theory can be extended such that self-selection of the buyer can be ensured, then the suppliers are more willing to offer this complex contract type which, in turn, increases supply chain performance. On the other hand, if the supplier has a mean to punish uncooperative behavior, he is more willing to trust the buyer, which in turn also results in an increased supply chain performance. The experiments presented in Chap.5 considers several supply chain configurations (serial and diverging) and, therefore, allow for discussing the effect of contract type complexity in different industrial settings. The empirical test of the screening theory in the present work bridges the gap between purely game-theory based model approaches and empirical support/testing of the assumptions of those theories.

4 A Behavioral Screening Model

Taking into account the findings of the laboratory experiments, a behavioral screening model was analyzed to shed more light on the impact of communication in this lot sizing context (see also [4]). For this purpose, the assumption that all communication

is deceptive and therefore ignored is relaxed by giving the supplier the opportunity to adjust the statistical forecast regarding the private information via Baye's rule. It is shown, that the key input into the suppliers decision support system is not the a-priori probability with respect to the private information, but the frequency with which the respective contract is chosen. The derivation of the optimal contracts relies on Operations Research methods of non-linear optimization and shows analytically that information sharing does not necessarily improve the supply chain performance, although a certain fraction of buyers always reveal their private information truthfully. Since simple comparative static analysis regarding the information sharing and information processing parameters is difficult due to the non-linearity of the contract schemes, critical levels of the interacting parameters for which communication improves (deteriorates) the overall supply chain performance are identified.

5 Conclusion

Summing up, the research on supply chain coordination via contracts relies on Operations Research methods that derive optimal contract schemes under certain basic assumptions. The present work analyses a strategic lotsizing framework and contributes to the existing literature by deriving optimal contracts even for situations in which the objective function loses its convexity property. Moreover, an empirical test of the underlying theory was performed, and critical assumptions have been identified. Building upon these insights, Methods of non-linear optimization have been applied to a behavioral screening model, in which the critical assumption that all information is ignored is relaxed. The laboratory experiments prove to be highly effective in identifying critical assumptions which, in turn, are an inspiring source for researchers to further modify and improve existing and established theories while dealing with new challenges from an Operations Research perspective, since favorable properties of the underlying optimization problems might vanish if critical assumptions are relaxed.

References

1. Voigt, G.: Information Sharing and Contracting in a Just-in-Time Environment. Lecture Notes in Economics and Mathematical Systems 650. Springer, Berlin (2011)
2. Voigt, G., Inderfurth, K.: Supply chain coordination and setup cost reduction in case of asymmetric information. OR Spectrum **33**, 99–122 (2011)
3. Inderfurth, K., Sadrieh, A., Voigt, G.: The impact of information sharing on supply chain performance in case of asymmetric information. Prod. Oper. Manage. (2012)
4. Voigt, G., Inderfurth, K.: Supply chain coordination with information sharing in the presence of trust and trustworthiness. IIE Trans. **44**(8), 637–654 (2012)

Part II
Applied Probability and Stochastic Programming, Forecasting

Financial Crisis, VaR Forecasts and the Performance of Time Varying EVT-Copulas

Theo Berger

1 Introduction

Even though financial portfolio research focused on risk measurement and risk management during the past decade, the recent financial crisis made evident that there is still a lack of reliable indicators for financial risk. In this paper, we address the accuracy of Value-at-Risk (VaR) predictions based on hybrid DCC-Copulas applicable to higher order portfolios.

Consequently, we add to the existing literature a broad empirical comparison of DCC approach and copulas feasible to higher order portfolios in turmoil market times.

2 Methodology

Based on the GARCH filtered i.i.d. residuals, additionally to parametric distributions, the EVT approach[1], known as *peaks over threshold approach*, will be applied in the connection of invariant dependency. The data that exceeds a predefined threshold at level α, is modeled via GPD[2]. So to say, the extreme observations in the tails of the empirical distribution are modeled via EVT (see [6]) and the cumulative distribution function F_ξ is given by:

[1] Due to the explicit separation between dependency structure and the modeling of univariate marginal distributions, semi-parametric GP distribution is additionally applied to the return series.

[2] In this survey, observations below $\alpha = 10\ \%$ are modeled via GPD.

T. Berger (✉)
University of Bremen, Bremen, Germany
e-mail: theoberger@uni-bremen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_6,

$$F(x)_{\xi;\beta,v} = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{\frac{-1}{\xi}}$$

with $x \geq 0$, $\beta > 0$ and $\xi > -0,5$. In this context, x represent the exceedances, ξ the tail index (shape) parameter and β the scale parameter respectively.

DCC

The DCC model of Engle [2] belongs to the class of multivariate GARCH models. The approach separates variance modeling from correlation modeling. Let the N x 1 vector r_t be a set of N asset log returns at time t. Volatilities are calculated in order to construct volatility adjusted residuals ε_t. For our research, we assume that each return follows a univariate GARCH(1,1) process. The correlations are estimated based on the standardized residuals. Let R_t denote the correlation matrix and D_t the diagonal matrix with conditional standard deviations at time t.

The full DCC setup is given by:

$$y_t | \Im_{t-1} \sim N(0, D_t R_t D_t),$$

$$D_t^2 = diag\{H_t\}, \quad H_t = V_{t-1} y_t,$$

$$H_{i,i,t} = w_i + \alpha_i y_{i,t-1}^2 + \beta_i H_{i,i,t-1},$$

$$\varepsilon_t = D_t^{-1} y_t,$$

$$R_t = diag\left\{Q_t^{1/2}\right\} Q_t diag\left\{Q_t^{1/2}\right\},$$

$$Q_t = \Omega + \alpha \varepsilon_{t-1} \varepsilon_{t-1}' + \beta Q_{t-1},$$

where as $\Omega = (1 - \alpha - \beta)\bar{R}$ and α and β are always positive and their sum is less than one.

Copula

The copula approach is based on Sklar's Theorem (1959):

Let $X_1, \ldots X_n$ be random variables, $F_1, \ldots, F_n$ the corresponding marginal distributions and H the joint distribution, then there exists a copula C: $[0, 1]^n \rightarrow [0, 1]$ such that:

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots F_n(x_n)).$$

Conversely if C is a copula and $F_1, \ldots, F_n$ are distribution functions, then H (as defined above) is a joint distribution with margins $F_1, \ldots F_n$.

The Gaussian and t copula belong to the family of elliptical copulas and are derived from the multivariate normal and t distribution respectively.

The setup of the Gaussian copula is given by:

$$C^{Ga}(x_1, \ldots, x_n) = \Phi_\rho(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_n)),$$
$$= \int_{-\infty}^{\Phi^{-1}(x_1)} \ldots \int_{-\infty}^{\Phi^{-1}(x_n)} \frac{1}{2(\pi)^{\frac{n}{2}} |\rho|^{\frac{1}{2}}} exp\left(-\frac{1}{2} z^T \rho^{-1} z\right) dz_1 \ldots dz_n$$

where as Φ_ρ stands for the multivariate normal distribution with correlation matrix ρ and Φ^{-1} symbolizes the inverse of univariate normal distribution.

Along the lines of the Gaussian copula, the t copula is given by:

$$C^t(x_1, \ldots, x_n) = t_{\rho,v}(t_v^{-1}(x_1), \ldots, , t_v^{-1}(x_n)),$$
$$= \int_{-\infty}^{t^{-1}(x_1)} \ldots \int_{-\infty}^{t^{-1}(x_n)} \frac{\Gamma\left(\frac{v+n}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{\frac{n}{2}} |\rho|^{\frac{1}{2}}} \left(1 + \frac{1}{v} z^T \rho^{-1} z\right)^{-\frac{v+n}{2}} dz_1 \ldots dz_n,$$

in this setup $t_{\rho,v}$ stands for the multivariate t distribution with correlation matrix ρ and v degrees of freedom (d.o.f.). t_v^{-1} stands for the inverse of the univariate t distribution and v influences tail dependency. For $v \to \infty$ the t distribution approximates a Gaussian.

Due to the fact that estimating parameters for higher order copulas might be computationally cumbersome, all parameters are estimated in a two step maximum likelihood method given by Joe and Xu [3][3]. The two steps divide the log likelihood into one term incorporating all parameters concerning univariate margins and into one term involving the parameters of the chosen copula.

DCC-Copulas

Due to the fact, that the classical copula approach does not ac- count for time varying dependency, we combine the conditional correlation parame- ters modeled by DCC with the copula setup. For elliptical copulas, this operation is straightforward and leads to the conditional DCC-Copula setup (Firstly addressed by Patton [7], Jondeau and Rockinger [4]).

[3] This approach is also known as inference for the margins (IFM).

Value-at-Risk and Backtesting

VaR is defined as the quantile at level α of the distribution of portfolio returns:

$$VaR_\alpha = F^{-1}(\alpha) = \int_{-\infty}^{VaR_\alpha} f(r)dr = P(r \leq VaR_\alpha).$$

A rolling window approach is applied to forecast the one-day ahead VaR thresholds based on the given dependence. The rolling window size is at 500 observations for all data sets.

In order to evaluate the different forecasting techniques we apply regulatory Basel II criteria[4]. Additionaly to the absolute amount of misspecifications, unconditional coverage (UC), independence (IND) and conditional coverage (CC) are applied. The unconditional coverage test, proposed by Kupiec [5], adds to the Basel II backtesting framework, and checks if the expected failure rate α of a VaR model is statistically different from its realized failure rate:

$$LR_{UC} = -2ln[(1-p)^{T-N}p^N] + 2ln[(1-(N/T))T - N(N/T)^N],$$

where p stands for the percent left tail level, T for the total days and N for the number of misspecifications. The likelihood ratio statistic for testing $H_0 = \hat{\alpha} = \alpha$ is $LR_{UC} \sim \chi^2(1)$. Due to the fact that the UC method exclusively tests the equality between VaR violations and the chosen confidence level, Christoffersen [1] developed a likelihood ratio statistic to test whether the VaR misspecifications are correlated in time. The likelihood ratio for testing time dependence versus time independence (H_0) is defined as:

$$LR_{IND} = -2ln[(1-\pi)^{(T_{00}+T_{01})}\pi^{(T_{01}+T_{11})}] + 2ln[(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}].$$

T_{ij} stands for the number of observed values i followed by j. Whereas 1 represents a misspecification and 0 a correct estimation. π represents the probability of observing an exception and π_i the probability of observing an exception conditional on state i.

Thus, this approach rejects a model that either creates too many or too few clustered VaR violations with $LR_{IND} \sim \chi^2(1)$ The CC test combines both test statistics with H_0 covering the empirical failure rate and time independence, whereas the likelihood ratio statistic is given by:

$$LR_{CC} = LR_{UC} + LR_{IND},$$

with $LR_{CC} \sim \chi^2(2)$ respectively.

[4] For 99 % VaR forecasts, the Basel II document allows for 4 misspecifications within 250 days.

Table 1 Backtesting 99 % VaR (2005–2009). The sample covers 1,000 days so that the expected failure rate for 99 % VaR forecasts is 10 days. The critical values for the LR tests are 3.89 for LR_{UC}, LR_{IND} and 5.99 for LR_{CC}

PF	Distribution	DCC-Copula	Errors	LR_{UC}	LR_{IND}	LR_{CC}
PF I (99)	EVT	Gauss	13	0.83	0.00	0.83
PF I (99)	EVT	t	12	0.38	0.00	0.38
PF II (99)	EVT	Gauss	14	1.44	0.00	1.44
PF II (99)	EVT	t	12	0.38	0.00	0.38
PF III (99)	EVT	Gauss	14	3.08	0.00	3.08
PF III (99)	EVT	t	13	0.83	0.00	0.83
PF IV (99)	EVT	Gauss	12	0.38	2.29	2.67
PF IV (99)	EVT	t	10	0.00	2.97	2.97

3 Results

Table 1 illustrates the backtesting results for DCC-EVT-Garch-Copula for the four portfolios consisting of five indices (portfolio I and II), five stocks (portfolio III) and five FX-rates (portfolio IV) in turmoil market times.

Due to the focus on tail risk, we modeled the univariate tails via GPD distribution linked via copulas. This approach outperformed both normal and t distributions for 99 % VaR forecasts, since it leads to adequate backtesting results througout all market times for all asset classes.

4 Conclusion

By applying univariate EVT, we showed that the choice of univariate marginal distribution is the key driver for adequate VaR forecasts. However, from an empirical as well as theoretical point of view and due to the theoretical accuracy of dependency modeled by the DCC-Copula approach, we showed that the hybrid DCC-EVT-GARCH-Copula approach represents a powerful tool regarding 99 % VaR forecasts feasible to higher order portfolios.

References

1. Christoffersen, P.F.: Evaluating interval forecasts. Int. Econ. Rev. **39**(4), 841–862 (1998)
2. Engle, R.: Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models. J. Bus. Stat. **20**, 339–350 (2002)
3. Joe, H., Xu, J.J.: The estimation method of inference functions for margins for multivariate models. Vol. 166. Tech. Rep. (1996)
4. Jondeau, E., Rockinger, M.: The copula-GARCH model of conditional dependencies: An international stock market application. J. Int. Money Finan. **25**, 827–853 (2006)

5. Kupiec, Paul H.: Techniques for verifying the accuracy of risk measurement models. J. DERIVATIVES, **3**(2) Winter 1995. Available at SSRN: URL http://ssrn.com/abstract=7065
6. Longin, F., Solnik, B.: Extreme correlation of international equity markets. J. Financ. **56**, 649–676, (2001). doi:10.1111/0022-1082.00340
7. Patton, A.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**, 527–556 (2006)

Part III
Continuous Optimization

The Linear Sum-of-Ratios Optimization Problem: A PSO-Based Algorithm

João Paulo Costa and Maria João Alves

1 Introduction

The linear sum-of-ratios optimization problem models situations where several rates (objectives) have to be optimized simultaneously. The numerators and denominators of the ratios may represent profits, capital, time, human resources, production output or input, etc. Furthermore, applying the weighted sum technique to multiobjective linear fractional programming problems turns the problem into a sum-of-ratios optimization problem. The weighted sum is probably the widest approach used in multiobjective programming to aggregate several objective functions in one function to be optimized. Having some disadvantages (which we will not explore in this paper) this approach continues to be the first one people think of when faced with a decision problem with several conflicting objectives.

Schaible and Shi [1] consider the sum-of-ratios optimization problem "one of the most difficult fractional programs encountered so far". Several attempts to find good algorithms, techniques and approaches to deal with it have been made, however these attempts can only deal with problems of small dimensions. Schaible and Shi [1] present a survey of existing algorithms until 2002. They emphasize the algorithm of Kuno [2] as the best developed so far. After that we can point out the works of Benson [3] and Costa [4]. In our opinion, the latter seems to be the algorithm that can deal with problems of higher dimensions (until ten ratios and hundreds of variables and constraints).

In this text we present a hybrid algorithm using particle swarm optimization [5] (PSO) techniques hoping to deal with problems of higher dimension, especially in

J. P. Costa · M. J. Alves (✉)
Faculty of Economics University of Coimbra—INESC Coimbra, Dias da Silva, 165, 3004-512 Coimbra, Portugal
e-mail: jpaulo@fe.uc.pt

M. J. Alves
e-mail: mjalves@fe.uc.pt

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_7,

what concerns the number of ratios. The algorithm has some variations with respect to the PSO [5] in order to guarantee the feasibility of the generated solutions. The algorithm also makes the most of some theoretical results in which the Branch & Cut technique (B&C) of Costa [4] is based on. Namely it also uses a special cut (presented in [4]) that reduces the search space where the solutions (particles) are generated and moved.

We formulate the linear sum-of-ratios problem in the following way:

$$\begin{array}{ll} \max & s = \sum_{k=1}^{p} z_k(x) = \sum_{k=1}^{p} \frac{c^k x + \alpha_k}{d^k x + \beta_k} \\ s.t. & x \in X = \{x \in R^n | Ax \leq b, x \geq 0\} \end{array} \tag{1}$$

where c^k, $d^k \in R^n$, $A \in R^{m \times n}$, $b \in R^m$ and α_k, $\beta_k \in R$, $k = 1, \ldots, p$ and $d^k x + \beta_k > 0$, $\forall_k$, $\forall x \in X$. We assume that X is non-empty and bounded.

Let $Z = \left\{ \left(z_1, \ldots, z_p \right) \in R^p : z_k = z_k(x), k = 1, \ldots, p, x \in X \right\}$.

The basic idea of the B&C technique [4] is to build a search region by computing the solutions that individually optimize each ratio. Note that the optimization of just one ratio turns out to be a linear programming problem if the variable transformation of Charnes and Cooper [6] is applied. The $z_k, k = 1, \ldots, p$, of these solutions organized in a table (pay-off table) give the upper and lower bounds of a search region to be explored. Then, the search region is divided (approximately by the 'middle') in two sub-regions which are analyzed. Their pay-off tables are computed and one is discarded if it can be proved that the maximum of s is in the other. The process is repeated with the remaining region(s). If it is not possible to discard one of the regions, the process must be repeated for both, building a search tree. The process ends when the remaining regions are so small that the differences among the maximum and minimum values of each ratio in each region are lower than a pre-defined error.

One region can be discarded when the value of s for the upper bounds of its pay-off table is worse than the value of s for a solution belonging to another region not yet discarded. In order to speed up the computations, an incumbent solution is maintained: the solution that has the best s computed so far. This incumbent solution is used to make the necessary comparisons with the upper bounds of the regions that can potentially be discarded. Meanwhile this condition to discard a region can be used in order to define some cuts and to trim the current search region. These cuts are linear constraints resulting from applying the discarding condition.

The algorithm has a good performance for problems of medium dimensions (less than roughly ten ratios) even considering a very small pre-defined error [4]. The integration of PSO techniques improves performance for problems of higher dimensions.

Wu, Sheu and Birbil [7] propose to solve the sum-of-ratios problem using a stochastic searching algorithm, specifically the Electromagnetism-like Mechanism (EM). Being a different stochastic technique it would be interesting to compare their results with the ones presented in this text. However, the EM technique [7] is applied

to a more general case, where the numerators and the denominators of the ratios can be non-linear. In [7] there are some reported tests with the linear sum-of-ratios case, but for the dimensions of the tested problems the B&C technique of Costa [4], working alone, has a fairly good performance and so there is no point in comparing the results. However, it must be pointed out that the stochastic search proposed in [7] operates in Z-space rather than in X-space, as the one presented in this text.

In our proposal, PSO is applied after running the B&C technique considering a big pre-defined error, for which a reasonable computing effort is enough. The swarm is then built upon the computed solutions of the remaining regions of the B&C technique. These solutions are fairly well distributed on the search region and would need further searching with the B&C technique in order to decrease the error. As the swarm (the set of potential solutions) is modified according to stochastic procedure outlined in Sect. 2, the new incumbent solutions that are generated are used to redefine cuts to further trim the search region. This strategy allowed to find better solutions with a reasonable computing effort for problems of higher dimensions than the previous ones from [4]. Computational tests were carried on, and are presented in Sect. 3, for problems with up to 25 ratios.

2 Algorithm Outline

The presentation of the outline of the algorithm will emphasize the adaptations of the general PSO technique [5] to the present case.

Framing the search region – The search region, Z, is framed by lower and upper bounds. The lower bounds are obtained through the cuts defined by [4], via the incumbent solution resulting after running the B&C technique with an *error* compatible with the available computing time. The upper bounds result from the optimization of each ratio individually over X and considering the lower bounds. The search region is reframed whenever a new incumbent solution is found.

Initializing the swarm – The set of solutions, or particles moving in the search region, is defined in Z. It is initialized with the solutions resulting from the B&C technique. The obtained incumbent solution is the first to be selected. Other solutions belonging to regions that would need further search with B&C, for decreasing the *error*, can also be selected, provided that their s-values are different from the s-values of the already selected solutions, by at least 10 % of the *error*. The selection process will stop when the predefined number of particles is achieved (population) or when there are no more already computed solutions available. If needed, a random process of generating particles is applied in order to complete the swarm.

Completing the swarm – A particle is randomly selected from the incomplete swarm to give raise to another particle. The differences between the $z_k \in R,\ k = 1, \ldots, p$, of this particle and the lower and upper bounds of the search region are computed. With a randomly generated number in $[-1, 1]$ together with these differences, new values of $z_k \in R,\ k = 1, \ldots, p$, are computed. A randomly selected ratio is then minimized over the framed search region constrained by these new values of

z. This implies to solve a linear programming (LP) problem. Sometimes this problem is infeasible. When this happens the process is repeated until a feasible solution is achieved, which will be the new particle.

Flying—shifting the particles – The movement of the particles is obtained through a shift vector. Each particle has its own shift vector. The shift of the last iteration is maintained (in the beginning it is the null vector). The best position that any particle ever attained is maintained (*global best*). The shift vector is obtained by adding two vectors to the shift vector of the last iteration multiplied by a factor of inertia (0.7 in our case). The first vector is obtained by subtracting the current position of the particle from the best position that the particle ever achieved (*personal best*) and multiplying the result by a uniformly random generated number between 0 and 1.4. The second vector is obtained by subtracting the current position of the particle from the *global best* (incumbent solution) and multiplying the result by a uniformly random generated number between 0 and 1.4. A randomly selected ratio is then maximized over the framed search region including these new values of z as constraints. This implies to solve an LP problem. If this problem is infeasible, the process of completing the swarm is used for this particle (the particle is reset). The used coefficients, 0.7, 1.4 and 1.4, for computing the shift vector are numbers based on the PSO literature and tuned for this case.

3 Computational Results

The algorithm was implemented with Delphi Pascal 2007 for Microsoft Windows. The code for solving the linear programming problems needed by the algorithm was the *lp_solve 5.5*. The tested problems were randomly generated: $c_j^k, d_j^k \in [0.0, 0.5]$ and $a_{ij} \in [0.0, 1.0]$, $\alpha_k, \beta_k \in [2.0, 100.0]$ are uniformly distributed numbers and b equals to 1000. All the problems have 50 constraints (m) and 100 variables (n). To simplify the coding for these tests, an absolute error was considered rather than a relative one.

From the set of generated problems we selected 5 problems, one for each number of ratios (5, 10, 15, 20 and 25). We selected the problems for which the B&C technique of [3] proved to have a lower performance. It must be remarked that the B&C does not perform well for problems with more than 10 ratios, considering relative *errors* lower than 0.1. To measure the performance of the algorithm we used the number of linear programming (LP) problems that were solved when applying the algorithms. Knowing that PSO is stochastic, we applied the algorithm 10 times for each problem and each set of parameters.

Applying the algorithm to the 5-ratios problem with an absolute *error* of 0.01 in each ratio, we obtained an improvement with PSO of about 0.03 % in all runs, considering 10 or 20 particles, 100 or 150 iterations. We obtained the same maximum by running the B&C technique alone with an absolute *error* of 0.0001. The 5-ratios problem was used to tuning the parameters of PSO and to test search options.

Table 1 Results for 20 ratios.

	100 iterations - 40 particles			100 iterations - 80 particles		
	No. of LPs	Infeasible	Improvement (%)	No. of LPs	Infeasible.	Improvement (%)
Max	10 142	5 677	3.600	20 177	12 186	3.590
Avg	9 493	5 163	3.591	19 495	11 657	3.590
Min	9 062	4 857	3.589	18 696	11 090	3.590
	150 iterations - 40 particles			*150 iterations - 80 particles*		
Max	14 394	7 769	3.591	29 179	17 137	3.593
Avg	13 463	7 128	3.590	27 571	15 727	3.591
Min	12 922	6 769	3.590	26 033	14 465	3.590

Applying the algorithm to the 20-ratios problem with an absolute *error* of 2 the B&C technique needed 440 420 LPs. Table 1 presents the obtained results after the PSO technique. The results are presented for several combinations of the number of iterations and the number of particles. The maximum (Max), average (Avg), and minimum (Min) of each measure (over 10 runs) are presented. The number of total LP problems (No. LPs), infeasible LP problems (Infeasible) and the Improvement on s, i.e. $\frac{s^{new}-s^{original}}{s^{original}}$, are presented. The 20-ratios problem was used to further experiment several combinations of the number of iterations and the number of particles in the swarm. We concluded that 100 iterations and $2p$ for the number of particle are enough to obtain good results.

Applying the algorithm to the problems with 10, 15 and 25 ratios, with large *errors* (absolute values of 1, 1 and 4, respectively), the B&C technique needed 40 890, 42 972 and 11 425 LPs, respectively. Table 2 presents the obtained results after the PSO technique.

Table 2 Results for 10, 15, 20 and 25 ratios.

	10 ratios - 100 iter. - 20 particles			15 ratios - 100 iter. - 30 particles		
	No. of LPs	Infeasible	Improvement (%)	No. of LPs	Infeasible	Improvement (%)
Max	5 838	3 825	4.079	6 791	3 386	3.770
Avg	7 488	5 524	4.075	6 228	2 930	3.765
Min	5 130	3 265	4.072	5 722	2 620	3.764
	20 ratios - 100 iter. - 40 particles			*25 ratios - 100 iter. - 50 particles*		
Max	10 142	5 677	3.600	13 243	7 579	10.17
Avg	9 493	5 163	3.591	12 308	6 811	10.16
Min	9 062	4 857	3.589	11 509	6 307	10.16

4 Final Considerations

The results show that using a combination of the PSO and B&C techniques, it is possible to achieve improvements in the computed solutions maintaining the needed computational effort in reasonable levels. Nevertheless, the sum-of-ratios optimization problem turns out to be a difficult case for the PSO because the problems constraints imply a lot of infeasible LPs.

Acknowledgments This work has been partially supported by FCT under project grant PEst-C/EEI/UI0308/2011 and by QREN, project EMSURE. We would also like to recognize the support of the *"40th Anniversary of Faculty of Economics of University of Coimbra"*.

References

1. Schaible, S., Shi, J.: Fractional programming: The sum-of-ratios case. Optim. Methods Softw. **18**, 219–229 (2003)
2. Kuno, T.: A branch-and-bound algorithm for maximizing the sum of several linear ratios. J. Glob. Optim. **22**, 155–174 (2002)
3. Benson, H.: A simplicial branch and bound duality-bounds algorithm for the linear sum of ratios problem. J. Optim. Theor. Appl. **112**, 1–29 (2007)
4. Costa, J.: A Branch & Cut technique to solve a weighted-sum of linear ratios. Pac. J. Optim. **6**(1), 21–38 (2010)
5. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of the IEEE international joint conference on, neural networks, pp. 1942–1948 (1995)
6. Charnes, A., Cooper, W.: Programming with linear fractional functions. Naval Res. Logistics Q. **9**, 181–186 (1962)
7. Wu, W.-Y., Sheu, R.-L., Birbil, S.: Solving the sum-of-ratios problem by a stochastic search algorithm. J. Glob. Optim. **42**, 91–109 (2008)

Part IV
Decision Analysis and Multiple Criteria Decision Making

Selection of Socially Responsible Portfolios Using Hedonic Prices

A. Bilbao-Terol, M. Arenas-Parra, V. Canal-Fernandez and C. Bilbao-Terol

1 Introduction

Various initiatives have been launched to assess the SEE dimensions in investment decisions [1–4]. The majority of the measures of responsible and irresponsible corporate behaviour presented in the literature are constructed from the databases of independent agencies (Vigeo, EIRIS, KLD ...) that to a large extent rely on publicly observable events (e.g. newspaper articles, Non-Governmental Organization reports, regulatory reports or company rankings) together with studies based on questionnaires to and interviews with stakeholder groups. The aim of this research is to provide a framework for constructing portfolios containing conventional and SRI assets, based on the application of the HPM [5] for the monetary valuation of socially responsible characteristics of financial assets [6]. We use multi-objective programming as a suitable mathematical technique for solving the portfolio selection problem, including several criteria in decision-making processes [7–10] in which the investment opportunities are described in terms of a set of attributes, with part of this set intended to capture and express the effects on society [4, 8, 11].

A. Bilbao-Terol · M. Arenas-Parra
Department of Quantitative Economics, University of Oviedo, Oviedo, Spain
e-mail: ameliab@uniovi.es

M. Arenas-Parra
e-mail: mariamar@uniovi.es

V. Canal-Fernandez (✉) · C. Bilbao-Terol
Department of Economics, University of Oviedo, Oviedo, Spain
e-mail: vcanal@uniovi.es

C. Bilbao-Terol
e-mail: cbilbao@uniovi.es

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_8,

2 Optimality of SRI Portfolios: A Multi-objective Programming Model

2.1 An Evaluation of Social Criteria for Portfolio Selection: Application of Hedonic Price Methodology

In this paper, a measurement strategy that evaluates the social responsibility of a portfolio based on the HPM [5] is constructed. The HPM breaks away from the traditional view that goods are the direct objects of utility; on the contrary it assumes that utility is derived from the properties or characteristics of the goods themselves. This method allows us to relate the price of SRI funds with their SEE characteristics and to obtain a monetary valuation for these features according to market. This is the so-called hedonic price of the characteristic. The theoretical foundations of the hedonic pricing model are based on what is known as the New Approach to Consumer Theory [13]. Our application of the hedonic pricing method to the financial market of SRI funds assumes that these funds are different due to their financial and social responsibility characteristics; in addition, the market values them on the basis of such characteristics. These prices are used to define the criteria of social responsibility for a portfolio of mutual funds as exposed below.

Let x_i denote the units to be allocated to the mutual fund i. If we consider a market of n mutual funds, a portfolio is represented by the n-dimensional vector $x = (x_1, \ldots, x_n)$. For the portfolio x, its exposure to a certain attribute f can be calculated as a weighted average of the attribute exposures of the individual mutual funds contained in this portfolio [11]. In our research this is done in the following way:

$$SR_f(x) = \sum_{i=1}^{n} pt_{if} P_{iT} x_i \tag{1}$$

where pt_{if} denotes the score of mutual fund i on the attribute f and P_{iT} denotes the price at the investment date T of the generic i-th mutual fund. The uni-dimensional functions (1) are combined in a linear aggregation using the hedonic price of each attribute f:

$$SR(x) = \sum_{f=1}^{F} h_f^* SR_f(x) \tag{2}$$

where h_f^* denotes the normalized hedonic price of attribute f. The expression (2) determines the objective of social responsibility that will be maximized in the portfolio selection.

2.2 An Evaluation of Financial Criteria for Portfolio Selection

The financial objective employed in this work is the final absolute wealth i.e. the value of the portfolio at the investment horizon. In this work, the certainty-equivalent is used as a preference index and the variance as risk measure, other risk measures as the CVaR (Conditional Value at Risk) [14] can be used. With regard to the constraints, the usual ones have been considered; the budget constraint and short sales are not allowed.

$$\left.\begin{array}{ll} max & EVE(x) = \sum_{i=1}^{n} E[P_i]x_i \\ min & RM(x) \\ max & SR(x) = \sum_{f=1}^{F} h_f^* SR_f(x) \\ s.t. & \\ & x \in X \end{array}\right\} \text{(PS-SRI)} \quad (3)$$

where $EVE(x)$ is the Expected Value of the portfolio x at the end of the investment horizon, $E[P_i]$ is the expected price of asset i and $RM(x)$ denotes a Risk Measure of the portfolio x.

3 A Procedure for SRI Portfolio Selection

A two-stage procedure based on a financial reference point and the maximization of SR criterion is built.

- Stage 1: *The Best Financial Performance Portfolio.* This stage is carried out in two steps: Step 1: An approximation of the $EVE - RM$ efficient frontier is obtained by applying ϵ-constraint method to the corresponding bi-objective problem.
  Step 2: The certainty-equivalent of final wealth is used to obtain the portfolio of maximum financial satisfaction [15]. The expected utility of each portfolio is calculated (Monte Carlo simulations) using sample prices, and on the basis of this information, the certainty-equivalent of each portfolio is estimated. The maximum certainty-equivalent provides the portfolio of maximum financial satisfaction for which EVE and RM values are denoted by EVE^* and RM^*, respectively.
- Stage 2: *The Best Socially Responsible Portfolio.* In this stage, a multi-objective problem is solved by the ϵ-constraint method. The objective is the maximization of the SR and the bounds of the EVE and RM values are closer to those of the maximum financial satisfaction portfolio that have been found in Step 2 of Stage 1, such bounds being denoted by EVE^*_{-1} and RM^*_{+1} . This problem is solved and then, we test whether the solution of problem is efficient with regards to the financial satisfaction and social responsibility criterion; otherwise, a better portfolio on the financial efficient frontier is selected.

4 Application to the Case of the Spanish Market

This study uses a data set consisting of 142 conventional funds and 18 SRI funds domiciled in Spain. Therefore, $n = 160$. We have daily prices from March 10, 2006 to December 31, 2009; thus, 995 observations for each mutual fund are available. Setting an estimation horizon of one week (i.e. 5 trading days), the investment time is $T = 196$ (December 31, 2009). The investment horizon has been set equal to one month (four weeks). As market invariants, we have chosen the non-overlapping weekly compounded. After describing our market, we will obtain the SRI portfolio. The SEE characteristics are included in the investment policy of SRI funds published by the Spanish National Securities Market (CNMV). There is a long list of SEE characteristics for SRI products. The principal components factor analysis technique is used by Bilbao and Canal [6] in order to reduce the number of these characteristics and correct the collinearity which exists between them prior to estimating the hedonic function. A factor is calculated as the un-weighted average value of the characteristics involved. The four resulting factors (*Product Responsibility Area*, *Labor Rights Area*, *Environmental Area* and *Gender Equality and Community Relations Area*) are used as the inputs for the multi-objective portfolio model. The prices obtained from estimating the hedonic regression for the four areas of social responsibility, expressed in millions of Euros are:

$$\left.\begin{array}{l} h_1 = 19,5393 \; h_2 = 19,5393 \\ h_3 = 22,3017 \; h_4 = -14,3475 \end{array}\right\} \text{(Hedonic Prices)} \tag{4}$$

To construct an objective that measures the Socially Responsible Quality of the portfolio according to a specific market, we use only the positive hedonic prices that consider a set of SEE characteristics for which the market is willing to pay.

4.1 Markowitz's Approach (EVE-Variance)

In order to determine the reference portfolio (with Initial wealth = 100), i.e. the maximum financial satisfaction portfolio, we have approximated the EVE-Variance financial efficient frontier and the certainty-equivalent has been used with an exponential utility function: $U = -exp(-(\frac{f(x)}{5}))$ where $f(x)$ is the final wealth and 0.2 is the Arrow-Pratt risk aversion (other types of utility functions can be used). The maximum certainty–equivalent on efficient portfolios determines the reference portfolio whose composition and characteristics are included in following tables (Tables 1, 2).

Once the portfolio of maximum satisfaction on the efficient frontier is obtained, we calculate the SRI optimal portfolio by solving the following uni–objective problem:

Table 1 Composition of the reference portfolio

Creacion de Cultura en Espanol FI	11.43
Banif global 3-98 FI	34.76
AC inversion selectiva FI	2.84
Metropolis renta FI	32.69
Prismafondo FI	10.00
Banif 2011 FI	8.28

Table 2 Characteristics of the reference portfolio

Variance = 0.526	EVE = 100.382
SR = 6.201	Satisfaction = 100.293

Table 3 Composition of the SRI–optimal portfolio

Creacion de Cultura en Espanol FI	11.40
Santander responsab. conserv. FI	44.95
Banif global 3-98 FI	32.45
AC inversion selectiva FI	4.13
Banif 2011 FI	7.06

Table 4 Characteristics of the SRI–optimal portfolio

Variance = 0.7520	EVE = 100.368
SR = 133.008	Satisfaction = 100.244

$$\left.\begin{array}{ll} max & SR(x) = \sum_{f=1}^{3} h_f^* SR_f(x) \\ s.t. & \\ & EVE(x) \geq 100.3582 \\ & Variance(x) \leq 0.752 \\ & x \in X \end{array}\right\} \tag{5}$$

in which the bounds of the EVE and the variance correspond to the "closest" point to that of the maximum financial satisfaction portfolio. The solution to the problem (5) is shown in the following tables (Tables 3, 4).

5 Conclusion

This paper presents an approach for portfolio selection based on the market valuation of the social responsibility of financial assets and multi–objective programming tools. The particular characteristics of our framework are: A multi-dimensional description of the investment opportunities together with the reduction data in order to identify the dimensions of SRI is used. A new measure of SRI which is not restricted to a particular investor, group of experts or stakeholders is proposed. The procedure is applicable to a broad public and should not require detailed a priori preference information from the investor. The method could be extended to any market because

it is based on public information. It provides information about the losses suffered by those investing in funds with SRI criteria and this supports the choice between conventional versus SRI investments. The model has been applied to a sample of 160 SRI and conventional funds domiciled and managed in Spain. Empirical results show that the financial sacrifice for investing in socially responsible funds is relatively small for "cautious" investors. These results could be good news for socially responsible investors.

Acknowledgments The authors wish to gratefully acknowledge the financial support from the Spanish Ministry of Education, Project ECO2011-26499.

References

1. Basso, A., Funari, S.: Measuring the performance of ethical mutual funds: a DEA approach. J. Oper. Res. Soc. **54**, 521–531 (2003)
2. Barrachini, C.: An ethical investments evaluation for portfolio selection. Electron. J. Bus., Ethics Org. Stud., 9(1), (2004).
3. Barnett, M., Salomon, R.: Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance. Strateg. Manage. J. **27**, 1101–1122 (2006)
4. Kempf, A., Osthoff, P.: SRI Funds: Nomen est Omen. J. Bus. Finance Account. **35**(9), 1276–1294 (2008)
5. Rosen, S.: Hedonic prices and implicit markets: Product differentiation in pure competition. J. Polit. Econ. **82**(1), 34–55 (1974)
6. Bilbao, C., Canal, V.: The market valuation of ethical assets in Spain: An application of the hedonic pricing method. Bus. Ethics: A, Eur. Rev. (2012). (in revision).
7. Zeleny, M.: Linear Multiobjective Programming. Springer, Berlin (1974)
8. Ignizio, J.P.: Linear Programming in Single and Multiple Objective Systems. Prentice-Hall, Englewood Cliffs, N.J. (1982)
9. Ballestero, E., Romero, C.: Portfolio selection: A compromise programming solution. J. Oper. Res. Soc. **47**(11), 1377–1386 (1996)
10. Arenas, M., Bilbao, A., Rodriguez, M.V.: A fuzzy goal programming approach to portfolio selection. Eur. J. Oper. Res. **133**(2), 287–297 (2001)
11. Hallerbach, W., Ning, H., Soppe, A., Spronk, J.: A framework for managing a portfolio of socially responsible investments. Eur. J. Oper. Res. **153**(2), 517–529 (2004)
12. Bilbao, A., Arenas, M., Canal, V.: Selection of socially responsible portfolios using goal programming and fuzzy technology. Inf. Sci. **189**, 110–125 (2012)
13. Lancaster, K.J.: A new approach to consumer theory. J. Polit. Econ. **7**(2), 132–157 (1966)
14. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
15. Meucci, A.: Risk and Asset Allocation. Springer Quantitative Finance, Berlin (2007)

Evaluation of Energy Efficiency Measures in Compressed Air Systems: A PROMETHEE Approach for Groups Facing Uncertainty

Simon Hirzel and Grit Walther

1 Introduction

The improvement of energy efficiency is a key pillar for realizing a more sustainable global energy system [9]. With an approximate share of 10 % in overall industrial electricity demand in Europe, compressed air systems are important contributors to energy demand. While numerous measures to improve their energy-efficiency are available (e.g. [13, 17]), the determination of the most suitable energy-efficiency measures is often challenging. A thorough evaluation requires the consideration of diverse aspects such as the cost-effectiveness of the measures, their technical risks, the complexity of the measures or their trialability. Especially the latter aspects are usually difficult to quantify and subject to imprecision, uncertainty and personal judgement. Furthermore, several decision-makers in a company are often affected by energy efficiency measures. This may lead to conflicts of interest and split-incentive problems, if the responsibility for implementing the energy-efficiency measures and the benefits from the resulting energy-savings are attributed to different parties (e.g. [13, 15]).

Multi-criteria methods are valuable tools to facilitate and structure decision making processes. Among the numerous available methods, the PROMETHEE family is considered as simple and well-suited for practical applications (see e.g. [1]). However, little attention has been paid to using these methods in conjunction with uncertain information in a multiple decision-maker environment, as when dealing with energy-efficiency measures. The aim of this paper is therefore to suggest a

S. Hirzel (✉)
Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Strasse 48, 76139 Karlsruhe, Germany
e-mail: simon.hirzel@isi.fraunhofer.de

G. Walther RWTH Aachen, Chair of Operations Management, Kackertstr. 7, 52072 Aachen, Germany
e-mail: walther@om.rwth-aachen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_9,

fuzzy version of PROMETHEE that explicitly takes multiple decision-makers and evaluations under uncertainty into account. An application to energy efficiency measures in compressed air systems illustrates the suggested method.

2 Extensions of the PROMETHEE Method Towards Uncertainty and Groups

When dealing with uncertainties in a decision problem, two main sources for uncertainty can be distinguished. A decision-maker will usually not have a clear perception of a decision-problem from the very beginning. His uncertainty concerning the selection of criteria, weights and preference functions brings uncertainty to the analysis. This type of uncertainty is mainly dealt with by the preference functions and by sensitivity analyses on criteria weights in the PROMETHEE family. The second source of uncertainty stems from uncertainty in the decision table, e.g. due to unquantifiable, incomplete, non-obtainable information or partial ignorance of the relevant data [4].

Numerous approaches deal therefore with a more explicit incorporation of uncertainty in the PROMETHEE method. Early attempts argue for example that results of PROMETHEE I and II are sometimes very close to each other due to the generalized criteria. Therefore they suggest calculating result intervals instead of crisp evaluations (see [3] on PROMETHEE III). Others introduce fuzzy decision-rules instead of preference functions [14]. According to [2], however, the first integration of fuzzy set theory into the PROMETHEE method is suggested by [16], who use interval-based input data. Subsequent works introduce triangular [6] or trapezoidal fuzzy numbers [5] into the algorithm. Other again deal with incorporating linguistic information in the PROMETHEE algorithms [7, 11] or suggest stochastic extensions [8], thus offering a broad range of different approaches.

While there are numerous publications dealing with uncertainty in PROMETHEE, explicitly dealing with groups has received less attention (see e.g. [18]). Even though decision-makers in a company may differ in their individual perceptions of a problem, they usually follow an overall goal as a team. In such a cooperative group, the group has to develop a common perception of the considered decision-making problem, based on the individual evaluations of its group members.

Depending on how the different group members articulate and aggregate their views, several types of aggregation models can be distinguished [12]: An often applied approach in group situations is the aggregation of individual views either prior or at an early stage of the MCDA algorithm (e.g. [18]). While these approaches are simple and require only comparatively little computational effort, they require the group members to agree on common input data. Thus, the individual group members lack the possibility to explore their individual analysis of the problem. This disadvantage is partially overcome when the aggregation takes place later and is done at the level of preferences per criterion. Then, the decision-makers may specify individual input information, but the results are only provided for the entire group. A full

appreciation of the individual problem is only possible by the third aggregation approach which is based on the individual results. Though this aggregation provides mathematically the same results as the former approach (see e.g. [10]), it complements the analysis by explicitly providing the individual evaluation results.

Algorithmic implementations for PROMETHEE allowing either aggregating preferences per criterion or individual results are suggested in [10]. For reasons of brevity, we only deal with the aggregation of individual results here. Given a decision problem with a alternatives and e decision-makers, with each decision-maker k possessing a weight $v_k > 0$, corresponding to his importance in the group, and where $\sum_{k=1}^{e} v_k = 1$. Using the PROMETHEE net flows $\Phi^{k,net}(A_i)$ for each decision-maker k and alternative A_i, the overall net flow of the group $\Phi^{G,net}(A_i)$ is calculated by:

$$\Phi^{G,net}(A_i) = \sum_{k=1}^{e} v_k \cdot \Phi^{k,net}(A_i) \quad \forall \quad i \in \{1, \ldots, a\}$$

Using this aggregation allows generating a PROMETHEE II ranking and, if applied to positive and negative flows separately, a PROMETHEE I ranking, as well.

3 A Fuzzy Group-Decision Approach

In the following, we suggest a PROMETHEE approach for groups facing uncertainty in the input data aggregating individual results. We assume that the group has sufficiently structured the decision problem, that a common list of a alternatives and c criteria has been developed, that each decision-maker k has provided a decision matrix with data in fuzzy format and a normalized list of crisp criteria weights $w_k^j > 0$, describing the relative importance of each criterion j with $\sum_{j=1}^{c} w_k^j = 1$. Furthermore, we assume that a normalized set of crisp weights $v_k > 0$, describing the importance of each decision-maker in the group, is available. Based on the fuzzy PROMETHEE approach for single decision-makers with trapezoidal fuzzy numbers as suggested in [5], combined with a set of normalized, crisp criteria weights as suggested in [6], we calculate the fuzzy net outranking flow $\tilde{\Phi}^k(A_i)$, obtaining:

$$\tilde{\Phi}^k(A_i) = (m_{l,i}^k, m_{u,i}^k, \alpha_i^k, \beta_i^k)_{LR} \quad \forall \quad i \in \{1, \ldots, a\}, k \in \{1, \ldots, e\}$$

For aggregating the e fuzzy net flows, we multiply each individual flow by the crisp weight v_k of each decision-maker and add up the results:

$$\tilde{\Phi}^G(A_i) = v_1 \cdot \tilde{\Phi}^1(A_i) \oplus v_2 \cdot \tilde{\Phi}^2(A_i) \oplus \ldots \oplus v_e \cdot \tilde{\Phi}^e(A_i) \quad \forall \quad i \in \{1, \ldots, a\}$$

This aggregation method allows both including data that is known with certainty as fuzzy singletons and data that is subject to uncertainty as intervals, triangular or

trapezoidal fuzzy numbers in the evaluation table. Perceived certainties and uncertainties from the different decision-makers compensate each other for each alternatives, based on the weights of the decision-makers v_k.

4 Illustration

For illustration, we assume that there are two decision-makers in a group, one representing the production department and one the facility management of an industrial company. They analyse three energy-efficiency measures for an existing compressed air system as alternatives, using individual decision matrices and a common criteria list with Gaussian preference functions and parameter σ_k^j (Table 1).

Applying the above proposed algorithm provides the results for $\tilde{\Phi}^{net}$ as illustrated in Fig. 1. The individual evaluation of the decision-maker 'production' indicates that he prefers the energy-efficiency measure 'compressor improvement' (CI) and the 'pressure reduction' (GPR) to the 'leakage reduction' (LR). However, it is not possible to clearly distinguish CI and GPR. For the 'facility management', however, CI is clearly the least attractive alternative, while alternatives GPR and LR are indifferent. Bringing together both individual evaluations leads to a group evaluation where alternative GPR is clearly better than alternatives LR and CI. However, no clear distinction can be made between LR and CI, though LR could be more preferable. So it is either possible to assume that the distinction between CR and LR is clear enough, or further information must be sought. Note further that alternative LR, which is the best alternative with regard to an economic evaluation, is only the second-best solution considering all criteria.

Table 1 Data of the decision-problem

	Investment	Cost reduction	Trialability
Criteria direction	Minimize	Maximize	Maximize
	Production ($v_1 = 0.6$)		
Criteria weights w_1^j	0.4	0.4	0.2
Preference parameter σ_1^j	500	5,000	0.6
Compressor improvement	0	[3,000; 5,000]	(0.9; 1.0; 0.2; 0.0)
Leakage reduction	1,000	[6,000; 8,000]	(0.5; 0.5; 0.2; 0.2)
Pressure reduction	0	[1,500; 2,000]	(0.9; 1.0; 0.2; 0.0)
	Facility management ($v_2 = 0.4$)		
Criteria weights w_2^j	0.5	0	0.5
Preference parameter σ_2^j	1,000	-	0.8
Compressor improvement	[2,500; 3,500]	0	(0.0; 0.1; 0.0; 0.2)
Leakage reduction	[0; 0]	0	(0.9; 1.0; 0.2; 0)
Pressure reduction	[0; 0]	0	(0.9; 1.0; 0.2; 0)

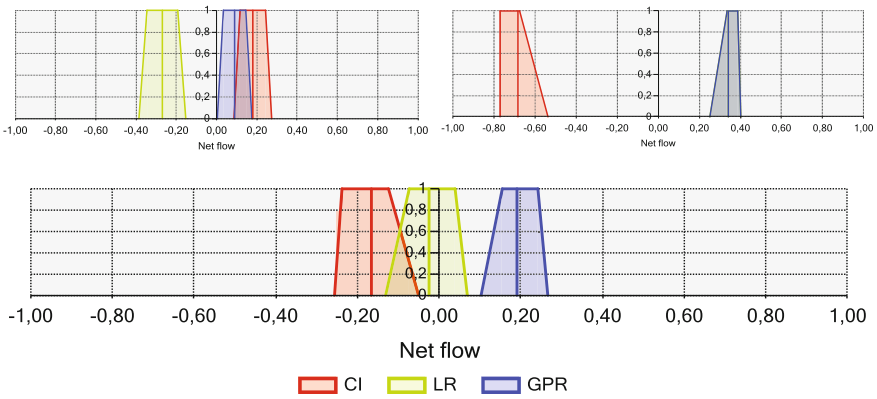


Fig. 1 Fuzzy outranking net flows $\tilde{\Phi}^{net}(A_i)$ for the individual decision-makers 'Production' (*upper left panel*) and 'Facility management' (*upper right panel*) and for the group (*lower panel*)

5 Discussion and Conclusions

As compared to implicit or input-based crisp PROMETHEE methods used for groups, the suggested method allows carrying out a full PROMETHEE analysis for each decision-maker. It can thus help the decision-makers to improve their individual understanding of the problem. By permitting to express uncertainties in the input data, it can furthermore reduce the hesitation a decision-maker might feel if he would have to provide precise values otherwise. This is especially relevant if the decision-maker deals with qualitative criteria or criteria that are difficult or expensive to evaluate or quantify. Even if the evaluations are not crisp, the method can help to identify alternatives that are clearly superior to others as illustrated above, and it allows different types of input formats. And finally, the need for sensitivity analyses is limited to either the weights of criteria or the weights of the decision-makers.

Nevertheless, there are also some drawbacks to the method: A fuzzy approach is per se more complicated, as the amount of processed information is larger. Furthermore, if input data is very fuzzy, a fuzzy approach might provide results that are too fuzzy for any reasonable decision-support. This might be dissatisfying for a decision-maker, even though the results reflect his input. Applying a defuzzifier might help overcome such problems, but defuzzified results should be analysed together with the underlying fuzzy results. Furthermore, there is little benefit in using a fuzzy approach if symmetric input data, preference functions that keep the symmetry and a symmetric defuzzifier is used (see also [6]).

The analysis of energy-efficiency measures is often a complex task, subject to uncertainty and requiring the consideration of multiple decision-makers. The suggested fuzzy group approach may help decision-makers to gain a better understanding of the various energy-efficiency measures as found in compressed air systems, but not limited to them.

Acknowledgments This work was partially funded by the German Federal Ministry of Economics and Technology under grant no. 0327484C as part of the integrated project EnEffAH.

References

1. Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdasi, M.: PROMETHEE: A comprehensive literature review on methodologies and applications. EJOR **200**, 198–215 (2010)
2. Bilsel, R.U., Gülçin, B., Ruan, D.: A Fuzzy Preference-Ranking Model for a Quality Evaluation of Hospital Web Sites. IJIS **21**, 1181–1197 (2006)
3. Brans, J.P., Mareschal, B., Vincke, P.: PROMETHEE: A new family of outranking methods in multicriteria analysis. In: Brans. J.P. (ed.) Operational Research'84, pp. 477–490. Elsevier, North-Holland (1984).
4. Chen, S.-J., Hwang, C.-L.: Fuzzy Multiple Attribute Decision Making. Methods and Applications. Springer, Berlin (1992)
5. Geldermann, J., Spengler, T., Rentz, O.: Fuzzy outranking for environmental assessment. Case study: iron and steel making industry. FSS **115**, 45–65 (2000)
6. Goumas, M., Lygerou, V.: An extension of the PROMETHEE method for decision making in fuzzy environment: ranking of alternative energy exploitation projects. EJOR **123**, 606–613 (2000)
7. Halouani, N., Chabchoub, H., Martel, J.-M.: PROMETHEE-MD-2T method for project selection. EJOR **195**, 841–849 (2009)
8. Hyde, K., Maier, H.R., Colby, C.: Incorporating uncertainty in the PROMETHEE MCDA method. JMCDA **12**, 254–259 (2003)
9. International Energy Agency (ed.): World Energy Outlook 2011. OECD/IEA, Paris (2011)
10. Macharis, C., Brans. J.-P., Mareschal, B.: The GDSS PROMETHEE procedure. JDS **7**, 283–307 (1998)
11. Martín, J.M., Fajardo, W., Blanco, A., Requena, I.: Constructing linguistic versions for the multicriteria decision support systems preference ranking organization method for enrichment evaluation I and II. IJIS **18**, 711–731 (2003)
12. Pedrycz, W., Ekel, P., Parreiras, R.: Fuzzy Multicriteria Decision-Making. Models, Methods and Applications. Wiley, Chichester (2011)
13. Radgen, P., Baustein, E. (eds.): Compressed Air Systems in the European Union. Energy, Emissions, Savings Potential and Policy Actions. LOG_X, Stuttgart (2001)
14. Radojevic, D., Petrovic, S.: A fuzzy approach to preference structure in multicriteria ranking. ITOR **4**, 419–430 (1997)
15. Sorrell, S., Mallet, A., Nye, S.: Barriers to industrial energy efficiency: A literature review. Working Paper 10/2011. UNIDO, Vienna (2011)
16. Le Téno, J.F., Mareschal, B.: An interval version of PROMETHEE for the comparison of building products' design with ill-defined data on environmental quality. EJOR **109**, 522–529 (1998)
17. United Nations Industrial Development Organization (ed.): Motor Systems Efficiency Supply Curves. UNIDO, Vienna (2010)
18. Weissfloch, U., Geldermann, J.: Dealing with conflicting targets by using group decision making with PROMETHEE. In: Klatte, D., Lüthi, H.-J., Schmedders, K. (eds.) Operation Research Proceedings 2011, pp. 115–120. Springer, Berlin (2012)

Competitive Ratio as Coherent Measure of Risk

Iftikhar Ahmad, Esther Mohr and Günter Schmidt

1 Introduction

Investors in financial markets are naturally exposed to risk. It is therefore useful to quantify the risk of a financial position in order to decide if it is acceptable or not. Among several risk measures proposed in literature *Value-at-Risk* (VaR) and coherent risk measures are most commonly used [1–3]. VaR is the most popular risk measure, especially in practice, but there are several criticisms (see, for example, [4]). The main critics are that (1) VaR is not sensitive to diversification, and (2) VaR disregards any loss beyond the VaR level. This led [1] to introduce an axiomatic definition of coherent measures of risk. The VaR turns out to be not coherent since it does not decrease when an investor diversifies (p. 10, [2]).

Recently only *Expected Shortfall* is suggested as practicable and sound alternative to VaR as it is coherent and takes into account losses beyond the VaR level (p. 1519, [3]). We prove the *Competitive Ratio* to be a further alternative as it satisfies the axioms of a coherent measure of risk.

The rest of the paper is organized as follows. Section 2 defines the concepts of online algorithms and risk management. In Sect. 3 it is shown that the *Competitive Ratio* is a coherent measure of risk. Section 4 concludes the paper.

I. Ahmad · E. Mohr
Saarland University, P.O. Box 151150, D-66041 Saarbrücken, Germany
e-mail: ia@orbi.uni-saarland.de

E. Mohr (✉)
e-mail: em@orbi.uni-saarland.de

G. Schmidt
Department of Statistical Sciences, University of Cape Town, Rondebosch, South Africa

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_10,

2 Online Algorithms and Risk Management

In their work [5] present a basic risk paradigm, and classify action under two categories, (1) a risk-free action which produces a certain outcome, and (2) a risky action where the outcome is not certain. Based on [5] we interpret the risk-free action as the one which guarantees a specific return. Similarly, a risky action does not guarantee some return.

2.1 Competitive Analysis

Competitive analysis is the main tool when analyzing online algorithms, and compares the performance of an online algorithm (ON) to that of an adversary, the optimal offline algorithm (OPT).

ON computes *online* if for each $t = 1, \ldots, T-1$, ON must compute an output for t before the input for $t+1$ is given. OPT computes *offline* if it can compute a feasible output given the entire input sequence I in advance. By definition, the return of OPT is $OPT(I) = \sup_{O \in F(I)} U(O, I)$, where $\mathcal{I}$ is a set of possible inputs I, and $F(I)$ is the set of feasible outputs O. U is a utility function such that for all I and $O \in F(I)$, $U(I, O) \in \mathbb{R}$. ON is c-competitive if for any input $I \in \mathcal{I}$ (p. 104, [6])

$$ON(I) \geq \frac{1}{c} \cdot OPT(I). \tag{1}$$

Any c-competitive ON is guaranteed at least the fraction $\frac{1}{c}$ of the optimal offline return $OPT(I)$ no matter how (un)fortunate the future will be. We consider a maximization problem, i.e. $c \geq 1$. The smaller c the more effective is ON.

If the *Competitive Ratio* is related to a *performance guarantee* it must be a worst-case measure. This certain (risk-free) value is denoted as worst-case competitive ratio c throughout the paper. When c is derived it is commonly assumed ON is confronted with the worst possible sequence of prices for each $I \in \mathcal{I}$, and thus achieves the worst possible return r_{ON} on each I. Whereas OPT achieves the best possible return r_{OPT}.

But in a real world scenario, an investor in a financial market might be willing to take some risk in order to gain a return $X \in [r_{ON}, r_{OPT}]$. To incorporate risk management to online algorithms [7] proposed a risk-reward framework based on the competitive ratio. The framework allows an investor to take risk for a (possibly) higher return (lower competitive ratio). The model is based on forecasts on future price movements. In case the forecast is true the investor obtains a competitive ratio $c_1 < c$. However, in case the forecast is not true the investor obtains competitive ratio $c_2 > c$. Further, assume the risk taken can be controlled by a certain factor. Let the acceptable level of risk for an investor be $a \in [1, c]$. Hence, a defines the minimum and maximum bound of returns where $a = 1$ reflects no risk and $a = c$ the maximum risk. If the forecast is true, the investor can thus achieve a competitive

ratio of $c_1 = c/a$. If the forecast is not true, the investor is guaranteed a competitive ratio not worse than $c_2 = a \cdot c$. Note that even if the assumptions based on which the investor is willing to take some risk are not true, the worst (possible) competitive ratio c_2 is still guaranteed.

Let the desired return of an investor be X, and the associated level of risk to achieve X be a. Consider an investor is willing to take some risk a to achieve a return $X \in [r_{ON}, r_{OPT}]$. Then the desired competitive ratio equals [cf. (1)]

$$c_X \leq \frac{r_{OPT}}{X}, \tag{2}$$

and the level of risk the investor is willing to take equals

$$\begin{aligned} a &= \frac{X}{r_{ON}} \\ &= \frac{c}{c_X}, \end{aligned} \tag{3}$$

where $a \in [1, c]$. Thus, the resultant investor return can vary in a range of $[\frac{r_{ON}}{a}, a \cdot r_{ON}]$.

2.2 Coherent Risk Measures

A risk measure determines the quantity of an asset that needs to be kept in reserve in order to make the risk taken by an investor acceptable. The notion of coherent risk measures arose from an axiomatic approach for quantifying the risk of a financial position, presented in the seminal paper of [1].

Definition 1 A risk measure ρ assigns a random variable X a non-negative real number R, i.e. $\rho : \mathcal{X} \rightarrow R$.

Consider a random return X viewed as an element of a linear space $\mathcal{X}$ of measurable functions, defined on an appropriate sample space. According to [1, 2, 8] a function $\rho : \mathcal{X} \rightarrow R$ is said to be a coherent risk measure for X if it satisfies the following set of axioms.

Axiom M: Monotonicity. For all random returns $\{X_1, X_2\} \in \mathcal{X}$, if $X_1 \geq X_2$ then

$$\rho[X_1] \geq \rho[X_2]. \tag{4}$$

Monotonicity implies that if a random return X_1 is always higher than a random return X_2, then the risk of X_1 should be greater than the risk of X_2.

Axiom S: Subadditivity. For all random returns $\{X_1, X_2\} \in \mathcal{X}$

$$\rho[X_1 + X_2] \leq \rho[X_1] + \rho[X_2]. \tag{5}$$

Subadditivity implies that the risk of two investments together cannot get any worse than adding the two risks separately.

Axiom PH: Positive Homogeneity. For all random returns $X \in \mathcal{X}$, if $\lambda > 0$ then

$$\rho[\lambda X] = \lambda \rho[X]. \tag{6}$$

Positive Homogeneity implies that if a random return X is increased by λ, then the risk associated is also increased by λ.

Axiom TI: Translational Invariance. For all random returns $X \in \mathcal{X}$, risk-free returns r, and $\alpha \in \mathbb{R}$

$$\rho[X + r] = \rho[X] - \alpha. \tag{7}$$

Translational Invariance implies that by adding a risk-free return r to a random return X the risk associated decreases by α.

A widely used coherent (and moreover a spectral) measure of financial portfolio risk is the *Conditional VaR* (for a formal definition see [8, p. 227]).

The *Conditional VaR* defines the expected loss of portfolio value given that a loss is occurring at or below a certain quantile-level q. Its effectiveness, however, depends on the accuracy of estimation (p. 999, [4]). In contrast, the (desired) competitive ratio does not depend on any estimates [cf. (1) and (2)].

3 Competitive Ratio as Coherent Measure of Risk

While considering *Competitive Ratio* as coherent risk measure, it is pertinent to note that the nature of the *Competitive Ratio* varies a great deal from the *Expected Shortfall*. The *Competitive Ratio* quantifies the maximum regret (possible loss) under worst-case assumptions. The *Expected Shortfall* quantifies the expected return in the worst q % of the cases.

Let us consider ON with return r_{ON}, and OPT with return r_{OPT}. As r_{ON} is risk-free, ON is guaranteed to achieve minimum r_{ON}. Further, consider an investor is willing to take some risk $a \geq 1$ for a higher reward, and wants to achieve a return $X \geq r_{ON}$. Then, the desired competitive ratio equals $c_X = \frac{r_{OPT}}{X}$ [cf.(2)], and the level of risk to achieve X, $\rho[X]$, equals $a = \frac{c}{c_X}$ [cf. (3)].

Axiom M: Monotonicity. From (4) we know for higher returns an investor has a higher risk level, and potentially greater losses: If desired return X_1 is greater than desired return X_2 $\forall X_1, X_2 \in [r_{ON}, r_{OPT}]$, then the associated risk (and thus the potential loss) of X_1 will be at least as high as that of X_2.

Proof If $X_1 = X_2$, it is trivial to show that $\rho[X_1] = \rho[X_2]$. If $X_1 > X_2$, using (3), we get

$$\begin{aligned} \frac{X_1}{r_{ON}} &> \frac{X_2}{r_{ON}} \\ \frac{c}{c_{X_1}} &> \frac{c}{c_{X_2}} \\ \rho[X_1] = a_1 &> a_2 = \rho[X_2]. \end{aligned} \tag{8}$$

□

Axiom S: Subadditivity. From (5) we know diversification never increases risk: If an investor wants to achieve a higher return $(X_1 + X_2)$ such that $r_{ON} < (X_1 + X_2) \leq r_{OPT}$, then risk associated is never greater than the sum of the individual risks associated with X_1 and X_2. hurt, i.e. $\rho[X_1 + X_2] \leq \rho[X_1] + \rho[X_2]$.

Proof For $\{X_1, X_2\} \in \mathcal{X}$, using (3), we get

$$\begin{aligned} \rho[X_1 + X_2] &= a_1 + a_2 \\ &= \frac{c}{c_{X_1}} + \frac{c}{c_{X_2}} \\ &= \rho[X_1] + \rho[X_2]. \end{aligned} \tag{9}$$

□

Axiom PH: Positive Homogeneity. From (6) we know that if a random return X is increased, then the risk associated is increased by the same factor: For a desired return λX, such that $r_{ON} < \lambda X \leq r_{OPT}$, the risk associated with λX is λ times greater than the associated risk for X.

Proof For $X \in \mathcal{X}$, if $\lambda > 0$, from (3) we get

$$\begin{aligned} \rho[\lambda X] &= \lambda a \\ &= \lambda \left(\frac{c}{c_X} \right) \\ &= \lambda \rho[X]. \end{aligned} \tag{10}$$

□

Axiom TI: Translational Invariance. From (7) we know that the introduction of a risk-free investment does never increase the level of risk. When considering the *Competitive Ratio* as a risk measure, Axiom TI needs to be redefined. We can state that in a risk-free environment there is no additional capital requirement to assure an investment decision since there is no uncertainty.

Assume an investor diversifies, and invests some amount risk-free. Then the desired return equals $Y = X + r$, and from (7) we get

$$\rho[Y] = \rho[X] - \alpha. \tag{11}$$

Further, let the risky return be r' be X, i.e. $Y = r' + r$. From (11), we get

$$\rho[Y] = \rho[r'] - \alpha. \tag{12}$$

Since a risky investment, resulting in r', never decreases the level of risk $\alpha = 0$. Thus, we have to show that

$$\rho[Y] \leq \rho[r']. \tag{13}$$

Proof For all random returns $X \in \mathcal{X}$, risk-free returns r, and risky returns r'

$$\begin{aligned} \rho[Y] &= \rho[X + r] \\ &= \rho[r' + r] \end{aligned} \tag{14}$$

by using Axiom S

$$\rho[Y] = \rho[r'] + \rho[r].$$

As $a = 1$ reflects no risk, and r is risk-free

$$\rho[Y] = \rho[r']. \tag{15}$$

□

4 Concluding Remarks

Two risk measures are well established and widely used: VaR and *Expected Shortfall*. As VaR is not coherent, it may underestimate risk under extreme asset price fluctuations or an extreme dependence structure of assets [4]. Information provided by VaR may mislead investors. In search for a suitable alternative to VaR, *Expected Shortfall* has been characterized as the smallest coherent risk measure to dominate VaR [3].

In this paper we showed the *Competitive Ratio* to be coherent. It is sensitive to diversification and thus also dominates VaR. Unfortunately, the *Competitive Ratio* is so far not established as a measure of risk, or even unknown to practitioners. But a risk measure that takes into account worst-case scenarios like crashes or situations of extreme stress on investor portfolios is essential.

We conclude that the use of a single risk measure should not dominate financial risk management, and suggest the *Competitive Ratio* as a further alternative to VaR. Existing coherent risk measures could complement one another to provide an effective way to facilitate a more comprehensive risk monitoring.

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Mathematical Finance **9**(3), 203–228 (1999). July
2. Szegö, G.: Measures of risk. European Jounal of Operational Research **163**(1), 5–19 (2005)
3. Tasche, D.: Expected shortfall and beyond. Journal of Banking and Finance **26**(7), 1519–1533 (2002)

4. Yamai, Y., Yoshiba, T.: Value-at-risk versus expected shortfall: A practical perspective. Journal of Banking and Finance **29**(4), 997–1015 (2005)
5. MacCrimmon, K., Wehrung, D., Stanbury, W.: Taking risks: the management of uncertainty. New York: Free Press; London: Collier Macmillan Publishers (1988).
6. El-Yaniv, R., Fiat, A., Karp, R., Turpin, G.: Optimal search and one-way trading algorithm. Algorithmica **30**(1), 101–139 (2001)
7. al-Binali, S.: A risk-reward framework for the competitive analysis of financial games. Algorithmica 25(1) (1999) 99–115.
8. Ahmed, S., Cakmak, U., Shapiro, A.: Coherent risk measures in inventory problems. European Journal of Operational Research **182**(1), 226–238 (2007)

Binomial Lattice Model: Application on Carbon Credits Market

Natália Addas Porto and Paulo de Barros Correia

1 Introduction

Clean Development Mechanisms (CDM) are part of an international legal regime, once they are provided by Kyoto Protocol. The main purpose of these legal instruments is to insert the participation of (NON- ANNEX I) countries in the reduction and stabilization of greenhouse gas emissions through practical measures for a sustainable development, either by investments and/or technologies transferring from (ANNEX I) countries or stakeholders.

The establishment of a carbon credit market, fundamentally the trade of Certified Emissions Reductions (CERs) obtained from CDM projects, fosters partnerships between certificated generators and parts that have goals to fulfill under Kyoto Protocol. Significant uncertainty in effecting a CDM project activity, followed by a price behavior analysis, are both key factors for not compromising the carbon credits contracts expected return.

In this sense, the present paper shows a theoretical basis for the evolution of CERs trade, emphasizing a contract pricing model based on binomial lattice.

2 Carbon Market

Flexible mechanisms are provided by Kyoto Protocol so ANNEX I countries can achieve their emission reduction targets, by generating credits related to emissions reduced in other countries: Join Implementation (JI), Clean Development Mechanism (CDM) and International Emission Trading (IET).

N. A. Porto (✉) · P. B. Correia
University of Campinas—UNICAMP, Campinas, Brazil
e-mail: natalia@fem.unicamp.br

P. B. Correia
e-mail: pcorreia@fem.unicamp.br

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_11,

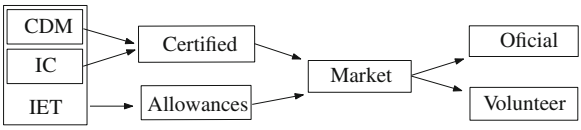


Fig. 1 Carbon market scheme

It is important to note that while JI and CDM create certifies based on projects, IET allows the transaction of such certificates. Thus, these three mechanisms enable international transactions of carbon credits, in which each carbon credit represents the reduction of one ton of carbon equivalent molecule (CO_2e).

In general, there are two types of markets facing carbon credits trade: the official market and the volunteer market. In the official market, credits are negotiated to facilitate the achievement of emission reduction targets under the climate regime. On the other hand, in the volunteer market there are no emission reduction targets to be fulfilled under the Kyoto Protocol, but only goals that were voluntarily established by companies and local governments.

It is noteworthy that in both markets, official and volunteer, it is possible to negotiate certified reduction emissions and emission allowances, as indicates Fig. 1.

The creation of a market focused on the trade of certified emissions and allowances, enables the use of financial instruments, both for the reduction of risk as to taking advantage of this risk. The carbon credits contracts are globally traded in OTC markets and securities, through future and options contracts or other derivative contracts.

The options contracts were, particularly created to be a protection instrument against price fluctuations in the spot market, providing to the holder the right, but not the obligation, to buy or to sell the given asset at a fixed price.

3 Pricing Model

The pricing procedure implemented is based on the binomial model developed by Cox [1]. Originally developed to deals with price assets, it is used in this paper to represent the random behavior of CERs' prices. Economic studies to evaluate the benefits and the risk exposure have been increasingly necessary to understand the long-term market, in which the idea is to create a credible commitment between firms and lenders, avoiding uncertainties, especially regarding CERs prices fluctuation during the option contract.

3.1 Multiplicative Model

The pricing multiplicative model is based on the dynamic relation $S_{t+1} = S_t u(t)$, where $u(t)$ are independent random variables, for $t = 0 \rightarrow T$. Taking the natural

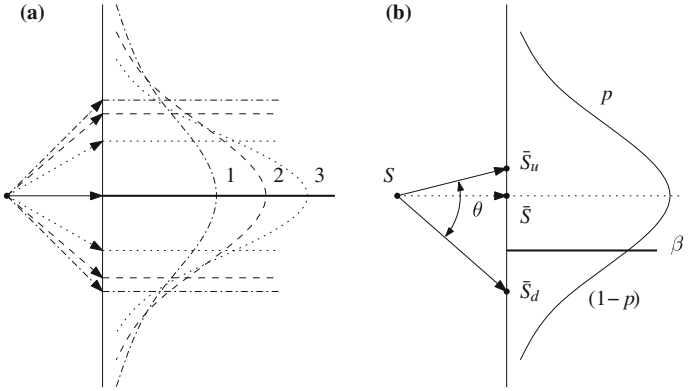


Fig. 2 Binomial lattice: risk

logarithm on its both sides results in $\ln[S_{t+1}] = \ln[S_t] + \ln[u(t)]$. The sequential application of this relation for $t = 0 \rightarrow k$ gives

$$\ln S_k = \ln S_0 + \sum_{t=0}^{k-1} \ln u(i). \tag{1}$$

If $\ln u(i)$ were independent normal random variables then $\ln S_k$ also is a normal random variable [2, 3]. Consequently S_k is a log-normal random variable.

However, the normal random variables are not the only ones to preserve their shapes under the sum operation. Others random distribution densities have this propriety, as the gamma distribution [3]. Thus, the paper investigated if the CERs prices random can be approximated with a random distribution that preserves its shape under the sum operation, allowing the use of the binomial lattice model.

3.2 Binomial Lattice

Despite its simplicity, the binomial lattice model adds a risk dimension to the expected behavior of a random variable, as shown if Fig. 2a, for three density functions with the same expected value.

$$\bar{S} = \int_{-\infty}^{\infty} Sf(s)ds. \tag{2}$$

Alternatively, for some β, $\bar{S}$ can be given as

$$\bar{S} = \int_{-\infty}^{\beta} Sf(s)ds + \int_{\beta}^{\infty} Sf(s)ds = p\bar{S}_u + (1-p)\bar{S}_d. \tag{3}$$

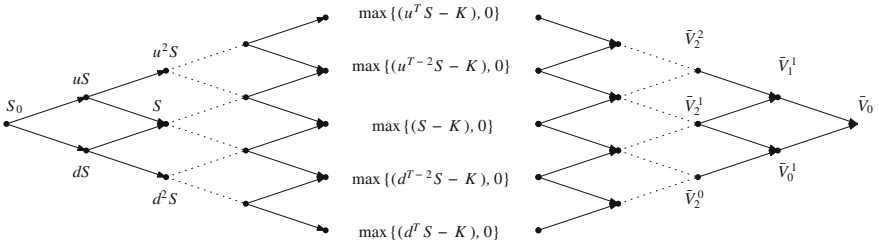


Fig. 3 Binomial lattice: proceeding

where the opening angle θ is an indication of risk as a dispersion of price, as shown in the Fig. 2b.

The parameters of the binomial lattice model are determined by solving the Eqs. 4 and 5, a system with two equations and three variables.

$$\mathrm{E}[\ln(S_{t+1}/S_t)] = p \ln u + (1 - p) \ln d \tag{4}$$

$$\mathrm{var}[\ln(S_{t+1}/S_t)] = p(1 - p)(\ln u - \ln d)^2 \tag{5}$$

Often the degree of freedom of system Eq. 4–5 is fixed doing $u = 1/d$. But if the prices tend to increase only, with $u > 1$ and $d > 1$, we could fix it by choosing $p \approx 1/2$.

The binomial lattice procedure has two phases: an expansion from the current price followed by a contraction to the expected price (Fig. 3). In the case of the pricing call options, the values at the lattice terminal nodes are given by

$$\max\{(S_i^j - K), 0\}. \tag{6}$$

4 Study of Case: Pricing Option Contract of CERs

By analyzing the price dynamics over time, the binomial lattice model can be used for pricing an option. Pricing options consists in determining a market value of an asset at the time of buying (for a call) or selling (for a put) an option. An option gives one the right, but not the obligation, to buy or sell an asset at a specified price K, or strike price, at a specified period of time T, paying for it a value called premium [4].

A call option, for example, gives to the holder the right to purchase an asset; if in the future, the stock price is greater than the strike price specified in the contract, the call option holder will exercise the option, and the seller has an obligation to fulfill the buyer as soon as requested.

Thus, the price of an option for a given year can be estimated and, with the possession of this reference, CERs investors can deal with the prices uncertainties and manage their risks.

Table 1 describes the parameters used for this case study.

We considered hypothetically a call option contract that had a strike price of 12.00 Euros/tCO_2e and the expiration date is early 2013. The variables u and d defines the relative change in price S between times t and $t + 1$; the probabilities of these possibilities are p and $(1 - p)$, respectively. Therefore, the value of the call option at expiration would be $\max\{(S_T - K), 0\}$ and at every period this value is discounted according to the risk-free rate i [2]. Moreover, σ^2 is the monthly variance in lognormal price.

This is computed with a commercial solver, the Lingo 11.0, which contemplates an integrated suite of optimization modules [5]. The dynamic programming is dealt with using backward recursion from its end nodes. As this is a call option, the end nodes of lattice are obtained as described in Eq. 6.

Figure 4 shows the procedures for expansion and contraction of the lattice

Given the current stock price S, the option price was estimated as 4.24 Euros/tCO_2e . This value is a reference of the premium to be paid by option. The premium paid

Table 1 Parameters

	Parameter	Value
T	Number of periods (months)	12
S	Current stock price (Euros/tCO_2e)	11.57
K	Strike price at option expiration (Euros/tCO_2e)	12.00
i	Yearly interest rate	0.10
σ^2	Monthly variance in lognormal price	0.12406
u	Up move factor	1.08076
d	Down move factor	0.96816
p	Probability of an up move	0.55

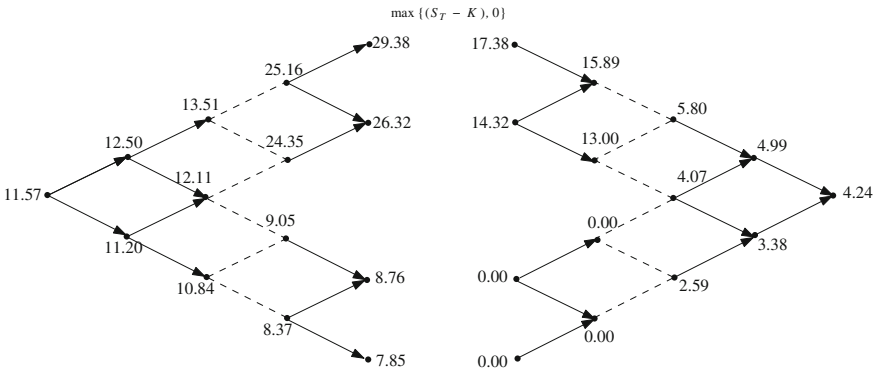


Fig. 4 Binomial lattice—*forward* and *backward*

to the contract must be less than 4.24 Euros/tCO_2e for the buyer to avoid losses and, similarly, the seller must charge a higher value to avoid losses.

5 Conclusions

The case study of this paper provided results for the certified market. We proposed an analysis of CERs contracts below the financial point of view, calibrating the binomial lattice model using historical data. However, one of the main obstacles to execute these contracts is due to large price variations associated with the economic crisis and recent historical prices, making it difficult to estimate a price that reflects the possibility of an encounter between suppliers and consumers. The proposed model through the use of binomial lattice demonstrated to be an appropriate tool for pricing options of CERs. Besides providing risk assessment of contracts with certain simplicity, these analyses where completed within the lattice.

References

1. Cox, J.C., Ross, S.A., Rubinstein, M.: Option pricing: A simplified approach. J. Financ. Econ. **7**, 229–263 (1979)
2. Luenberger, D.G.: Investment Science, p. 494. Oxford University Press, New York (1998)
3. Ross, S.: A First Course in Probability, 7th edn, p. 565. Pearson, Prentice Hall, New Jersey (2006)
4. Hull, J.: Options, Futures and Other Derivatives, 5th edn. Prentice Hall, New Jersey (2002)
5. Lindo Systems Inc. http://www.lindo.com

Sustainability Assessment of Concepts for Energetic Use of Biomass: A Multi-Criteria Decision Support Approach

Nils Lerche, Meike Schmehl and Jutta Geldermann

1 Introduction

A usage of biomass for energy supply is attracting attention among a variety of stakeholders, such as politicians, power suppliers or the public. This energetic use of biomass is thereby often associated with sustainability, e.g. due to its less pronounced contribution to climate change or the preservation of fossil energy reserves [5]. Contrary to other renewable sources, as wind or solar, biomass can be stored and used for a steady and reliable production of electricity. Moreover, it could be used not only for electricity production, but also for the production of heat, cooling, fuels or lubricants and other products of petrochemistry. This makes biomass an especially versatile option for further energy challenges [12].

There arises also an increasing social awareness towards sustainability, which can be subdivided into the three dimension environmental quality, economic prosperity and social equity [15, 16]. This development is of growing importance for technologies and infrastructure with respect to energy supply [16]. Concerning the major advantages as renewable energy source or as resource for the chemical industry in combination with a growing demand for sustainable concepts, biomass represents a suitable and interesting option. Therefore, it seems to be of special interest to identify efficient concepts for biomass usage on a regional scale under sustainable criteria.

But, despite the mentioned advantages, the public perception of the energetic use of biomass has recently been characterized by decreasing acceptance. Expanding

N. Lerche (✉)
Chair of Production and Logistics, Georg-August-Universität Göttingen, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany
e-mail: nils.lerche@wiwi.uni-goettingen.de

M. Schmehl
e-mail: schmehl@wiwi.uni-goettingen.de

J. Geldermann
e-mail: geldermann@wiwi.uni-goettingen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_12,

monocultures are viewed critically, as they would result in considerable land-use changes and affect biodiversity negatively. Additionally, the associated increasing transport activity results in noise disturbance as well as air pollution. Moreover, the use of food crops for energy production is also a controversial issue from an ethical perspective. This is a matter of particular interest in cases where areas are used for food production, nature conservation or grassland [5, 8].

Hence, it becomes apparent that the assessment of potential concepts for the energetic use of biomass is of multi-criteria nature and should imply not only economic, but also environmental, social and technical aspects. For that reason, methods of Multi-Criteria Decision Analysis (MCDA) seem to be adequate for assessing potential concepts of biomass-usage. Since the given problem concerns both energy supply and sustainability, it is necessary to develop and apply an appropriate method based on MCDA for the assessment of alternative concepts for the energetic use of biomass with respect to sustainable development.

2 Assessment of Concepts for Biogas in a Rural Area

As mentioned before, a suitable method for the sustainability assessment of biomass potentials on a regional scale is to be developed. One initial point is an already conducted study, where the MCDA-method Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) is applied on the sustainability assessment of different concepts for the use of biomass [5, 6]. PROMETHEE was chosen as approach due to its consideration of discrete alternatives, as well as its transparency and ability of generating additional information through the whole decision process [13, 14, 16]. In this study, PROMETHEE was applied exemplarily on small-scale and large-scale biogas plants, as well as on a new concept of bioenergy villages in a rural area in Lower Saxony in Germany. Thus, the focus of the application was the comparison of different concepts for the use of biogas on a local scale [5, 6].

Moreover, a criteria-hierarchy has been developed to cover the different dimensions of sustainability. From an ecological perspective, the choice of criteria represents the impact of the use of biomass for energy supply not only on air, soil and water, but also on biodiversity and the preservation of resources. The economic criteria arise from the different stakeholders as the operating company, employees, heat clients, farmers and the region itself. At this point it should be mentioned that the determination of values for some criteria is challenging due to conflict of interest. An identical criteria could be assessed in a contrary way, as the following example points out. While farmers prefer high prices, the operating company and the villagers are better served with cheap biomass. The social aspects of sustainability had been categorized by sub-categories, such as acceptance, participation, psychological consequences and employment. Additionally, a fourth dimension, the technical dimension, was introduced, because the efficiency of the technical conversion of biomass is seemed to be of particular importance and is closely connected to the corresponding alternative [5, 6].

Based on the presented approach, further adjustment concerning the application of PROMETHEE on a broader scale shall be discussed, in particular to address the special needs of assessing sustainability and considering conflictive stakeholders.

3 The PROMETHEE Approach

The PROMETHEE approach belongs to the methods of Multi-Criteria Decision Analysis (MCDA). The goal of any method of MCDA is to deliver a decision support for at least one decision maker, based on his or her individual preferences due to a set of multiple criteria. Moreover, the application of the different methods should make the whole decision process more transparent and elicit further information from the involved decision makers. Thereby, methods of MCDA facilitate a better understanding of the decision problem as such [2, 7].

From the two kinds of MCDA-methods the approaches of Multiple-Attribute Decision Making (MADM) seem to be appropriate, because they are used to analyse a set of discrete alternatives. One subset of MADM methods are Outranking approaches, to which the PROMETHEE approach belongs. PROMETHEE was developed by Brans [3] and has already been used in various applications for assessing sustainability [1, 14]. It is based on pairwise comparisons of alternatives due to the multiple criteria and enables the inclusion of weak preferences and incomparabilities. For these pairwise comparisons six specific kinds of preference functions are recommended. There are several enhancements of PROMETHEE from PROMETHEE I up to VI, where as PROMETHEE I and II are the most frequently applied versions. The aim of PROMETHEE I is to find a partial preorder, which can be transformed in PROMETHEE II through the determination of net outranking flows into a complete preorder [4]. With respect to the given problem, PROMETHEE is considered as an appropriate approach for different reasons. Firstly, as MADM-method it analysis a set of discrete alternatives. Furthermore, it is possible to consider criteria not only of quantitative, but also of qualitative nature simultaneously [3, 4]. Since its application improves the transparency of the decision process, it also helps both to structure that process and to prompt the decision maker to rethink about the whole decision problem, whereby new findings could arise [2, 4]. In addition, since the outranking approaches are additive methods, it is possible that under certain conditions an application of PROMETHEE and the Multi-Attribute utility theory (MAUT) leads to identical results [10]. A further advantage of PROMETHEE is, that a broad spectrum of sensitivity analysis can be applied for revealing additional information concerning the decision problem and process, respectively [9].

Thus, PROMETHEE is an approach which attempts to deal with vague preferences by the decision maker and criteria of different nature. It helps also to structure the decision process and makes it more transparent. Furthermore, with respect to MCDA-methods, it is also recommended to apply not just one single approach, but to combine different approaches. As every approach has its own advantages and drawbacks, this gives the opportunity to compensate especially undesirable drawbacks [13].

4 Outlook on Potential Enhancements

Further adjustments to the application of PROMETHEE shall be discussed to make it more suitable for the assessment of concepts for biomass usage with regard to sustainability aspects. The adjustments have their origin not only in issues which come along with sustainable development or energy supply, but also in the application on a broader scale with a variety of stakeholders. As energy supply and sustainable development are especially emotional topics in combination with the consideration of stakeholders with possibly very conflictive positions, it is typically a field of cognitive biases [11]. For that reason, the adjustments of the application of PROMETHEE are aimed both at the inherent integration of stakeholders with contrary positions and the consideration of cognitive biases through the decision process.

As mentioned in section two, one drawback of the original approach was that stakeholders could express contrary preferences to some criteria, as in the case of the price for biomass. It can be expected that the differences between stakeholders widen with increasing scale, so that on regional scale with a lot of stakeholders, the consideration of their individual evaluation due to different criteria is of special interest. Therefore, the inherent implementation of individual preferences, whether a criteria should be maximized or minimized, in a decision support approach based on PROMETHEE, shall be facilitated. On a regional scale it is also necessary to consider a greater variety of potential alternatives. Conceivable alternatives in addition to biogas plants or bioenergy villages could also be industrial plants as biomass-to-liquids plants, bio-refineries or plants producing biocoal. This broader spectrum comes along with a growing number of potential stakeholders, too. Moreover, in that context it would be meaningful not only to assess biomass usage concepts on their own separated from other potential options. The combination of the decision with respect to biomass usage with already existent or further potential solutions, e.g. in combination with wind power or photovoltaic, represents therefore another reasonable enhancement.

In order to deal with cognitive biases in the decision process, the approach of structured decision making (SDM) [11] represents an interesting basic concept. SDM can hereby described as a code of practice how to improve and structure the decision process. It explicitly implies approaches to reduce cognitive biases, such as the phenomena of representativeness due to the weighting, or the availability of information by assigning values to the alternatives for different criteria for example. Based on that idea, it seems reasonable to add further procedural steps based on behavioural sciences, through which cognitive biases are reduced when determining weights and values or facts, respectively. While the originally developed SDM-approach combines MAUT with methods of decision analysis, it would be interesting to use PROMETHEE as an initial method. This seems reasonable, as Løken argues that PROMETHEE is appropriate as basis for decision support, as it helps to structure the decision problem and makes it more transparent. Thus, it is particularly suitable for combination with a second method for decision support, which could be applied afterwards [13]. Additionally, the assumption of Outranking methods, that

the decision maker is not entirely aware of his or her preferences, supports rather the application of PROMETHEE compared to MAUT, if one wants to deal with cognitive biases. Another aspect of SDM which could be added for a sustainability assessment of biomass usage concepts is the construction of alternatives, rather than the assessment of ex ante determined alternatives [11]. As it is the aim to assess different concepts for the use of regional biomass potentials, the construction of alternatives could represent an active process to reveal new innovative concepts.

It can be concluded, that the application of MCDA-methods enhances the transparency of subjective decision processes and helps further to identify potential conflicts of objectives through the implementation of sensitivity analysis.

Acknowledgments This research is funded by a grant of the "Ministry of Science and Culture of Lower Saxony" and "Volkswagenstiftung". It is part of the research sub-project "A. 1.1: Multi-Criteria decision support for the choice of biomass usage concepts" of the joint research project "Sustainable use of bioenergy: bridging climate protection, nature conservation and society" with a duration from 2009 till 2014.

References

1. Behzadian, M., Kazemzadeh, R.B., Albdavi, A., Agdhasi, M.: PROMETHEE: a comprehensive literature review on methdologies and applications. Eur. J. Oper. Res. **200**, 198–215 (2010)
2. Belton, V., Stewart, T.: Multiple Criteria Decision Making: An Integrated Approach. Kluwer Academic Press, Boston (2002)
3. Brans, J.P., Vincke, P.: A Preference Ranking Organisation Method. Manage. Sci. **31**, 647–656 (1985)
4. Brans, J.P., Mareschal, B.: PROMETHEE Methods. In: Figueira, J., Greco, S., Ehrgott, M. (eds.) Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, New York (2005)
5. Eigner-Thiel, S., Schmehl, M., Ibendorf, J., Geldermann, J.: Assessment of different bioenergy concepts regarding sustainable development. In: Ruppert, H., Kappas, M. (eds.) Sustainable Bioenergy Production: An Integrated Approach. Springer, Berlin (in preparation (2013)
6. Eigner-Thiel, S., Geldermann, J., Schmehl, M.: Soziale Kriterien zur Bewertung der Nachhaltigkeit unterschiedlicher Biomassepfade. In: Böttcher (eds.): Biogas, Springer, Berlin, (in preparation, 2013).
7. Figueira, J., Greco, S., Ehrgott, M. (eds.): Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, New York (2005)
8. Fritsche, U.R., Hennenberg, K., Hermann, A., Hnecke, K., Schulze, F., Wiegmann, K., Fehrenbach, H., Roth, E., Hennecke, A., Giegrich, J.: Sustainable Bioenergy: Current Status and Outlook, ifeu. Darmstadt, Heidelberg (2009)
9. Geldermann, J.: Mehrzielentscheidungen in der industriellen Produktion. Universitätsverlag Karlsruhe, Karlsruhe (2005)
10. Geldermann, J., Schöbel, A.: On the similarities of some MCDA methods. J. Multi-Criteria Decis. Anal.s 18, 219–230 (2011).
11. Gregory, R., Failing, L., Harstone, M., long, G., McDaniels, T., Ohlson, D.: Structured Decision making A Practical Guide to Environmental Management Choices, Wiley, New York (2012)
12. Kaltschmitt, M., Hartmann, H., Hofbauer, H.: Energie aus Biomasse Grundlagen, Techniken und Verfahren, 2nd edn. Springer, Heidelberg (2009)
13. Løken, E.: Use of multicriteria decision analysis methods for energy planning problems. Renew. Sustain. Energy Rev. **11**, 1584–1595 (2007)

14. Oberschmidt, J., Geldermann, J., Ludwig, J., Schmehl, M.: Modified PROMETHEE approach for assessing energy technologies. Int. J. Energy Sect. Manage. **4**, 183–212 (2010)
15. von Gleich, A., Gling-Reisemann, S.: Industrial Ecology—Erfolgreiche Wege zu nachhaltigen industriellen Systemen. Vieweg + Teubner, Wiesbaden (2008)
16. Wang, J.-J., Jing, Y.-Y., Zhang, C.-F., Zhao, J.-H.: Review on multi-criteria decision analysis aid in sustainable energy decision making. Renew. Sustain. Energy Rev. **13**, 2263–2278 (2009)

A Comparison of Two Visualisation Methods for Decision Support in MCDM Problems

Bastian Schmidtmann, Genoveva Uskova, Harald Uhlemair and Jutta Geldermann

1 Introduction

By using outranking-methods like Preference Ranking Organisation METHod for Enrichment Evaluations (PROMETHEE) [1], it is possible to include the alternatives, the relevant criteria and the personal preferences of each decision maker in the decision process. PROMETHEE is based on the pairwise comparison of the alternatives for each single criterion. With respect to personal preferences and weightings of the criteria by the decision maker an order of the alternatives is proposed. As an illustrative case study, the simplified comparison of different mobility concept is taken as an example. Rising energy prices, a strong dependency on fossil fuels and climbing CO_2 emissions are just some aspects that can have a great impact on future mobility behaviour. Table 1 shows exemplary data on mobility concepts of electric vehicles, hybrid vehicles and fuel cell vehicles which are compared to a conventional fossil fuel vehicle.

PROMETHEE calculates positive and negative preference flows which represent the strengths and weaknesses of each alternative and allow for a partial ranking in PROMETHEE I. For the complete ranking of all alternatives PROMETHEE II is applied by generating the difference between the negative and positive preference

B. Schmidtmann (✉)
Leibniz Universitaet Hannover, Koenigsworther Platz 1,
30167 Hannover, Germany
e-mail: schmidtmann@controlling.uni-hannover.de

G. Uskova · H. Uhlemair · J. Geldermann
Georg-August-Universitaet Goettingen,
Platz der Goettinger Sieben 3, 37073 Goettingen, Germany
e-mail: guskova1@uni-goettingen.de

H. Uhlemair
e-mail: harald.uhlemair@wiwi.uni-goettingen.de

J. Geldermann
e-mail: geldermann@wiwi.uni-goettingen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_13,

Table 1 Exemplary data and parameters of PROMETHEE

	Fossil fuel vehicle	Electric vehicle	Hybrid vehicle	Fuel cell vehicle	Min/Max	Type	q	p	δ	Weights
Costs of purchase	18,000	26,000	30,000	36,000	Min	5	6,000	12,000	-	0,2
Variable costs	2,000	1,500	600	2,100	Min	3	-	2,400	-	0,2
CO2 Emissions	149	89	80	110	Min	6	-	-	30	0,2
Charging time	5	5	420	7	Min	2	10	-	-	0,2
Range	900	100	150	400	Max	5	200	400	-	0,2

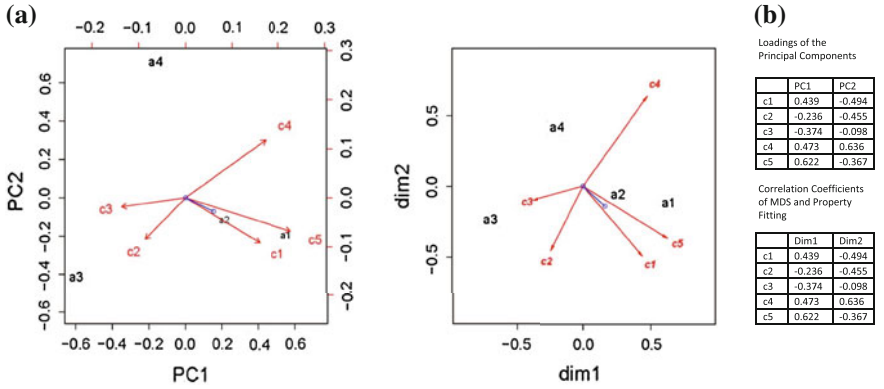
flows. The resulting information is represented in a matrix (ϕ) containing the net flows that is used as starting point for the visualisation methods.

$$\phi = \begin{pmatrix} 0.156 & 0.022 & -0.067 & -0.111 \\ -0.05 & 0.005 & 0.106 & -0.061 \\ -0.158 & 0.069 & 0.091 & -0.002 \\ 0.067 & 0.067 & -0.2 & 0.066 \\ 0.133 & 0.133 & -0.15 & -0.117 \end{pmatrix}$$

To get a better understanding of the results, [2] developed a visualisation method based on the principal component analysis (PCA). The so called GAIA method not only places the alternatives in the plane but also shows the weighted criteria and their impact on the alternatives in the plane. The GAIA method is not self-explanatory if the decision maker is not familiar with the PCA. To evaluate if the GAIA method itself or visualization methods themselves are a difficult to handle for the decision maker a new visualisation method for PROMETHEE results is developed and compared with the GAIA method. The method is based on the MDS technique that uses a distance matrix to visualize the similarities of the alternatives in relation to the weighted criteria. The shorter the distances between the alternatives, the more alike the alternatives are, regarding the criteria. By using the technique of Property Fitting the criteria are placed into the MDS plane. In the following, the GAIA method, which is based on the PCA, and the MDS technique are being applied and compared.

2 Visualization of PROMETHEE Using PCA

For the application of the PCA it is necessary to transform the matrix of the net flows into a new matrix. For an easy understanding and interpretation of the MCDM model a representation with two PCs is suitable. The alternatives are placed in the plane as points. Based on the composition of the PCA, Mareschal [2] developed the GAIA method to visualize MCDM models. In addition to placing the alternatives in the PCA plane the criteria are also visualized. Considering the eigenvectors and the weights the criteria are projected in the plane. The following can be said about the length and arrangement of the criteria: Criteria that strongly differentiate the alternatives show a large variance. If two criteria have similar preferences they will have a positive covariance. Two criteria that have independent preferences will be represented by orthogonal arrows in the plane and conflicting criteria show negative covariance and their arrows point in opposite directions. The decision stick is used as decision support in the GAIA plane. It is defined as a projection of a weighted vector in the plane. The stick helps to identify the best alternative as the direction of the decision stick points towards the same direction as the best alternative. The decision stick is the visualization of PROMETHEE II. The first two PCs span the plane for the placement of the four alternatives and the five criteria in the two dimensional space. The visualization of the results is shown in Fig. 1a. The cumulated proportion



	PC1	PC2
c1	0.439	-0.494
c2	-0.236	-0.455
c3	-0.374	-0.098
c4	0.473	0.636
c5	0.622	-0.367

	Dim1	Dim2
c1	0.439	-0.494
c2	-0.236	-0.455
c3	-0.374	-0.098
c4	0.473	0.636
c5	0.622	-0.367

Fig. 1 **a** Visualization with PCA (left) and **b** with MDS (right)

of variance shows how much of the origin information is explained by the first two PCs. In literature a cumulated proportion of variance between 60 and 80% is expected to provide a proper decision support for the decision maker [3]. In the example used in this paper a variance of 88.33% of the origin data is explained by the first two PCs. The loadings explain how strong the relationship between the PCs and the origin data in the net flow matrix is.

That relationship is calculated by the correlation coefficient between the PCs and each single criterion. A high/low correlation coefficient explains that the criterion on the PC has a high/low loading. That means the criterion is strongly/weakly represented by the PC. Without the further information of the loadings there are a few things that can be said on the visualization. The decision stick shows that under the preferences of the decision maker alternative a2 is best, followed by a1. Both alternatives have positive characteristics on criteria c1 (cost of purchase) and c5 (range). Alternative a3 shows positive characteristics on the criteria c3 (CO_2 emissions) and c2 (variable costs). But those criteria are conflicting to c1 and c5. The fourth alternative is dominated by all other alternatives and has no positive characteristics on the chosen criteria.

3 Visualization of PROMETHEE Using MDS

By applying the MDS the objects in the configuration represent the a-alternatives. To show how similar the alternatives are, the distances between the objects are essential. Two alternatives that are placed beneath each other possess similar characteristics whereas large distances imply dissimilarities. The most applicable method for the representation of the distances is the Euclidean method because of its good interpretability. To place the alternatives as points in the configuration the distance matrix is transformed to a scalar product matrix [4]. In contrast to the GAIA method, the

criteria are not projected into the configuration by the eigenvectors but by the application of the Property fitting method. Property fitting is a vector model that provides the information which the decision maker gives by evaluating the criteria of the alternatives [5]. The further information are expressed by the inter- and intracriterial preferences. The starting point of the visualisation is the same as in the PCA, the net flow matrix. Due to the pairwise comparison of the alternatives in PROMETHEE the net flow matrix can be interpreted as a similarity matrix and therefore as the starting point for the visualisation with the MDS method. In Fig. 1b the visualization is shown. It can be seen that a1 and a2 are located beneath each other and therefore a possibility of clustering exists. We use the Property Fitting to perform a regression analysis for each single criterion [5]. The two dimensions are visualized as independent variables with which the variance of the characteristics of the criteria is explained. If they are compared to the loadings of the PCA it is obvious that the values are identical. The projection of the criteria is carried out by the weighted vector. The Interpretation is similar to the one in the PCA visualization. The best alternative is a2 followed by a1. If we take a look at alternatives a3 and a4 there is no possibility to form a cluster, as both alternatives have a large distance between them.

4 Procrustes Analysis for Comparing the Two Methods

To compare the result of two or more visualization methods there exist a variety of methods . The best fit for the comparison is the Procrustes Analysis (PA) which shows the similarity of both methods in the least-square-sum criterion [6]. The process of the PA is to take one of the visualizations as the target and to transform the other one with operations like rotating, stretching, shrinking, sliding and mirroring as good as possible into the shape of the target. These operations can be applied and do not change the results of the MCDM model because of the Euclidean metric that is used to set up the visualizations. By applying the PA, no information of the MCDM model gets lost. As it can be seen in Fig. 1a and b the two visualizations are similar. On the first sight the positions of the alternatives seem to be different. But that is due to the two different scales used in the biplot. By comparing the scores of the alternatives it becomes obvious that the alternatives are nearly identical. The result of the PA is the sum of least squares between the scores. Therefore, the smaller the sum the more identical the two visualizations are. For the developed MCDM model and the visualizations with the PCA and the MDS the sum of least squares after the PA is 0.02. That shows both visualizations are similar and almost identical. After comparing the result we now compare the course of action of both methods. The single steps are shown in Fig. 2.

The initial point for GAIA and MDS is the net flow matrix as the result of PROMETHEE. In the first step the PCA transforms the net flow matrix into the covariance matrix to make the relationship of the criteria visible. The MDS starts by converting the net flow matrix to a distance matrix to measure the distance between the alternatives. Secondly the PCA calculates the explained variance of each PC and

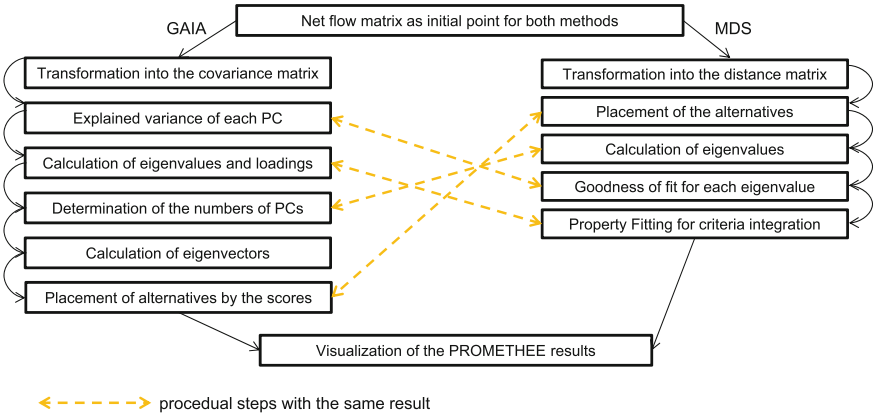


Fig. 2 Methodological comparison of the visualization methods

the cumulative variance whereas in the MDS the eigenvalues are calculated by a spectral decomposition to determine the number of dimensions. Step three is different as well. In the PCA the eigenvalues and loadings are detected to identify how good the single PCs explain the single criteria. The MDS holds placement of the alternatives on the results of step two. The number of PCs ist determined in the fourth step of the PCA whereas the MDS explains the extend of explained variance of the original data. To integrate the criteria into the plane the PCA uses the eigenvectors in step five. The MDS does integrate the criteria by the application of the Property Fitting and the visualization of the MDS is completed. To finish the PCA visualization the sixth step is necessary. The alternatives are placed into the plane by the scores. In the end both methods provide a visualization of the PROMETHEE results that are identical.

5 Summary

In this paper two visualization methods for additive MADM problems were are presentedcompared. Both methods are identical if the results are compared and provide similar information and decision support for the decision maker.

The PCA as originally proposed for PROMETHEE as the GAIA analysis is a transparent and easily interpretable visualization method as long as just the alternatives are projected into the plane. By integrating the criteria in the PCA plane more knowledge of interpretation is demanded by the decision maker. With the loadings it is possible to extract the correlation of single criterions with the Principal Components and analyse them in detail. The interpretation of the MDS visualization is just possible from a subjective point of view. The method produces no values on the relationship between the alternatives and criteria. Only by applying the prop-

erty fitting technique the interpretability of both methods seems to be comparable. The projection of the decision stick visualizes the rank of all alternatives concerning PROMETHEE II. Without the decision stick it is not obvious which alternative is best under the subjective preferences of the decision maker. Different authors indicate that the handling of graphical representations is not a matter of course and especially not native for human beings but has to be learned carefully [7] Graphical representations can easily been misunderstood, further research on the psychological aspects of the perception of evaluation results of MCDM problems is necessary.

References

1. Brans, J.-P., Mareschal, B.: How to select and how to rank projects: The PROMETHEE method. Eur. J. Oper. Res. **24**(2), 228–238 (1986)
2. Mareschal, B., Brans, J.-P.: Geometrical representations for MCDA. Eur. J. Oper. Res. **34**(1), 69–77 (1988)
3. Brans, J.-P., Mareschal, B.: The PROMCALC and GAIA decision support system for multicriteria decision aid. Decis. Support Syst. **12**, 292–310 (1994)
4. Chatfield, C., Collins, A.: Introduction to multivariate analysis. Chapman and Hall, London and New York (1980)
5. Homburg, C., Krohmer, H.: Marketing management: Strategie - Instrumente -Umsetzung - Unternehmensfhrung. Gabler, Wiesbaden (2009)
6. Gower, J., Dijksterhuis, G.: Procrustes problems, Oxford statistical science series. Oxford University Press, Oxford (2004)
7. Weidenmann, B.: Wissenserwerb mit Bildern. Bern, Switzerland, Huber (1994)

Part V
Discrete and Combinatorial Optimization, Graphs and Networks

Energy Efficiency in Extensive IP-Over-WDM Networks with Protection

Andreas Betker, Dirk Kosiankowski, Christoph Lange, Frank Pfeuffer, Christian Raack and Axel Werner

1 Introduction

Telecommunication networks are large nation-wide distributed systems that consume significant amounts of electricity. From a national economy's point of view the sustainability aspect is important to consider, especially in the light of climate change. Furthermore, these networks' power consumption causes considerable energy bills for operators. Therefore, energy efficiency of telecommunication networks has gained increasing interest in research and industry in the recent past.

Conventionally, network capacity is planned and provided based on peak traffic—with capacity reserves included—constantly over time. However, demands show significant temporal fluctuation. Therefore, a promising direction for network energy efficiency improvements is a load-adaptive network operation coupling the provided capacity to the real—varying—traffic demands.

The work presented in this article is focused on the potential of power saving in a nation-wide telecommunication network based on IP-over-DWDM and Ethernet

A. Betker · D. Kosiankowski · C. Lange
Deutsche Telekom AG, Telekom Innovation Laboratories, Deutsche, Berlin
e-mail: andreas.betker@telekom.de

D. Kosiankowski
e-mail: dirk.kosiankowski@telekom.de

C. Lange
e-mail: christoph.lange@telekom.de

F. Pfeuffer · C. Raack · A. Werner (✉)
Zuse-Institut,Zuse, Berlin
e-mail: pfeuffer@zib.de

C. Raack
e-mail: raack@zib.de

A. Werner
e-mail: werner@zib.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_14,

architecture. Earlier work in this direction has been presented, for instance, in [4] and [5]. We refer to [6] for a thorough discussion of relevant literature and [2] for a list of frequently used acronyms. [2] and [3] collect extensive material regarding the power consumption of multilayer networks, which serves as a numerical basis.

In this paper we consider the following question: Given an existing network with state-of-the-art hardware and traffic demands for busy hour as well as low-traffic time, what is the minimal amount of energy that is needed to route the occurring traffic, under the assumption that the routing can be reconfigured according to the situation? To answer this, we solve a suitable mixed integer program to obtain an energy-optimal routing in both the IP-layer, as well as the optical layer.

Assumptions on network topology, hardware, and routing conditions, as well as the demand matrices used in this study are described in detail in the next section. The most notable features we address are the following:

- A realistic, nation-wide network with a core and regional sections.
- Traffic originating at regional nodes, plus peering and international transit traffic.
- Computations based on the power consumption of the complete network equipment, assuming state-of-the-art hardware.
- Network failure protection and diverse routing schemes in the IP domain.

2 Modelling the Problem

In this section we provide a detailed description of the input—network topology and architecture, demand scenarios—and the used mathematical model. We also illuminate practical side constraints, such as routing schemes and network protection.

Network topology and architecture. Basis of our model is a hierarchical nation-wide optical network, which coincides with real existing structures to a great extent; see Fig. 1. The precise equipment we assumed for the computations, as well as its power consumption, is given in Table 1. The network consists of 918 nodes in total and is subdivided into ten regional parts with 898 nodes and the backbone with 20 nodes. The backbone, comprising 29 links, is responsible for long-distance traffic transport. There are always two nodes per city, in different locations, for resiliency purposes. Each regional node is connected to the backbone over one of 180 half rings, where every half ring disjointly connects five regional nodes to both of their corresponding backbone nodes.

In the backbone we assume an IP-over-DWDM architecture as described in detail in [4] and [6] with a partly meshed IP layer, cf. Fig. 1. Backbone nodes are equipped with IP routers and optical cross connects (OXC), where channels can be terminated and handed over to the IP router or optically bypassed. Regional nodes contain Ethernet aggregation switches and reconfigurable optical add/drop multiplexers (ROADM), over which traffic is directly transmitted to the associated backbone node. Channel granularity in the whole network is 10 Gbps, with a 40 channel system in the regional section and 80 channels in the backbone. For the IP layer we either assume

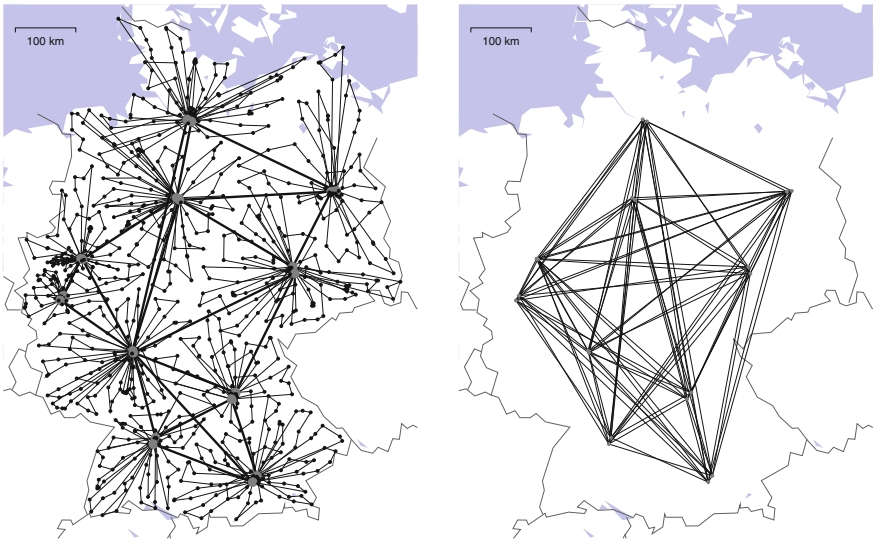


Fig. 1 Used network structure with regional (*thin lines*) and backbone (*thick lines*) sections; physical network (*left*) and partly meshed backbone IP network (*right*)

Table 1 Network equipment and power consumption taken from tables in [2]

Section	Equipment	Table in [2]	Power consumption (W)
Ethernet	Chassis: Juniper EX8216	4,5	1080
Aggregation	Line cards: Juniper EX8200-8XS, 8×10G	4,5	299
WDM	Cisco 15454 MSTP WDM terminal	12	230
Regional	Cisco 15454 MSTP ROADM 40ch	13	436
Aggregation	OLA Cisco 15454 MSTP EDFA	11	200
40λ	Transponder card Cisco 15454 MSTP 10G	9	35
IP/MPLS	Cisco CRS-3 16-slot shelf	2,3	2401
Backbone	Slot/Port Card CRS-3 MSC, 14×10G	3	536
	CRS-1 Fabric Card Shelf	3	8100
WDM	Cisco 15454 MSTP OXC degree N	13	$(150+N\times160)$
Backbone	OLA Cisco 15454 MSTP EDFA	11	200
80λ	Transponder card Cisco 15454 MSTP 10G	9	35

OSPF (open shortest path first), which results in packet streams routed on single IP paths from source to destination, or MPLS (multiprotocol label switching), where traffic may be routed along multiple paths.

To protect against single link failures in the backbone, we assume that every IP link must be realized by two link-disjoint paths in the optical layer. Furthermore, we assume that each half ring maintains enough capacity so that traffic from each regional node can be routed in both directions; additionally, each of the two backbone nodes in a city must be able to take over the complete traffic originating within the

adjacent half rings. In this way, the network is protected against link or node failures in regional rings, but also to some extent against backbone node failures.

Traffic demands. The basis of the traffic model is a combination of the forecasts [7] and [8] and other non-public forecasts. In 2012 about 40 million private and business access lines will generate a peak data flow of 3 Tbps within the regional networks. A share of 90% of the traffic remains in the region, but is managed over a backbone node, the remaining 10% are guided over the backbone to other regions. In addition to customer traffic there are IP peering traffic (1.32 Tbps) and IP transit traffic (0.35 Tbps). Every backbone node serves as both a peering and a transit point, and the distribution of this IP traffic share inside the backbone depends on the peering/transit point utilization.

Traffic fluctuates significantly over a day. Measurements have shown that during night-times the customer traffic demand drops down to 20% of the peak flow, while IP peering/transit traffic drops down to about 25% [1]. For the model a temporal correspondence between the regional and the backbone traffic curves is assumed.

Mathematical model. To compute an energy-optimal routing in the IP-over-WDM backbone network, we solve the following mixed integer program (MIP):

$$\min \quad \text{(total power consumption)} \qquad \sum_{i\in V}(E_P\, p_i + E_L r_i + E_S q_i + E_F w_i) + \sum_{e\in E} E_e z_e$$

$$\text{s.t.} \sum_{\{i,j\}\in L} (f^k_{i,j} - f^k_{j,i}) = \delta^k_i d_k \quad \forall i \in V,\ k \in K, \quad \sum_{\{u,v\}\in E} (g^\ell_{v,u} - g^\ell_{u,v}) = \delta^\ell_v y_\ell \quad \forall v \in V,\ \ell \in L$$

$$\sum_{k\in K}(f^k_{i,j} + f^k_{j,i}) \le C_\ell\, y_\ell \quad \forall \ell = \{i,j\} \in L, \sum_{\ell\in L}(g^\ell_{u,v} + g^\ell_{v,u}) \le C_e\, z_e \quad \forall e = \{u,v\} \in E$$

$$\sum_{\{i,j\}\in L} y_{\{i,j\}} = p_i,\ p_i \le C_P\, q_i, r_i \le C_L w_i + 1 q_i \le C_S\, r_i,\ w_i \le 1 \quad \forall i \in V$$

$$f \ge 0,\ g, z, y, p, q, r, w \ge 0,\ \text{integer}$$

Routing of demands K in the IP layer (graph (V, L), variables f, demand values d) and optical channels in the optical layer (graph (V, E), variables g) is optimized simultaneously, minimizing the power consumed by devices. Variables p_i, q_i, r_i, w_i, z_e count the number of ports, cards, line card and fabric card shelves, and optical equipment on links, respectively; constants C_* and E_* denote the corresponding capacities and energy consumption values. The first two constraints describe the flow in the IP and optical layer, respectively (we set δ^k_i to 1 or -1 if i is source or target of $k \in K$, otherwise $\delta^k_i = 0$, similar for δ^ℓ_v), followed by the IP link and fibre link capacity constraints. The remaining constraints correspond to node equipment capacity. The model can be adapted easily for the protection and single-path scenarios.

Power consumption in the regional sections is computed independently from the backbone. Since no rerouting of traffic in the half rings is possible, no optimization is necessary apart from switching off hardware that provides exceeding capacity.

Table 2 Total network power consumption (W) in different scenarios

Network section	Scenario	Reference	Peak	Night
Backbone	Split IP traffic (MPLS)	382,034	353,086	222,676
	Single-path IP traffic (OSPF)	382,034	361,838	225,660
	Unprotected (OSPF)	284,428	268,172	205,692
Regional			1,861,549	1,847,530

3 Results

Table 2 gives a summary of some results. Stated are total power consumption values for the different network sections under various routing schemes, at peak traffic times and during the night (cf. Sect. 2). In the reference scenario all connections are realized via shortest paths in the graph. The MIP for backbone routing was solved using CPLEX 12.4 and Gurobi 5, with remaining optimality gaps below 1 %. Notice that the table states the power savings in watts, not the energy savings, which would be obtained by integrating the load-adaptive power over time.

The computed power consumption of the backbone network during low-traffic times decreases by about one third compared to peak hours, even by more than 40 % in comparison to the (straight-forward operating) reference network. Furthermore, forcing single-path IP routing increases power consumption only marginally, while protection constraints result in considerably higher increases of about 30 %, with less effect in the low-traffic scenario (below 10 %). Figure 2 illustrates network operations in the single-path scenario. Dark connections indicate higher utilization of links, which is achieved in the optimized solutions.

By far the most power is consumed in the regional sections; at the same time the potential for savings is negligible here. The main power consumers at regional nodes are the Ethernet switches, which cannot be turned off even at low-traffic times. A way to utilize energy saving potential in this situation is to develop switches adapting their base power consumption (e.g., by fans and internal modules) to the traffic load. Alternatively, using finer channel granularity in the regional sections significantly increases the gap between power consumption at low-traffic and peak hours—at the expense of increased CAPEX due to the higher number of ports needed.

Outlook. We plan to extend the work presented in this paper in several directions:

- Consider OTN technology, which is frequently used in backbone networks and can have interesting effects on routing flexibility and power consumption.
- Use more refined demand scenarios, describing traffic over a complete time-period (day/week/month).
- In practice some devices, such as transponders, would rather be left running during low-traffic hours, since they need much time to recalibrate after having been switched on; this results in a more fine-grained power and device modelling.
- Extend the modelling to guarantee full node-protection in the backbone.

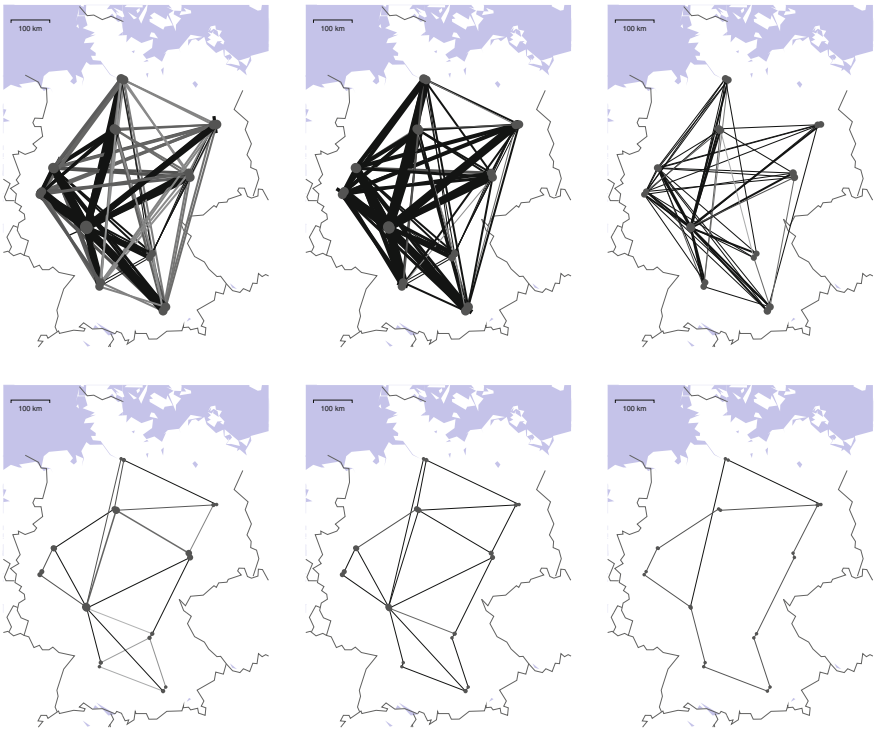


Fig. 2 Backbone IP layer (*above*) and WDM layer (*below*) in the reference network (*left*), and in peak and night-time traffic scenarios (*middle* and *right*); dark connections indicate high utilization

Acknowledgments This work was carried out within the project DESI, supported by the IT2Green programme of the German Federal Ministry of Economics and Technology (BMWi) and the DFG Research Center MATHEON (See www.desi-t2green.de, www.it2green.de, and www.matheon.de.

References

1. German internet exchange traffic statistics, www.de-cix.net/about/statistics (2012).
2. Van Heddeghem, W.,Idzikowski, F.: Equipment power consumption in optical multilayer networks-source data. Technical report, Ghent University, Ghent, (2012). http://powerlib.intec.ugent.be
3. Van Heddeghem, W., Idzikowski, F.,Vereecken, W., Colle, D., Pickavet, M., Demeester, P.: Power consumption modeling in optical multilayer networks. Photonic Network Communications, (2012). published online: 26 Jan 2012.
4. Hülsermann, R., Lange, Ch,. Kosiankowski, D., Gladisch, A.: Analysis of the energy efficiency in IP over WDM networks with load-adaptive operation. In Photonische Netze, vol. 228 of ITG-Fachbericht, pp. 162–168. VDE-Verlag, (2011).

5. Idzikowski, F., Chiaraviglio, L., Portoso, F.: Optimal design of green multi-layer core networks. In Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet, pp. 15:1–15:9. ACM, (2012).
6. Idzikowski, F., Orlowski, S., Raack, Ch., Woesner, H., Wolisz, A.: Dynamic routing at different layers in IP-over-WDM networks—Maximizing energy savings. Optical Switching and Networking, Special Issue on Green Communications **8**(3), 181–200 (2011)
7. Cisco Systems. Cisco visual networking index 2011–2016, 2012. www.cisco.com/web/solutions/sp/vni/vni_forecast_highlights/index.html
8. Deutsche Telekom. Third quarter report, 2011.www.telekom.com/investor-relations/publications/financial-reports/19304

Simultaneous Optimization of Berth Allocation, Quay Crane Assignment and Quay Crane Scheduling Problems in Container Terminals

Necati Aras, Yavuz Türkoğulları, Z. Caner Taşkın and Kuban Altınel

1 Introduction

There has been a considerable growth in the share of containerized trade in the world's total dry cargo during the last 30 years. Therefore, the efficient management of seaport container terminals has become a crucial issue [2]. In this work, we concentrate on the integrated planning of seaside operations, which includes the berth allocation problem (BAP), quay crane assignment problem (CAP) and quay crane scheduling problem (CSP). Generally, BAP deals with the determination of the optimal berthing times and positions of vessels in container terminals. The focus of CSP, on the other hand, is mainly on the problem of determining an optimal handling sequence of vessels for the available cranes at the terminal. However, as can be realized, the assignment of the cranes to vessels has a direct effect on the processing times of the vessels. As a result, crane assignment decisions can be embedded within either BAP or CSP models.

In this work we formulate two new MILP formulations integrating first BAP and CAP (BACAP), and then BAP, CAP, and CSP (BACASP). Both of them consider a continuous berth layout where vessels can berth at arbitrary positions within the range of the quay and dynamic vessel arrivals where vessels cannot berth before the expected arrival time. The crane schedule found by solving the BACASP formulation determines the specific crane allocation to vessels for every time period. These MILP models are the first models solved exactly rather than heuristically in the literature for relatively large instances.

N. Aras (✉) · Y. Türkoğulları · Z. C. Taşkın · K. Altınel
Boğaziçi University, Istanbul, Turkey
e-mail: arasn@boun.edu.tr

Y. Türkoğulları
e-mail: turkogullari@boun.edu.tr

Z. C. Taşkın
e-mail: caner.taskin@boun.edu.tr

K. Altınel
e-mail: altinel@boun.edu.tr

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_15,

2 Model Formulation

The underlying assumptions of our models are given as follows. The planning horizon is divided into equal-sized time periods. The berth is divided into equal-sized berth sections. Each berth section is occupied by no more than one vessel in each time period. Each quay crane can be assigned to at most one vessel per time period. Each vessel has a minimum and maximum number of quay cranes that can be assigned to it. The service of a vessel by quay cranes begins upon that vessel's berthing at the terminal, and it is not disrupted until the vessel departs. The number of quay cranes assigned to a vessel does not change during its stay at the berth, which is referred to as a time-invariant assignment [1]. Furthermore, the set of specific cranes assigned to a vessel is kept the same. By letting i the index of vessels, g the index of crane groups, j the index of berth sections, k the index of number of cranes, t the index of time periods, c_l^g the index of the leftmost crane in group g, c_r^g the index of the rightmost crane in group g, and $C(g)$ the index set of cranes in group g, we define the following parameters: B = the number of berth sections, G = the number of crane groups, N = the number of available quay cranes, T = the number of time periods in the planning horizon, V = the number of vessels, d_i= due time of vessel i, e_i = arrival time of vessel i, $\underline{k}^i$ = lower bound on the number of cranes that can be assigned to vessel i, $\overline{k}^i$ = upper bound on the number of cranes that can be assigned to vessel i, ℓ_i = the length of vessel i measured in terms of the number of berth sections occupied, p_i^k = processing time of vessel i if k cranes are assigned to it, s_i = desired berth section of vessel i, ϕ_{i1} = cost of one unit deviation from the desired berth section for vessel i, ϕ_{i2} = cost of berthing one period later than the arrival time for vessel i, ϕ_{i3} = cost of departing one period later than the due time for vessel i.

Let us define a binary variable X_{ijt}^k, which is equal to one if vessel i starts berthing at section j in time period t, and k quay cranes are assigned to it, and zero otherwise. Constraint (1) ensures that each vessel berths at a unique section and time period, and the number of quay cranes assigned to it lies between the minimum and maximum allowed quantities.

$$\sum_{j=1}^{B-\ell_i+1} \sum_{k=\underline{k}^i}^{\overline{k}^i} \sum_{t=e_i}^{T-p_i^k+1,} X_{ijt}^k = 1 \qquad i = 1, \ldots, V. \tag{1}$$

Constraint set (2) guarantees that each berth section is occupied by at most one vessel in each time period. To put it differently, there should not be any overlap among the rectangles representing vessels in the two-dimensional time-berth section space, which are located between $\max\left(e_i, t - p_i^k + 1\right)$ and $\min\left(T - p_i^k + 1, t\right)$ on the time dimension, and between $\max\left(1, j - \ell_i + 1\right)$ and $\min\left(B - \ell_i + 1, j\right)$ on the berth section dimension.

$$\sum_{i=1}^{V} \sum_{j'=\max(1,j-\ell_i+1)}^{\min(B-\ell_i+1,j)} \sum_{k=\underline{k}^i}^{\overline{k}^i} \sum_{t'=\max(e_i,t-p_i^k+1)}^{\min(T-p_i^k+1,t)} X_{ij't'}^k \le 1 \qquad j=1,\ldots,B; t=1,\ldots,T \tag{2}$$

We next discuss how quay crane availability can be handled in the BACAP model. Let us denote the number of available quay cranes by N. Constraint set (3) ensures that in each time period the number of active quay cranes is less than or equal to the available number of cranes:

$$\sum_{i=1}^{V} \sum_{j=1}^{B-\ell_i+1} \sum_{k=\underline{k}^i}^{\overline{k}^i} \sum_{t'=\max(e_i,t-p_i^k+1)}^{\min(T-p_i^k+1,t)} k X_{ijt'}^k \le N \qquad t=1,\ldots,T \tag{3}$$

The objective function (4) of our model minimizes the total cost, whose components for each vessel are: (1) the cost of deviation from the desired berth section, (2) the cost of berthing later than the arrival time, and (3) the cost of departing later than the due time. Our integer programming formulation for BACAP can be summarized as follows:

$$\min \sum_{i=1}^{V} \sum_{k=\underline{k}^i}^{\overline{k}^i} \sum_{j=1}^{B-\ell_i+1} \sum_{t=e_i}^{T-p_i^k+1} \left\{ \phi_{i1}|j-s_i| + \phi_{i2}(t-e_i) + \phi_{i3}\left(t+p_i^k-1-d_i\right)^+ \right\} X_{ijt}^k \tag{4}$$

subject to constraints (1), (2), (3)

$$X_{ijt}^k \in \{0,1\} \quad i=1,\ldots,V; j=1,\ldots,B-\ell_i+1; k=\underline{k}^i,\ldots,\overline{k}^i; \\ t=e_i,\ldots,T-p_i^k+1.$$

Recall that although the availability of quay cranes is considered in constraint set (3) in BACAP, a schedule is not generated for each quay crane. To develop a mathematical programming formulation for BACASP we extend the formulation for BACAP by including the constraint sets (1)–(3) and defining new variables and constraints so that feasible schedules are obtained for quay cranes, which do not incur setup due to the change in the relative order of cranes. We should remark that if quay cranes $i-1$ and $i+1$ are assigned to a vessel in a time period, then quay crane i has to be assigned to the same vessel as well since quay cranes are located along the berth on a single railway. Hence, we define a *crane group* as a set of adjacent quay cranes and let the binary variable Y_{it}^g denote whether crane group g assigned to vessel i starts service in time period t. Constraint set (5) relates the $\mathbf{X}$ and $\mathbf{Y}$-variables. It ensures that if k quay cranes are assigned to vessel i, it must be served by a crane group g that is formed by $|C(g)| = k$ cranes, where $C(g)$ is the index set of cranes in group

g and $|\cdot|$ denotes the cardinality of a set. Moreover, G is the total number of crane groups.

$$\sum_{j=1}^{B-\ell_i+1} X_{ijt}^k - \sum_{\substack{g=1 \\ |C(g)|=k}}^{G} Y_{it}^g = 0 \quad i = 1,\ldots,V; k = \underline{k}^i,\ldots,\overline{k}^i; t = e_i,\ldots,T-p_i^k+1 \tag{5}$$

It should be emphasized that each crane can be a member of multiple crane groups. However, each crane can operate as a member of at most one group in each time period. The next set of constraints (6) guarantees that this condition holds:

$$\sum_{i=1}^{V} \sum_{\substack{g=1 \\ c\in C(g)}}^{G} \sum_{t=\max(e_i,t-p_i^k+1)}^{\min(T-p_i^k+1,t)} Y_{it'}^g \le 1 \quad c = 1,\ldots,N; t = 1,\ldots,T \tag{6}$$

Even though constraints (5) and (6) make sure that each quay crane is assigned to at most one vessel in any time period, they do not guarantee that quay cranes are assigned to vessels in the correct sequence. In particular, let the quay cranes be indexed in such a way that a crane positioned closer to the beginning of the berth has a lower index. Since all cranes perform their duty along a rail at the berth, they cannot pass each other or stated differently their order cannot be changed. The next four constraint sets help to ensure preserving the crane ordering. Here, Z_{ct} denotes the position of crane c in time period t.

$$Z_{ct} \le Z_{(c+1)t} \quad c = 1,\ldots,N-1; t = 1,\ldots,T \tag{7}$$

$$Z_{Nt} \le B \quad t = 1,\ldots,T \tag{8}$$

$$Z_{c_i^g \bar{t}} + B(1 - Y_{it}^g) \ge \sum_{j=1}^{B-\ell_i+1} \sum_{k=\underline{k}^i}^{\overline{k}^i} j X_{ijt}^k \quad i = 1,\ldots,V; g = 1,\ldots,G; \quad t = e_i,\ldots,T-p_i^k+1; t \le \bar{t} \le t+p_i^k-1 \tag{9}$$

$$Z_{c_i^g \bar{t}} \le \sum_{j=1}^{B-\ell_i+1} \sum_{k=\underline{k}^i}^{\overline{k}^i} (j+\ell_i-1) X_{ijt}^k + B(1 - Y_{it}^g) \quad i = 1,\ldots,V; g = 1,\ldots,G; \quad t = e_i,\ldots,T-p_i^k+1; t \le \bar{t} \le t+p_i^k-1 \tag{10}$$

Constraint set (7) simply states that the positions of the cranes (in terms of berth sections) are respected by the index of the cranes. This means that the position of crane c is always less than or equal to the position of crane $c+1$ during the planning horizon. Constraint set (8) makes sure that the last crane (crane N) is positioned within the berth. By defining c_l^g and c_r^g as the index of the crane that is, respectively, the leftmost and rightmost member of crane group g, constraint set (9) guarantees that if crane group g is assigned to vessel i and vessel i berths at section j, then the position of the leftmost member of crane group g is greater than or equal to j. Similarly, constraint set (10) ensures that if crane group g is assigned to vessel i and vessel i berths at section j, then the position of the rightmost member of crane group g is less than or equal to $j + \ell_i - 1$, which is the last section of the berth occupied by vessel i.

3 Solution

As can be observed, BACASP formulation is significantly larger than our BACAP formulation with which we can solve instances up to 60 vessels. Hence, it should be expected that only small BACASP instances can be solved exactly using CPLEX 12.2. This fact has motivated us to make use of the formulation for BACAP in solving larger sized BACASP instances to optimality. By carefully analyzing the optimal solutions of BACAP and BACASP in small sized instances, we have figured out that an optimal solution of BACASP can be generated from an optimal solution of BACAP provided that the condition given in Proposition 1 is satisfied. This condition is based on the notion of complete sequence of vessels (with respect to their occupied berthing positions), which is defined as follows.

Definition 1 A vessel sequence $v_1, v_2, \ldots, v_n$ is *complete* if (1) v_1 is the closest vessel to the beginning of the berth, (2) v_n is the closest vessel to the end of the berth, (3) v_i and v_{i+1} are two consecutive vessels with v_i closer to the beginning of the berth, and (4) two consecutive vessels in this sequence must be at the berth during at least one time period.

A complete sequence is said to be *proper* when the sum of the number of cranes assigned to vessels in this sequence is less than or equal to N. Otherwise, it is called an *improper* complete sequence.

Proposition 1 *An optimal solution of BACASP can be obtained from an optimal solution of BACAP by a post-processing algorithm if and only if every complete sequence of vessels is proper.*

The proof of this proposition can be found in [3]. If there is at least one improper complete sequence of vessels in an optimal solution of BACAP, then we cannot apply the post-processing algorithm given as Algorithm 1 to obtain an optimal solution of BACASP from an optimal solution of BACAP.

Algorithm 1 Post-processing algorithm

Initialization: Let $\mathcal{V}_{NA} \leftarrow \{1, 2, \ldots, V\}$
WHILE $\mathcal{V}_{NA} \neq \varnothing$
Select vessel $v \in \mathcal{V}_{NA}$ that berths in the leftmost berth section
Find the vessels in $\mathcal{V}_A$ that are in the berth with v in at least one time period. Among the cranes assigned to these vessels, find the crane $c_{\max}$ that is in the rightmost berth section
IF $\mathcal{V}_A = \varnothing$ or $\nexists$ any vessel in $\mathcal{V}_A$ that is at the berth with v in at least one time period

$c_{\max} \leftarrow 0$

ENDIF
Assign cranes indexed from $c_{\max} + 1$ to $c_{\max} + \theta_v$ to vessel v, where θ_v is the number of cranes assigned to vessel v

$\mathcal{V}_{NA} \leftarrow v_{NA} \setminus v$

ENDWHILE

In Algorithm 1, $\mathcal{V}_A$ ($\mathcal{V}_{NA}$) denotes the set of vessels to which cranes (no cranes) are assigned yet. Clearly, $\mathcal{V}_{NA} \cup \mathcal{V}_A = \{1, 2, \ldots, V\}$. Notice that the way the vessels are picked up from $\mathcal{V}_{NA}$ and added to the set $\mathcal{V}_A$ implies that the order of the vessels forms one or more complete sequences in the set $\mathcal{V}_A$. It is also ensured that these complete sequences are proper.

If there exists a complete sequence where the sum of the number of cranes assigned to vessels is larger than N, then it is possible to add the cut given in (11) corresponding to an improper complete sequence into the formulation of BACAP, where IS refers to an improper complete sequence and $|IS|$ is the total number of vessels involved in that complete sequence. Note that this cut is used to eliminate feasible solutions that involve IS.

$$\sum_{i \in IS} X^{k(i)}_{ij(i)t(i)} \leq |IS| - 1 \tag{11}$$

The left-hand side of (11) consists of the sum of the X^k_{ijt} variables which are set to one for the vessels involved in IS. In other words, there is only one $X^k_{ijt} = 1$ for each vessel $i \in IS$. The j, k, and t indices for which $X^k_{ijt} = 1$ related to vessel i are denoted as $j(i)$, $k(i)$, and $t(i)$ in (11). Upon the addition of this cut, BACAP is solved again. The addition of these cuts is repeated until the optimal solution of BACAP does not contain any improper complete sequences. At that instant, Algorithm 1 can be called to generate an optimal solution of BACASP from the existing optimal solution of BACAP.

Acknowledgments We gratefully acknowledge the support of IBM through a open collaboration research award #W1056865 granted to the first author.

References

1. Bierwirth, C., Meisel, F.: A survey of berth allocation and quay crane scheduling problems in container terminals. Eur. J. Oper. Res. **202**, 615–627 (2010)
2. Stahlbock, R., Voß, S.: Operations research at container terminals: A literature update. OR Spectrum **30**, 1–52 (2008)
3. Türokoğulları, Y.B., Aras, N., Taşkın Z.C., Altınel, İ.K.: Simultaneous Optimization of Berth Allocation, Quay Crane Assignment and Quay Crane Scheduling Problems in Container Terminals, Research Paper, http://www.ie.boun.edu.tr/~aras

A Genetic Algorithm for the Unequal Area Facility Layout Problem

Udo Buscher, Birgit Mayer and Tobias Ehrig

1 Introduction

This paper addresses the facility layout problem (FLP) with unequal areas, where a certain number of departments of specified areas has to be allocated on a given rectangular region with fixed dimensions. Since the facility layout and the material handling design have an important influence on the operating costs, minimizing the material handling costs is critical for the facility layout planner. Whereas the problem with equal areas can be modeled easily as a quadratic assignment problem (QAP), this does not hold for the problem with unequal areas. Hereby, the centroids are critical to calculate the distance between areas. Unfortunately, the centroids depend on the exact configuration selected and therefore, corresponding QAP formulations are less tractable than their equal-area counterparts [8].

Two main classes of FLP could be distinguished: Continual Layout Problems allow the placement of departments anywhere within the planar site as long as the departments do not overlap. In this paper the focus is on the Discrete Layout Problem, where the plant size is covered by a grid of rectangular blocks of equal size (sometimes called unit area [1]) in order to overcome the limitations of the original quadratic assignment problem which ignores area requirements. The location of a department may comprise one or more adjacent blocks, so that unequal department areas can be considered. In this paper we follow an approach with space filling curves (SFC) to determine the final layout.

U. Buscher (✉) · B. Mayer · T. Ehrig
Chair of Industrial Management, TU Dresden, Dresden, Germany
e-mail: udo.buscher@tu-dresden.de

B. Mayer
e-mail: birgit.mayer@tu-dresden.de

T. Ehrig
e-mail: tobias.ehrig@mailbox.tu-dresden.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_16,

The paper is organized as follows. In Sect. 2, the concept of SFC is introduced and one representative—the Peano curve—is discussed in detail. This is followed by a short overview of the chosen genetic algorithm in Sect. 3, which determines the placement sequence of departments and some computational results are reported. Section 4 concludes the paper with a discussion on this method's extensions and potential for future work.

2 Peano Curve as Space Filling Curve

We assume a rectangular plant size which is covered by a grid of N unit areas and the space filling curve is constructed in a way that all unit areas are visited without any overlapping [4]. The shape of one department does not need to be a rectangular. Instead, it is sufficient to express the space requirements of a department as an integer number of unit areas. We have to ensure only that the unit areas are adjacent to another unit area that has been assigned to the same department in order to guarantee continuous department areas. The application of the SFC has two advantages. Firstly, a SFC ensures feasible solutions, because the shape requirements of departments are fulfilled automatically. Secondly, since the SFC is fixed the problem reduces to find the optimal sequence of departments. As a consequence the complexity to arrange departments on the plant size is significantly reduced [1]. Let m_j denote the number of unit areas allocated to department j for $j = 1, \ldots, M$, the number of layout alternatives of the problem is reduced from $N!/\prod_{j=1}^{M} m_j!$ to $M!$.

Evidently, the particular choice of the space filling curve has a great impact on the placement of departments and is therefore critical. Traditionally, sweeping, spiral or Hilbert-type patterns are proposed [6]. Following the space filling curve the visited unit squares are filled with the corresponding number of department unit areas according to the department sequence. In Fig. 1 this procedure is illustrated by a small 5 x 4 plant area including 20 unit areas which are connected with a specific SFC. Eight departments each with a given number of department unit areas are arranged on the plant subject to the department sequence H, B, E, F, D, G, A, C. In this example the SFC reduces the number of layout alternatives from $20!/82.944 = 2.93 \cdot 10^{13}$ to 40.320.

To the best of our knowledge one specific SFC – the so-called Peano curve – has not been applied to the FLP so far. The Peano-curve is the first SFC and was

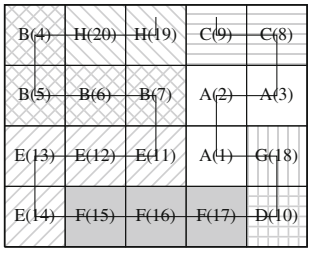


Fig. 1 Unequal area facilty layout with SFC

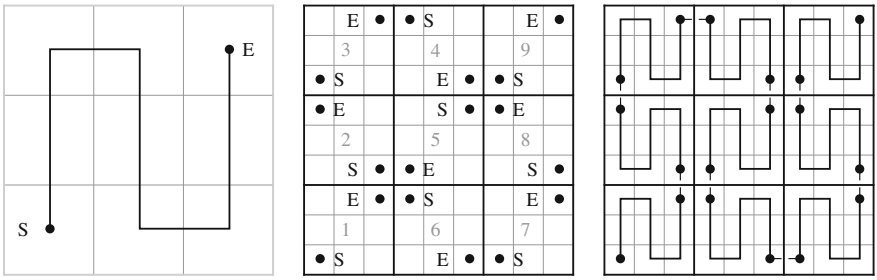


Fig. 2 Construction of the Peano curve

already developed by Giuseppe Peano in 1890 [7]. The principle idea is to connect iteratively the same basic pattern, which is generated on a plant area of 3 x 3 square units. The curve starts in the south-west edge, follows the way shown in the left part of Fig. 2 and ends up in the opposite edge. The basic pattern can be scaled up requiring a plant area of $3^{n-1} \times 3^{n-1}$, with $n \geq 2$. For $n = 3$ the generation of the Peano curve is depicted in the middle and the right part of Fig. 2. Following this basic principle the subsequent basic pattern is connected in the way that the starting point is in the adjacent unit square of that unit square in which the previous basic pattern ends. As a result the starting point is in the south-west edge and the end point is in the opposite (north-east) edge.

In practice the plant floor area does not necessarily correspond with the space requirements of the Peano curve. Therefore, the Peano-curve has to be modified to meet the given space conditions. The construction of the modified Peano curve in pseudo-code is depicted in Algorithm 1.

To illustrate the construction of the peano curve a plant area of 20×12 square units is chosen. Following the algorithm, we set $n = 2$ in this example. Hence, a Peano curve is drawn in the south-west edge of the plant area (see area A' in Fig. 3). Subsequently, area A' is mirrored at the northern edge which leads to area A. To fill the northern part of area A (equals area B), the algorithm starts from the beginning considering only area B. Thereby, the algorithm ensures that the curve ends in the north-east edge of area B. After that the areas A', A, and B are mirrored twice at the eastern edge. In case of a remaining area C, this area is filled again by restarting the algorithm considering area C. To fulfill the continuity of the Peano-curve, area C has to be rotated as long as the curve in area C starts in the north-western edge.

3 Genetic Algorithm and First Numerical Results

The focus of the paper is on the applicability of the Peano curve as a SFC. However, the technique used to determine the order of departments has an important influence on the objective function. The objective is to minimize the material flow factor

Algorithm 1 Modified Peano Curve

```
1: procedure PRK(r × c)                                   ▷ r ≥ c; otherwise rotate plane
2:     find max k, such that (3^k ≤ c)                    ▷ define areas A,A',B,C (Fig. 3(a))
3:     find max m, such that (3^(m−1) ≤ r)
4:     n = min (k, m)
5:     if n = 0 then
6:         if c = 1 then
7:             curve equals straight line return
8:         else
9:             area has a width of 2, apply Hilbert-curve with side of 2^1 return
10:        end if
11:    end if
12:    if n = 1 & r > 6 then                              ▷ avoid unnecessary long-drawn pattern
13:        draw Peano curve with side of 3^n, starting in south-west
14:        co = ⌊r/(2 · 3^n)⌋                             ▷ determine number of mirror images
15:        if r mod (2 · 3^n) = 0 then                    ▷ Ensure sufficient place
16:            co = ⌊r/(2 · 3^n)⌋ − 1                     ▷ determine number of mirror images
17:        end if
18:        mirror basic pattern 2 · co − 1 times to the north
19:        PRF (r − co · 2 · 3^n × 3^n)
20:    else
21:        draw Peano curve with side of 3^(n−1), starting in south-west
22:        mirror A' at the northern edge                 ▷ area A results
23:        PRK (r − 2 × 3^(n−1))                          ▷ recursive call to determine remaining area B
24:        mirror A+B two times at the eastern edge
25:    end if
26:    if c > 3^n then
27:        PRK (r × c − 3^n)                              ▷ determine area C
28:        rotate area C in a way that the starting point is in the north-west
29:    end if
30: end procedure
```

cost (MFFC) between departments. Let C_{ij} denote the transportation cost for a unit material for a unit distance between departments i and j, f_{ij} the material flow from department i to j and d_{ij} the rectilinear distance between the centroids of departments i and j. The objective can be written as follows [9]:

$$MFFC = \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} f_{ij} d_{ij}$$

Without limiting the generality, all parameters C_{ij} are assumed to be one. If the concept of the SFC is used, the task of the genetic algorithm is to determine the order of departments. In this paper we apply Kochhar's so-called heuristically operated placement evolution (HOPE) technique [6]. Since HOPE is mainly adopted without major changes, we limit our considerations to the main characteristics. To transform the above noted cost function into a suitable fitness function, a simple linear relationship is used: $fitness = -k\ MFFC$. In the version applied in this paper (HOPE Large), each population has 100 individuals. The parents for the recombination are selected in proportion to their fitness with a simple roulette wheel. For the recombination a modified Uniform Order-Based Crossover Operator is used. Further, we apply a simple swap-mutation-operator which operates in the way that two genes

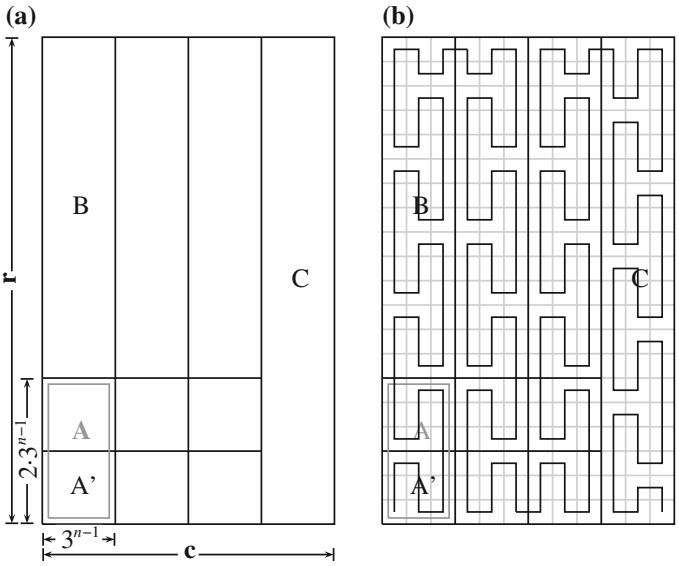


Fig. 3 Construction of the modified Peano curve

on the chromosome are randomly selected and then swapped. Thereby, a constant mutation rate of 0.1 is assumed.

The question arises, whether the Peano curve is a favourable alternative to other SFC. To answer this question first numerical tests are presented. In the following we consider three benchmark instances with different sizes. The first one goes back to Banerjee et al. [3] and was adopted by Islier [5], whereas the second and third instances are introduced by Bozer et al. [4] and Armour and Buffa [2]. The plant sizes and the number of departments which have to be placed can be seen in Table 1.

In order to provide an insight of whether the Peano curve is an interesting alternative to well-established SFC (e.g. Hilbert curve), a Java-application was programmed. Both, Peano and Hilbert curves are included and the genetic algorithm from Kochhar (HOPE) is adopted. To compare the suitability of Peano and Hilbert curves, each test instance was solved 50 times and the best solution as well as the average value are presented. We can see that the best solutions of the test instances one (5,926.60) and three (38,226.30) presented by Wang et al. [9] can be improved (see Table 1). Thereby, the Peano curve performs better than the Hilbert curve. This applies for both the best and the average value. For the second test instance it is exactly the

Table 1 Computational results

Problem	Problem size	Peano/Best	Peano/Ave.	Hilbert/Best	Hilbert/Ave.
Armour/Buffer	$30x20$; $M = 20$	5,713.72	6,146.60	5,809.70	6,210.51
Bozer	$15x15$; $M = 15$	33,270.00	37,466.00	33,190.00	36,487.40
Banerjee/Islier	$19x14$; $M = 12$	37,500.00	39,646.00	38,400.00	40,038.00

opposite. These first numerical results may allow the assumption that the Hilbert curve performs better for plant floors which resemble more or less squares, whereas the Peano curves delivers good results for rectangular plant floors.

4 Conclusions

The first numerical results indicate that (modified) Peano curves could be an interesting alternative to traditionally used SFC in order to solve single-floor facility layout problems with departments of unequal sizes. However, we observe that there are a lot of interesting open questions. First of all, additional numerical tests have to show, for which form of plant layout the Peano curve fits best. Secondly, besides the material handling costs further factors might influence the facility layout. For example the shape ratio factor (SRF) reflects the fact that a more regular individual department often leads to lower cost of the department layout and arrangement compared to a non-regular formed department [9]. This has to be borne in mind especially for the practical realization of the layout. In particular for such problems which consider the SRF it might be interesting to apply more advanced metaheuristics compared to the HOPE-algorithm.

References

1. Al-Araidah, O., Krishnamurthy, A., Malmborg, C.J.: A comparative study of single-phase and two-phase approaches for the layout problem with material handling costs. Int. J. Prod. Res. **45**, 951–970 (2007)
2. Armour, G., Buffa, E.: A heuristic algorithm and simulation approach to relative location of facilities. Manage. Sci. **9**, 294–309 (1963)
3. Banerjee, P., Montreuil, B., Moodie, C.L., Kashyap, R.L.: A modelling of interactive facilities layout designer reasoning using qualitative patterns. Int. J. Prod. Res. **30**, 433–453 (1992)
4. Bozer, Y., Meller, R., Erlebacher, S.: An improvement-type layout algorithm for single and multiple-floor facilities. Manage. Sci. **40**, 918–932 (1994)
5. Islier, A.A.: A genetic algorithm approach for multiple criteria facility layout design. Int. J. Prod. Res. **36**, 1549–1569 (1998)
6. Kochhar, J.S., Foster, B.T., Heragu, S.S.: HOPE: a genetic algorithm for the unequal area facility layout problem. Comput. Oper. Res. **25**, 583–594 (1998)
7. Peano, G.: Sur une courbe, qui remplit toute une aire plane. Mathematische Annalen **36**, 157–160 (1890)
8. Tate, D.M., Smith, A.E.: Unequal-area facility layout by genetic search. IIE Trans. **27**, 465–472 (1995)
9. Wang, M.-J., Hu, M.H., Ku, M.-Y.: A solution to the unequal-area facilities layout problem by genetic algorithm. Comput. Ind. **56**, 207–220 (2005)

A New Theoretical Framework for Robust Optimization Under Multi-Band Uncertainty

Christina Büsing and Fabio D'Andreagiovanni

1 Introduction

A central assumption in classical optimization is that all parameters describing the problem are known exactly. However, many real-world problems consider data that are uncertain or not known with precision (for example, because of measurement methodologies which introduce an error or because of approximated numerical representations). Neglecting the uncertainty may have dramatic effects and turn optimal solutions into infeasible or very costly solutions. Since the groundbreaking investigations by Dantzig [11], many works have thus tried to find effective ways to deal with uncertainty (see [2] for an overview). During the last years, Robust Optimization (RO) has attracted a lot of attention as a valid methodology to deal with uncertainty affecting optimization problems. A key feature of RO is to take into account uncertainty as hard constraints, which are added to the original formulation of the problem in order to cut off solutions that are not *robust*, i.e. protected from deviations of the data. For an exhaustive introduction to the theory and applications of RO, we refer the reader to the book by Ben-Tal et al. [1], to the recent survey by Bertsimas et al. [2] and to the Ph.D. Thesis [6].

An approach to model uncertain data that has been highly successful and has been adapted to several applications is the so-called Γ-scenario set (BS) by Bertsimas and Sim [3]. The uncertainty model for a Linear Program (LP) considered in BS assumes that, for each coefficient a we are given a nominal value $\overline{a}$ and a maximum deviation d and that the actual value lies in the symmetric interval $[\overline{a} - d, \overline{a} + d]$. Moreover,

C. Büsing (✉)
Department of Operations Research, RWTH Aachen University, Kackertstrasse 7,52072Aachen, Germany
e-mail: buesing@or.rwth-aachen.de

F. D'Andreagiovanni
Department of Optimization, Zuse-Institut Berlin (ZIB), Takustrasse 7, 14195 Berlin, Germany
e-mail: d.andreagiovanni@zib.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_17,

a parameter Γ is introduced to represent the maximum number of coefficients that deviate from their nominal value and to control the conservativeness of the robust model. A central result of BS is that, under the previous characterization of the uncertainty set, the robust counterpart of an LP can be formulated as a compact linear problem. However, the use of a single and symmetric deviation band may greatly limit the power of modeling uncertainty, as it becomes evident when the deviation probability sensibly varies within the band: in such a case, if we neglect the inner-band behavior and we just consider the extreme values like in BS, we obtain a rough estimation of the deviations and thus an unrealistic uncertainty set. Reaching a higher modeling resolution would therefore be very desirable, as also highlighted and requested by our industrial partners. This can be accomplished by breaking the single band into multiple and narrower bands, each with its own Γ value. Such model is particularly attractive when historical data on the deviations are available, a very common case in real-world settings. Thus, a multi-band uncertainty set can effectively approximate the shape of the distribution of deviations built on past observations, guaranteeing a much higher modeling power than BS. The multi-band idea was first exploited by Bienstock for the special case of Robust Portfolio Optimization [4]. However, a general definition of the multi-band model applicable also in other contexts and a deep theoretical study of its properties have not yet been done. The main objective of our original study is to fill such a gap.

We remark that, while the present work was under revision, we have refined and extended our results, realizing a new paper [8] that include additional results about dominance among uncertainty scenarios, uncertain Binary Programs and probability bounds of constraint violation.

2 Multi-Band Uncertainty in Robust Optimization

We study the robust counterpart of a Linear Programming Problem (LPP) whose coefficient matrix is subject to multi-band uncertainty. The deterministic Linear Program that we consider is of the form:

$$\begin{aligned} \max & \sum_{j \in J} c_j \, x_j \quad \text{(LPP)} \\ & \sum_{j \in J} a_{ij} \, x_j \leq b_i \quad i \in I \\ & x_j \geq 0 \quad j \in J \end{aligned} \tag{1}$$

where $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$ denote the set of constraints and variable indices, respectively. We assume that the value of each coefficient a_{ij} is uncertain and that such uncertainties are modeled through a set of scenarios $\mathcal{S}$. A scenario $S \in \mathcal{S}$ specifies the deviation d_{ij}^S experienced by each coefficient of the problem

with respect to its nominal value $\overline{a}_{ij}$. The actual value a_{ij} is thus equal to $a_{ij} = \overline{a}_{ij} + d_{ij}^S$. The robust counterpart of (LPP) is the optimization problem providing *robust solutions* that are protected against deviations of a specified scenario set $\mathcal{S}$. A natural formulation of the robust counterpart of (LPP) can be obtained by replacing each constraint (1) with its counterpart which considers the deviations allowed by $\mathcal{S}$, namely $\sum_{j\in J} \overline{a}_{ij}\ x_j + d_{ij}^S\ x_j \leq b_i, i \in I, S \in \mathcal{S}$.

One of the purposes of our work is to characterize the robust counterpart of (LPP) when the set of scenarios corresponds to what we call a *multi-band uncertainty set*. This set is denoted by $\mathcal{S}_M$ and generalizes the Bertsimas-Sim uncertainty model. Specifically, we assume that, for each coefficient a_{ij}, we are given its nominal value $\overline{a}_{ij}$ and maximum negative and positive deviations $d_{ij}^{K^-}$, $d_{ij}^{K^+}$ from $\overline{a}_{ij}$, such that the actual value a_{ij} lies in the interval $[\bar{a}_{ij} + d_{ij}^{K^-}, \bar{a}_{ij} + d_{ij}^{K^+}]$ for each scenario $S \in \mathcal{S}_M$. Moreover, we derive a generalization of the Bertsimas-Sim model by partitioning the single deviation band $[d_{ij}^{K^-}, d_{ij}^{K^+}]$ of each coefficient a_{ij} into K bands, defined on the basis of K deviation values:

$$-\infty < d_{ij}^{K^-} < \cdots < d_{ij}^{-2} < d_{ij}^{-1} < d_{ij}^0 = 0 < d_{ij}^1 < d_{ij}^2 < \cdots < d_{ij}^{K^+} < +\infty.$$

Through these deviation values, we define: (1) a set of positive deviation bands, such that each band $k \in \{1, \ldots, K^+\}$ corresponds to the range $(d_{ij}^{k-1}, d_{ij}^k]$; (2) a set of negative deviation bands, such that each band $k \in \{K^- + 1, \ldots, -1, 0\}$ corresponds to the range $(d_{ij}^{k-1}, d_{ij}^k]$ and band $k = K^-$ corresponds to the single value $d_{ij}^{K^-}$ (the interval of each band but $k = K^-$ is thus open on the left). With a slight abuse of notation, in what follows we indicate a generic deviation band through the index k, with $k \in K = \{K^-, \ldots, -1, 0, 1, \ldots, K^+\}$ and the corresponding range by $(d_{ij}^{k-1}, d_{ij}^k]$.

Additionally, for each band $k \in K$, we define a lower bound l_k and an upper bound u_k on the number of deviations that may fall in k, with $l_k, u_k \in \mathbb{Z}$ satisfying $0 \leq l_k \leq u_k \leq n$. In the case of band k = 0, we assume that $u_0 = n$, i.e. we do not limit the number of coefficients that take their nominal value. We also assume that $\sum_{k\in K} l_k \leq n$, so that there exists a feasible realization of the coefficient matrix.

The robust counterpart of (LPP) under a multi-band uncertainty set defined by $\mathcal{S}_M$ can be equivalently written as:

$$\begin{aligned} \max & \sum_{j\in J} c_j \quad x_j \\ & \sum_{j\in J} \overline{a}_{ij}\ x_j + DEV_i(x, \mathcal{S}_M) \leq b_i \quad i \in I \\ & x_j \geq 0 \quad j \in J, \end{aligned} \tag{2}$$

where $DEV_i(x, \mathcal{S}_M)$ is the maximum overall deviation allowed by the multi-band scenario set $\mathcal{S}_M$ for a feasible solution x when constraint i is considered. Finding the value $DEV_i(x, \mathcal{S}_M)$ can be formulated as a 0-1 linear maximization problem, whose optimal solution defines a distribution of the coefficients among the bands that maximizes the deviation w.r.t. the nominal values, while respecting the bounds on the number of deviations of each band (see [7] for details about this problem). As a consequence, the robust counterpart (2) is actually a max-max problem. However, we prove that (2) can be reformulated as a compact and linear problem (we refer the reader to [7] for the complete proofs of the results presented in this section).

Theorem 1 *The robust counterpart of problem (LPP) under the multi-band scenario set $\mathcal{S}_M$ is equivalent to the following compact Linear Program:*

$$
\begin{aligned}
&\max \sum_{j \in J} c_j \, x_j \quad \textit{(Rob-LP)} \\
&\sum_{j \in J} \bar{a}_{ij} \, x_j - \sum_{k \in K} l_k \, v_i^k + \sum_{k \in K} u_k \, w_i^k + \sum_{j \in J} z_i^j \leq b_i \quad i \in I \\
&- v_i^k + w_i^k + z_i^j \geq d_{ij}^k \, x_j \quad i \in I, j \in J, k \in K \\
&v_i^k, \; w_i^k \geq 0 \quad i \in I, k \in K \\
&z_i^j \geq 0 \quad i \in I, j \in J \\
&x_j \geq 0 \quad j \in J.
\end{aligned}
$$

In comparison to (LPP), this compact formulation uses $2 \cdot K \cdot m + n \cdot m$ additional variables and includes $K \cdot n \cdot m$ additional constraints.

As an alternative to the direct solution of (Rob-LP), we also investigated the possibility of adopting a cutting-plane approach [16]: in this case, given a solution $\overline{x} \in \mathbb{R}^n$ we want to test if $\overline{x}$ is robust feasible, i.e. $a_i^S \overline{x} \leq b_i$ for every scenario $S \in \mathcal{S}_M$ and $i \in I$. Specifically, our solution strategy is the following: we start by solving the nominal problem (LPP) and then we check if the optimal solution is robust. If not, we add a cut that imposes robustness (*robustness cut*) to the problem. This initial step is then iterated as in a typical cutting plane method [16].

In the case of the Bertsimas-Sim model, the problem of separating a robustness cut for a given constraint is very simple and essentially consists in sorting the deviations in increasing-order and choose the worst $\Gamma > 0$ (see [12] for details). In the case of multi-band uncertainty, this simple approach does not guarantee the robustness of a computed solution. However, we prove that the separation can be done in polynomial time by solving a *min-cost flow problem* (see [7] for the the detailed description of how we build the instance of this problem and structure the corresponding proof), as formalized in the following theorem.

Theorem 2 *Let $x \in \mathbb{R}_+^n$ and let $\mathcal{S}_M$ be a multi-band scenario set. Moreover, let $(G, c)_x^i$ be the min-cost flow instance corresponding to a solution x and a constraint*

$i \in I$ of (LPP). The solution x is robust for constraint i w.r.t. $\mathcal{S}_M$ if and only if $\bar{a}_i' x - c_i^*(x) \leq b_i$, where $c_i^*(x)$ is the minimum cost of a flow of the instance $(G, c)_x^i$.

The proof is based on showing the existence of a one-to-one correspondence between the integral flows and the non-dominated feasible solutions of the binary program expressing the maximum deviation allowed by the multi-band uncertainty set.

3 Application to Wireless Network Design

We applied our new theoretical results about Robust Optimization to the design of wireless networks, considering the *Power Assignment Problem* (PAP): this is the problem of dimensioning the power emission of each transmitter in a wireless network, to provide service coverage to a number of users, while minimizing the overall power emission. The PAP is particularly important in the (re)optimization of networks that are updated to new generation digital transmission technologies. For a detailed introduction to the PAP and the general problem of designing wireless networks, we refer the reader to [9, 10, 15].

A classical LP formulation for the PAP can be defined by introducing the following elements: (1) a vector of non-negative bounded continuous variables p that represent the power emissions of the transmitters; (2) a matrix A of the coefficients that represent signal attenuation (*fading coefficients*) for each transmitter-user couple; (3) a vector of r.h.s. δ (signal-to-interference thresholds) that represents the minimum power values that guarantee service coverage. If the objective is to minimize the overall power emission, the PAP can be written as:min $\mathbf{1}' p$ s.t. $Ap \geq \delta$, $0 \leq p \leq P^M$, where exactly one constraint $a_i' p \geq \delta_i$ is introduced for each user i to represent the corresponding service coverage condition.

Each fading coefficient of the matrix A summarizes the different factors which influence propagation (e.g., distance between transmitter and receiver, terrain features) and is classically computed by a propagation model. However, the exact propagation behavior of a signal cannot be evaluated and thus each fading coefficient is naturally subject to uncertainty. Neglecting such uncertainty may lead to unexpected coverage holes in the resulting plans, as devices may be actually uncovered for bad deviations affecting the fading coefficients. For a detailed presentation of the technical aspects of propagation, we refer the reader to [17]. Following the ITU recommendations (e.g., [13]), we assume that the fading coefficients are mutually independent random variables and that each variable is log-normally distributed. The adoption of the Bertsimas-Sim model would provide only a very rough modeling of the deviations associated with such distribution. Contrarily, the multi-band uncertainty model provides a much more refined representation of the fading coefficient deviations. In our computational study, we considered realistic instances corresponding to region-wide networks based on the Terrestrial Digital Video Broadcasting technology (DVB-T) [13] and were taken as reference for the design of the new Italian DVB-T national

network. We built the uncertainty set taking into account the ITU recommendations [13] and discussions with our industrial partners in past projects about wireless network design. For a description of our benchmark instances and the corresponding computational results, we refer the reader to [7]. The main purpose of our tests was to compare the efficiency of solving directly the compact formulation (Rob-LP) with that of a cutting-plane method based on the robustness cuts presented in the previous section. For the Bertsimas-Sim model, such comparison led to contrasting conclusions in past works (e.g., [12, 14]). In our case, we found that the cutting-plane approach produced optimal solutions in less time for the bigger instances, while in all the other cases the compact formulation performed better. Concerning the price of robustness, we noted that imposing robustness with the multi-band model led to a sensible increase in the overall power emission, that was anyway lower than that of the Bertsimas-Sim model in all but two cases. On the other hand, the power increase under multi-band uncertainty was compensated by a very high average protection factor against deviations. The good overall performance encourages further investigations. In particular, future research will be focused on refining the cutting plane method and enlarging the computational experience to other relevant real-world problems.

Acknowledgments This work was partially supported by the *German Federal Ministry of Education and Research* (BMBF), project *ROBUKOM*, grant 03MS616E [5], and by the DFG Research Center MATHEON, Project B23 - *Robust optimization for network applications*.

References

1. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Springer, Heidelberg (2009)
2. Bertsimas, D., Brown, D., Caramanis, C.: Theory and applications of robust optimization. SIAM Rev. **53**(3), 464–501 (2011)
3. Bertsimas, D., Sim, M.: The price of robustness. Oper. Res. **52**(1), 35–53 (2004)
4. Bienstock, D.: Histogram models for robust portfolio optimization. J. Comp. Finance **11**(1), 1–64 (2007)
5. Bley, A., D'Andreagiovanni, F., Hanemann, A.: Robustness in Communication Networks: Scenarios and Mathematical Approaches. In: Proc. of the ITG Symposium on Photonic Networks 2011, pp. 10–13. VDE Verlag, Berlin (2011)
6. Büsing, C.: Recoverable Robustness in Combinatorial Optimization. Ph.D. Thesis, Technische Universität Berlin, Berlin (2011)
7. Büsing, C., D'Andreagiovanni, F.: New Results about Multi-band Uncertainty in Robust Optimization. In: Klasing, R. (ed.) Experimental Algorithms - SEA 2012, LNCS, vol. 7276, pp. 63–74. Springer, Heidelberg (2012)
8. Büsing, C., D'Andreagiovanni, F.: Robust Optimization under Multi-band Uncertainty. Submitted for publication (2012)
9. D'Andreagiovanni, F.: Pure 0–1 Programming approaches to wireless network design. 4OR-Q. J. Oper. Res. (2011) doi:10.1007/s10288-011-0162-z
10. D'Andreagiovanni, F., Mannino, C., Sassano, A.: GUB covers and power-indexed formulations for wireless network design. Management Sci. (2012). doi:10.1287/mnsc.1120.1571
11. Dantzig, G.: Linear programming under uncertainty. Manage. Sci. **1**, 197–206 (1955)
12. Fischetti, M., Monaci, M.: Cutting plane versus compact formulations for uncertain (integer) linear programs. Math. Prog. C **4**(3), 239–273 (2012)

13. International Telecommunication Union (ITU): Terrestrial and satellite digital sound broadcasting to vehicular, portable and fixed receivers in the VHF/UHF bands (2002)
14. Koster, A.M.C.A., Kutschka, M., Raack, C.: Robust Network Design: Formulations, Valid Inequalities, and Computations. ZIB Tech. Rep. 11–34, ZIB Berlin, Berlin, Germany (2011)
15. Mannino, C., Rossi, F., Smriglio, S.: The network packing problem in terrestrial broadcasting. Oper. Res. **54**(6), 611–626 (2006)
16. Nemhauser, G., Wolsey, L.: Integer and Combinatorial Optimization. Wiley, Hoboken (1988)
17. Rappaport, T.S.: Wireless Communications: Principles and Practice, 2nd edn. Prentice Hall, Upper Saddle River (2001)

Relevant Network Distances for Approximate Approach to Large p-Median Problems

Jaroslav Janacek and Marek Kvet

1 Introduction

The p-median problem represents the core problem for designing the optimal structure of most public service systems [2, 7, 8]. The available commercial IP-solvers usually fail in the cases of mediate instances described by the location-allocation model [5]. At least two approximate approaches have been developed to avoid introducing the vast structure of allocation variables for each pair of the possible facility location and customer. Both approaches are based on the so called radius model. In the model a customer is not assigned to a specific facility location but the model only recognizes whether some facility is located in the given radius from the customer. The first approach [3] forms the particular finite system of radii for each customer according to the distances from the customer to possible facility locations. The second approach [4] discussed in this paper forms one common system of radii based on so called dividing points, which are determined in accordance to the estimated distance relevance. The deployment of dividing points impacts the deviation of the approximate objective function from the original objective function of the p-median problem. The dividing points can be determined so that the expected deviation is minimized [5, 6] and the deviation is expressed using so called distance value relevance. The distance value relevance expresses the strength of expectation that the distance belongs to the optimal solution of the p-median problem. There are several ways to express the distance value relevance. In previous works [5, 6] we based our research on the idea that the distance value relevance dropped exponentially with the distance value. In this contribution we base our approach to the relevance estimation

J. Janacek (✉) · M. Kvet
Faculty of Management Science and Informatics, University of Zilina, Univerzitna 8215/1,Zilina010 26, Slovak Republic
e-mail: Jaroslav.Janacek@fri.uniza.sk

M. Kvet
e-mail: Marek.Kvet@fri.uniza.sk

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_18,

on so called column ranking, where not only value but also the order of the distance is taken into account. We suggest the formulation of the ranking relevance and study the influence of the associated dividing point deployment on the effectiveness of the approximate approach.

2 Radial Model of the p-Median Problem

To formulate the p-median problem, we denote the set of serviced customers by J and the set of possible facility locations by I. It is necessary to determine at most p locations from I so that the sum of the network distances from each element of J to the nearest located facility is minimal. The distance between the possible location i and the customer j from J is denoted as d_{ij}. The general radial model of the p-median problem can be formulated by further introduced decision variables. The variable $y_i \in \{0, 1\}$ models the decision about the facility location at the place $i \in I$. The variable takes the value of 1 if the facility is located at i, and it takes the value of 0 otherwise.

The approximate approach is based on the relaxation of the unique assignment of a facility location to a customer. In the radius approach, the distance between the customer and the nearest facility is approximated, unless the facility must be assigned. Each customer j from J is provided with the increasing sequence of distance values $0 = D_{j0}, D_{j1}, \ldots, D_{jr(j)}, D_{jr(j)+1} = max\{d_{ij} : i \in I\}$ and zero-one constant a_{ij}^k for each triple $[i, j, k]$ where $i \in I$, $j \in J$ and $k \in \{0, \ldots, r(j)\}$ is introduced. The constant a_{ij}^k is equal to 1, if the distance between the customer j and the possible location i is less or equal to D_{jk}, otherwise a_{ij}^k is equal to 0. In addition, the auxiliary zero-one variable x_{jk} for $k = 0, \ldots, r$ is introduced. The variable takes the value of 1, if the distance of the customer j from the nearest located facility is greater than D_{jk}, and this variable takes the value of 0 otherwise. The zone k of the customer j corresponds with the interval $(D_{jk}, D_{jk+1}]$. The width of the k-th interval is denoted by e_{jk} for $k = 0, \ldots, r(j)$. Then the expression $e_{j0}x_{j0} + e_{j1}x_{j1} + \cdots + e_{jr}x_{jr}$ constitutes the upper approximation of d_{ij}. If the distance d_{ij} belongs to the interval $(D_{jk}, D_{jk+1}]$, it is estimated by the upper bound D_{jk+1} with a possible deviation e_{jk}. Then a covering-type model can be formulated as follows:

$$\text{Minimize} \sum_{j \in J} \sum_{k=0}^{r(j)} e_k x_{jk} \tag{1}$$

$$\text{Subject to:} \quad x_{jk} + \sum_{i \in I} a_{ij}^k y_i \geq 1 \qquad \text{for } j \in J \text{ and } k = 0, \ldots, r(j) \tag{2}$$

$$\sum_{i \in I} y_i \leq p \tag{3}$$

$$x_{jk} \geq 0 \qquad \text{for } j \in J \text{ and } k = 0, \ldots, r(j) \tag{4}$$

$$y_i \in \{0, 1\} \qquad \text{for } i \in I \tag{5}$$

In this model, the objective function (1) gives the upper bound of the sum of the original distances. The constraints (2) ensure that the variables x_{jk} are allowed to take the value of 0, if there is at least one center located in the radius D_{jk} from the customer j. The constraint (3) limits the number of the located facilities by p.

The first approach [3] forms the particular finite system of $D_{j0}, D_{j1}, \ldots, D_{jr(j)+1}$ for the customer j by the selection from the j-th column of the matrix $\{d_{ij}\}$. The second approach [4] forms one common system of dividing points $D_0, D_1, \ldots, D_{r+1}$, which are determined in accordance to the estimated relevance, where the dividing points are deployed in the range of distance values contained in $\{d_{ij}\}$.

3 Relevance Based Deployment of Dividing Points

Obviously, the number r of dividing points $D_1, D_2, \ldots, D_r$ influences the size of the covering model (1)–(5) which concerns either the number of variables x_{jk} or the number of constraints (2). That is why the number r must be kept in a mediate extent to achieve the resulting solution quickly enough. On the other hand, the smaller is the number of dividing points, the bigger inaccuracy afflicts the approximate solution. The elements of the distance matrix $\{d_{ij}\}$ form the finite ordered sequence of values $d_0 < d_1 < \cdots < d_m$, where $D_0 = d_0$ and $D_m = d_m$. Let the value d_h have the frequency N_h of its occurrence in the matrix $\{d_{ij}\}$ reduced by the $p - 1$ largest distances from each column. The distance d between a customer and the nearest located center can only be estimated taking into account that it belongs to the interval $(D_k, D_{k+1}]$ given by the pair of dividing points. Let us denote $D_k^1, D_k^2, \ldots, D_k^{r(k)}$ the values of the sequence $d_0 < d_1 < \cdots < d_m$, which are greater than D_k and less then D_{k+1}. The maximal deviation of the upper estimation D_{k+1} from the exact value d is $D_{k+1} - D_k^1$. If we were able to anticipate the frequency n_h of each d_h in the optimal solution, we could minimize the deviation using the dividing points obtained by solving the following problem:

$$\text{Minimize} \sum_{t=1}^{m} \sum_{h=1}^{t} (d_t - d_h) n_h z_{ht} \tag{6}$$

$$\text{Subject to:} \quad z_{(h-1)t} \leq z_{ht} \qquad \text{for } t = 2, \ldots, m \text{ and } h = 2, \ldots, t \tag{7}$$

$$\sum_{t=h}^{m} z_{ht} = 1 \qquad \text{for } h = 1, \ldots, m \tag{8}$$

$$\sum_{t=1}^{m-1} z_{tt} = r \tag{9}$$

$$z_{ht} \in \{0, 1\} \qquad \text{for } t = 1, \ldots, m \text{ and } h = 1, \ldots, t \tag{10}$$

If the distance d_h belongs to the interval ending with d_t, then the variable z_{ht} takes the value of 1. The link-up constraints (7) ensure that the distance d_h belongs to the interval ending with d_t, only if each distance between d_{h-1} and d_t belongs to this interval. The constraints (8) assure that each distance d_h belongs to some interval, and the constraint (9) enables only r dividing points. After the problem (6)–(10) is solved, the nonzero values of z_{tt} indicate the distances d_t corresponding with the dividing points.

4 Distance Value Relevancies

The former approach to the estimation of the relevance follows the hypothesis that the relevance decreases exponentially with the increasing distance value [6]. The hypothesis can be mathematically expressed by (11), where T is a positive parameter and N_h is the above-mentioned d_h occurrence frequency.

$$n_h = N_h e^{\frac{-d_h}{T}} \tag{11}$$

Hereby we present a new approach to the relevance, where the column ranking evaluation $L_j^{ts}(d_{ij})$ of the distance d_{ij} is used to define the relevance n_h according to

$$n_h = L^{tsh} = \sum_{j \in J} \sum_{\substack{i \in I \\ d_{ij} = d_h}} L_j^{ts}(d_{ij}) \tag{12}$$

The linear column ranking function $L_j^{ts}(d_{ij})$ is defined as follows: let $P_j(d_{ij})$ be the position of d_{ij} in the ascending sequence of the jth column items of the distance matrix $\{d_{ij}\}$ and let a denote the cardinality of J. Then $L_j^{ts}(d_{ij}) = a + s * (1 - P_j(d_{ij}))$ for $P_j(d_{ij}) < a + 1 - t$ and $L_j^{ts}(d_{ij}) = 0$ otherwise. The parameters t and s represent a threshold and a step respectively. The threshold influences the number of $t - 1$ largest distances of the j-th column, which are not taken into account, and the step gives the difference between the contributions of k-th and $k - 1$th item of the ascending sequence of the jth column items. The parameter t can vary over the range $[p..a - 1]$

of integers, and the step s can take values from the interval $[0, a/(a-t)]$. At the end of this section we introduce the third approach to the relevance, which combines both exponential and ranking approaches. The associated relevance n_h is defined in accordance with (13).

$$n_h = L^{tsh} e^{\frac{-d_h}{T}} \tag{13}$$

5 Numerical Experiments

We performed the sequence of numerical experiments to test the effectiveness of the suggested relevancies applied to the approximate method for the p-median problem. The solved instances of the problem were obtained from OR-Lib set of the p-median problem instances [1]. This pool of benchmarks was enlarged by the instances obtained from the Slovak road network for the cardinality of I above 900 possible facility locations [5]. The cardinality of J is the same as the cardinality of I, and the maximal number p of the located facilities was set at the quarter cardinality of I. An individual experiment is organized so that the initial sequence of the dividing points is obtained according to the model (6)–(10), where the relevance n_h of the distance d_h is computed in accordance to the formulae (12) or (13) for the particular setting of parameters t, s and T. Then the associated constants a_{ij}^k and e_{jk} are defined from the computed sequence of r dividing points, and the problem (1)–(5) is solved. The resulting values of y_i are used in (14) to obtain the real objective function value of the resulting p-median problem solution.

$$\sum_{j \in J} min\{d_{ij} : i \in I, y_i = 1\} \tag{14}$$

All experiments were performed using the optimization software FICO Xpress 7.2 (64-bit, release 2011). The associated code was run on a PC equipped with the Intel Core i7 2630 QM processor with the parameters: 2.0 GHz and 8 GB RAM. The preliminary experiments were performed for $T = 1$ and $s = a/(a-t)$, which causes the maximal difference between the contributions L_j^{ts} of the neighboring distances in the ascending sequence of the column items. The threshold t was set at one quarter, one half and three quarters of its possible range, i.e. at values $t_1 = (3p+a-1)\ div\ 4$, $t_2 = (p+a-1)\ div\ 2$ and $t_3 = (p+3a-3)\ div\ 4$ respectively. The number r of the dividing points was set to the value of 20.

The obtained results for the relevance computed according to (12) and (13) are reported in Table 1, which contains a gap in percent between the real objective function of the approximate solution and the exact one. The exact solution value was taken as hundred percent. The computational times necessary for the individual instance solving vary from 1 to 56 s for the largest instance.

Table 1 Gaps in % of the exact solution-relevance computed according to (12) and (13)

	$\|I\|$	100	200	300	400	500	800	1100	1400	1700
	t_1	6.32	4.21	5.45	5.39	5.02	12.52	84.06	112.41	133.35
(12)	t_2	5.76	2.92	4.93	2.59	2.76	7.33	61.37	84.67	83.25
	t_3	5.03	2.39	3.07	1.48	2.21	2.29	32.37	31.74	31.47
	t_1	15.13	9.17	2.20	0.53	0.12	0.00	0.00	0.00	0.00
(13)	t_2	17.67	9.17	2.20	0.41	0.12	0.00	0.00	0.00	0.00
	t_3	17.67	9.17	2.20	0.41	0.12	0.00	0.00	0.00	0.00

6 Conclusion

We have presented the ranking method of the distance relevance estimation in the approximate approach to the p-median problem. The performed preliminary numerical experiments show that the ranking method constitutes a promising way of solving large instances of the p-median problem, especially when combined with the former exponential reduction. As far as the ranking estimation without the exponential reduction is concerned, it can be found that it performs better, only if the instance size does not exceed three hundred possible locations but it fails when larger instances are solved. This phenomenon deserves further research, and its explanation will possibly open the door to further approaches based on ranking.

Acknowledgments This work was supported by the research grants VEGA 1/0296/12 Public Service Systems with Fair Access to Service and APVV-0760-11 Designing of Fair Service Systems on Transportation Networks.

References

1. Beasley, J.E.: OR library: Distributing test problems by electronic mail. J. Oper. Res. Soc. **41**(11), pp. 1069–1072 (1990)
2. Current, J., Daskin, M., Schilling, D.: Discrete network location models. In: Drezner Z (ed) et al. Facility Location. Applications and Theory, pp. 81–118. Springer, Berlin (2002)
3. Garcia, S., Labbe, M., Marin, A.: Solving large p-median problems with a radius formulation. INFORMS J. Comput. **23**(4), 546–556 (2011)
4. Janacek, J.: Approximate Covering Models of Location Problems. In: Proceedings of the International Conference ICAOR 08, Vol. 1: Yerevan. Armenia 2008, 53–61 (Sept 2008)
5. Janacek, J., Kvet, M.: Sequential zone adjustment for approximate solving of large p-median problems. In: OR proceedings, : selected papers, Aug 30–Sept 2, 2011. Springer, Zurich, Switzerland, Berlin Heidelberg (2011). ISBN 978-3-642-29209-5, pp. 269–274 (2012)
6. Janacek, J., Kvet, M.: Dynamic Zone Adjustment for Approximate Approach to the p-Median Problem. In: ORP3, : conference proceedings, July 16–20 2012, pp. 555–562. Linz, Austria (2012)
7. Janosikova, L.: Emergency medical service planning. In: Commun. Sci. Lett. Univ. Zilina **9**(2), pp. 64–68 (2007)
8. Marianov, V., Serra, D.: Location Problems in the Public Sector. In: Drezner Z (ed.) et al. Facility Location. Applications and Theory, pp. 119–150. Springer, Berlin (2002)

Impact of Utility Function to Service Center Location in Public Service System

Michal Kohani, Lubos Buzna and Jaroslav Janacek

1 Introduction

To design the public service systems, such as the fire brigade system or the emergency medical system [2], [3], [5] and [7], the weighted p-median model is usually used, in which the total social costs of served customers are minimized. The customer's disutility is often considered to be proportional to the distance between the customer and the nearest located service center. In our numerical experiments the customers are represented by municipalities populated by inhabitants. Therefore we assign a weight corresponding to the number of inhabitants to each municipality. In general, this problem can be seen as an example of the resource allocation problem, with a central planner regardless the fact that not all customers have the same access to the service. The public service system design is then obtained by solving the associated weighted p-median problem, where p service centres are located at nodes of the transportation network, so that the sum of weighted distances from the aggregate customers to the nearest service centres is minimal. The solution of this problem gives us a so-called system optimum. Nevertheless, some customers from sparsely populated dwelling places may consider this solution to be unfair because their weight is small and, consequently, their distance from the nearest service center is much bigger than the average distance in the service system. A common opinion is that the public system should serve all involved customers, and it should consider some equity of customers in the service accessibility. Therefore a solidarity criterion

M. Kohani (✉) · L. Buzna · J. Janacek
Faculty of Management Science and Informatics, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia
e-mail: kohani@frdsa.fri.uniza.sk

L. Buzna
e-mail: Lubos.Buzna@fri.uniza.sk

J. Janacek
e-mail: Jaroslav.Janacek@fri.uniza.sk

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_19,

should be applied to obtain fairer public service design [1].We have formulated the classical model of the public service system design problem, which minimizes overall social costs and introduce the associated model of the public service system design problem with the min-max fair criterion. Then we will discuss possible definitions of utility function in public service systems, and suggest two utility functions enabling us to use tools of the fairness theory. These preliminary steps will be followed by the proposal of solving techniques for both, the system and fair approach. At the end of the paper we will focus on the impact of used utility functions on the service centres location. We will compare the distribution of the centres to former approaches, and explore their impact on the price of fairness.

2 Public Service System Design Problem

The public service system design problem belongs to the family of location problems [2]. With these problems, it must be decided where to locate the service center. The served region consists of dwelling places situated in the nodes of the transportation network. These dwelling places form a finite set J. The number of inhabitants at the node $j \in J$ is denoted as b_j. The public service system design problem can be formulated as the decision about the location of at most p service centres at some places from the set I of their possible locations, so that the value of chosen criterion is minimal. Let j be a served node and i be the possible location of a service center, then d_{ij} denotes the distance between i and j. The distance can be used to express the disutility of a customer from node j, if the nearest service center is located at i. Then the total discomfort of all inhabitants situated at location j is expressed as $c_{ij} = d_{ij} \cdot b_j$. A model of the total discomfort minimization is as follows:

$$\text{Minimize } C = \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \tag{1}$$

subject to

$$\sum_{i \in I} z_{ij} = 1 \qquad \text{for } j \in J \tag{2}$$

$$z_{ij} \leq y_i \qquad \text{for } i \in I \text{ and } j \in J \tag{3}$$

$$\sum_{i \in I} y_i \leq p \tag{4}$$

$$z_{ij} \in \{0, 1\} \qquad \text{for } i \in I \text{ and } j \in J \tag{5}$$

$$y_i \in \{0, 1\} \qquad \text{for } i \in I \tag{6}$$

A decision on a service center location at place $i \in I$ is modelled by the zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1 if the center is located at i. The variable

$z_{ij} \in \{0, 1\}$ assigns the node j to the possible center location $i (z_{ij} = 1)$. Using the above introduced decision variables and some of the formulated constraints, the model of the first-level min-max public system design problem known as the p-center problem can be described as follows:

$$\text{Minimize } h \tag{7}$$

subject to

$$\sum_{i \in I} d_{ij} z_{ij} \leq h \qquad \text{for } j \in J \tag{8}$$

$$h \geq 0 \tag{9}$$

and constraints (2–6).

An optimal solution of (7–9) and (2–6) gives an upper bound of discomfort imposed on the worst situated inhabitants. The discomfort of these inhabitants cannot be diminished but there is a possibility to improve the discomfort for other customers. In the following steps the model can be repeatedly used together with fixing some location variables y_i and reducing set J by some customers whose utility cannot be further improved. When p locations are fixed a generalized min-max fair solution is obtained. For more detailed description of the solving, see the references [4] and [6].

3 Utility and Disutility in Public Service Systems

Traditionally, social costs are estimated by the direct distance between a service center and a customers-community, and multiplied by some customers' weight (e.g. the size of the population). This way, the type of disutility rather than utility is employed. No natural notion of utility is known for the case of the public service system. The utility for the individual participant of the public service system is matter of their individual feelings or attitude to the distributed service. In the following definitions we are trying to design such utility functions which may possibly reflect the boundary opinions of average inhabitants. Nevertheless, they must comply with the theoretical properties of utility as it was suggested in previous works [1, 8], i.e. the utility for each participant must be positive, the utilities of various participants can be added and possibly they should be of the same range for individual customers. Let $\{d_{ij}\}$ denote the matrix of distances between dwelling places and possible locations. We denote d_i^{max} the maximal distance in j-th column of matrix and similarly d_i^{min} denotes the minimal column distance. Then (10) and (11) correspond to the utility for an individual participant that j assigns to the service center located at i.

$$u_{ij} = d_j^{max} - d_{ij} + \varepsilon \tag{10}$$

$$u_{ij} = \frac{d_j^{max} - d_{ij}}{d_j^{max} - d_j^{min}} + \varepsilon \tag{11}$$

The associated disutilities are expressed by (12) and (13) respectively.

$$u_{ij} = d_{ij} - d_j^{min} + \varepsilon \tag{12}$$

$$u_{ij} = \frac{d_{ij} - d_j^{min}}{d_j^{max} - d_j^{min}} + \varepsilon \tag{13}$$

We include a small positive constant ϵ in each definition to avoid zero utilities. This will turn out practical also later when comparing these results with a proportionally fair solution which requires positive utilities.

4 Numerical Experiments

We have performed numerical experiments to test the suggested utility and disutility functions. Hereby we are presenting the results for four instances of the problem, selected from three sets of real topological data. The number of customers in each set ranges from 150 to 450. The number p_{max} varies from 3 to 15.

In the resulting tables we show the solutions obtained for the dataset ($ZA300$), which has $|I| = 30$ possible locations of service centres and $|J| = 300$ locations of

Table 1 Numerical experiments—$ZA300$

$\|p_{max}\| = 3$					$\|p_{max}\| = 5$				
Function	(10)	(11)	(12)	(13)	Function	(10)	(11)	(12)	(13)
SYSTEM	349399	349399	353026	349508	*SYSTEM*	305659	305659	313026	307833
FAIR	352969	353437	353905	353437	*FAIR*	307475	333491	320905	316875
PoF[%]	1.02	1.16	0.25	1.12	*PoF*[%]	0.59	9.11	2.52	2.93
Time[*s*]	1.36	10.02	7.89	9.72	*Time*[*s*]	2.42	9.19	2.18	9.75

Table 2 Numerical experiments—$ZA300$.

$\|p_{max}\| = 11$					$\|p_{max}\| = 15$				
Function	(10)	(11)	(12)	(13)	Function	(10)	(11)	(12)	(13)
SYSTEM	294574	295618	311513	307033	*SYSTEM*	286404	291815	302332	286528
FAIR	297332	310235	316502	310235	*FAIR*	295091	295964	309755	295729
PoF[%]	0.94	4.94	1.6	1.04	*PoF*[%]	3.03	1.42	2.46	3.21
Time[*s*]	2.48	10.44	1.99	11.76	*Time*[*s*]	2.9	13.51	2.0	15.74

customers. Both approaches presented in Sect. 2 were used to obtain the system and fair optimum for each test instance. Instead of d_{ij} in the model (1–6) and in the model (7–9) and (2–6) the utility or disutility u_{ij} was used. The associated optimization process was converted to maximization, when utilities (10) and (11) were used. This way, four solutions were obtained, corresponding to definitions (10–13). Each instance and each obtained solution was evaluated in terms of 'social costs', i.e. in accordance to the objective function (1). These values are reported in Tables 1 and 2 in the rows denoted as '$SYSTEM$' and '$FAIR$'. The number 'p_{max}' denotes maximum number of located centres and row denoted as '$Time$' is the computational time of the model (7–9) and (2–6) in seconds.

The relation between system optimum and fair optimum is expressed by so-called price of fairness (PoF) and this value in % is given in the row 'PoF'. For utility functions (10) and (11) the price of fairness is calculated by (14).

$$PoF = \frac{FAIR - SYSTEM}{SYSTEM} \tag{14}$$

All experiments were performed on a PC with Intel Core i7, 3 GHz processor and 8 GB RAM using the IP-solver XPRESS Optimization Suite 7.1.

5 Conclusion

We have presented various approaches to design the utility or disutility functions in solving the public service system design problem. We performed numerical experiments to test suggested utility and disutility functions on real topological data.

When comparing the utility functions (10) and (11) according to the natural objective function with '$SYSTEM$' criterion, in Tables 1 and 2 it can be seen that we obtain slightly better solutions using (10) than (11). In the cases of the disutility functions (12) and (13) we obtain slightly better solutions using (13) than (12).

When we compare the utility functions (10) and (11) to the the disutility functions (12) and (13) from the viewpoint of the natural objective function with the '$SYSTEM$' criterion, it seems to be better to use the utility functions for solving this type of problems.

Furthermore, this algorithm enables to study the price of fairness which must be paid for applying the fair solution instead of the system utility solution. The price of fairness compares higher costs for building the public service system with the fair criterion only unilaterally. It might be an interesting idea for further research to evaluate these costs in relation to a measure expressing the increase of fairness which is obtained for these costs.

Acknowledgments This work was supported by the research grant VEGA 1/0296/12 *Public Service Systems with Fair Access to Service* and APVV-0760-11 *Designing of Fair Service Systems on Transportation Networks*

References

1. Bertsimas, D., Farias, V.F., Trichakis, N.: The price of fairness. Oper. Res. **59**(1), 17–31 (2011)
2. Current, J., Daskin, M., Schilling, D., et al.: Discrete network location models. In: Drezner, Z. (ed.) Applications and Theory, pp. 81–118. Springer, Berlin (2002)
3. Janacek, J., Linda, B., Ritschelova, I.: Optimization of municipalities with extended competence selection. Prager Econ. Pap. Q. J. Econ. Theor. Policy **19**(1), 21–34 (2010)
4. Janacek, J., Kohani, M.: Waiting time optimization with IP-solver. Commun. Sci. Lett. Univ. Zilina **12**(3A), 199–206 (2010)
5. Janosikova, L.: Emergency medical service planning. Commun. Sci. Lett. Univ. Zilina **9**(2), 64–68 (2007)
6. Kohani, M., Buzna, L., Janacek, J.: Proportionally fairer service system design. In: Proceedings of International Conference on Operations Research. IFOR, Zurich, 2011
7. Marianov, V., Serra, D., et al.: Location problems in the public sector. In: Drezner, Z. (ed.) Facility Location. Applications and Theory, pp. 119–150. Springer, Berlin (2002)
8. Nace, D., Doan, L.N., Klopfenstein, O., Balshllari, A.: Max-min fairness in multicommodity flows. Comput. Oper. Res. **35**, 557–573 (2008)

How Does Network Topology Determine the Synchronization Threshold in a Network of Oscillators?

Lubos Buzna, Sergi Lozano and Albert Díaz-Guilera

1 Introduction

The electrical power grid attained considerable attention of the complex systems community due to its topological properties, cascading effects, self-organized criticality as plausible explanation for the distribution of the blackouts size, its interaction with other systems and also synchronization phenomena. A power system is formed by a large number of generators and loads interconnected in a complex pattern to fulfil its function. Considering the complete set of variables and characterizing the system in detail on a large scale would require a significant modelling effort and a still unavailable computer power. Here we follow the research direction of reducing the complexity of individual elements in favour of the simplicity in capturing the synchronization phenomena. This way we aim for a compromise between the technical perspective (electrical engineering) involving very detailed models and aÂ more statistical view (network science), where the main goal is to understand the global behaviour of the system.

L. Buzna (✉)
Department of Transportation Networks, University of Zilina,
Univerzitna 8215/5, 01026 Zilina, Slovakia
e-mail: buzna@frdsa.uniza.sk

S. Lozano
IPHES, Institut Català de Paleoecologia Humana i Evolució Social,
43007 Tarragona, Spain
e-mail: slozano@iphes.cat

A. Díaz-Guilera
Departament de Física Fonamental, Universitat de Barcelona,
08028 Barcelona, Spain
e-mail: albert.diaz@ub.edu

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_20,

2 Oscillatory Dynamics

We consider simple dynamical behaviour of generators (e.g. not including complete dynamical description of electromagnetic fields and flow equations) considering only basic mechanisms that ensure one of their most important dynamical properties, namely synchronization. We assume a set of oscillating units organized in a graph $\mathcal{G}$ composed from a set of nodes N and from a set of links L. Each node $i \in N$ is characterized by a natural frequency ω_i and by a phase angle φ_i. Adopting assumptions and deductions introduced in [1] and neglecting inertia component we model synchronization dynamics by the following set of Kuramoto equations [2]:

$$\frac{d\varphi_i}{dt} = \omega_i + \sigma \sum_j a_{ij} \sin(\varphi_j - \varphi_i), \tag{1}$$

where a_{ij} are the elements of the adjacency matrix and σ is parameter characterizing the coupling strength of links which are assumed to be uniform across the entire network. In order to distinguish between generators and loads, we split the nodes into two subsets. The elements in the first subset, $N_+ \subseteq N$, represent generators and thus $\omega_i = \omega_+ > 0$ for $i \in N_+$. Similarly, loads are represented by the subset $N_- \subseteq N$, where $N_- \cup N_+ = N$, $N_- \cap N_+ = \emptyset$ and $\omega_i = \omega_- < 0$ for $i \in N_-$.

From the synchronization point of view, it is relevant how Eq. (1) behaves with respect to the parameter σ. When $\sigma = 0$ the movement of all oscillators is driven by the natural frequency ω_i. As σ grows, the oscillators influence each other more and more until, for a high enough value, they end up synchronizing their frequencies. The smallest value of σ for which this synchronization can occur is called a critical (threshold) value (σ_c) and, following the analytical deductions described in [3, 4], it can be estimated as:

$$\sigma_c = \max_S \frac{\left|\sum_{i \in S} \omega_i\right|}{\sum_{i \in S, j \notin S} a_{ij}}, \tag{2}$$

where $S \subset N$ is non empty subset.

3 Mathematical Formulation of the Optimization Problem

The problem of identifying the subset of nodes maximizing Eq. (2) can be efficiently solved by using binary linear programming. First, we formalise it as a combinatorial optimization problem defined as follows: we decide which nodes are and are not included in the subset S. Such decisions can be modelled by a set of binary variables $y_i \in \{0, 1\}$ for $i \in N$. The variable $y_i = 1$ iff $i \in S$ and otherwise $y_i = 0$. In the denominator of Eq. (2), we consider the number of links connecting nodes belonging to the set S with nodes which do not belong to the set S. To count them, we introduce for each link connecting nodes i and j a binary variable $x_{i,j} \in \{0, 1\}$. The variable

$x_{i,j} = 1$ iff either $i \in S$ and $j \in N - S$ or $j \in S$ and $i \in N - S$ and $x_{i,j} = 0$ otherwise. Then we can formulate the following combinatorial optimization problem:

$$\text{Maximize } f = \frac{\sum_{i \in N} \omega_i y_i}{\sum_{(i,j) \in L} x_{i,j}} \tag{3}$$

subject to

$$\sum_{i \in N} y_i \geq 1 \tag{4}$$

$$x_{i,j} \geq y_j - y_i \qquad \text{for } (i, j) \in L \tag{5}$$

$$x_{i,j} \geq y_i - y_j \qquad \text{for } (i, j) \in L \tag{6}$$

$$y_i,\ x_{i,j} \in \{0,\ 1\} \qquad \text{for } i \in N, (i, j) \in L \tag{7}$$

The constraint (4) ensures that the set S is non-empty. The set of constraints (5) and (6) makes sure that for each link connecting nodes i and j variable $x_{i,j} = 1$ if $y_i = 1$ and $y_j = 0$ or when $y_i = 0$ and $y_j = 1$. Note that the objective function (3) is forcing variables $x_{i,j}$ to take value of zero whenever it is possible. Therefore, $x_{i,j} = 0$ if $y_i = 0$ and $y_j = 0$ or when $y_i = 1$ and $y_j = 1$.

This way, we obtain an optimization problem with the non linear objective function (3) and the linear set of Eqs. (4)–(6). This problem can be easily turned into a feasibility problem, which is linear. Then for the value σ_c, it can be found arbitrarily tight lower σ_l^A and upper σ_u^A bounds by the bisection method. Finally, the interesting

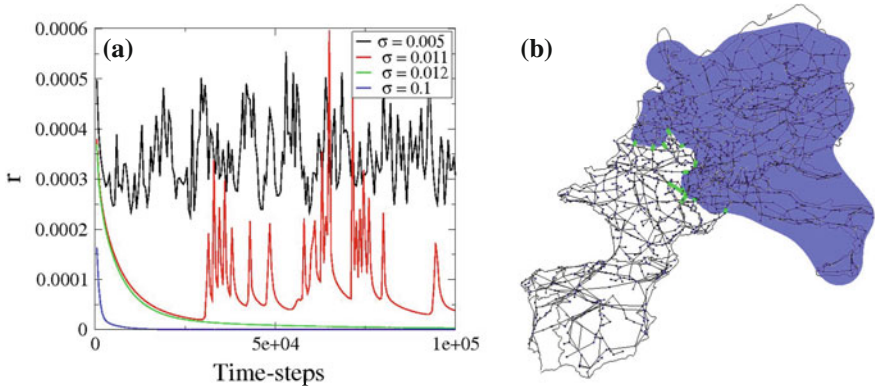


Fig. 1 **a** Temporal evolution of the frequency synchronization in the European network for different coupling values. The frequency dispersion r is used to show the effect of coupling σ below (0.005 and 0.011) and above (0.012 and 0.1) the critical value. All simulations start with the phases set to 0. As the dynamics are purely deterministic, only one realization per coupling value was needed. **b** The *shadowed area* corresponds to the partition defining the critical value of coupling (σ_l^A) when the European network is complete, and the links marked in *green* connect it to the rest of the network. Notice that the complementary partition (i.e. the whole network except the shadowed area) presents the same critical coupling, so either of the two could be considered as critical

side-product of this calculation is the network partition defined by the subset of nodes S, which topologically determines the critical value σ_c.

4 Numerical Experiments

The goal of our numerical experiments was twofold. Firstly, we tested the goodness of the estimate of critical coupling obtained by solving the model (3)–(6) by putting it to the contrast with numerical simulations. Secondly, we did some numerical experiments to investigate the sensitivity of the σ_c value to the removal of single links.To fulfil these tasks, we used approximate model of the European interconnected power system [5]. We started by solving the system of differential equations (1) using lsoda solver implemented in the scipy library [6]. Then, we measured the order parameter r defined by Eq. (8), which enables us to recognize whether the system is synchronized or not. Figure 1a shows four examples of temporal evolution of the order parameter r. If σ is large enough ($\sigma = 0.012$ and $\sigma = 0.1$) r relaxes towards zero indicating frequency synchronization. Using numerical calculations we measured σ_c^N values listed in Table 1.

Table 1 Network size and the values of σ_c^N, σ_A^l and σ_A^u for the whole European high-voltage electrical network and the national sub-networks. σ_c^N was obtained numerically from simulations, as described in the main text. Values σ_A^l and σ_A^u have been calculated by solving the model (3)–(6) by the XPRESS solver in time t_{XPRESS} using PC equipped with processor Intel Core(TM) 2CPU 6700, 2.66 GHz and 3 GB RAM

Country Name	N	L	σ_c^N	σ_l^A	σ_u^A	t_{XPRESS} [s]
Europe	1254	1943	0.0119	0.00818157	0.00818253	56.4344
Austria	36	429	0.1125	0.112499	0.1125	0.0992
Belgium	22	21	0.116305	0.116666	0.116667	0.0619
Croatia	17	20	0.251064	0.25	0.250001	0.0414
Czech Republic	34	52	0.090916	0.090909	0.09091	0.1013
Denmark	8	8	0.465897	0.466666	0.466667	0.0484
France	318	519	0.034278	0.034482	0.034483	5.2009
Germany	229	313	0.035254	0.0294113	0.0294123	1.0567
Hungary	27	36	0.166107	0.166666	0.166667	0.0658
Italy	139	204	0.047949	0.0476074	0.0476084	0.7216
Luxembourg	3	2	1.0	0.999999	1	0.0494
Netherlands	22	24	0.301843	0.290598	0.290599	0.0569
Poland	99	140	0.053808	0.043478	0.043479	0.3019
Portugal	24	44	0.196379	0.188889	0.18889	0.0609
Slovakia	25	30	0.166107	0.166666	0.166667	0.0659
Slovenia	8	8	0.333091	0.333333	0.333334	0.0416
Spain	193	316	0.030372	0.03125	0.031251	1.3661
Switzerland	47	76	0.115329	0.115384	0.115385	0.1450

$$r = \sqrt{\frac{1}{N}\sum_{i\in N}[\frac{d\varphi_i}{dt} - \langle\omega\rangle]^2}. \tag{8}$$

In order to find lower and upper bounds (σ_l^A and σ_u^A) for our analytical estimate of σ_c given by Eq. (2), we solve the problem (3)–(6) using universal solver XPRESS [7]. Comparing these values with σ_c^N values, we find very good agreement (see Table 1). This outcome indicates that Eq. (2) is good estimator of the critical coupling σ_c. Therefore, it can help to simplify studies of frequency synchronization in networks without the necessity to solve iteratively the system of differential equations (1).

We used Eq. (2) to study the effect of link removals on the synchronization stability. Figure 1b shows the sssnetwork partition determining the σ_A^l value for the complete network. Then for each one of the links e, we take it out from the network, recalculate $\sigma_A^l(e)$ and put it back to its position. As shown in Fig. 2a only few links in the whole network have an impact on the global σ_c value. Among them, a minority decreases it

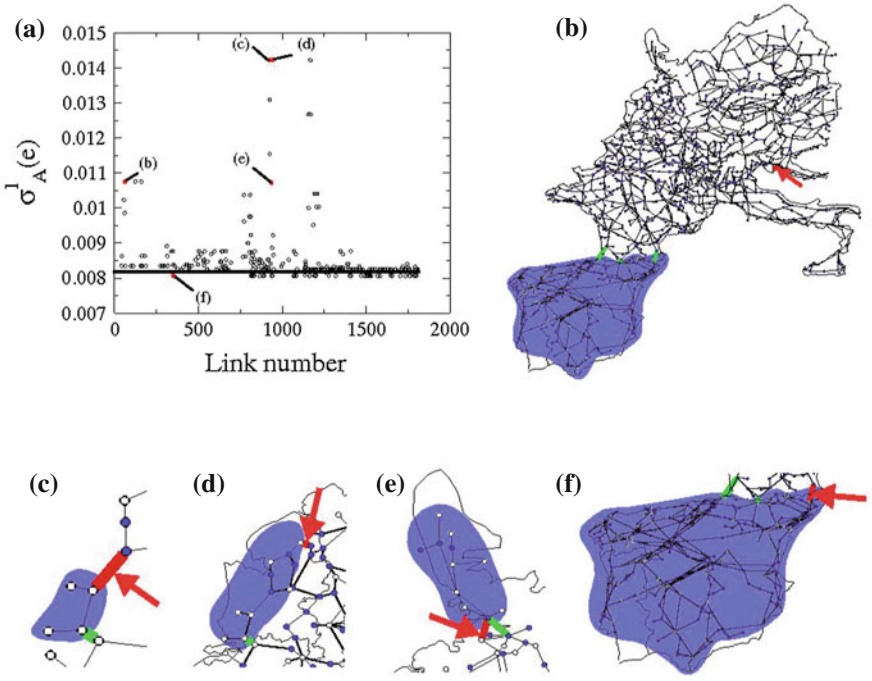


Fig. 2 **a** Impact on the frequency synchronization caused by removing each edge of the European network. New critical coupling values $\sigma^l{}_A(e)$ were obtained by solving problem (3)–(6). The network is robust to removal of all but very few edges. **(b–f)** Examples of link removals modifying the critical partition and value σ_A^l. In cases (c)–(f), the link removal increases σ_A^l, while it is reduced in the case (b). For all cases, the removed links are marked in red and indicated with an arrow. The new critical partition is marked as a shadowed area, and the links connecting it to the rest of the network are marked in green.

(i.e. makes the synchronization easier), while the rest have the opposite effect when removed.These results can be explained by analysing the local structure around the removed links. Figures 2b–f provide detailed view of the local structure for five selected links the deletion of which changes σ_c. Figure 2b illustrates the case when the σ_c is reduced. Here the erased link was connecting a leaf node to the network and, therefore, its removal reduces the numerator in Eq. (2) for the critical partition. Figures 2c–f show the cases where the link deletion increases the coupling strength. Here the deletion of the link reduces the number of connections between the critical partition and the rest of the network, and, consequently, the denominator in Eq. (2) is reduced.

5 Conclusions

We studied the system of dynamical equations which can be interpreted as a simple model of synchronization in an electrical power grid. We tested the analytical formula describing the minimal value of the parameter representing the coupling strength which is required to ensure frequency synchronization. This formula represents a combinatorial optimization problem. By numerical experiments we assessed how good is the estimate of the synchronization threshold obtained by solving this optimization problem, and we investigated the impact of link removals on its value.

Acknowledgments L.B. was supported by projects VEGA 1/0296/12 and APVV-0760-11.

References

1. Filatrela, G., Nielsen, A., Pedersen, N.: Analysis of a power grid using a Kuramoto-like model. Eur. Phys. J. B. **61**(4), 485–491 (2008)
2. Kuramoto, Y.: Chemical Oscillations, Waves, and Turbulence. Springer-Verlag, New York (1984)
3. Buzna, L., Lozano, S., Díaz-Guilera, A.: Synchronization in symmetric bipolar population networks. Phys. Rev. E **80**(6), 066120 (2009)
4. Lozano, S., Buzna, L., Díaz-Guilera, A.: Role of network topology in the synchronization of power systems. Eur. Phys. J. B. **85**(7), 231–238 (2012)
5. Zhou, Q., Bialek, J.: Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades. IEEE Trans. Power Syst. **20**(3), 782–788 (2005)
6. http://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.odeint.html
7. http://www.fico.com

Approximating Combinatorial Optimization Problems with Uncertain Costs and the OWA Criterion

Adam Kasperski and Paweł Zieliński

1 Preliminaries

Let $E = \{e_1, \ldots, e_n\}$ be a finite set of elements and let $\Phi \subseteq 2^E$ be a set of *feasible solutions* composed of some subsets of the element set E. In a deterministic case, each element $e_i \in E$ has some nonnegative cost c_i and we seek a solution $X \in \Phi$ whose total cost $F(X) = \sum_{e_i \in X} c_i$ is minimal. We will denote such a deterministic problem by $\mathscr{P}$. This formulation describes a wide class of combinatorial optimization problems. In particular, the important class of network problems, where E is a set of arcs of a given graph $G = (V, E)$ and Φ contains some object in G such as paths, trees, matchings or cuts, can be viewed in this setting. In many applications, more than one cost vector associated with the elements of E may be specified. This is the case, for example, under uncertainty, where the element costs are not known in advance and a set of all possible cost vectors which may appear with a positive probability is a part of the input.

Let $\Gamma = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_K\}$ contain $K \geq 1$ vectors of the element costs, called *scenarios*, where $\boldsymbol{c}_j = (c_{1j}, \ldots, c_{nj})$ for $j = 1, \ldots, K$. Now $F(X, \boldsymbol{c}_j) = \sum_{e_i \in X} c_{ij}$ is the cost of solution X under scenario $\boldsymbol{c}_j$. In order to choose a solution, we need to apply some criterion which allows us to aggregate the K possible values of the cost function into a single one. In this paper, we use one of the most popular criterion in decision theory, called *Ordered Weighted Averaging* aggregation operator (shortly OWA) proposed in [9]. Let us introduce weights $w_1, \ldots, w_K \in [0, 1]$ such that $w_1 + w_2 + \cdots + w_K = 1$. Given $X \in \Phi$, let $\sigma = (\sigma(1), \ldots, \sigma(K))$ be a permutation

A. Kasperski (✉)
Institute of Industrial Engineering and Management, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: adam.kasperski@pwr.wroc.pl

Paweł Zieliński
Faculty of Fundamental Problems of Technology, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: pawel.zielinski@pwr.wroc.pl

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_21,

of $[K] = \{1, \ldots, K\}$, such that $F(X, \boldsymbol{c}_{\sigma(1)}) \geq \cdots \geq F(X, \boldsymbol{c}_{\sigma(K)})$. Then the OWA operator is defined as follows:

$$\text{OWA}(X) = \sum_{j \in [K]} w_j F(X, \boldsymbol{c}_{\sigma(j)}).$$

It is worth pointing out that using some particular weight distributions in $[0, 1]$, we can obtain some well known aggregation operators. Let us list some of them. If $w_1 = 1$ and $w_j = 0$ for $j = 2, \ldots, K$, then $\text{OWA}(X) = \max_{j \in [K]} F(X, \boldsymbol{c}_j)$. If $w_K = 1$ and $w_j = 0$ for $j = 1, \ldots, K-1$, then $\text{OWA(X)} = \min_{j \in [K]} F(X, \boldsymbol{c}_j)$. If $w_j = 1/K$ for $j = 1, \ldots, K$, then $\text{OWA}(X) = (1/K) \sum_{j \in [K]} F(X, \boldsymbol{c}_j)$ is simply the average cost. If $w_1 = \alpha$, $w_K = 1 - \alpha$ and $w_j = 0$ for $j = 2, \ldots, K-1$, $\alpha \in [0, 1]$, then $\text{OWA}(X)$ is the well known Hurwicz criterion.

Our goal, in this paper, is to find a solution $X \in \Phi$ that minimizes $\text{OWA}(X)$. We will denote such problem by OWA $\mathscr{P}$. The complexity of OWA $\mathscr{P}$ depends on the way in which the weights $w_1, \ldots, w_K$ are specified. The complexity for various weight distributions will be investigated in the next section.

2 Complexity and Approximation Results

Consider first the case when $K = 2$. Then it is well known that most classical problems, such as: the shortest path, minimum spanning tree, minimum assignment and selecting items are NP-hard when $w_1 = 1$ and $w_2 = 0$, i.e. when the OWA operator becomes the maximum [2, 6]. On the other hand, it is easy to verify that the case of uniform weights, i.e. $w_1 = w_2 = 1/2$ is polynomially solvable, if only the deterministic problem $\mathscr{P}$ can be solved in polynomial time. It is thus interesting to check how the complexity of problem $\mathscr{P}$ changes when the weights are slightly different from the uniform distribution. We prove the following result:

Theorem 1 *Let $K = 2$. Then for any $\varepsilon \in (0, 1/2]$ and weights $w_1 = 1/2 + \varepsilon$, $w_2 = 1/2 - \varepsilon$, the* OWA SHORTEST PATH *problem is weakly NP-hard, even for series-parallel graphs.*

Proof We show a reduction from the PARTITION problem, in which: we are given a collection of positive integers $A = (a_1, \ldots, a_n)$ such that $\sum_{i=1}^{n} a_i = 2S$. We ask if there is a subset $I \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in I} a_i = S$. Given an instance of PARTITION, we construct a graph shown in Fig. 1.

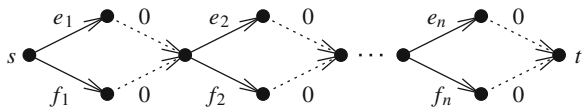


Fig. 1 The graph in the reduction. The dummy (*dashed*) arcs have zero costs under $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$.

We also form two scenarios. Under the first scenario $\boldsymbol{c}_1$, the costs of the arcs $e_1, \ldots, e_n$ are $a_1, \ldots, a_n$ and the cost of all the remaining arcs are 0. Under the second scenario $\boldsymbol{c}_2$, the costs of the arcs $f_1, \ldots, f_n$ are $a_1, \ldots, a_n$ and the costs of all the remaining arcs are 0. We claim that the answer to PARTITION is yes if and only if there is a path X from s to t such that $\mathrm{OWA}(X) \leq S$. Indeed, if the answer is yes, then we form the path X by choosing arcs e_i for $i \in I$ and f_i for $i \notin I$ and complete it by dummy arcs. Then $F(X, \boldsymbol{c}_1) = F(X, \boldsymbol{c}_2) = S$ and $\mathrm{OWA}(X) = S$. On the other hand, suppose that the answer to PARTITION is no. Then for each path X either $F(X, \boldsymbol{c}_1) = S_1 > S$ or $F(X, \boldsymbol{c}_2) = S_2 > S$. Assume that the first case holds (the second one is symmetric). Then $F(X, \boldsymbol{c}_2) = 2S - S_1$ and $\mathrm{OWA}(X) = (\frac{1}{2} + \varepsilon)S_1 + (\frac{1}{2} - \varepsilon)(2S - S_1) = S + 2\varepsilon(S_1 - S)$ and so $\mathrm{OWA}(X) > S$ since $S_1 > S$ and $\varepsilon > 0$. □

By modifying the reductions proposed in [2, 3, 6], we can obtain similar results for other problems such as: the selecting items, minimum spanning tree or minimum assignment.

Proposition 1 *Let $\alpha \geq 1$ and X, Y be solutions such that $F(Y, \boldsymbol{c}_j) \leq \alpha F(X, \boldsymbol{c}_j)$ for all $j \in [K]$. Then* $\mathrm{OWA}(Y) \leq \alpha\, \mathrm{OWA}(X)$.

Proof Let σ and ρ be two permutations of $[K]$ such that $F(Y, \boldsymbol{c}_{\sigma(1)}) \geq \cdots \geq F(Y, \boldsymbol{c}_{\sigma(K)})$ and $F(X, \boldsymbol{c}_{\rho(1)}) \geq \cdots \geq F(X, \boldsymbol{c}_{\rho(K)})$. If we prove that $F(Y, \boldsymbol{c}_{\sigma(j)}) \leq \alpha\, F(X, \boldsymbol{c}_{\rho(j)})$ for all $j \in [K]$, then the proposition follows from the definition of the OWA criterion. Suppose, by contradiction, that $F(Y, \boldsymbol{c}_{\sigma(j)}) > \alpha\, F(X, \boldsymbol{c}_{\rho(j)})$ for some $j \in [K]$. If there is $\boldsymbol{c}$ such that $\boldsymbol{c} \in \{\boldsymbol{c}_{\sigma(1)}, \ldots, \boldsymbol{c}_{\sigma(j)}\}$ and $\boldsymbol{c} \in \{\boldsymbol{c}_{\rho(j+1)}, \ldots, \boldsymbol{c}_{\rho(K)}\}$, then $F(Y, \boldsymbol{c}) \geq F(Y, \boldsymbol{c}_{\sigma(j)}) > \alpha\, F(X, \boldsymbol{c}_{\rho(j)}) \geq \alpha\, F(X, \boldsymbol{c})$, a contradiction. Hence, each scenario which is in $\{\boldsymbol{c}_{\sigma(1)}, \ldots, \boldsymbol{c}_{\sigma(j)}\}$ must also be in $\{\boldsymbol{c}_{\rho(1)}, \ldots, \boldsymbol{c}_{\rho(j)}\}$ and, in consequence, $\{\boldsymbol{c}_{\sigma(1)}, \ldots, \boldsymbol{c}_{\sigma(j)}\} = \{\boldsymbol{c}_{\rho(1)}, \ldots, \boldsymbol{c}_{\rho(j)}\}$. Thus, in particular $\boldsymbol{c}_{\rho(j)} \in \{\boldsymbol{c}_{\sigma(1)}, \ldots, \boldsymbol{c}_{\sigma(j)}\}$ and $F(Y, \boldsymbol{c}_{\rho(j)}) \geq F(Y, \boldsymbol{c}_{\sigma(j)}) > \alpha\, F(X, \boldsymbol{c}_{\rho(j)})$, a contradiction. □

We now apply Proposition 1 to construct a fully polynomial time approximation scheme (FPTAS) for some particular cases of OWA $\mathscr{P}$, when K is a constant. Let $\varepsilon > 0$ and $P_\varepsilon(\Phi)$ be the set of solutions such that for all $X \in \Phi$, there is $Y \in P_\varepsilon(\Phi)$ such that $F(Y, \boldsymbol{c}_j) \leq (1 + \varepsilon)\, F(X, \boldsymbol{c}_j)$ for all $j \in [K]$. We recall the definition of *exact problem* associated with problem $\mathscr{P}$ [7]. Given a vector $(v_1, \ldots, v_k)$. We ask if there is a solution $X \in \Phi$ such that $F(X, \boldsymbol{c}_j) = v_j$ for all $j \in [K]$. Basing on results obtained in [8], it was shown in [7] that if the exact problem associated with problem $\mathscr{P}$ can be solved in pseudopolynomial time, then for any $\varepsilon > 0$, the set $P_\varepsilon(\Phi)$ can be determined in time polynomial in the input size and $1/\varepsilon$. This implies the following result:

Theorem 2 *If the exact problem associated with problem $\mathscr{P}$ can be solved in pseudopolynomial time, then* OWA $\mathscr{P}$ *admits an FPTAS.*

Proof Let Y be a solution of the minimal value of $\mathrm{OWA}(Y)$ among all the solutions in $P_\varepsilon(\Phi)$. From results given in [7, 8], it follows that we can find Y in time polynomial

in the input size and $1/\varepsilon$. Furthermore, for any solution $X \in \Phi$, it holds $F(Y, \boldsymbol{c}_j) \leq (1+\varepsilon)\, F(X, \boldsymbol{c}_j)$ for all $j \in [K]$. So, by Proposition 1, $\text{OWA}(Y) \leq (1+\varepsilon)\,\text{OWA}(X)$ for all $X \in \Phi$. We have thus obtained an FPTAS for OWA $\mathscr{P}$. □

It turns out that the exact problem associated with problem $\mathscr{P}$ can be solved in pseudopolynomial time for many particular problems $\mathscr{P}$, when the number of scenarios K is a constant. This is the case, for example, for the shortest path, minimum spanning tree and some other problems described in [1].

We now give an approximation algorithm whose idea is to aggregate the costs of each element $e_i \in E$ by using the OWA operator and compute an optimal solution for the aggregated costs. Consider element $e_i \in E$ and let $\hat{c}_{i1} \geq \hat{c}_{i2} \geq \cdots \geq \hat{c}_{iK}$ be the ordered costs of e_i. Let $\hat{c}_i = \sum_{j\in[K]} w_j \hat{c}_{ij}$ be the aggregated cost of e_i and $\hat{C}(X) = \sum_{e_i \in X} \hat{c}_i$. Let $\hat{X}$ be a solution minimizing $\hat{C}(X)$. Of course, $\hat{X}$ can be obtained in polynomial time if only $\mathscr{P}$ is polynomially solvable. We now wish to explore how far from the optimal value, the value $\text{OWA}(\hat{X})$ may be. Assume that the weights are nonincreasing, i.e. $w_1 \geq w_2 \geq \cdots \geq w_K$. Notice that this case contains both the maximum and average as special cases and $w_1 \geq 1/K$.

Proposition 2 *If the weights are nonincreasing, then* $\text{OWA}(\hat{X}) \leq w_1 K \cdot \text{OWA}(X)$ *for any* $X \in \Phi$ *and the bound it tight.*

Proof Let $F(\hat{X}, \boldsymbol{c}_{\sigma(1)}) \geq \cdots \geq F(\hat{X}, \boldsymbol{c}_{\sigma(K)})$. From the definition of the OWA operator and the assumption that the weights are nonincreasing, we have:

$$\text{OWA}(\hat{X}) = \sum_{j\in[K]} w_j \sum_{e_i\in\hat{X}} c_{i\sigma(j)} = \sum_{e_i\in\hat{X}} \sum_{j\in[K]} w_j c_{i\sigma(j)} \leq \sum_{e_i\in\hat{X}} \sum_{j\in[K]} w_j \hat{c}_{ij} = \hat{C}(\hat{X}). \tag{1}$$

It also holds

$$\hat{C}(\hat{X}) \leq \hat{C}(X) = \sum_{e_i\in X} \sum_{j\in[K]} w_j \hat{c}_{ij} \leq w_1 \sum_{e_i\in X} \sum_{j\in[K]} c_{ij} \tag{2}$$

and

$$\text{OWA}(X) \geq \sum_{j\in[K]} \frac{1}{K} F(X, \boldsymbol{c}_{\sigma(j)}) = \frac{1}{K} \sum_{e_i\in X} \sum_{j\in[K]} c_{ij} \tag{3}$$

Combining (1, 2) and (3) yields: $\text{OWA}(\hat{X}) \leq w_1 K \cdot \text{OWA}(X)$. In order to prove that the bound is tight consider the problem where $E = \{e_1, \ldots, e_{2K}\}$ and $\Phi = \{X \subseteq E : |X| = K\}$. The cost vectors are shown in Table 1.

Observe that all the elements have the same aggregated costs. Hence, we may choose any feasible solution as $\hat{X}$. If $\hat{X} = \{e_1, \ldots, e_K\}$, then $\text{OWA}(\hat{X}) = w_1 K^2$. But if $X = \{e_{2K+1}, \ldots, e_{2K}\}$, then $\text{OWA}(X) = \sum_{j\in[K]} w_j K = K$ and so $\text{OWA}(\hat{X}) = w_1 K \cdot OWA(X)$. □

Let us focus on some consequences of Theorem 3. Since the weights are nonincreasing, $w_1 \in [1/K, 1]$. Thus, if $w_1 = 1$ (the worst possible case), i.e. when the OWA becomes the maximum, we get a K-approximation algorithm, which is known

Table 1 A hard example.

	$\boldsymbol{c}_1$	c_2	$\boldsymbol{c}_3$	...	$\boldsymbol{c}_K$
e_1	0	0	0	...	K
e_2	0	0	0	...	K
$\vdots$					
e_K	0	0	0	...	K
e_{K+1}	K	0	0	...	0
e_{K+2}	0	K	0	...	0
$\vdots$					
e_{2K}	0	0	0	...	K

from the literature. If $w_1 = 1/K$ (the best possible case), i.e. when the OWA becomes the average, $\hat{X}$ minimizes OWA(X). Therefore, when w_1 is close to $1/K$ (a distribution of the weights is close the uniform), the approximation ratio of the proposed algorithm is better. Obviously the algorithm has a good approximation ratio, if the number of scenarios K is small.

Assume now that the weights are nondecreasing , i.e. $w_1 \leq w_2 \leq \cdots \leq w_K$. Notice that this case contains both the minimum and average as special cases. The following theorem shows that this case is much harder than the one with nonincreasing weights:

Theorem 3 *Assume that the weights are nondecreasing and K is unbounded. Then for any function $f(n)$ computable in polynomial time, there is no $f(n)$-approximation algorithm for the* OWA SHORTEST PATH *problem unless $P = NP$.*

Proof We make use of the following MIN 3- SAT problem which is known to be NP-complete [5]. We are given $x_1, \ldots, x_n$ boolean variables and a collection of clauses $C_1, \ldots, C_m$, where each clause is a disjunction of at most three literals. We ask if there is a truth assignment to the variables which satisfies at most L clauses. Given an instance of MIN 3- SAT, we construct the graph shown in Fig. 1 – the same graph as in the proof of Theorem 1. The arcs $e_1, \ldots, e_n$ correspond to literals $x_1, \ldots, x_n$ and the arcs $f_1, \ldots, f_n$ correspond to literals $\overline{x}_1, \ldots, \overline{x}_n$. There is one-to-one correspondence between paths from s to t and truth assignments to the variables. We fix $x_i = 1$ if a path chooses e_i and $x_i = 0$ if a path chooses f_i. The set Γ is constructed as follows. For each clause $C_j = (l_1^j \vee l_2^j \vee l_3^j)$, $j \in [m]$, we form the cost vector $\boldsymbol{c}_j$ in which the costs of arcs corresponding to l_1^j, l_2^j and l_3^j are set to 1 and the costs of the remaining arcs are set to 0. We fix $w_1 = \cdots = w_L = 0$ and $w_{L+1} = \cdots = w_K = 1/(K-L)$. Notice that the weights are nondecreasing. Suppose that the answer to MIN 3- SAT is yes. Then there is a truth assignment satisfying at most L clauses. Consider the path X corresponding to this truth assignment. From the construction of Γ it follows that the cost of X is positive under at most L scenarios. In consequence OWA$(X) = 0$. On the other hand, if the answer to MIN 3- SAT is no, then any truth assignment satisfies more than L clauses and each path X has positive cost for more than L cost vectors. This implies OWA$(X) \geq 1$ for all $X \in \Phi$.

Accordingly to the above, we have: the answer to MIN 3- SAT is yes if and only if there is a path X such that $\mathrm{OWA}(X) = 0$. Thus any $f(n)$-approximation algorithm for the problem would solve the MIN 3- SAT problem which implies P=NP. □

In some applications, we may wish to minimize the median of the costs. Median is a special case of the OWA operator when $w_{\lfloor K/2 \rfloor+1} = 1$ and $w_j = 0$ for $j \neq \lfloor K/2 \rfloor + 1$. Notice that in this case the weights are not monotone. However, the problem of minimizing the median is also hard to approximate.

Theorem 4 *Let K be unbounded and $w_{\lfloor K/2 \rfloor+1} = 1$ and $w_j = 0$ for $j \neq \lfloor K/2 \rfloor + 1$. Then for any function $f(n)$ computable in polynomial time, there is no $f(n)$-approximation algorithm for the* OWA SHORTEST PATH *problem unless $P = NP$.*

Proof The reduction is very similar to that in the proof of Theorem 4. It suffices to modify it as follows: if $L < \lfloor m/2 \rfloor$, then we add to Γ additional $m-2L$ scenarios with the costs equal to 1 for all the arcs; if $L > \lfloor m/2 \rfloor$, then we add to Γ additional $2L-m$ scenarios with the costs equal to 0 for all the arcs. Finally, we set $w_{\lfloor |\Gamma|/2 \rfloor+1} = 1$ and $w_j = 0$ for $j = 1, \ldots, |\Gamma|$, $j \neq \lfloor |\Gamma|/2 \rfloor + 1$. The rest of the proof runs as the one of Theorem 4. □

Theorems 4 and 5 hold for OWA SHORTEST PATH. However, it is easily seen that the same results hold also for the minimum spanning tree problem. Since there is a cost preserving reduction from the shortest path to the minimum assignment problem, both negative results remain valid also for the minimum assignment problem.

Acknowledgments The work was partially supported by Polish Committee for Scientific Research, grant N N206 492938.

References

1. Aissi, H., Bazgan, C., Vanderpooten, D.: General approximation schemes for minmax (regret) versions of some (pseudo-)polynomial problems. Discrete Optim. **7**, 136–148 (2010)
2. Averbakh, I.: On the complexity of a class of combinatorial optimization problems with uncertainty. Math Program. **90**, 263–272 (2001)
3. Kasperski, A., Zieliński, P.: On the approximability of minmax (regret) network optimization problems. Inf. Process. Lett. **109**, 262–266 (2009)
4. Kasperski, A., Zieliński, P.: A randomized algorithm for the min-max selecting items problem with uncertain weights. Ann. Oper. Res. **172**, 221–230 (2009)
5. Kohli, R., Krishnamurti, R., Mirchandani, P.: The minimum satisfiability problem. SIAM J. Discrete Math. **7**, 275–283 (1994)
6. Kouvelis, P., Yu, G.: Robust Discrete Optimization and its applications. Kluwer Academic Publishers, Dordrecht (1997)
7. Mittal, S., Schulz, A.S.: A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. In: APPROX-RANDOM 2008, Lecture Notes in Computer Science, vol. 5171, pp. 179–192. Springer-Verlag (2008)
8. Papadimitriou, C.H., Yannakakis, M.: On the Approximability of Trade-offs and Optimal Access of Web Sources. In: FOCS 2000, pp. 86–92 (2000)
9. Yager, R.R.: On Ordered Weighted Averaging Aggregation Operators in Multi-Criteria Decision Making. IEEE Trans. Syst. Man Cybern. **18**, 183–190 (1988)

Recoverable Robust Combinatorial Optimization Problems

Adam Kasperski, Adam Kurpisz and Paweł Zieliński

1 Introduction

Let $E = \{e_1, \ldots, e_n\}$ be a finite ground set and let $\Phi \subseteq 2^E$ be a set of subsets of E called the set of the *feasible solutions*. A nonnegative cost c_e is given for each element $e \in E$. A *combinatorial optimization problem* $\mathscr{P}$ with a linear objective function consists in finding a feasible solution X whose total cost, $C(X) = \sum_{e \in X} c_e$, is minimal, namely:

$$\mathscr{P}: \quad \min_{X \in \Phi} C(X). \tag{1}$$

Formulation (1) encompasses a large variety of the classical combinatorial optimization problems. In practice, the precise values of the element costs c_e, $e \in E$, in (1) may be ill-known. This uncertainty can be modeled by specifying a set of all possible realizations of the element costs (states of the world) called *scenarios*. We denote by $\mathscr{S}$ the set of all scenarios. Formally, a scenario is a vector $S = (c_e^S)_{e \in E}$, that represents an assignment of costs to the elements of E. Let $C^S(A) = \sum_{e \in A} c_e^S$, where $A \subseteq E$. A popular approach to combinatorial optimization problems $\mathscr{P}$ for hedging against the uncertainty of the element costs, modeled by scenarios, is a *robust approach*, in which we seek a solution minimizing a worst case performance over all scenarios (see, e.g. [9]):

$$\text{ROB}\,\mathscr{P}: \quad OPT_{Rob} = \min_{X \in \Phi} C_{Rob}(X) = \min_{X \in \Phi} \max_{S \in \mathscr{S}} C^S(X). \tag{2}$$

A. Kasperski (✉)
Institute of Industrial Engineering and Management, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: adam.kasperski@pwr.wroc.pl

A. Kurpisz · P. Zieliński
Faculty of Fundamental Problems of Technology, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: adam.kurpisz@pwr.wroc.pl

P. Zieliński
e-mail: pawel.zielinski@pwr.wroc.pl

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_22,

In this paper, we investigate two *Recoverable Robust* (RR) models for combinatorial optimization problems with uncertain element costs under the scenario uncertainty representation. These models were originally proposed in [4] for the shortest path problem. Here, we generalize them to the combinatorial optimization problem (1).

In the *Rent-Recoverable Robust* model, we are given a *rental factor* $\alpha \in (0, 1)$ and an *inflation factor* $\beta \geq 0$. Let $C_R^S(X) = \alpha C^S(X)$ be the *rent cost* of solution $X \in \Phi$ under scenario S and $C_I^S(X) = \min_{Y \in \Phi}\{(1-\alpha)C^S(Y) + (\alpha+\beta)C^S(Y \setminus X)\}$ be the *implementation cost* of solution $X \in \Phi$ under scenario S. Define $C_{Rent}(X) = \max_{S \in \mathscr{S}}\{C_R^S(X) + C_I^S(X)\}$. In the RENT- RR $\mathscr{P}$ problem we wish to find a solution $X \in \Phi$ minimizing $C_{Rent}(X)$, namely:

$$\text{RENT- RR } \mathscr{P}: \quad OPT_{Rent} = \min_{X \in \Phi} C_{Rent}(X) = \min_{X \in \Phi} \max_{S \in \mathscr{S}}\{C_R^S(X) + C_I^S(X)\}. \quad (3)$$

In the *k-Distance-Recoverable Robust* model, we are given the *first stage element costs* c_e^1, $e \in E$, and a *recovery parameter* $k \in \mathbb{N}$. For a given $X \in \Phi$ and k, we will denote by Φ_X^k the set of feasible solutions Y such that $|Y \setminus X| \leq k$. Let $C^1(X) = \sum_{e \in X} c_e^1$ and $C_{Rec}(X) = \max_{S \in \mathscr{S}} \min_{Y \in \Phi_X^k} C^S(Y)$ be the first stage and recovery costs, respectively. Define $C_{Dist}(X) = C^1(X) + C_{Rec}(X)$. In the k-DIST-RR $\mathscr{P}$ problem we seek a solution $X \in \Phi$ minimizing $C_{Dist}(X)$, namely:

$$k - \text{DIST-RR}\,\mathscr{P}: \quad OPT_{Dist} = \min_{X \in \Phi} C_{Dist}(X) = \min_{X \in \Phi}\{C^1(X) + C_{Rec}(X)\}. \quad (4)$$

In this paper we consider two methods of describing the set of scenarios $\mathscr{S}$. In the *discrete scenario uncertainty representation*, the scenario set, denoted by $\mathscr{S}_D$, is defined by explicitly listing all possible scenarios. So, $\mathscr{S}_D = \{S_1, \ldots, S_K\}$ is finite and contains exactly $K \geq 1$ scenarios. We distinguish the *bounded case*, where the number of scenarios K is bounded by a constant and the *unbounded case*, where the number of scenarios K is a part of the input. In the *interval uncertainty representation* the element costs are only known to belong to closed intervals $[\underline{c}_e, \overline{c}_e]$. Thus, the set of scenarios, denoted by $\mathscr{S}_I$, is the Cartesian product of these intervals, i.e. $\mathscr{S}_I = \times_{e \in E}[\underline{c}_e, \overline{c}_e]$.

2 Rent-RR Combinatorial Optimization Problems

In this section we discuss the RENT- RR $\mathscr{P}$ problem. We provide some new complexity and approximation results for various problems $\mathscr{P}$. We now focus on the discrete scenario uncertainty representation. Consider first the case when $\mathscr{P}$ is the MINIMUM SPANNING TREE. Then E is the set of edges of a given undirected graph $G = (V, E)$ and Φ contains all spanning trees of G (a spanning tree is a subset of exactly $|V| - 1$ edges that form an acyclic subgraph of G).

Proposition 1 *There is a polynomial time approximation preserving reduction from* ROB MINIMUM SPANNING TREE *to* RENT- RR MINIMUM SPANNING TREE.

Proof Let $(G = (V, E), \mathscr{S}_D = \{S_1, \ldots, S_K\})$ be an instance of ROB MINIMUM SPANNING TREE. We build a graph $G' = (V', E')$ by adding an additional node v' to V and additional parallel edges $e_v^1, \ldots, e_v^K$ of the form $\{v', v\}$ for each node $v \in V$. We form the scenario set $\mathscr{S}'_D = \{S'_1, \ldots, S'_K\}$ as follows. If $e \in E$, then the cost of e under S'_k is the same as under S_k. The cost of additional edge e_v^j, $v \in V$, $j \in [K]$, under S'_k equals 0 if $j = k$ and M otherwise, where $M = |E| \max_{e \in E} \max_{S \in \mathscr{S}_D} c_e^S$. Finally, we add the *distinguished edge*, denoted by f, that connects v' with any node of V. The edge f has zero costs under all scenarios $S \in \mathscr{S}'_D$. Now it is easy to check that every solution X', to the RENT- RR MINIMUM SPANNING TREE in graph G' and with scenario set $\mathscr{S}'_D$, whose cost is $C_{Rent}(X') < \alpha M$ (at least one such solution always exists) is of the form $X' = X \cup \{f\}$, where X is a spanning tree in G (X is a solution to ROB MINIMUM SPANNING TREE). Furthermore $C_I^S(X') = 0$ for all $S \in \mathscr{S}'_D$. So, $C_{Rent}(X') = \alpha \max_{S \in \mathscr{S}'_D} C^S(X \cup \{f\}) = \alpha \max_{S \in \mathscr{S}_D} C^S(X) = \alpha C_{Rob}(X)$. Therefore, it is evident that the reduction becomes approximation preserving one. □

We now examine the case when $\mathscr{P}$ is MINIMUM S- T CUT. We are given a graph $G = (V, E)$ with distinguished two nodes s and t and Φ consists of all s-t-cuts in G, that is the subset of the edges whose removal disconnects s and t.

Proposition 2 *There is a polynomial time approximation preserving reduction from* ROB MINIMUM S- T CUT *to* RENT- RR MINIMUM S- T CUT.

Proof Let $(G = (V, E), \mathscr{S}_D = \{S_1, \ldots, S_K\}, s, t)$ be an instance of ROB MINIMUM S- T CUT. We form graph $G' = (V', E')$ by adding to V additional nodes $v^1, \ldots, v^K$ and edges $e^1 = \{t, v^1\}, e^2 = \{v^1, v^2\}, \ldots, e^K = \{v^{K-1}, v^K\}$. Furthermore $s' = s$ and $t' = v^K$. We form the scenario set $\mathscr{S}'_D = \{S'_1, \ldots, S'_K\}$ in the following way. If $e \in E$, then the cost of e under S'_k is the same as under S_k. The cost of additional edge e^j, $j \in [K]$, under S'_k equals 0 if $j = k$ and M otherwise, where $M = |E| \max_{e \in E} \max_{S \in \mathscr{S}_D} c_e^S$. The rest of the proof runs similarly as the one of Proposition 1. □

Assume now that $\mathscr{P}$ is MINIMUM SELECTING ITEMS, where E is a set of n items and $\Phi = \{X \subseteq E : |X| = p\}$, where p is a given integer between 1 and n.

Proposition 3 *There is a polynomial time approximation preserving reduction from* ROB MINIMUM SELECTING ITEMS *to* RENT- RR MINIMUM SELECTING ITEMS.

Proof Given an instance $(E, \mathscr{S}_D = \{S_1, \ldots, S_K\}, p)$ of ROB MINIMUM SELECTING ITEMS, we form E' by adding to E additional items $e_1^j, \ldots, e_p^j$ for each $j \in [K]$. We form the scenario set $\mathscr{S}'_D = \{S'_1, \ldots, S'_K\}$ in the following way. If $e \in E$, then the cost of e under S'_k is the same as under S_k. The cost of additional item $e_i^j, i \in [p], j \in [K]$, under S'_k equals 0 if $j = k$ and M otherwise, where $M = |E| \max_{e \in E} \max_{S \in \mathscr{S}_D} c_e^S$. The reasoning is then similar to that in the proof of Proposition 1. □

Assume now that $\mathscr{P}$ is MINIMUM ASSIGNMENT, so we are given a bipartite graph $G = (V, E)$ and Φ consists of all perfect matchings in G.

Proposition 4 *There is a polynomial time approximation preserving reduction from* RENT- RR SHORTEST PATH *with a discrete scenario set to* RENT- RR MINIMUM ASSIGNMENT *with a discrete scenario set.*

Proof In [2] it has been proposed an approximation preserving reduction from ROB SHORTEST PATH with $\mathscr{S}_D$ to ROB MINIMUM ASSIGNMENT with $\mathscr{S}_D$. A reduction from RENT- RR SHORTEST PATH to RENT- RR MINIMUM ASSIGNMENT is almost the same. □

From some complexity results for the robust versions of the problems under consideration with a discrete scenario set [1, 3, 6–9] and Propositions 1-4, we obtain the following two theorems:

Theorem 1 *For the bounded case,* RENT- RR MINIMUM SPANNING TREE, RENT- RR MINIMUM ASSIGNMENT *and* RENT- RR MINIMUM SELECTING ITEMS *are weakly NP-hard,* RENT- RR MINIMUM S- T CUT *is strongly NP-hard even for two scenarios.*

Theorem 2 *For the unbounded case,* RENT- RR MINIMUM S- T CUT *and* RENT- RR MINIMUM ASSIGNMENT *are not approximable within* $\log^{1-\varepsilon} K$ *for any* $\varepsilon > 0$, *unless NP* $\subseteq$ *DTIME*$(n^{\text{poly}\log n})$, RENT- RR MINIMUM SPANNING TREE *is not approximable within* $O(\log^{1-\varepsilon} n)$ *for any* $\varepsilon > 0$, *where n is the input size, unless NP* $\subseteq$ *DTIME*$(n^{\text{poly}\log n})$ *and* RENT- RR MINIMUM SELECTING ITEMS *is not approximable within constant factor* $\gamma > 1$, *unless P=NP.*

We now show some positive results, which are generalizations of the results given in [4], for the shortest path problem. We consider first the ROB $\mathscr{P}$ and RENT RR $\mathscr{P}$ problems with the same discrete scenario set $\mathscr{S}_D$.

Theorem 3 *Suppose that there exists an approximation algorithm for* ROB $\mathscr{P}$ *with a performance ratio of* γ. *Let* $X_{Rob} \in \Phi$ *be a solution constructed by this algorithm. Then* $C_{Rent}(X_{Rob}) \leq \min\{\gamma + 1 + \beta, \gamma/\alpha\} \cdot OPT_{Rent}$.

Proof The following bounds can be concluded directly from (2) and (3):

$$OPT_{Rent} \geq \alpha \min_{X\in\Phi} \max_{S\in\mathscr{S}_D} C^S(X) = \alpha OPT_{Rob}, \quad OPT_{Rent} \geq \max_{S\in\mathscr{S}_D} \min_{Y\in\Phi} C^S(Y), \tag{5}$$

$$\begin{aligned} C_{Rent}(X) &= \max_{S\in\mathscr{S}_D} \{\alpha C^S(X) + \min_{Y\in\Phi}\{(1-\alpha)C^S(Y) + (\alpha+\beta)C^S(Y \setminus X)\}\} \\ &\leq \max_{S\in\mathscr{S}_D} \{\alpha C^S(X) + (1-\alpha)C^S(X)\} = C_{Rob}(X) \text{ for all } X \in \Phi. \end{aligned} \tag{6}$$

$$C_{Rent}(X_{Rob}) = \max_{S\in\mathscr{S}_D}\{\alpha C^S(X_{Rob}) + \min_{Y\in\Phi}\{(1-\alpha)C^S(Y) + (\alpha+\beta)C^S(Y\setminus X_{Rob})\}\} \leq$$

$$\gamma\alpha\, OPT_{Rob} + \max_{S\in\mathscr{S}_D}\min_{Y\in\Phi}\{(1-\alpha)C^S(Y) + (\alpha+\beta)C^S(Y)\} \overset{(5)}{\leq} (\gamma+1+\beta)\, OPT_{Rent}.$$

$$C_{Rent}(X_{Rob}) \overset{(6)}{\leq} C_{Rob}(X_{Rob}) \leq \gamma\, OPT_{Rob} \overset{(5)}{\leq} (\gamma/\alpha)\, OPT_{Rent}.$$ □

We now consider the interval uncertainty representation.

Theorem 4 *An optimal solution to* RENT- RR $\mathscr{P}$ *with scenario set* $\mathscr{S}_I$ *can be obtained by computing an optimal solution of its deterministic counterpart* $\mathscr{P}$ *with the costs* $\overline{c}_e$, $e \in E$.

Proof Let $\overline{X} \in \Phi$ be an optimal solution for $\mathscr{P}$ with the costs $\overline{c}_e$, $e \in E$. Then for every $X \in \Phi$ it holds: $C_{Rent}(X) = \max_{S\in\mathscr{S}_I}\{\alpha\sum_{e\in X} c_e^S + \min_{Y\in\Phi}\{(1-\alpha)\sum_{e\in Y} c_e^S + (\alpha+\beta)\sum_{e\in Y\setminus X} c_e^S\}\} \geq \alpha\sum_{e\in X}\overline{c}_e + (1-\alpha)\min_{Y\in\Phi}\sum_{e\in Y}\overline{c}_e \geq \min_{Y\in\Phi}\sum_{e\in Y}\overline{c}_e = \sum_{e\in\overline{X}}\overline{c}_e$. A trivial verification shows that $C_{Rent}(\overline{X}) = \sum_{e\in\overline{X}}\overline{c}_e$. □

3 k-Dist-RR Spanning Tree Problem

In this section, we prove hardness and inapproximability results for k-Dist-RR MINIMUM SPANNING TREE with scenario set $\mathscr{S}_D$.

Theorem 5 *The* k*-Dist-RR* MINIMUM SPANNING TREE *problem with scenario set* $\mathscr{S}_D$ *is weakly NP-hard in series-parallel graphs, even for two scenarios and any constant* k.

Proof Consider an instance of 2- PARTITION [5] in which we are given a set $A = \{a_1, \ldots, a_n\}$ and an integer size $s(a)$ for each $a \in A$ such that $\sum_{a\in A} s(a) = 2b$. We ask if there is a subset $A' \subset A$ such that $\sum_{a\in A'} s(a) = \sum_{a\in A\setminus A'} s(a)$. We construct an instance of k-Dist-RR MINIMUM SPANNING TREE as follows: graph $G = (V, E)$ is a series composition of $n+k$, 4-edge subgraphs, $G_1, \ldots, G_n, G'_1, \ldots, G'_k$, where G_i corresponds to element $a_i \in A$ and k is a constant. The costs of each edge $e \in E$ are given by a triple $(c_e^1, c_e^{S_1}, c_e^{S_2})$, where c_e^1 is the first stage cost and $c_e^{S_1}$ and $c_e^{S_2}$ are the costs under scenarios S_1 and S_2, respectively. The reduction is depicted in Fig. 1, $M > 2b$.

It is not difficult to show that a 2-partition exists if and only if there exists an optimal spanning tree X in G such that $C_{Dist}(X) = b$ (see Fig. 1). □

Theorem 6 *For the unbounded case, the* k*-Dist-RR* MINIMUM SPANNING TREE *with scenario set* $\mathscr{S}_D$ *is strongly NP-hard and not at all approximable unless P=NP.*

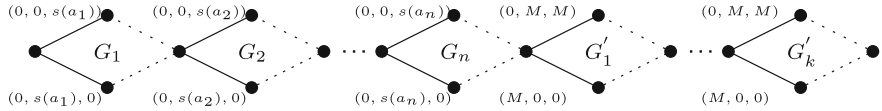


Fig. 1 A reduction from 2- PARTITION to k-Dist-RR MINIMUM SPANNING TREE. All the dummy edges (the dashed edges) have costs (0, 0, 0)

Proof We show a gap-introducing reduction from a decision version of MINIMUM DEGREE SPANNING TREE [5]. We are given a graph $G = (V, E)$ and $d \in \mathrm{IN}$, $d < |V|$. We ask if there is a spanning tree in G such that its maximum node degree is not greater than d. For each $e = \{i, j\} \in E$, we add to E a *recovery edge*, $e^r = \{i, j\}^r$, that connects nodes i and j. The resulting graph $G' = (V, E')$ is a multigraph such that $|E'| = 2|E|$. All the edges in E have zero first stage costs and all the recovery edges have the first stage costs equal to $|V|$. The scenario set $\mathcal{S}_D = \{S_1, \ldots, S_{|V|}\}$. The cost of edge $\{u, v\} \in E$ under scenario S_j equals 1 if $u = j$ or $v = j$ and 0 otherwise; the cost of recovery edge $\{u, v\}^r$ under scenario S_j equals 0 if $u = j$ or $v = j$ and $|V|$ otherwise. Finally, we set $k = d$. Suppose that the maximum node degree of some spanning tree X of G is at most d. Clearly, X is also a spanning tree of G' and does not use any recovery edge. Under each scenario $S_j \in \mathcal{S}_D$, we can decrease the cost of X to zero be replacing at most $k = d$ edges incident to node j with their recovery counterparts. Thus, $C_{Dist}(X) = 0$. On the other hand, if $C_{Dist}(X) = 0$, then the spanning tree X of G' cannot use any recovery edge (because its first stage cost is positive) and at most d edges incident to each node j, which can be replaced by at most $k = d$ recovery edges. Thus X is a spanning tree of G with the maximum node degree at most d. □

Acknowledgments The third author of the paper was partially supported by Polish Committee for Scientific Research, grant N N206 492938.

References

1. Aissi, H., Bazgan, C., Vanderpooten, D.: Complexity of the min-max (regret) versions of cut problems. In: ISAAC 2005, Lecture Notes in Computer Science, vol. 3827, pp. 789–798. Springer-Verlag (2005).
2. Aissi, H., Bazgan, C., Vanderpooten, D.: Complexity of the min-max and min-max regret assignment problems. Operation Research Letters **33**, 634–640 (2005)
3. Averbakh, I.: On the complexity of a class of combinatorial optimization problems with uncertainty. Mathematical Programming **90**, 263–272 (2001)
4. Büsing, C.: Recoverable robust shortest path problems. Networks **59**, 181–189 (2012)
5. Garey, M.R., Johnson, D.S.: Computers and Intractability. A Guide to the Theory of NP-Completeness. W. H, Freeman and Company (1979)
6. Kasperski, A., Zieliński, P.: On the approximability of minmax (regret) network optimization problems. Information Processing Letters **109**, 262–266 (2009)

7. Kasperski, A., Zieliński, P.: On the approximability of robust spanning tree problems. Theoretical Computer Science **412**, 365–374 (2011)
8. Kasperski, A., Kurpisz, A., Zieliński, P.: Approximating the minmax (regret) selecting items problem. Information Processing Letters **113**, 23–29 (2013)
9. Kouvelis, P., Yu, G.: Robust Discrete Optimization and its applications. Kluwer Academic Publishers (1997).

Part VI
Energy and Environment

A MILP-Based Approach for Hydrothermal Scheduling

Dewan Fayzur Rahman, Ana Viana and João Pedro Pedroso

1 Introduction

The Hydrothermal Scheduling Problem (HSP) is one of the challenging optimization problems in power systems that optimally determines which of the thermal and hydro units must be committed/decommitted over a planning horizon (that for short-term planning lasts from 1 day to 2 weeks, generally split into periods of 1 hour each) and aims at minimizing the overall operation cost of thermal units while satisfying practical thermal and hydro constraints. Main decision variables are: (1) which thermal generators must be committed in each period of the planning horizon, as well as their production level; and (2) the discharge rates of each hydro generator.

A wide range of optimization techniques has been researched to solve the HSP ranging from exact approaches and Lagrangian Relaxation, to very elaborate heuristics and metaheuristics. Lagrangian Relaxation and augmented Lagrangian relaxation techniques are among the most popular techniques but convergence due to non-linearity of the problem is a major issue. Furthermore, the combinatorial nature of the problem and its multi-period characteristics prevented exact approaches from being successfully used in practice: they resulted in very inefficient algorithms that were only capable of solving small size instances of no practical interest. Metaheuristics

D. F. Rahman
INESC TEC (formerly INESC Porto), Porto, Portugal
e-mail: dewan.m.rahman@inescporto.pt

A. Viana (✉)
School of Engineering, Polytechnic Institute of Porto and INESC TEC (formerly INESC Porto), Porto, Portugal
e-mail: aviana@inescporto.pt

J. P. Pedroso
Department of Computer Science, Faculty of Sciences, University of Porto and INESC TEC (formerly INESC Porto), Porto, Portugal
e-mail: jpp@fc.up.pt

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_23,

algorithms have also been used. A survey on different optimization techniques and modeling issues is provided in [2].

Lately, improvements made on Mixed Integer Linear Programming (MILP) solvers led to increasing research on MILP techniques for the thermal problem [1, 4, 7, 8]. In [8] the authors proposed a MILP-based procedure for thermal unit commitment that is able to converge to optimality, even for large size problems.

In this paper we extend the procedure proposed in [8] to solve the HSP. Furthermore, we use the concept of "Matheuristics", an optimization technique that integrates (meta)heuristics and MILP strategies to try to speed up convergence of the base iterative method. Two matheuristic approaches are explored: one based on "Local Branching" [3] and another where a "Particle Swarm Optimization" (PSO) algorithm cooperates with the MILP solver.

All the proposed approaches were tested on several benchmark instances always converging to the optimal solution. Also worth mentioning that the CPU times necessary for the matheuristics to converge to optimality were not smaller than the time associated to the base method, showing the effectiveness of this simple method.

2 MILP Solution Approaches for the HSP

The objective of the HSP is to minimize total production costs over a given planning horizon. They are expressed as the sum of fuel costs (quadratic functions that depend on the production level of each unit) and start-up costs. Shut-down costs are assumed to be zero. In this work, the following thermal constraints are included in the model: ramps, minimum up and down times and production limits. For the hydro model the constraints considered are: water continuity constraints, hydro units discharge limits, hourly reservoir inflows, reservoir storage balance, and spillage limits.

A first requirement to solve the HSP with MILP solvers is to sort out the many non-linear constraints that the problem has. This point is thoroughly discussed in [1, 4, 7, 8]. Furthermore, for efficiency purposes [8] discusses a procude that to accurately represents the quadratic fuel cost by an "equivalent" linear function. The procedure is described in Sect. 2.1.

2.1 Iterative Piecewise Linear Algorithm

To circumvent the difficulty brought up by the quadratic fuel cost function ($\mathbf{Q}$), an iterative method based on a dynamic piecewise linear approximation of $\mathbf{Q}$ is designed. The method starts by approximating $\mathbf{Q}$ by two linear functions, one tangent to $\mathbf{Q}$ in $P^{\min}$ and the other in $P^{\max}$ ($P^{\min}$ and $P^{\max}$ being the minimum and maximum production levels of a unit). Afterwards, at each iteration an additional segment tangent to $\mathbf{Q}$ in P^{x} is added for each unit (P^{x} being the production level obtained by the MILP formulation in that iteration). The method stops when the difference in

cost between **Q** and its current linear approximation is smaller than a given error, for all the units. Details on this algorithm can be found in [8].

2.2 Matheuristics for the HSP

"Matheuristics" refers to the idea of interoperation of mathematical programming and (meta)heuristic approaches in order to obtain good and robust results. Some approaches using mathematical programming combined with metaheuristics can be followed, e.g.: (1) some good quality incumbents of metaheuristics can be provided to branch-and-bound to speed up the search; (2) MILP solvers may be used to search neighborhoods within a metaheuristic; (3) solutions of relaxed problems can be used as a starting point for metaheuristics search; etc. In this work we will explore two such approaches: Local Branching and a problem-suited PSO algorithm that will provide a MILP solver with high quality initial solutions. In both cases we use the iterative piecewise linear approximation mentioned in Sect. 2.1 in the MIPL solver.

HSP and Local Branching: Local Branching is based on the k-opt concept of local search that starts with an initial feasible solution and then iteratively moves to a neighbor solution by changing up to k components of the solution. The feasible region of an optimization problem is divided into smaller subregions (also called neighborhoods) and a MILP solver is applied to find the best solution in each of the subregions.

For a given positive integer parameter k, the k-opt neighborhood of an initial feasible solution $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ of a MILP problem with associated binary variables $x = (x_1, x_2, \ldots x_n)$ can be defined by local branching constraints (2.1) where $\bar{\mathcal{F}}$ is the index set of the variables that are set to 1 in $\bar{x}$, and the two terms on the right side represent the binary variables that will flip their values (with respect to $\bar{x}$).

$$\Delta(x, \bar{x}) = \sum_{j \in \bar{\mathcal{F}}} (1 - x_j) + \sum_{j \in (1,2,\ldots,n) \setminus \bar{\mathcal{F}}} x_j \tag{2.1}$$

The feasible search space associated with the current branching node **C** will be partitioned into two disjunct sub-regions, $\mathscr{P}_1$ and $\mathscr{P}_2$; $\mathscr{P}_1$, called left branch, is the optimization problem that results from adding the constraint $\Delta(x, \bar{x}) \leq k$ to the current problem; and $\mathscr{P}_2$, called right branch, is the problem that results from adding the constraint $\Delta(x, \bar{x}) \geq k + 1$. A $\mathcal{P}_1$ must be solved completely before a $\mathscr{P}_2$ starts. If a better solution $\bar{x_A}$ is found in $\mathcal{P}_1$, it becomes the new incumbent solution. The process is then reapplied to $\mathscr{P}_2$. If the solution is not improved, $\mathscr{P}_2$ is explored by tactical branch-and-bound [3].

In this implementation Local Branching builds additional constraints to the model only for the binary variables (continuous variables are determined by the MILP solver accordingly). The vector x will represent the state of each thermal unit in each time period, i.e. $x = (y_{11}, y_{12}, \ldots, y_{UT})$ where U and T are the number of thermal

units and the length of planning horizon, with one period per hour. Initially, the Local Branching idea is applied to the MILP formulation described in [8] with only two segments, the ones corresponding to $P^{\min}$ and $P^{\max}$; additional segments are dynamically generated until convergence is achieved.

HSP solution by PSO with MILP solver: In the third methodology PSO [6] is coupled with a MILP solver and works as follows. First PSO is run until a stoping criterion is met (maximum number of iterations is reached). Once this optimization step is concluded, the results obtained are used to provide the MILP solver with high quality incumbent solutions. Contrary to the Local Branching implementation, where the piecewise linear approximation of the quadratic cost function of each thermal unit is initially represented by two segments, in this implementation the production levels of each unit on the *pbest* solutions found by PSO are used to add *pbest* additional segments to the initial linear cost function, besides the segments associated to $P^{\min}$ and $P^{\max}$. After that, the iterative piecewise linear algorithm starts solving the model with (*pbest* + *2*) segments. If necessary, new segments can be added to the model until a user pre-defined precision is reached.

3 Computational Results

The proposed algorithms have been used to solve seven sets of benchmark instances using the generator in [5], the number of thermal units ranging from 20 to 200 and the number of hydro units ranging from 10 to 100, for 24-hour planning horizons with one period per hour. Ten instances of each size were generated (all instances can be provided to the interested reader upon request). The parameters of Local Branching and PSO are considered as follows: the neighborhood size k is set to 30 and the number of solutions of PSO is set to 10. All the other parameters are set to the default values suggested in the corresponding reference papers.

CPU times were obtained with Microsoft Visual C++ 6.0 compiler and CPLEX 12.1 on a computer with a AMD Athlon(tm) 64 X2 Dual Core Processor at 2.01 GHz with 4.0 GB RAM running Windows 7. Only one thread was assigned to this experiment. Models were written in the AMPL language and the default CPLEX parameters were used for solving the linear programs. All approaches stop when the relative error on the difference between the quadratic and linear functions is less than 10^{-6}.

Table 1 reports the computational results associated to the three implementations: "base" iterative piecewise linear method; Local Branching and PSO with CPLEX. As all problems were solved to optimality, only CPU times are displayed. Columns **Best**/**Worst** report the minimum/maximum computational time required to solve one particular instance from a set of ten instances of the same size. **Avg** reports the average computational time required to solve all the ten instances and **Std** reports the standard deviation of computational time required to solve all the ten instances. In all cases, the time necessary to solve each problem to optimality is within a reasonable range, allowing the effective use by decision maker of the information provided by the methods. Unexpectedly, the use of hybrid methods did not improve the CPU time

Table 1 Computational time (in seconds) required to solve the HSP by proposed approaches.

		Iterative linear algorithm				Local branching				PSO and CPLEX			
u	h	Best	Worst	Avg	Std	Best	Worst	Avg	Std	Best	Worst	Avg	Std
20	10	37.1	105.3	63.29	19.12	39.1	107.3	67.29	23.42	65.5	180.65	90.87	75.68
50	20	195.7	1078.3	423.18	279.80	198.5	1078.2	455.67	293.91	300.25	1435.15	687.36	486.35
75	35	305.6	1063.7	713.37	225.41	315.9	1085.2	725.37	231.41	405.6	1585.3	842.7	283.49
100	50	379.8	2462.3	1141.71	730.05	410.28	2552.3	1165.71	749.86	472.5	2621.9	1275.64	793.52
150	75	751.1	3792.5	2162.28	832.43	760.5	3882.5	2862.28	839.34	817.25	4011.56	2942.75	857.3
200	100	1331.6	5108.2	3409.04	920.97	1481.1	5417.8	3610.27	935.56	1598.6	7564.64	4256.45	994.8

values. A more attentive analysis of the MIP solver report showed that for this set of instances the intermediate optimal solutions, i.e. the optimal solutions obtained in each iteration of the base method, are found within very short CPU time. Therefore, more elaborate techniques are not necessary.

4 Conclusions

The main contribution of this paper are algorithms capable of solving the HSP to optimality, for instances of practical size.

The reason for having higher CPU times in Local Branching is that it needs some extra CPU time to check local branching constraints, even though the optimal solution exist at the top of the tree and is easily achieved at an early stage by MILP solver. Whereas PSO provides incumbent solutions and their upper bounds to MILP solver which could lead to CPU time saving, PSO also needs to run for some pre-defined number of iterations increasing total CPU time. Computational results show that the "base" iterative piecewise linear method is very powerful at handling these instances and recourse to more sophisticated methods does not seem to be necessary.

As future work, we will test these approaches on HSP test instances where the valve point effect is present in the objective function. We will also extend the algorithm in order to solve Wind-Hydrothermal Scheduling considering the stochastic nature of the wind power system.

Acknowledgments Financial support for this work was provided by the Portuguese Foundation for Science and Technology (under Project PTDC/EGEGES/099120/2008) through the "Programa Operacional Temático Factores de Competitividade (COMPETE)" of the "Quadro Comunitário de Apoio III", partially funded by FEDER. The authors would like to thank Prof. Andrea Lodi for providing with the Local Branching code used in this work.

References

1. Carrion, M., Arroyo, J.: A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. IEEE Trans. Power Syst. **21**, 1371–1378 (2006)
2. Farhat, I.A., El-Hawary, M.E.: Optimization methods applied for solving the short-term hydrothermal coordination problem. Electr. Power Syst. Res. **79**, 1308–1320 (2009)
3. Fischetti, M., Lodi, A.: Local branching. Math. Program. **98**, 23–47 (2003)
4. Frangioni, A., Gentile, C., Lacalandra, F.: Tighter approximated milp formulations for unit commitment problems. IEEE Trans. Power Syst. **24**, 105–113 (2009)
5. Frangioni, A., Gentile, C., Lacalandra, F.: Sequential Lagrangian-MILP approaches for Unit commitment problems. Electr. Power Energy Syst. **33**, 585–593 (2011)
6. Kennedy, J., Eberhart, R.: Particle swarm optimization. IEEE International Conference on Neural Networks **4**, 1942–1948 (1995)
7. Simoglou, C.K., Biskas, P.N., Bakirtzis, A.G.: Optimal selfscheduling of a thermal producer in short-term electricity markets by milp. IEEE Trans. Power Sys. **25**, 1965–1977 (2010)
8. Viana, A., Pedroso, J.: A new MILP-based approach for unit commitment in power production planning. Electr. Power Energy Sys. **44**, 997–1005 (2013)

Mixing Behavioral and Technological Data in Mathematical Programming Framework

Roman Kanala and Emmanuel Fragnière

1 Introduction

Tools of operations research are suitable to resolve a very large domain of problems. Issues of energy planning have been assessed since using OR since four decades [1, 3]. MARKAL (MARKet ALlocation) model stems from seventies [1] and, as a successor of earlier models, was jointly developed at Brookhaven National Laboratory (BNL) and Kernforschungsanlage (KFA) Jülich, under aegis of the International Energy Agency (IEA) Energy Technology Systems Analysis Project (ETSAP). From its early formulations in FORTRAN, it was rewritten several times using more modern programming languages such as OMNI, GAMS, AMPL, OSeMOSYS with user interfaces that got more user friendly at each iteration. Recently, it fusioned with EFOM model from which it got several advanced features such as variable time step and better description of technologies, to form The Integrated Markal Efom System (TIMES). MARKAL / TIMES has become a standard tool for long-term multi-objective energy-environment planning with a growing user base. Modern MARKAL is not even a model anymore as it became just a convention how to name variables and their classes in different model implementations for quick and easy communication within the modeling community.

The MARKAL model is an optimisation model and can be formulated as follows:

$$minimise \sum_i (c_i X_i)$$

R. Kanala (✉)
Institut des sciences de l'environnement, Université de Genève, 7 route de Drize, CH-1227 Carouge, Switzerland
e-mail: roman.kanala@unige.ch

E. Fragnière
Hautes Etudes de Gestion de Genève, 7 route de Drize, CH-1227 Carouge, Switzerland

S. Helber et al. (eds.), *Operations Research Proceedings 2012*, Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_24,

$$\sum_i a_{ji} X_i \leq b_j \quad \text{and} \quad X_i \geq 0$$

where the coefficients c_i for the objective function, technical coefficients a_{ji} (investment costs, availability factors, load factors, yields and efficiencies, etc.) and demands b_j (useful demands) are known parameters and variables X_i (activities of technologies and energy agents), represent the solution vector of the problem. The objective function to be minimised is the global system cost with contributions from each technology, each extracted or imported energy agent, each operation and maintenance cost. MARKAL is a dynamic multi-period model usually with 9 periods with 5 years step, so the contributions have to be reckoned for every year and discounted in time. The constraints are physical (energy conservation, Carnot cycle efficiency), technical (energy transformation is not even close to maximum theoretical efficiency and technologies do not always run at maximum potential), operational (peak demands are different in winter and in summer, technologies need maintenance that reduces availability), environmental constraints (emissions such as global constraint on CO_2, local constraints on SO_2, NO_x) and even political constraints (such as abandon of the nuclear power). Every active constraint increases the global cost of the energy system and shadow prices are interesting results associated with dual variables Fig.1.

2 Methodology

The MARKAL model in its classical formulation is based on a number of assumptions. Some of them, such as inelastic demands have been addressed in [9] and the connection with the rest of the economy instead of soft-linking MARKAL through exogenous demands, these can be included as part of the optimisation problem in MARKAL-MACRO [8]. Issues with small numbers when changing the scale from a country down to a region or a city, when some variables may become integer or even binary, as well as decision under incertitude, were addressed using stochastic mixed integer programming [4]. Local effects, such as influence of centralised / decentralised heating, road network and industrial infrastructure can be assessed through coupling MARKAL with an atmospheric dispersion model and a GIS [2], [5]. However the hypothesis of the perfect information and of perfect economic rationality so far did not get attention of the modelers. Indeed, these become increasingly important when modeling the Demand Side Management (DSM) measures in detail.

When moving on the model scale from country or regional models toward city or community models, in addition to objective and optimised centralised decisions on the production side, decentralised decisions on the demand side are also getting importance. But these frequent, even everyday penny-worth investment decisions are made by a huge number of decision makers with an increased role of subjectivity. Information and economic rationality are only few of the number of drivers that determine the investment decision. Depending on the technology, there may be other

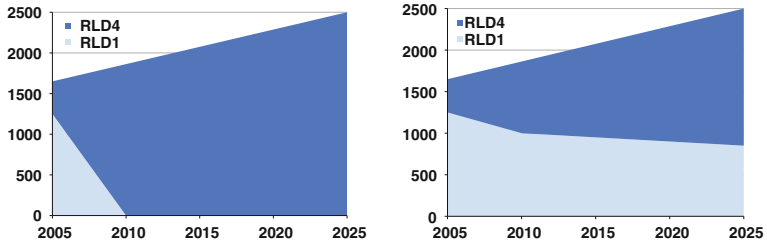
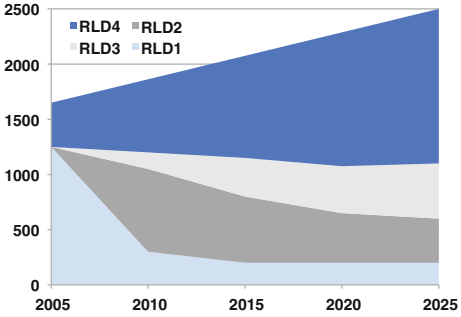


Fig. 1 An unbound system will immediately replace all incandescence bulbs by economically more viable low-consumption bulbs. However the observed behaviour of consumers is very different. Many people are attached to incandescent bulbs. User-introduced bounds will certainly improve this systematic error, but the model will become subjective and the results will depend on the erudition and opinion of the modeler [10]

Fig. 2 Use of virtual process technologies can remedy on this flaw. In addition to classical MARKAL competition, we get two more technologies that may penetrate to optimal solution if the investment cost to their upstream marketing and information campaign remains competitive



criteria, be they rational or irrational. Among rational ones, one can count, for example, medical reasons such as macular degenerescence of elderly people or flicker sensibility of an important part of population. These consumers will consequently refuse electroluminiscent bulbs as long as they have choice. Irrational criteria may include for example fear of obscurity, thus a need to turn on all the lights in a room or in the entire house. When modeling the DSM measures on a small-scale model, thus extending the original model beyond its domain of validity, all these irrational behavioural parameters may not be negligible anymore and have to be taken into account. The usual way to address this issue was so far to design bounds on penetration of some demand technologies and DSM.

We propose a method to circumvent this issue consists in soft-linking data from sociological surveys that determine technical coefficients for MARKAL model. Given the size of MARKAL community using standard tools, an important criterion was to remain entirely compatible with the entire modeling framework created so far under the aegis of the International Energy Agency, such as ANSWER and VEDA interfaces to MARKAL and TIMES models. The approach consists of modeling behaviour as virtual technologies associated with moderate use or technology switch. All the usual attributes of a technology such as investment cost, efficiency, availabil-

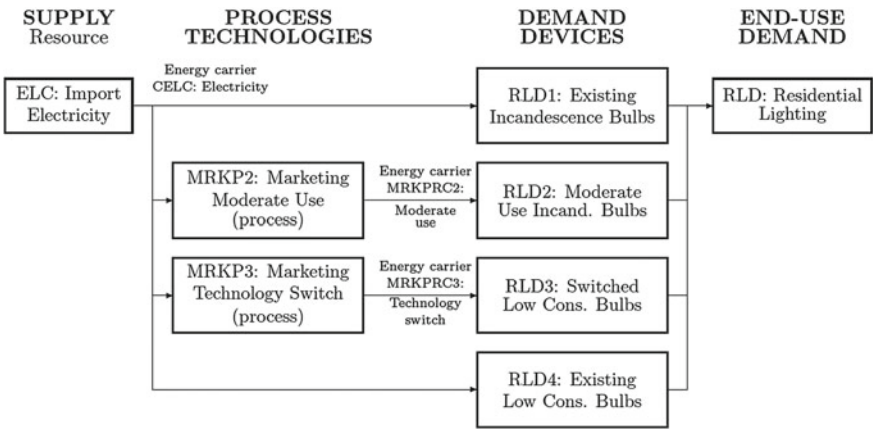


Fig. 3 Reference energy system for Nyon city residential lighting

ity factor etc. can be described with parameters coming from sociological surveys that measure the real behaviour of consumers.

As the underlying model used was the Reference Energy System (RES) of the municipal MARKAL model of Nyon, a small city half-way between Geneva and Lausanne. The particularity of this RES is that there is no extraction and no transformation of energy. All the energy is imported and consumed without processing, so there are no interdependences between technologies and energy forms. Energy flows can be separated into independent streams where final energies, such as natural gas, electricity, gasoline, Diesel fuel, etc., are imported and directly consumed in demand devices.

In order to assess the consumer behaviour, two virtual technologies associated are introduced:

1. moderate use of a demand device
2. technology switch between "old" and "new" demand device

These two types of behavioural patterns are covering all the possible situations. Note that like there may be more than one "real" technology influencing the useful demand of one type, the same stands for "virtual" process technologies describing behaviour. However, for simplicity we consider the minimal case. The corresponding Reference Energy System (RES) is drawn in Fig. 3.

The investment cost in behavioural virtual technologies (savings, technology switch) is zero as they depend only on user behaviour. However the upstream virtual process technologies have a non-zero cost that corresponds to the cost of a marketing and information campaign.

The bounds that have to be introduced are:

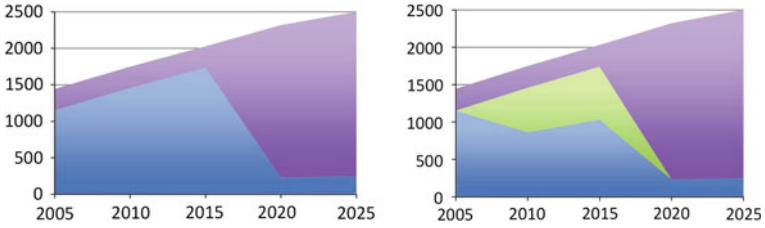


Fig. 4 If the cost of marketing and information campaign in favour of energy savings and of technology switch is too high, results obtained correspond to a classical MARKAL competition of two technologies. With lowering the cost of the triggering virtual technologies of marketing, one or both behavioural technologies may appear in the optimal solution—from [6]. For the example with both virtual technologies active, see Fig. 2

1. a bound on capacity of energy savings and a bound on technology switch so that it cannot exceed the capacity of the existing "old" technology, otherwise there might be more savings or technology switch than installed real technology
2. a bound on installed capacity of existing low consumption bulbs in order to prevent it to fall under residual capacity level and a bound on the maximum proportion of incandescence bulbs—even though these constraints are unlikely to ever become active
3. a constraint on proportion of installed capacity of energy savings virtual process technology that should not exceed 71.2 % of the sum of RLD1 and RLD2—proportion of YES answers to Question 3 of the survey: Do you systematically turn off light when leaving a room?
4. a constraint on capacity of technology switch RLD3 that should not exceed 69 % of capacity of RLD1—comes from Question 12: If you were better informed about advantages of low-consumption bulbs, would you be ready to completely abandon the incandescent bulbs?

3 Implementation and conclusion

Detailed description of the methodology and discussion of the first results have been published in [6] and [10]. Generation and testing of hypotheses as well as detailed methodology of integration of data from sociological surveys into MARKAL have been described in [7].

Social reality, as measured by special-crafted sociological surveys, can serve as compatible input to energy models and create virtual technologies that compete with real technologies at the same level of importance. Inclusion of behavioural parameters may improve the model by eliminating a systematic error up to 30 % on demand side. At the same time, this new approach opens a way how to include social changes into long-term energy planning scenarios.

Acknowledgments The authors wish to acknowledge the contributions of Dr. Denis Lavigne and Dr. Francesco Moresino to this topic.

References

1. Abilock, H., Fishbone, L.G.: User's Guide for MARKAL (BNL Version). Brookhaven National Laboratory, Upton, New York (1979). BNL 27075
2. Fedra, K., Greppin, H., Haurie, A., Hussy, C., Dao, Hy, Kanala, R.: GENIE: An integrated environmental information and decision support system for Geneva. Part I: Air Quality. Arch. Sci. Geneva **49**(Fasc.3), 247–263 (1996)
3. Fishbone, L.G., Abilock, H.: MARKAL, a linear-programming model for energy systems analysis: Technical description of the BNL version. Int. J. Energy Res. **5**(4), 353–375 (1981)
4. Fragnière, E., Haurie, A.: A stochastic programming model for energy/environment choices under uncertainty. Int. J. Environ Pollut **6**(4–6), 587–603 (1996)
5. Fragnière, E., Haurie, A., Kanala, R.: A GIS-based regional energy-environment policy model. Int. J. Global Energy Issues **12**(1–6), 159–167 (1999)
6. Fragnière, E., Kanala, R., Lavigne, D., Moresino, F., Nguene, G.: Behavioral and technological changes regarding lighting consumptions: A MARKAL case study. Low Carbon Econ. **1**, 8–17 (2010)
7. Fragnière, E., Kanala, R., Turin, N.: An energy optimization framework for sustainability analysis: Inclusion of behavioral parameters as a virtual technology in energy optimization models. In: Mechanism Design for Sustainability, Springer, Netherland, 2012
8. Hamilton, L.D., Goldstein, G.A., Lee, J., Manne, A.S., Marcuse, W., Morris, S.C., Wene, C.O.: MARKAL-MACRO: An Overview. Brookhaven National Lab., Brookhaven, New York (1992)
9. Loulou, R., Lavigne, D.: MARKAL model with elastic demands : Application to greenhouse gas emission control. In: Carraro, C., Haurie, A. (eds.) Operations Research and Environmental Management, pp. 201–220. Kluwer Academic Publishers, Dordrecht (1996)
10. Nguene, G., Fragnière, E., Kanala, R., Lavigne, D., Moresino, F.: Socio-Markal (Somarkal): First modeling attempts in the Nyon residential and commercial sectors taking into account behavioural uncertainties. Energy Sustain. Dev. **15**(1), 73–83 (2011)

Profit Optimization of the Cross-Border Trade Between the Nordic and Russian Electricity Markets

Olga Gore, Satu Viljainen, Kalevi Kylaheiko and Ari Jantunen

1 Introduction

Most European electricity markets clasify as energy-only markets with a zonal or uniform marginal pricing principle. It means that generators only earn revenues when they produce electricity. In contrast, in a so-called energy + capacity market generators earn revenues both by producing electricity and for being available to produce. At present, there is an ongoing discussion in Europe that energy-only market alone may not provide enough resources to cover the fixed costs of the generators (especially with high penetration rate of renewable energy). This may lead to the introduction of capacity payments in addition to energy revenues in many European markets in the future [1, 8]. In energy + capacity market, the energy market is designed to cover the variable costs of the generators, while capacity market is designed to cover the fixed costs of the generators. The demand side in the energy + capacity market pays separately for the electricity and capacity.

Different market designs create challenges to market integration. Usually efficient market integration of two energy-only markets ensures efficient use of interconnections, increase in social welfare and price convergence in both electricity markets [7]. However, integration of energy only and energy + capacity markets may lead to the non optimal use of transmission lines and the emergence of "dead bands" or price intervals when it is not rational to transmit electricity between the two markets. This has proven to be the case between the Russian energy + capacity and Nordic energy-only market. Even though the electricity price in Nordic market has been higher that the electricity price in Russian market, electricity flow at the cross-border interconnection has been reduced as a result of the profit maximization task of the

O. Gore (✉) · S. Viljainen · K. Kylaheiko · A. Jantunen
Lappeenranta University of Technology, Skinnarilankatu 34, Lappeenranta, Finland
e-mail: olga.gore@lut.fi

S. Viljainen
e-mail: satu.viljainen@lut.fi

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_25,

cross-border trader. In this paper, the optimal strategy of the cross-border trader operating in two markets with different designs will be shown by calculating the optimal electricity flows at different price levels.

2 Cross-Border Trade Optimization

In this paper, we will demonstrate the optimal behaviour of the cross-border trader operating in energy-only and energy + capacity markets.

The cross-border trader operates in two markets: B is the energy + capacity market and A is the energy-only market. Market A is a deficit area and market B is a surplus area. Prices in market A are higher than prices in market B and, therefore, it is profitable to transfer the electricity from B to A. Cross-border trader buys electricity in market B and sells it in market A. In addition to electricity procurement from the market B, cross-border trader has to pay for the capacity in market B. The profit of the trader is electricity sold to the market A minus electricity plus capacity bought in B. In our optimization problem, we assume that the capacity payment depends on the electricity export volumes during specifit peak hours (which corresponds to the situation in Russia, as discussed in detail in [3]).

The daily profit of the cross-border trader can be derived as follows:

$$\sum_{1}^{k} V_k \cdot \lambda_k^A - \sum_{1}^{k} V_k \cdot \lambda_k^B + \sum_{1}^{n} V_n \cdot \lambda_n^A - \sum_{1}^{n} V_n \cdot \lambda_n^B - CP \cdot V_n^{max} \qquad (1)$$

where k is a number of off-peak hours in a day and n is a number of peak hours in a day; λ_k^A, λ_n^A, λ_k^B, λ_n^B are market clearing prices in off-peak and peak hours in the market A and B; $V_n^{max} = max(V_n)$ is maximum volume of electricity transferred within the peaking hours; CP is daily capacity price, determined by the results of the capacity market (Euro/MW·day).

From the Eq. (1) we can see that the profit of the cross-border trader strongly depends on the price difference in export and import markets and the capacity payment. The trader can avoid the capacity payment by reducing the volume of capacity bought in the capacity market. This leads into reduced volume of electricity to be delivered during the peak hours. Thus, the optimal volume of electricity to be transferred in peak hours can be obtained by estimating the possible savings from capacity costs, and effects of capacity withholding (reduction of export) on the electricity prices.

For the simplicity, let's assume that the expected profit of the cross-border trader for the next day depends on the expected prices. Then the formula [1] can be simplified and the expected daily profit can be derived as follows:

$$\pi = k \cdot V_k \cdot \lambda_k^A - k \cdot V_k \cdot \lambda_k^B + n \cdot V_n \cdot \lambda_n^A - n \cdot V_n \cdot \lambda_n^B - CP \cdot V_n \qquad (2)$$

We assume that cross-border trader reduces the export during the peak hours up to V^*. The extent of capacity withholding ΔV can be derived as:

$$\Delta V = V - V^* \tag{3}$$

Reduction of the export reduced supply in import market A, which may result in the rise of the market clearing price to $\lambda_n^{A^*}$ [5].

$$\lambda_n^{A^*} = \lambda_n^A + \Delta\lambda_n^A = \lambda_n^A + \frac{\Delta\lambda_n^A}{\Delta V} \cdot \Delta V \tag{4}$$

The extent of the price rise depends the extent of capacity withholding ΔV and on the slope of the supply curve $\frac{\Delta\lambda_n^A}{\Delta V}$.Thus, the profit of the cross-border trader in case of exported volume V_n^* is:

$$\pi^* = k \cdot V_k \cdot \lambda_k^A - k \cdot V_k \cdot \lambda_k^B + n \cdot V_n^* \cdot (\lambda_n^{A^*} - n \cdot V_n^* \cdot \lambda_n^B - CP \cdot V_n^*) \tag{5}$$

In order to estimate the profitability effects of the export capacity withholding, we will define the change in profit $\Delta\pi = \pi^* - \pi$ taking into account the Eqs. (3) and (4).

$$\Delta\pi = -n \cdot \lambda_n^A \cdot \Delta V + n \cdot \lambda_n^B \cdot \Delta V + CP \cdot \Delta V + n \cdot \frac{\Delta\lambda_n^A}{\Delta V} \cdot \Delta V \cdot (V - \Delta V) \tag{6}$$

The Eq. (6) represents the profitability effects of the withholding export capacity. The first component in Eq. 6 represents the lost profit due to reduced electricity sale in import market A. Second component represents cost savings due to reduced electricity buy from the export market B. Third component represents cost savings due to reduced capacity buy in export capacity market B. Fourth component represents the revenue gain due to selling the remaining electricity volume by increased market clearing price in import market A.

The optimal extent of capacity withholding can be defined when maximizing the change of the profit $max(\Delta\pi)$. Solving the equation $\frac{\delta\Delta\pi}{\delta\Delta V} = 0$ the optimal capacity limit can be derived as:

$$\Delta V = \frac{-n \cdot \lambda_n^A + n \cdot \lambda_n^B + CP + n \cdot \frac{\Delta\lambda_n^A}{\Delta V} \cdot V}{2n \cdot \frac{\Delta\lambda_n^A}{\Delta V} \cdot V} \tag{7}$$

Subject to constraints $0 \leq \Delta V \leq V_{max}$. Where V_{max} is maximum transmision capacity of the cross-border line.

3 Case Study on Russian-Finnish Interconnection

Using the methodology described in the previous section, we will estimate the optimal capacity exports of the cross-border trader who is operating simultaneously in Nordic and Russian electricity markets. Current monopoly structure of the cross-border trade creates the possibilities for the cross-border trader to decide how much electricity is imported to Finland.

3.1 Current Operational Principles

Nordic energy-only and Russian energy + capacity market are interconnected by the transmission line with maximum capacity 1,400 MW [4].

The Nordic market is energy-only market. Nordic prices are highly volatile resulting from high seasonal and daily variations in supply/demand. The import of Russian electricity covers around 10 percent of the total consumption in Finland. Nordic electricity prices have always been higher than the Russian electricity prices [6].

Russian market is energy + capacity market. In the capacity market, the System Operator defines the possible peak demand for electricity in a particular month and selects the capacities required to cover that demand. The last accepted capacity offer sets the capacity price. The principles of capacity market are discussed in detail [2]. Capacity price is currently about 3,000–4,000 /MW/month and is likely to increase in the future.

After introduction of the capacity market in Russia,the cross-border trader has repeatedly limited the export capacity to Finland during peak hours, which is a result of his profit optimization.

3.2 Calculation of Optimal Capacity

In this example, we will look at the cross-border trader's optimal export schedule from Russia to the Nordic market in February 2012.In other words we will estimate how much electricity the cross-border trader is willing to sell in the Nord Pool (the Nordic Power Exchange) in peak hours. The number of peak hours for each month is assigned by the System Operator [9]. In February, there are eight peak hours in each working day. According to the expression (7) the extent of capacity withholding depends on the expected electricity prices for the next day, capacity price and the slope of the supply curve. Capacity price is 155 €/MW*day. We assume that the expected electricity prices in Russian electricity market are deterministic and equalling to 25 €/MWh. However, expected prices in Nord Pool depend on the slope of the supply curve and the expected demand. Assuming that the supply curve has the same

Table 1 Optimal export capacity and benefit from export limits

i	Demand	λ_n^A	slope	Δ V	V^*	Revenue loss	Electricity cost savings	Capacity cost savings	Revenue gain	Total benefit $\Delta\pi$
	MW	€/MWh		MW	MW	€	€	€	€	€
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
1	39,500	30	1.85E-3	1,400	0	−336, 000	280,000	217,000	0	161,000
2	42,200	35	0.96E-3	1,400	0	−392, 000	280,000	217,000	0	105,000
3	47,400	40	3.0 E-3	1,400	0	−448, 000	280,000	217,000	0	49,000
4	47,600	41	3.0 E-3	1,262	138	−413, 936	252,400	195,610	4,180	38,254
5	47,900	42	3.0 E-3	1,095	305	−367, 920	219,000	169,725	8,015	28,820
6	48,200	43	3.0 E-3	930	470	−319, 920	186,000	144,150	10,490	20,720
7	48,500	44	3.0 E-3	762	638	−268, 224	152,400	118,110	11,668	13,954
8	49,000	45	5.1 E-3	669	731	−240, 840	133,800	103,695	39,123	35,778
9	49,500	50	0,003	606	794	−242, 400	121,200	93,930	115,479	88,209
10	50,500	80	0,1	522	878	−334, 080	104,400	80,910	366,653	217,883
11	51,000	130	0,1	272	1128	−282, 880	54,400	42,160	245,453	59,133

$[3] = ([2]-[2](-1))/([1]-[1](-1))$ $[4] = (-8\cdot[2]+8\cdot 25+155+8\cdot[3]\cdot 1,400)/(2\cdot 8\cdot[3])$, $0 \le [4] \le 1,400$ $[5] = 1,400-[4]$; $[6] = -8\cdot[2]\cdot[4]$; $[7] = 8\cdot 25\cdot[4]$; $[8] = 155\cdot[4]$ $[9] = 8\cdot[3]\cdot[4]\cdot[5]$; $[10] = [6]+[7]+[8]+[9]$

shape for the whole month, we can define the prices as an intersection point of the supply curve and the mean demand. The expected market clearing prices estimated for a particular demand level in the absence of import reductions from Russia to Nord Pool are presented in the Table 1. With varying demand from 39,500 to 51,000 the expected prices vary from 30 to 130 €/MWh. The slopes of the supply curve are calculated to estimate the effect of withholding on market-clearing prices:

$$\frac{\delta \lambda_i}{\delta V} = \frac{\lambda_{i+1} - \lambda_i}{Demand_{i+1} - Demand_i} \tag{8}$$

Thus, using Eq. (7) subject to constraints the optimal export capacity is presented for each expected price. The benefit of export capacity withheld was estimated using the Eq. (6).

The calculations were made for the risk taking trader whose decision is driven only by the expected prices. From the result of the calculation we can observe that if expected prices in Nord Pool in peak hours are 30–40 €/MWh, it is not profitable to export any electricity during the peak hours. The profit of the cross-border trader is simply the profit received by exporting electricity at full capacity during the off-peak hours. If the expected prices are 40–50 €/MWh, it is profitable to export 100–800 MW out of available 1,400 MW, and the optimal export volume is sensitive to the expected prices. When the expected prices are 50–130 €/MWh then it is profitable to export 800–1,200 MW. In this case the profit is driven by the extra revenue gain obtained through the sharp increase in Nord Pool prices due to the capacity withholding and the high slope of the supply curve.

4 Conclusion

The results show that the optimal amount of electricity to be transferred through the cross-border interconnection depends on the expected price and the slope of the supply curve in the import market and electricity and capacity prices in the export market. Drawback of current operating principle of the cross-border trade between the energy-only Nordic market and the energy + capacity Russian market is the decreased use of the cross-border interconnection. In our future work we will estimate how the current cross-border trade influences the welfare distribution between consumers and producers in both markets. To solve the problem and ensure the efficient use of transmission capacity, new trading arrangements are needed between the two markets.

References

1. Chaigneau, M., Lannez, S., Passelergue, J., Hesamzadesh, M.: Forward capacity markets: Maintaining system reliability in Europe. In: Proceedings of IEEE Conference Publications, 2012
2. Gore, M., Viljainen, S., Makkonen, M., Kuleshov, D.: Russian electricity market reform: deregulation or re-regulation. Energy Policy **41**, 676–685 (2011)
3. Gore, M., Viljainen, S.: Barriers and opportunities for new market players to entry the Russian power sector. In: Proceedings of IEEE Conference Publications, 2012
4. Fingrid.: Web pages. Available via. http://www.fingrid.fi/
5. Joscow, P., Kahn E.: A quantitative analysis of pricing behavior in California's wholesale electricity market during summer 2000. NBER Working paper series, 2001
6. Nord Pool Spot.: Web pages. Available via. http://www.nordpoolspot.com/
7. Parisio, L., Bosco, K.: Electricity prices and cross-border trade:volume and strategy effets. Energy Econ. **30**, 1760–1775 (2008)
8. Pfeifenberger, J., Spees K., Schumarcher.: A comparison of PJM's RPM with alternative energy and capacity market designs. The Brattle group working paper, 2009
9. SO-UPS Russia: planned monthly peak hours 2012. In: System Operator. Available via. http://www.so-ups.ru/fileadmin/files/company/markets/pik_chas2012.pdf Cited Jan 2011

Dynamic Portfolio Optimization for Power Generation Assets

B. Glensk and R. Madlener

1 Introduction

In finance, Markowitz's mean-variance portfolio theory has been a heavily used workhorse, which over the last years has also been increasingly applied to the valuation of power generation assets (for useful reviews, see e.g. [2, 10]); the first application in the energy domain dates back to 1976 [1]. Although standard MVP analysis has been successfully applied to asset allocation, mutual fund construction, financial planning, and even power generation mixes, it suffers from several limitations. One of them is its single-period character. Such models cannot capture the investors' stakes in long-term investment decisions. In contrast, multi-period models, if properly formulated, can solve these limitations and take advantage of volatility in values by rebalancing the asset mix. In order to address these important issues, investors from financial markets as well as energy utilities need such a multi-period optimization model for portfolio selection and investment planning. Specifically, they need to capture the uncertain changes of different market parameters over time. In this respect, multi-stage models provide a suitable framework.

The development of dynamic portfolio models and solving methods for financial markets date back to the late 1960s and early 1970s [13]. Unfortunately, the direct application of the mean-variance portfolio selection model to multi-stage portfolio problems is not trivial [9]. Different dynamic portfolio optimization approaches can be found in the literature. The backward-recursive procedure as a solution method for multi-period portfolio selection problems, presented by Mossin [13], was one of the first attempts at a transformation of the Markowitz mean-variance methodology

B. Glensk · R. Madlener (✉)
Institute for Future Energy Consumer Needs and Behavior (FCN), School of Business and Economics / E.ON Energy Research Center, RWTH Aachen University, Mathieustrasse 10, 52074 Aachen, Germany
e-mail: BGlensk@eonerc.rwth-aachen.de

R. Madlener
e-mail: RMadlener@eonerc.rwth-aachen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_26,

into a multi-period model. Recursive solution algorithms for portfolio selection were later presented by Östermark [15] and Steinbach [18]. Elton and Gruber [3] propose to solve a series of single-period portfolio problems in order to obtain the solution of a multi-period model. The planning processes of portfolios are extremely complex, multi-period, multiobjective ones, subject to uncertainty. Regarding all these points, different studies (e.g. [4, 6, 8] or [14]) put forward models which try to combine these aspects and propose possible solutions. Furthermore, the complexity of such models needs a special numerical method for finding the optimal solution, such as e.g the global optimization algorithm proposed by Maranas [11].

Concerning the application of dynamic approaches for energy planning, the literature is still scarce. Pereira and Pinto [16, 17] present an algorithm which can be used for the allocation of hydrothermal power systems. The authors describe the methodology, which allows for determining the generation target of power plants and minimizing the expected operating costs. The algorithm introduced is based on stochastic programming and Benders' decomposition. Another application of multi-period portfolio optimization is presented in Kleindorfer and Li [7], who characterize a multi-period, VaR-constrained portfolio problem and apply this model to real and contractual assets in the energy sector.

The aim of our paper is the application of a dynamic approach to power generation assets, in order to capture the continuously changing values of the economic and technical parameters considered when evaluating investments in power plants. Section 2 introduces the model specification applied, followed by an illustrative case study in Sect. 3 and some conclusions in Sect. 4.

2 Model Specification

The model presented below is based on the multi-stage stochastic optimization problem introduced by Mulvey et al. [14], which they applied to the portfolio of an insurance company. The application of this model allows for the reallocation of the portfolio assets. However, the specific character of portfolio selection problems for power generation assets requies some adjustments in comparison to the original model. Based on the model description provided in Mulvey et al. [14] and Maranas et al. [11], the dynamic portfolio optimization problem adopted for power generation assets can be described as follows.

For each power plant or real asset $i \in N$, time $t \in T$, and scenario $s \in S$, the following parameters and decision variables are defined.

Parameters:

$r_{i,t}^{s}$: uncertain return of technology i in period t, given scenario s,
α : parameter indicating the relative importance of variance as compared to the expected value; $\alpha \in \langle 0, 1 \rangle$,
$x_{i,max}$: maximal share of technology i in the portfolio,

q^s: probability that scenario s occurs among all possible scenarios S ($\sum_{s=1}^{S} q^s = 1$).

Decision variables:

$x_{i,t}^s$: percentage of technology i in time t, given s,
$a_{i,t}^s$: proportion of technology i added to the portfolio in time t, given s,
$d_{i,t}^s$: proportion of technology i removed from the portfolio in time t, given s.

Based on these definitions of parameters and variables, the model can be specified, with the following objective function:

$$\alpha R_{p,T} - (1-\alpha) Var(R_{p,T}) \rightarrow \max \tag{2.1}$$

subject to:

$$\sum_{i=1}^{N} x_{i,0}^s = 1, \quad \forall s \in S \tag{2.2}$$

$$\sum_{i=1}^{N} x_{i,t}^s = 1, \quad \forall s \in S, \quad t = 1, \ldots, T \tag{2.3}$$

$$x_{i,t}^s = x_{i,t-1}^s + a_{i,t}^s - d_{i,t}^s, \quad \forall s \in S, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N \tag{2.4}$$

$$0 \leq x_{i,t}^s \leq x_{i,max}, \quad \forall s \in S, \quad t = 0, \ldots, T, \quad i = 1, \ldots, N \tag{2.5}$$

$$x_{i,t}^s = x_{i,t}^{s'}, \quad \forall s \in S, \quad t = 0, \ldots, T, \quad i = 1, \ldots, N, \tag{2.6}$$

for all scenarios s and s' ($s \neq s'$ and $s, s' \in S$) with identical past up to time t.

In the objective function (2.1), the portfolio return $R_{p,T}$ is determined across all scenarios at the end of the planning horizon T, and given as follows:

$$R_{p,T} = \sum_{s=1}^{S} q^s R_{p,T}^s, \tag{2.7}$$

where $R_{p,T}^s$ defines the portfolio return for scenario s at the end of the planning horizon T, defined as:

$$R_{p,T}^s = \prod_{t=0}^{T} \left(\sum_{i=1}^{N} r_{i,t}^s x_{i,t}^s \right). \tag{2.8}$$

In the second part of the objective function (1), the portfolio's variance $Var(R_{p,T})$ denotes the variance across all scenarios at the end of the planning horizon T; it is specified as follows:

$$Var(R_{p,T}) = \sum_{s=1}^{S} q^s (R_{p,T}^s - R_{pT})^2. \tag{2.9}$$

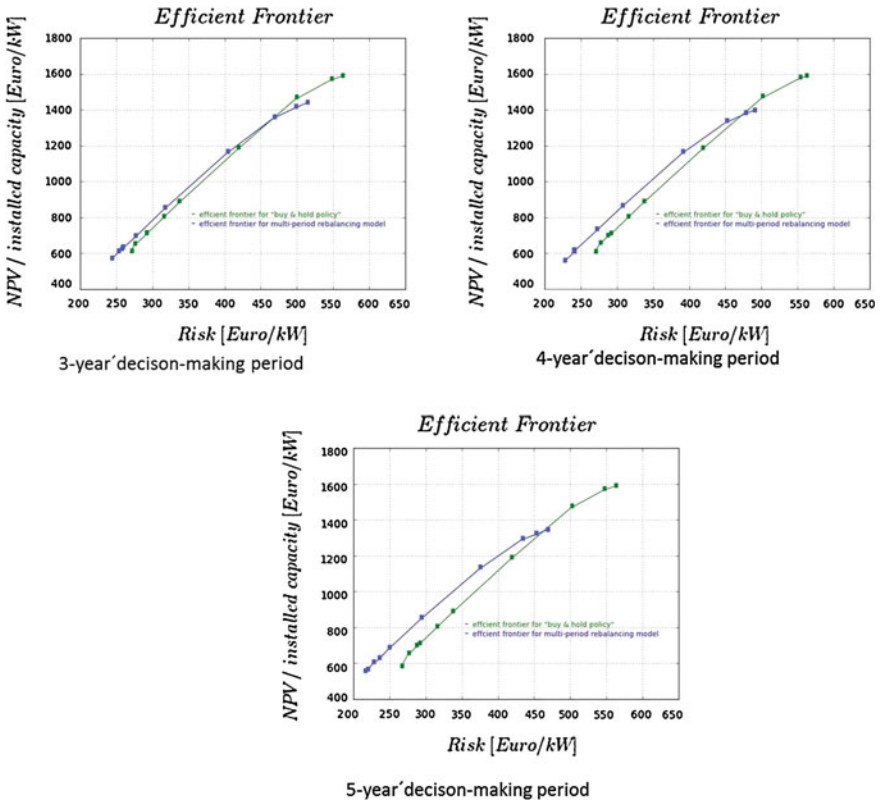


Fig. 1 Efficient frontiers for conventional and new offshore wind and CSP technology, comparison static versus multi-period rebalancing model (cf. [5])

3 Results of a Case Study

The case study presented here considers different power generation technologies already in use in Germany today (nuclear power plants, conventional fossil fuel power plants, as well as renewables including concentrating solar power/CSP and onshore and offshore wind power). The technical characteristics as well as the economic parameter assumptions are based on literature research (cf. [5]).

The analytical procedure used to compute efficient portfolios based on the multi-period portfolio selection model is composed of the following three steps:

(1) the historic time series of electricity, fuel, and CO_2 prices were used to calculate the volatility and cross-correlations of these parameters;
(2) a Monte Carlo simulation (conducted with the software Crystal Ball®) was run (100,000 runs) to compute the distribution of the NPV (in EUR/kW) for all technologies and the cross-correlations between technologies;

(3) the multi-period portfolio selection problem was used to generate the efficient portfolios for the analyzed market.

Because of the nonlinear and noncovex character of the presented model, the efficient frontier was obtained through the implementation of nonlinear programming algorithms in the dynamic object-oriented programming language Python 2.7. Finally, for analytical reasons, the efficient frontier for the so-called 'buy-and-hold' policy (static MVP optimization) was calculated and compared with the efficient frontier obtained for the multi-period model (see Fig. 1).

Figure 1 illustrates the efficient frontiers obtained for the static buy-and-hold policy and the multi-period rebalancing model considering a 3, 4, and 5-year decision-making period with rebalancing in each year. We observe an increasing distance between the efficient frontiers for the buy-and-hold policy in comparison to the multi-period rebalancing model (comparison of all three situations). It means considering dynamic aspects of portfolio selection problems for the same risk level, but where a higher return can be obtained. Furthermore, we notice the shift of the efficient frontier for the multi-period rebalancing model not only to the left but also downwards (i.e. portfolios that feature a smaller risk and smaller return). Moreover, comparing the three decision situations presented here, we notice that the distance between efficient frontiers for the buy-and-hold policy and the multi-period rebalancing model is larger when the number of periods considered increases. Nevertheless, because of the early stage of this study, the conclusion regarding a possible relation between the increasing number of periods and the location of the efficient frontier is preliminary only, and will be further investigated.

4 Conclusions

Optimization plays a very important role in both financial planning and risk management. As many practical studies have shown, optimization can be successfully applied. Applications in new areas, such as the energy sector, aim at making good use of already existing methods and models, and contributions to their further development. The results obtained in our study show that the consideration of portfolio rebalancing has a beneficial impact on the decision-making process and could improve the expected goals. Application of the multi-period rebalancing model allows for changes in the portfolio structure during the periods taken into consideration, and for interactively adding new technologies to the portfolio. Especially new power plant investments, with their often very long construction times, can be successfully integrated in the portfolio construction process using the multi-period rebalancing model. The proper restrictions built into the model allow for considering different maximum share constraints for each technology. That means that the rebalancing model can very well capture the new investments considered and measure their impact on the portfolio mix.

References

1. Bar-Lev, D., Katz, S.: A portfolio approach to fossil fuel procurement in the electric utility industry. J. Finance. **31**(3), 933–947 (1976)
2. Bazilian, M., Roques, F. (eds.): Analytical methods for energy diversity and security: a tribute to Shimon Awerbuch. Elsevier, Amsterdam (2008)
3. Elton, E., Gruber, M.: On the optimality of some multi-period portfolio selection criteria. J. Bus. **47**(2), 231–243 (1974)
4. Frauendorfer, K., Siede, H.: Portfolio selection using multistage stochastic programming. CEJOR **7**(4), 277–289 (1999)
5. Glensk B., Madlener R.: Dynamic portfolio selection methods for power generation assets. FCN Working Paper No.16/2011, RWTH Aachen University, November, (2011).
6. Gülpinar, N., Rustem, B.: Worst-case decisions for multi-period mean-variance portfolio optimization. Eur. J. Oper. Res. **183**(3), 981–1000 (2007)
7. Kleindorfer, P., Li, L.: Multi-Period VaR-Constrained Portfolio Optimization with Application to the Electric Power Sector. Energy J. **26**(1), 1–25 (2005)
8. Korhonen, A.: A dynamic bank portfolio planning model with multiple scenarios, multiple goals and changing priorities. Eur. J. Oper. Res. **30**(1), 13–23 (1987)
9. Li, D., Hg, W.: Optimal dynamic portfolio selection: Multi-period mean-variance formulation. Math. Finance **3**(10), 387–406 (2000)
10. Madlener, R.: Portfolio Optimization of Power Generation Assets. In: Rebennack, S., Pardalos, P.M., Pereira, M.V.F., Iliadis, N.A., Zheng, Q.P. (eds.) Handbook of CO_2 in power systems, pp. 275–296. Springer, Berlin (2012)
11. Maranas, C., Andrulakis, I., Floudas, C., Berger, A., Mulvey, J.: Solving long-term financial planning problems via global optimization. J. Econ. Dyn. Control **21**(7–8), 1405–1425 (1997)
12. Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. Blackwell, Cambridge MA and Oxford UK (1959)
13. Mossin, J.: Optimal multi-period portfolio policies. J. Bus. **41**(2), 215–229 (1968)
14. Mulvey, J., Rosenbaum, D., Shetty, B.: Strategic financial management and operations research. Eur. J. Oper. Res. **97**(1), 1–16 (1997)
15. Östermark, R.: Vector forecasting and dynamic portfolio selection: empirical efficiency of recursive multi-period strategies. Eur. J. Oper. Res. **55**(1), 46–56 (1991)
16. Pereira, M., Pinto, L.: Multi-stage stochastic optimization applied to energy planning. Math. Program. **52**(2), 359–379 (1991)
17. Pereira, M., Pinto, L.: Stochastic optimization of multi-reservoir hydroelectric system: A decomposition approach. Water Resour. Res. **21**(6), 779–792 (1985)
18. Steinbach, M.: Recursive direct algorithms for multistage stochastic programs in financial engineering, pp. 98–23. Konrad-Zuse-Zentrum fur Informationstechnik Berlin, Preprint SC (1998)

Optimizing Storage Placement in Electricity Distribution Networks

J. M. van den Akker, S. L. Leemhuis and G. A. Bloemhof

1 Introduction

With a growing focus on 'green' energy, the function of the electricity grid also changes. In the classical power grid, there were only central electricity producers (coal/gas/oil fired plants, nuclear plants) and consumers (homes and companies). Nowadays, decentral power generation systems, such as Combined Heat Power (CHP) systems, photovoltaic (PV) systems, windmills, also produce a part of the required electricity. While the production of the classical production facilities could easily be controlled, this cannot be done for the new energy sources. Especially PV systems, which are often applied in residential areas, will produce most of their energy during the middle of the day, at which time power consumption in these areas usually is low.

This causes large amounts of current transportation through the network, which may lead to overloading. In case of decentralized generation while demand is low, the voltages might be higher than the tolerated maximum. On the other hand, during times of high demand, the voltage at the location of the consumer may drop to a level that is below an acceptable level. A storage system can alleviate the problems since it can decrease the voltage by charging, or raise the local voltage by discharging.

J. M. van den Akker (✉)
Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
e-mail: J.M.vandenAkker@uu.nl

S. L. Leemhuis
Department of Information and Computing Sciences, Utrecht University and DNV KEMA Energy and Sustainability, 3584 CC Utrecht, The Netherlands
e-mail: S.J.Leemhuis@gmail.com

G. A. Bloemhof
DNV KEMA Energy and Sustainability, Utrechtseweg 310, 9035, 6800 ETArnhem, The Netherlands
e-mail: Gabriel.Bloemhof@dnvkema.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_27,

Besides prevention of overloading and of violation of the voltage limits, possibilities for trading, i.e., buy at low prices and sell at high prices, are positive effects of storage systems.

Energy networks with storage systems have been analyzed in [1] by a multivariate time series model. Models to calculate the optimal power flow for a given configuration of storage systems, based on three phase alternating current loadflow calculations, have been developed in [2, 3]. KEMA has already developed a software tool called PLATOS to optimize the numbers, types and locations of storage systems through a genetic algorithm (see [4]). Improving the computational efficiency of the included optimization technique is considered to be a next step.

In [5], we developed a model to determine the effects of decentral power generation. In particular, the model was designed to determine the maximum amount decentral power generation units that could be placed in a distribution network without causing a line to overload. The loadflow calculations were formulated as a deterministic linear programming problem. As a consequence the model could be solved very efficiently.

In this paper a method consisting of two hierarchical models is proposed to solve the storage location problem. The first model is called the *loadflow series model* and is based on the model developed in [5]. It will try to find a *storage operation plan* to minimize overloading of lines and to respect voltages bounds as much as possible. Also, it will use the excess capacity of the storage systems and the network to trade. By doing so, the model will show the added value of a storage configuration in a given network, i.e., with fixed storage systems.

Building upon this, we have a model called the *SLOPER model*, which is designed to optimize storage configurations with respect to added value, i.e., it computes an optimal or very good location plan of storage systems in the network. This model is based on a *local search* algorithm to find promising storage location plans. To evaluate such a plan, the SLOPER model will use the loadflow series model as a subroutine It will look for a balance between investment cost and added value to determine the best solution to the storage location problem.

2 Loadflow Series Model with Storage

In this section we introduce a linear programming model for calculating a loadflow series for a network with fixed storage systems. We model the power network as a graph. The set of vertices V consists of the set of locations where electricity is produced/consumed and possibly storage units are located (mainly households) together with the set of connection points. The set of edges E represents the pairs of vertices directly connected by a cable or line. We divide time into a discrete set of intervals called timeframes.

Variables

The variables are

- $v_{i,t}$: the voltage in vertex i in timeframe t
- $f_{i,j,t}$: the flow of current from i to j during timeframe t, flow in one direction is negative flow in the opposite direction so $f_{i,j,t} = -f_{j,i,t}$.
- $q_{i,t}$ is the actual current production by decentral power generation and power consumption at vertex i during timeframe t
- $s^c_{i,t}$: the stored amount of energy in vertex i at the end of timeframe t
- $s^+_{i,t}$: the amount of current used for charging the storage system at vertex i during timeframe t.
- $s^-_{i,t}$: the amount of current used for discharging the storage system at vertex i during timeframe t.
- $g_{i,t}$ the amount of current produced at the grid connection, i.e., the amount of current taken from outside the network (usually from the medium voltage grid), at vertex i during timeframe t. A negative amount means that current is sent to the outside grid.

If the vertex does not contain a storage system, the storage variables are zero by definition. The same is true for the grid connection variables for vertices not located at the grid connection. We are also going to use some penalty variables to represent constraint violations, but we will introduce these together with the corresponding constraints.

Constraints

The total net production of current of a vertex is the sum of the net production of current of the load (decentral generation and consumption), grid connection and storage at this vertex. We assume constant current load, which implies that the load production at vertex i equals $\frac{P_{i,t}}{V_N}$, where $P_{i,t}$ is the produced power and V_N is the nominal voltage. So we have

$$\forall i \in V, \forall t \in T : q_{i,t} = g_{i,t} - s^+_{i,t} + s^-_{i,t} + \frac{P_{i,t}}{V_N}$$

We have Kirchoff's current law, i.e., the sum of the currents leaving the vertex should be equal to the total net production of current of the vertex:

$$\forall t \in T, k \in V : \sum_{i \in V} f_{k,i,t} = q_{k,t}$$

The voltage drop along each edge is the product of the flow of current along the edge and the resistance of the edge. In our model this ensures Kirchoff's voltage law

which states that the net voltage drop around a loop in the network should be zero. We obtain the following:

$$\forall (i, j) \in E, \forall t \in T : v_{j,t} - v_{i,t} = f_{i,j,t} \times R_{i,j}$$

The total energy stored in a storage system at a vertex at the end of timeframe t is the total energy stored at the end of timeframe t-1 increased with the energy charged during timeframe t and decreased with the energy discharged during timeframe t, taking into account losses in conversion of charging and discharging, and losses for storing the energy during timeframe t, represented by the factors e_i^+, e_i^- and e_t^C, respectively. L_t denotes the length of timeframe t. Hence

$$\forall i \in V, \forall t \in T - \{T_0\} : s_{i,t}^c = (e_i^c)^{L_t} \times s_{i,t-1}^c + (e_i^+ \times s_{i,t}^+ - \frac{1}{e_i^-} \times s_{i,t}^-) \times L_t \times V_N$$

In a similar way, we formulate the constraint that the total energy stored in a storage system at the end of the simulated time period should be the same as the total energy stored at the beginning of the simulated time period.

Bounds

The flow of current along each of the edges may not be larger than the rated current of the edge. However, we allow that the constraint is violated at a very large cost. This is expressed by the variables x_{ij} and is formulated as follows:

$$\forall (i, j) \in E, \forall t \in T : -I_{i,j}^{max} - x_{i,j} \leq f_{i,j,t} \leq I_{i,j}^{max} + x_{i,j}$$

The voltage at each of the vertices may never be larger than the maximum nominal voltage and never be smaller than the minimum nominal voltage. Again, we allow a violation of the constraint at at very large cost modeled by the variables y_i and obtain:

$$\forall i \in V, \forall t \in T : V_N^{min} - y_i \leq v_{i,t} \leq V_N^{max} + y_i$$

Finally, a storage system can never be charged more than its maximum charging power, it can never be discharged over its maximum discharge power and can never contain more or less stored energy than the minimum and maximum capacity of the storage system. Clearly, this results in bounds on the storage variables.

Objective

The objective consists of two parts. The first part is to minimize the energy $g_{i,t} \times V_N$ taken from the grid connection multiplied by their prices $k_{i,t}$. The second part is to minimize the cost involved with violating the constraints: for each edge minimizing

the violation of the rated current, and for each vertex minimizing the violation of the voltage bounds. Our objective is

$$\min \sum_{i \in V} \sum_{t \in T} g_{i,t} \times k_{i,t} \times V_N + \sum_{(i,j) \in E} X \times x_{i,j} + \sum_{i \in V} Y \times y_i$$

3 Optimization of Storage Locations

In the previous section a linear programming model for the calculation of the loadflow series with fixed storage systems was explained. In this section, we discuss two models to optimize storage locations that build upon the linear programming model.

A first idea is to extend the linear programming model to an Integer Linear Programming (ILP) model, which solves the storage location problem to optimality. This is achieved by creating a binary variable $d_{s,i}$ for each combination of a storage type and a location where a storage system can be placed to represent the decision that we do or dot not place a storage system of type s at location i. We enforce that there will be at most one storage type at a position. Moreover, we can express the storage capacity and the maximum charge and discharge rates as linear combinations of the variables $d_{s,i}$. Drawback of this model is that we cannot incorporate storage losses in a linear way. Under the assumption that there are no storage losses, the model can be solved for smaller networks, but takes a lot of computation time for larger instances.

To solve problems for larger networks, we developed a *simulated annealing* algorithm, which is called the SLOPER model. In each iteration of the algorithm, the locations of the storage systems are fixed, and the solution is evaluated through our loadflow series model. In this model, we simply adapt the bounds of the variables for the charge, discharge and capacity. To be able to solve the model faster, we reduced the number of variables by expressing the voltage and current variables in term of the storage variables through a matrix transformation. Further details are omitted for reasons of brevity.

The simulated annealing algorithm starts with an initial solution which can be either provided by the user or, by default, is a network without any storage systems.

The neighborhood space is defined by a number of simple actions:

- **Add**: A random type of storage system is added to a random position without storage.
- **Remove**: A storage system is removed from a random location.
- **Change**: The type of a storage system is changed randomly at a random location.
- **Swap**: Two locations are randomly selected and swap their storage systems.

In each iteration of simulated annealing, the LP model is slightly changed by adapting the bounds of variables, and hence the solution to the new LP can be calculated efficiently using the previous solution.

4 Computational Results and Conclusion

We have implemented our loadflow series modes, the Integer Linear Programming extension of this model, and the SLOPER model using Java and ILOG CPLEX 12.2.

Our loadflow series model is based on direct current approximations, which is quite different from the usual loadflow calculations methods based on three-phase alternating current physics. It was developed in this way to increase the efficiency of the calculation, such that more solutions can be examined by the SLOPER model during a period of time. The loadflows calculated by the loadflow series model were validated with Plexos, a tool for production planning for power plants, and PowerFactory, a power system analysis tool. Both are well accepted tools in practice. The developed loadflow series model proves to be an interesting and efficient approach to calculating optimal storage behaviour.

We tested the ILP model and the SLOPER model on small example networks. We tested different parameter settings of simulated annealing to identify the best one. It turned out that simulated annealing quite frequently found the exact optimum. For larger instances, we only applied the SLOPER model. SLOPER was able to find good solutions rather quickly. Although tests on real life networks still have to be executed, the approach of using a linear loadflow series model combined with a local search heuristic seems to be an efficient tool for solving investment problems for storage systems in electricity networks.

In our current model, overload is allowed at a very large cost in all time frames. However, the durations of overloads and the aging of the lines are very important. To include a more complex overload model is an interesting topic for further research. Other extensions are the inclusion of aging effects of storage systems and net losses. The computational efficiency could be improved by restricting the model to critical timeframes.

Finally, it would be very interesting to extend the model with more traditional methods of solving network problems, such as the addition of lines and voltage changes at the grid connection point.

References

1. Klockl, B.: Multivariate Time Series Models Applied to the Assessment of Energy Storage in Power Systems. Proceedings of Probabilistic Methods Applied to, Power Systems(PMAPS) (2008)
2. Mani Chandy, K., Low, S.H., Topcu, U., Xu, H.: A Simple Optimal Power Flow Model with Energy Storage. Proceedings of the 49th IEEE Conference in Decision and Control (CDC), pp. 1051–1057, 2010.
3. Gayme, D., Topcu, U.: Optimal power flow with distributed energy storage dynamics. The American Control Conference, San Francisco (2011)
4. Cremers, R., Bloemhof, G.: Storage optimization in distribution systems. Proceedings of CIRED, 21st International Conference on Electricity Distribution, Frankfurt (2011).
5. van den Akker, M., Bloemhof, G., Bosman, J., Crommelin, D., Frank, J., Yang, G.: Optimal distributed power generation under network load constraints Proceedings of the 72nd European Study Group Mathematics with Industry, Amsterdam (2010)

Optimizing Electrical Vehicle Charging Cycle to Increase Efficiency of Electrical Market Participants

Y. Hermans, S. Lannez, B. Le Cun and J.-C. Passelergue

1 Electric Vehicle Fleet Optimizer

In this section, we describe an Electric Vehicle Fleet Optimizer under a mathematical programming formulation which can be used to schedule electric vehicles charge at a minimal cost. First we present the data then the variables, the objective function and finally the model.

The optimizer requires a schedule of cars transportation reservation as data. Each reservation have a beginning date, an ending date and the minimal charge expected in the battery. Let's consider $V = \{1, \ldots, n\}$ where n is the number of electric vehicles considered and let's the schedule to be defined during p time intervals ($I_1 = [t_1;\ t_2], \ldots,\ I_p = [t_p;\ t_{p+1}]$). Each time interval has a duration τ. To simplify time notations let's define $K = \{1, \ldots, p\}$ and $\overline{K} = K \cup \{p + 1\}$. To help defining future notations, let's considering the following sample schedule:

Y. Hermans (✉)
PRiSM Laboratory, CNRS/UMR-8144, University of Versailles-SQY, Versailles, France

ECONOVING International Chair in Eco-Innovation, Paris, France
e-mail: yann.hermans@prism.uvsq.fr

S. Lannez
Alstom Grid Network Management Solutions, Paris, France
e-mail: sebastien.lannez@alstom.com

B. Le Cun
University Paris Ouest Nanterre La Défense, Paris, France

PRiSM Laboratory, CNRS/UMR-8144, University of Versailles-SQY, Versailles, France
e-mail: bertrand.lecun@prism.uvsq.fr

J.-C. Passelergue
Alstom Grid Network Management Solutions, Paris, France
e-mail: jean-christophe.passelergue@alstom.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_28,

v	t_2	t_3	t_4	t_5	t_6	t_7	
1			18				10
2		4			6		
3			24				
4				12			

where line i indicates when vehicle i is used for transportation and its minimal expected charge in the battery. In this example, we consider vehicles with 25 kWh batteries ($\forall v \in V,\ C_v^{\max} = 25$). From this schedule, we define some new data to help to model the problem. First we define $D_{v,k}$ to be equal to 1 if vehicle v is available for charging during the time interval $[t_k;\ t_{k+1}]$ and otherwise $D_{v,k} = 0$. Then we define $R_{v,k}$ to be equal to 0 except if vehicle v is booked for transportation at time t_k. In that case $R_{v,k}$ is equal to the quantity of energy requested by the associated reservation. Values for $D_{v,k}$ and $R_{v,k}$ associated to the previous sample are :

$D_{v,k}$	1	2	3	4	5	6	7
1	1	1	0	0	1	1	0
2	1	0	1	1	0	0	1
3	1	0	0	0	1	1	1
4	1	1	1	0	0	1	1

$R_{v,k}$	1	2	3	4	5	6	7	8
1	0	0	18	0	0	0	10	0
2	0	4	0	0	6	0	0	0
3	0	24	0	0	0	0	0	0
4	0	0	0	12	0	0	0	0

Some more data are required to express the charging speed constraints. $c^+_{v,\tau}$ is the maximal quantity of energy that could be added to the vehicle v battery during a time τ (i.e. during two consecutive times). Similarly $c^-_{v,\tau}$ is the maximal quantity of energy that could be extracted from the battery during the same time period. This data is not necessarily equal to zero since we allow the energy exchanges between the vehicles fleet. Finally we define λ_k to be the electricity unit price during the time interval I_k.

We now define the variables used in our model:

- $C^{\text{eff}}_{v,k}$ represents the state of charge of vehicle c at time t_k;
- $E^{\text{A2V}}_{v,k}$ expresses the quantity of energy transfered to vehicle v during I_k. This quantity could de negative when vehicle v restitute energy to others;
- E^{G2A}_k stands for the energy bought to charge vehicles during I_k.

Our goal is to minimize our energy consumption bill (by time-shifting vehicles charges). This bill depends of the energy consumption on each interval I_k quantified by E^{G2A}_k and of the electricity unit price on this interval: λ_k. We formulate the bill value to be $f(E^{\text{G2A}}_k) = \sum_{k \in K} E^{\text{G2A}}_k.\lambda_k$. This function will be modified in the next section to allow our model to interact with a BRP.

All theses definitions lead us to the following mathematical program:

$$\min\ f\left(E^{\text{G2A}}_k\right)$$

$$\forall k \in K, \quad E_k^{\text{G2A}} \geq 0 \tag{1}$$

$$\forall k \in K, \quad E_k^{\text{G2A}} = \sum_{v \in V} E_{v,k}^{\text{A2V}} \tag{2}$$

$$\forall (v,k) \in V \times K, \quad C_{v,k+1}^{\text{eff}} = C_{v,k}^{\text{eff}} + E_{v,k}^{\text{A2V}} - R_{v,k} \tag{3}$$

$$\forall (v,k) \in V \times \overline{K}, \quad R_{v,k} \leq C_{v,k}^{\text{eff}} \leq C_v^{\max} \tag{4}$$

$$\forall (v,k) \in V \times K, \quad E_{v,k}^{\text{A2V}} \leq c_{v,\tau}^{+}.C_{v,k}^{\text{eff}}.D_{v,k} \tag{5a}$$

$$\forall (v,k) \in V \times K, \quad E_{v,k}^{\text{A2V}} \geq -c_{v,\tau}^{-}.C_{v,k}^{\text{eff}}.D_{v,k} \tag{5b}$$

where:

- (1) imposes that the fleet only consume energy;
- (2) ensures that all the bought energy goes in a vehicle battery and that vehicles can exchange energy;
- (3) is used to describe the state of charge evoluction law;
- (4) ensures the state of charge to be greater than transportation needs and lower than the battery capacity;
- (5a) and (5b) impose a vehicle not to have energy transfers when its used for transportation $\left(D_{v,k} = 0 \Rightarrow E_{v,k}^{\text{A2V}} = 0\right)$ and to respect batteries charging speed when $D_{v,k} = 1$, $\left(\Rightarrow -c_{v,\tau}^{-} \leq E_{v,k}^{\text{A2V}} \leq c_{v,\tau}^{+}\right)$.

2 Demand-Response Scheme with a BRP Optimizer

In this section we present how the BRP and the EVRS Optimizers interact. The demand-response scheme is explicited by Fig. 1 where exponents (i) indicate the demand-response iteration number.

The main idea is that EVRS, based on its energy buying contract and the reservations of its vehicles, uses EVFO to obtain a first load forecast of the vehicle fleet. Then the BRP Portfolio Generation Optimizer takes into account the EFVO load forecast and some possible operational bias (due to power generation failures for example) to schedule and dispatch its generation portfolio. We made the assumption that the BRP Portfolio Generation Optimizer knows its marginal generation price for each production period. After the scheduling phase, the BRP determines its marginal generation price. For each time period, the BRP determines a price reward offered to all shifted loads that are willing to increase their benefits. The price reward is calculated in such a way that it ensures load reduction when needed under the constraints defined in a Rebates Program contract. In this study, when consi-dering a load shifting from a time period I_d to I_e, we define the EVRS benefits due to its load shifting to be the difference between the rebates provided for reducing a load during I_d and the one during time I_e (for each energy unit shifted). The main idea

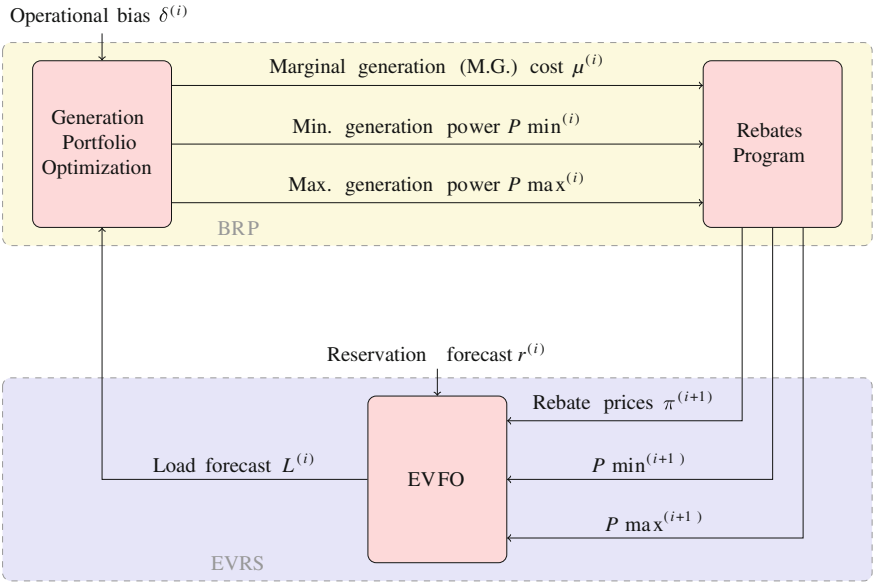


Fig. 1 Interaction scheme

is to motivate the EVRS not only to shift unwanted loads from I_d, but also to shift it to the best (economically speaking) time period I_e.

Moreover the BRP may want to limit the EVRS energy consumption variation. Indeed, high consumption during some time periods may lead the BRP Portfolio Generation Optimizer to turn on a new power plant which may prove to be very expensive. On the contrary, to increase power plants profitability, the BRP may request to have a high consumption in a time period. This is mainly due to the fact that power plants are less efficient when producing low energy. To help the BRP controlling those aspects, we defined $P\,\max_k$ to be the maximal EVFO energy which could be added to the initial consumption during $I_k = [t_k; t_{k+1}]$. By analogy, $P\,\min_k$ is the maximal EVFO energy which could be removed. This results in $P\,\max = (P\,\max_1, \ldots, P\,\max_p)$ and $P\,\min = (P\,\min_1, \ldots, P\,\min_p)$.

Then $P\,\min$, $P\,\max$ and rebates π, are communicated to the EVRS which calls EVFO to reajust its load forecast. To take into account the information provided by the BRP, we have to modify EVFO :

- To take into account rebate prices, we modify the objective function to be :

$$f\left(E^{\mathrm{G2A}}\right) = \sum_{k\in K} F_k^{\mathrm{G2A}}.\lambda_k + \sum_{k\in K}(E_k^{\mathrm{G2A}} - F_k^{\mathrm{G2A}}).\pi_k$$

  where π_k is the rebate price offered by the BRP and F_k^{G2A} the initial EVFO forecast during I_k. This amounts to minimize $\sum_{k\in K} E_k^{\mathrm{G2A}}.\pi_k$.

- To integrate P min and P max values, we add the following constraint set:

$$\forall k \in K,\ -P\min_k \leq E_k^{\mathrm{G2A}} - F_k^{\mathrm{G2A}} \leq P\max_k$$

This ensures the quantity of energy added and/or removed from the initial load forecast F_k^{G2A} to satisfy to the BRP wishes about consumption range.

3 Demand-Response Scheme Convergence and Benefits for Everyone

In this section we explain and demonstrate that under some hypothesis we can guarantee the convergence of the demand-response scheme explicited in the previous section. We impose the rebates to be proportional to the BRP benefits ($\pi_k = \alpha.\mu_k$ with $\alpha \in]0, 1[$).

First we consider that after a certain number of iterations, no modifications in the operational bias or changes in the reservation forecast occurs. Then we hypothesize that the BRP always have the same behaviour when having the same data. Finally we impose $P\min_k$ and $P\max_k$ to be positive.

Under these hypothesis, any EVFO execution has in its feasible domain the previously decided load forecast. This is due to the fact that since $P\min_k$ and $P\max_k$ are positive, constraint 2 allows F_k^{G2A} to be equal to E_k^{G2A}. As a consequence, an EVFO execution cannot provide a more expensive value than the one decided at the previous iteration.

Concerning the BRP side, we assume that the BRP has given to the EVRS its marginal costs as rebates and that $P\min_k$ and $P\max_k$ are the ranges of marginal costs validity. In that case, we ensure that without operational bias, the BRP Portfolio Generation Optimizer may keep its decisions without affecting the promised rebates and its marginal costs. As a consequence, the BRP can conserve its decisions. If it does not, it implies that the new decisions are better for him in terms of marginal costs and as a consequence better for the EVRS side since $\pi_k = \alpha.\mu_k$.

In each case the EVRS and the BRP increase their benefits together since the BRP benefits and the rebates are proportional. Since the quantity of benefits are always limited, those inscreasing sequences converge.

4 Conclusions

We present an new approach to improve coordination of BRP activities and large loads in a smart grid context. The proposed approach uses concepts from Game Theory to design welfare improving mechanism which is proved to converge to a Nash equilibria, but offers some information about marginal production costs to the customers.

References

1. Hermans, Y., Le Cun, B., Bui, A.: Modéle d'optimisation basé sur le Vehicle-to-grid pour limiter l'impact des pics de consommation électrique sur la production. Archives HAL (2011). http://hal.archives-ouvertes.fr/hal-00577915/fr/
2. Passelergue, J.-C.: Near real-time economic scheduling and dispatch within the new generation control organizations and rules. EPCC - International Workshop On Electric Power Control Centers, Dublin, Ireland (2009). http://www.epcc-workshop.net/archive/2009/final_tech_programme.html
3. Hermans, Y., Le Cun, B., Bui, A.: Individual decisions & schedule planner in a Vehicle-to-grid context. Electric Vehicle Conference (IEVC), 2012 IEEE International. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6183275
4. Spees, K., Lave, L.: Demand Response and Electricity Market Efficiency. Electr. J. **20**(3), (2007). doi: 10.1016/j.tej.2007.01.006
5. Albadi, M.H., El-Saadany, E.F.: A summary of demandresponse in electricity markets. Electr. Power Syst. Res. **78**(11), (2008). doi: 10.1016/j.epsr.2008.04.002

Part VII
Financial Modeling, Banking and Insurance

Near Term Investment Decision Support for Currency Options

Rouven Wiegard, Cornelius Köpp, Hans-Jörg von Metthenheim and Michael H. Breitner

1 Introduction

Our paper focuses on foreign exchange (FX) futures and options on this future contract (also called currency future options). An FX future (or forward) is a binding contract to exchange one currency for another at a specified date in the future. Currency options are often traded OTC rather than on exchanges. Interbank deals are common. There are two basic types of options. A call option gives its owner the right to buy the underlying at or up to a certain date at a specified price. A put option gives its owner the right to sell the underlying at or up to a certain date at a specified price. Derivatives on the OTC market have no standardized contract specifications and are designed manually and tailored to customers' needs [2]. Generally, derivative experts use information systems, in particular DSS, to determine the value of derivatives. Typically, a derivate trader must be able to come up with a price within a few minutes. Otherwise the customer will look elsewhere. The decision to buy or sell a currency option and at what price therefore often happens very quickly. Our research question is: "How can a Financial Decision Support System (FDSS) help to approximate options' market prices, and determine suitable trading times for options in the short term?" Decision making is the process of developing and analyzing alternatives, and then selecting from the available alternatives. The paper focuses on finding the optimal time to buy or sell an option. We develop an FDSS based on

R. Wiegard (✉)
Institut für Wirtschaftsinformatik, Leibniz Universität Hannover, Hannover, Germany
e-mail: wiegard@iwi.uni-hannover.de

C. Köpp
e-mail: koepp@iwi.uni-hannover.de

H. von Metthenheim
e-mail: mettenheim@iwi.uni-hannover.de

M. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_29,

artificial neural networks [5] that indicates the optimal time to trade. Some models primarily address perceived weaknesses of models that are in use today by (financial) decision makers [1, 3, 4, 6]. Our approach aims to extend existing models by not only calculating the fair option price, but also determining the optimal time at which this option should be bought or sold.

2 Decision Problem and Data

Our decision problem is to find the optimal trading time in the near future (+5 to +30 min). The lower limit allows a time frame of several minutes for decision making. The trading time decision depends on the dynamics of the midpoint, so a forecast of the midpoint supports the decision making. We assume that all tick data up to the current time is available. The forecast can be updated each time it receives a new data tick. Our decision problem is therefore of interest for buyers and sellers of options. Both are interested in finding good prices within a short time-frame.

We collected data from the Chicago Mercantile Exchange between Wednesday, October 12th, 2011 and Thursday, October 20th, 2011. This method yields individual prices for every last completed trade of the December-future underlying for 125,000 € per contract in the specified time frame. 8,084 bid prices and 6,725 ask prices for the future option were collected. Every observation was marked with a time stamp with millisecond accuracy. We focus on futures and options on futures. For these two assets, different kinds of data types are generally available from exchanges or OTC traders:

Bid and ask: prices at which the counterparty is instantaneously willing to trade. The bid/ask is the price at which the counterparty will buy/sell an asset. The bid price is less than the ask price.

Midpoint: to avoid unwanted artifacts in the data, we compute the mean of bid and ask. Otherwise, the time-series tends to oscillate between bid and ask prices. The observed price changes are not real, but are a manifestation of the bid-ask bounce.

Last: the option market is not very liquid, but the underlying future market is very actively traded. The last transactions provide better data than bid or ask prices. Taking last prices in the option market is not useful because there are not enough trades to provide a meaningful basis for learning.

We have data sources for ask, bid and last ticks. Each contains an assignment of the form "time → value". We merge them into one data structure with a common time base. This results in an assignment of the form "time → tick type → value, time". As ticks of different types occur asynchronously in nearly all times, the data structure is incomplete. We fill the gaps with the value of the last previous tick of the same type and copy the time occurrence in order to obtain the original incoming time and allow a distinction of originally occurring tick values. So the data structure is complete starting at the time when a tick of each type has occurred at least once. For each time, it contains the latest information available for each tick data type at the same time. We calculate the midpoint for each time t as following:

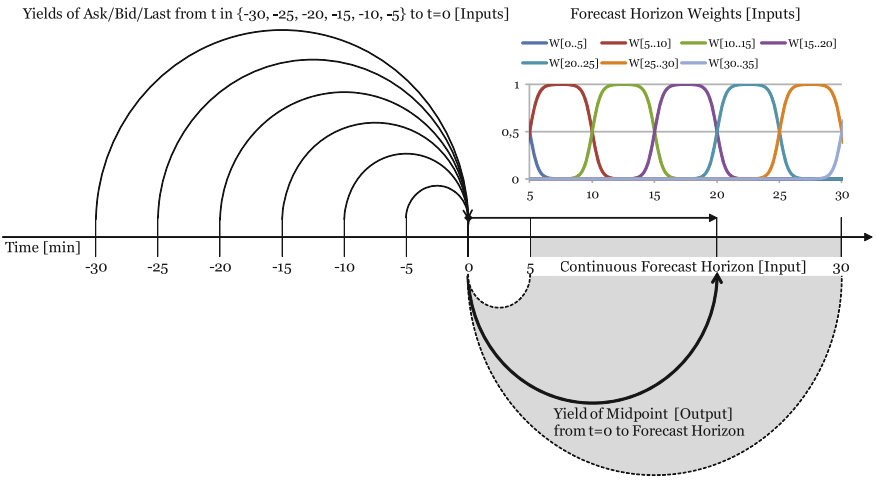


Fig. 1 Neural network inputs and outputs

$$midpoint(t) := \frac{ask(t) + bid(t)}{2}$$

This value changes for every new incoming tick of the types ask and bid. We insert the midpoint value and the time of last change in the created data structure.

3 Neural Network Forecast Model and Training

We trained a neural network forecast model (three-layer perceptron with one hidden neurons) based on the processed data as previously described. The model uses the available tick data information to forecast the future value of the midpoint of the option. A special feature of the model is the continuous forecast time horizon. Unlike other models, the forecast is not set to a fixed time period. We set the forecast horizon to a time range of five to 30 min in future, as described in the Sect. 2.

For input data from the past, we used a time horizon from -30 to -5 min. This time span was selected according to the forecast horizon. To deal with the tick data that occurs at different times, we divided the past time into a frame of five minutes. In the following we used the last tick information of each type available at the points in this time frame. Instead of using the raw data we compute the yield to make our series stationary:

$$yield_{ticktype}(-t) := \frac{value_{ticktype}(0)}{value_{ticktype}(-t)} - 1$$

We created input for the forecast horizon in two ways: first we use the forecast horizon as a direct input in a millisecond time base. This is the same time base that is used for tick data collection and data processing. Next we used a fuzzy weighted coding of time segments in the intended forecast horizon range. The time segmentation is derived from the five minutes raster of past data. So there are five time segments of five minutes in length within the intended forecast horizon range. We added two segments just before and after the intended forecast horizon range. Each of the seven forecast horizon weights is calculated by a combination of tanh()-functions:

$$W[t_{begin}..t_{end}](t) := \frac{\tanh((t - t_{begin}) \cdot 1.5) - \tanh((t - t_{end}) \cdot 1.5)}{2}$$

We used two hours of the processed data for training and validation and selected approximately the last 20 % of the data for validation and control network quality with early stopping: when the validation error increases twice in error, training stops. For output target generation we considered all tick times within the intended forecast horizon range after the selected tick and randomly selected 2.5 % of the considered future ticks to restrict the amount of training and validation data to a usable size.

4 Evaluation and Discussion

We select the best five networks from neural network training according to the total error on the training set. Initially, we trained 10,000 networks with randomly initialized weights. Figure 2 shows a typical forecast result on the out-of-sample data set. The upper part of the figure illustrates the forecast error as a heat-map of the error value. Ideally, forecast error should be zero. The lower part of the figure shows the actual data. Bid and ask is recorded data, the midpoint (shown as line) is calculated as above. The horizontal axis depicts ongoing time, and the vertical axis measures the forecast horizon. We notice that the forecast error often meanders around zero. In most cases, the forecast accurately hits the target midpoint. Figure 2 also shows what happens when a shock hits the system: approximately around the 80th minute (in the lower part of the figure) the ask price explodes. The effect is that the midpoint increases correspondingly. This is not an outlier (or a measurement error). The ask price indeed moved by more than 25 %. As the option market is not very liquid, short-term imbalances can lead to such spikes and are quite common. This very sharp price spike caused significant forecast errors, as the four distinct lines in the upper part of the figure show. The diagonal line is a pre-forecast error. The neural network could not anticipate this particular price spike and forecasts the 80th minute wrong until the event. After the event, the price spike becomes history and negatively influences forecast performance every five minutes (when the event is present in the rolling window). After 30 min (not shown) the event disappears from history and the neural networks forecasts smoothly again. When we use our model as an FDSS, we can expect that such price spikes will eventually happen. It is important

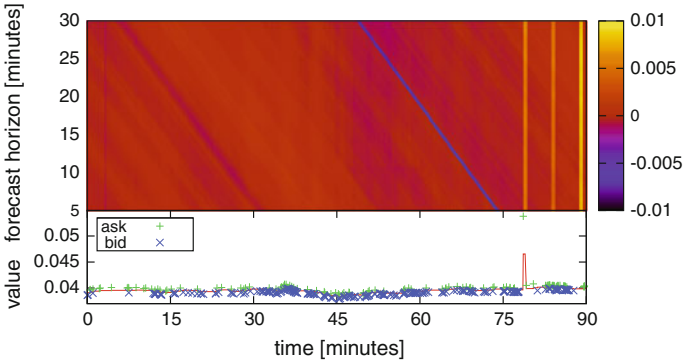


Fig. 2 Forecasting error on out-of-sample data set by time and forecast horizon

to be aware of filter-techniques to mitigate the effects of these spikes. We could, for example, only "accept" a data point into the historical price series if it is confirmed by at least two neighboring points. While filter-rules are arguably very important for practical application of FDSS, the rules are also somewhat arbitrary. They increase the complexity of the system and obscure the underlying core (a neural network). In the first step, our goal is to evaluate whether the results are sensible in an FDSS context. Nevertheless further investigations should deal with spike handling.

5 Conclusion and Further Research

We present steps towards a FDSS for currency option pricing. We investigate how an FDSS can help in pricing options. Our FDSS is useful in the sense that it provides accurate forecasts for the realized EUR|USD currency option prices within a continuous forecast horizon of five to 30 min. With the help of the FDSS forecast an option trader can choose whether to transact now or later (within the next thirty minutes). At the core of our FDSS is a neural network that we train on tick data. It only uses price data of the options and the underlying future. Especially, we do not use volatility or interest rate data like other approaches towards option pricing. This leads to a manageable model, even with tick data. The continuous forecast horizon is a highlight of our model: the desired forecast into the future is an input to model. This allows a fine-grained view on the probable development of the currency option price. We evaluate our FDSS on a single-tick data price series. While the results are encouraging, we also have to take the following limitations into account. It is important to note that a single price series cannot validate the FDSS for general use. Therefore an evaluation with several other time series is of interest. We should at least analyze more different currency pairs, with different expirations of the underlying futures at different points in time. Preliminary tests show that the model generalizes well to other time series. We are therefore confident that using other assets will lead

to satisfactory results. Like in every modeling effort, the result depends significantly on the inputs. We try to keep the inputs as simple as possible and only use price information. While this approach reduces complexity, it may also be that it reduces model performance. We explicitly do not use typical input parameters from other option pricing models like volatility and interest rates. It would be interesting to gauge results with these additional input factors. Also, it is possible to incorporate different (but related) price series.

References

1. Bookstaber, R.A.: Demon of our own Design: Markets, Hedge Funds, and the Perils of Financial Innovation. Wiley, Hoboken, NJ (2007)
2. Breitner, M.H., Burmester, T.: Optimization of european double-barrier options via optimal control of the Black-Scholes equation. In: Chamoni, P. (ed.) Operations Research Proceedings, pp. 167–174. Springer, Heidelberg (2002)
3. Khandani, A.E., Lo, A.W.: In: Journal of Investment Management. What happened to the quants in August 2007? **5**(4), 5–54 (2007)
4. Laidi, A.: Currency Trading and Intermarket Analysis: How to Profit from the Shifting Currents in Global Markets. Wiley, Hoboken, NJ (2009)
5. Li, Y., and Ma, W.: Applications of Artificial Neural Networks in Financial Economics: A Survey. In: Proceedings of the International Symposium on Computational Intelligence and Design (1), pp. 211–214. (2010).
6. Turban, E., Sharda, R., Delen, D., Aronson, J.E., Liang, T.-P., King, D.: Decision Support and Business Intelligence Systems. Prentice Hall, New Jersey (2010)

An Efficient Method for Option Pricing with Finite Elements: An Endogenous Element Length Approach

Tomoya Horiuchi, Kei Takahashi and Takahiro Ohno

1 Introduction

The price of a financial derivative is formulated as a solution of a partial differential equation (PDE). Since it is difficult to solve for most of the PDEs explicitly, they are implemented using several numerical techniques. A typical approach is a numerical analysis of the PDEs such as the finite difference method (FDM) [4] and the finite element method (FEM) [2]. However, they have a drawback in that the actual computation time becomes exponentially longer as the dimension of the PDE increases. Furthermore, there occurs a discretization error, the difference between the approximate solution and the exact solution when using these methods. Therefore, the domain with high curvature may have a large error and then need to be divided into smaller meshes. In option pricing, this issue corresponds to evaluating options which have high curvature, such as a barrier option.

Since the FDM uses the static domains to discretize, the user has to divide the entire domain into smaller ones. However, because the FEM employs arbitrary elements, the user can divide only a domain with high curvature into smaller elements to reduce computation time.

In this study, we provide an endogenous method that determines element lengths depending on the curvature of the PDE. By using this algorithm, we are able to refine the element with high curvature and coarsen the element length with low curvature at each time step.

T. Horiuchi (✉)
Department of Business Design and Management, Graduate School of Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
e-mail: thoriuchi0531@gmail.com

K. Takahashi · T. Ohno
Department of Industrial and Management Systems Engineering, School of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
e-mail: k-takahashi@aoni.waseda.jp
e-mail: ohno@waseda.jp

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_30,

2 Method

2.1 Option Pricing with Finite Elements

The Black-Scholes PDE can be written as

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r-q)\, SV_S - rV = 0, \tag{1}$$

where $V = V(t, S)$ denotes option value, S is the price of the underlying asset, t is time, σ is volatility, r is the interest rate and q is the dividend rate. After using the log transformation $x = \log(S/K)$, the value function $w(t, x) = V(t, S)$ is

$$w_t - \frac{1}{2}\sigma^2 w_{xx} + \left(r - q - \frac{1}{2}\sigma^2\right) w_x - rw = 0, \tag{2}$$

where K denotes the strike price. In this study, only the spatial domain is discretized by the FEM; the time domain is discretized by the FDM (the Crank-Nikolson method). We assume linear interpolation function between two nodes:

$$w(x) = w_1 \left(1 - \frac{x}{L}\right) + w_2 \left(\frac{x}{L}\right), \tag{3}$$

where w_1, w_2 are the option values on each node and L is the length of the element. Equation (3) means that option values between the nodes are linearly approximated. The Galerkin method requires each element to satisfy

$$\int_0^L \left(w_t - \frac{1}{2}\sigma^2 w_{xx} + \left(r - q - \frac{1}{2}\sigma^2\right) w_x - rw\right) \phi(x)\, \mathrm{d}x = 0, \tag{4}$$

where $\phi(x)$ is the weighting function. By integrating the second-order derivative w_{xx} to eliminate, Eq. (4) as a matrix form results in

$$\left(\frac{1}{\Delta t}\mathbf{B} + \frac{1}{2}\mathbf{A}\right) \{u\}_{t=t-\Delta t} = \left(\frac{1}{\Delta t}\mathbf{B} - \frac{1}{2}\mathbf{A}\right) \{u\}_{t=t}, \tag{5}$$

$$\mathbf{A} \equiv \frac{\sigma^2}{2L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{r-q}{2} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} - \frac{rL}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \mathbf{B} \equiv \frac{L}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{6}$$

where Δt is the time step. By assembling the above elemental equation and inserting the boundary conditions, we obtain the global equations. These equations are described as a system of equations. In order to derive the option prices at $t - \Delta t$ from those at t, the calculation of the inverse matrix of the left-hand side is needed. In this study, we employ the projected SOR method to valuate the American options.

2.2 Mesh Regeneration

In this section, we introduce endogenous mesh regeneration at each time step. Although there are many studies on option pricing using the FEM (e.g., [3, 5, 6]), they use a priori-determined element lengths. Therefore, attempting various combinations of element lengths is needed to gain more precise solutions. Some studies employ mesh adaption approach; however, it takes more computation time because of its iteration [1]. The regeneration consists of two algorithms, element coarsening and element refinement.

Mesh coarsening is used to expand the element length if the local domain has low curvature. In this study, we calculate the sum of the curvature R of the three neighboring nodes. The curvature of node R_i is given by

$$R_i = \frac{w_{xx}}{\left(1 + (w_x)^2\right)^{3/2}}. \tag{7}$$

If the curvature is lower than the coarsening threshold R_1^*, the center node is extinguished at that time step. The extinguished node will no longer used in the calculation until it is refined by the mesh refinement we mention below.

The mesh refinement is used to redivide elements with high curvature. Since the Black-Scholes PDE is a type of diffusion equation, the domain with high curvature changes as time progresses. Hence, the extinguished node is refined if the curvature is higher than the refinement threshold R_2^*.

3 Numerical Experiment

We apply our method to valuing a European up-and-out call option and an American put option by using the parameters shown in Table 1 and 2. Accuracy is measured to compare with the Feynman-Kac solution[1] [4] in the European option, and with the

Table 1 Parameters for the European up-and-out call option

Parameter	Value
Strike price	50
Up-and-out barrier	110
Rebate	0
Interest rate	0.05
Dividend rate	0.0
Volatility	0.2
Maturity	0.5 year

[1] The cumulative distribution function of the Gaussian is calculated using the Gauss-Lugendre algorithm.

Table 2 Parameters for the American put option [2]

Parameter	Value
Strike price	10
Interest rate	0.12
Dividend rate	0.0
Volatility	0.5
Maturity	0.5 year

Table 3 Results for the European up-and-out call option given the time step $\Delta t = 0.001$, the initial number of elements $N = 500$, $R_1^* = 0.0001$, and $R_2^* = 0.001$

	FEM without regeneration	FEM with coarsening	FEM with coarsening and refinement
Computation time (s)	12.89	0.654	5.675
RMSE	0.007	4.037	0.302

FDM solution in the American option. The difference is defined as the root-mean-square error (RMSE). The dynamics of the underlying asset are given by geometric Brownian motion.

First, the European up-and-out call option is computed. Table 3 shows the effect of the mesh regeneration.[2] By implementing the coarsening, the experiment time is reduced drastically while the accuracy becomes worse. However, the refinement displays improvement in the precision. This is because the method relocates the nodes to the domain with high curvature. Secondarily, the American put option is computed. The mesh regeneration reduces the computation time along with the barrier option, while the accuracy remains at a low level.

Figure 1 shows the lengths of elements at each time step with the regeneration. As seen in Fig. 1, the domain near the strike price and the upper barrier remain fine lengths, while other areas are coarsened.

Figures 2 and 3 show the change of the RMSE and the computation time, respectively, as the number of elements increases. The RMSE is improved with the increase of the elements at a certain level. However, too many elements lead to deterioration of the RMSE. This is because the discretization error of Eq. (3) decreases, while the error due to the regeneration increases. The actual computation time increases as number of the elements are incremented (Fig. 3). In this study, the calculation of the inverse matrix using projected SOR features the most CPU-expensive mathematics. Hence, experiment time rises as a function of the elements.

[2] We used the GCC version 4.4.4 on an Intel(R) Xeon(R) with 3.47 GHz under Linux, and O3 as an optimize option.

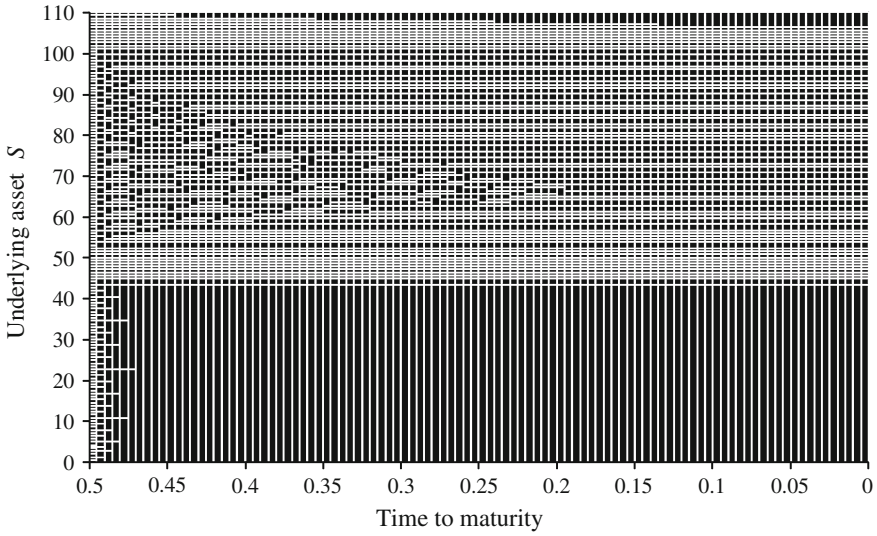


Fig. 1 The element regeneration for the European up-and-out call option using coarsening and refinement given by Table 1 when $\Delta t = 0.005$, $N = 150$, $R_1^* = 0.001$, and $R_2^* = 0.005$. *The black object*, *the horizontal line*, and *the vertical line* represent one element, one node, and the time step, respectively

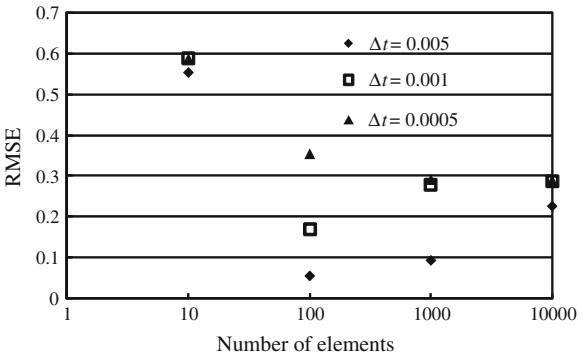


Fig. 2 The RMSE versus the number of elements with mesh regeneration

4 Conclusion

We introduced an endogenous method of the FEM to the option pricing to determine an efficient element length at each time step. In order to implement mesh regeneration, we employed two remeshing algorithms, coarsening and refinement. By using our method, the model can alter the length of the elements depending on its curvature.

The characteristics of the PDE change depending on whether diffusion or convection dominated; our approach worked better for the former. Hence, further inves-

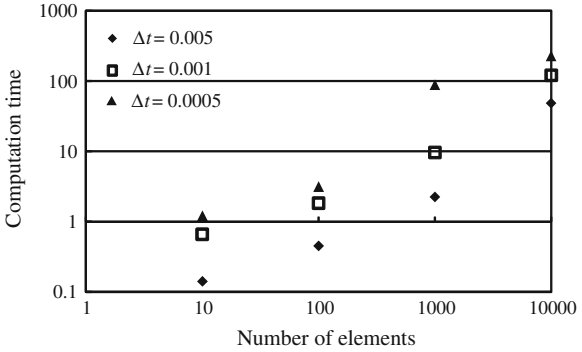


Fig. 3 The actual computation time versus the number of elements with mesh regeneration

Table 4 Results for the American put option given the time step $\Delta t = 0.001$, the initial number of elements $N = 500$, upper boundary $S_\infty = 100$, $R_1^* = 0.001$, and $R_2^* = 0.005$

	FEM without regeneration	FEM with coarsening	FEM with coarsening and refinement
Computation time (s)	9.727	0.129	0.924
RMSE	0.006	0.284	0.128

tigations into the convection-dominated problems are needed to obtain more precise outcomes and stability.

References

1. Christara, C., Dang, D.M.: Adaptive and high-order methods for valuing American options. J. Comput. Fin. **14**, 73–113 (2011)
2. Topper, J.: Financial Engineering with Finite Elements. Wiley, New York (2005)
3. Topper, J.: Finite Element Modeling of Exotic Options. Department of Economics, University of Hannover, Technical report (1999)
4. Wilmott, P.: Quantitative Finance. Wiley, New York (2003)
5. Winkler, G., Apel, T., Wystup, U.: Valuation of Options in Heston's Stochastic Volatility Model Using Finite Element Methods, pp. 283–303. Foreign Exchange Risk (2001)
6. Zvan, R., Vetzal, K.R., Forsyth, P.A.: PDE methods for pricing Barrier options. J. Econ. Dyn. Control **24**, 1563–1590 (2000)

Part VIII
Game Theory and Experimental Economics

Simulation of Bribes and Consequences of Leniency Policy. Results from an Experimental Study

Alexandra Christöfl, Ulrike Leopold-Wildburger and Arleta Rasmußen

1 Introduction: The Problem of Bribery

Bribery is commonplace in global business today so much that e.g. the United States Justice Department is waging an aggressive campaign against it. Also in Europe most countries have to deal with the problem of corruption and several activities are started battling corporate bribery. In US the foreign Corrupt Practices Act prohibits American companies and foreign companies whose securities are traded on exchanges in the United States from bribing foreign officials to attract or to keep business.

For many years, there were few prosecutions under the act. In 2003, for instance, not a single person was charged. However, during the last few years, a large number of companies have paid huge amounts to settle such corruption charges. For US the estimated amount is approx. 3.75 billion for the last four years. One of the biggest cases under the Foreign Corruption Practices Act involved Siemens/Germany. It was accused in 2008 of bribing officials in a number of countries. Siemens paid 800 million to regulators in the United States and another 800 million in Germany to settle the case. The World Bank estimates that 1 trillion in bribes is paid annually to government officials, according to Transparency International, which tracks corruption.

We are able to list some available data about the economic damage caused by corruption: damage in Austria: 25 millions Euro in 2009 (according to [6]). The estimates of corruption and bribery in the building industry in Germany are the following: large-scale projects: 1.5–3 % of contract amount and small-scale projects:

A. Christöfl (✉)
Schererstrasse 11c/6, 8052Graz, Austria
e-mail: alexandra.christoefl@gmail.com

U. Leopold-Wildburger · A. Rasmußen
Department of Statistics and Operations Research, Universitätsstraße 15, A-8010Graz, Austria
e-mail: ulrike.leopold@uni-graz.at

A. Rasmußen
e-mail: arleta.mietek@uni-graz.at

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_31,

up to 10 % of contract amount [14]. According to a study of the World Bank: around 1 billion US each year.

2 Theoretical Considerations

Empirical data are vague and there is still little theoretical research. The concept of corruption involves connections with fairness and honesty and it is an issue that provokes the interest of various disciplines. Especially experiments of the Zurich Group around Ernst Fehr and Simon Gächter have shown that self-interest explicitly increases as soon as money is included. The fact that individuals who show unfair behavior have more success in the long run as a fair partner applies often (Fehr and Gächter (2000)).

The expected payoffs for the employers and the bidders are dependent on the available budget and the offer, but also on the bribe and the probability of detection and in this connection with the potential fine. For more details see the following papers: [8–11].

Theoretical considerations on the concept of corruption are discussed in several contexts. In economics acts of corruption are seen as abnormal transactions. The calculations of costs of corruption should be calculated in an analogous way as the calculations of the hypothetical monopolist do. A polypoly becomes under certain circumstances a temporary monopoly. Of course there is a certain welfare loss as a result of higher prices.

3 The Experimental Study

The experimental study of a contract award with the possibility of corruption is based on different aspects and ideas of the described experiments in the papers of [1–3], [4, 7] and [5]. Thereby the bribery game by [7] formed the basis. These ideas were extended and include now the active participation of an agent and also a simulated principal. The treatments got originally influenced by [13]. Next, on the one hand the moonlighting game by [2] and on the other hand the leniency of the experiment developed by [4] influenced the development of our version of the experiment. The resultant developed contract award allows the study of the willingness of bribery and corruption, using different detection probabilities and the analysis of leniency as a preventive measure to combat corruption for the first time. Figure 1 shows the decision tree for the participants and demonstrates the sequential arrangement within the experiment.

We created a computer simulation of a contract award with the following treatments: (Table 1)

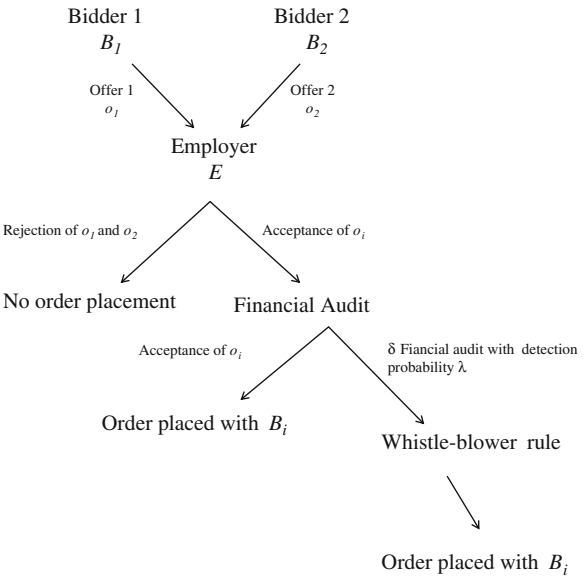


Fig. 1 Decision tree

Table 1 Treatments: detection probability TRlow/TRhigh, without (CO)/with whistle blow WB.

Detection Probability	Whistle-Blow Without(Control)	WB Rule With
10 %	X	X
40 %	–	X

We distinguish between different detection probabilities, namely low (10 %) and high (40 %) and further on, we distinguish whether the possibility of a leniency policy/whistle blow (WB) is available or not.

4 Research Questions and Hypotheses

We run two different settings: 8 contract awards with different partners (stranger design) with detection and 5 contract awards with the same partners (partner design) without detection. The computer experiment is programmed in z-Tree [12] a software that enables to program and to conduct economic experiments. We recruited 120 participants for the experiment. They all were drawn from the University of Graz campus and included undergraduate, graduate and post-graduate students. The participation and all participants' responses remain completely anonymous. Participants know only that they are dealing with another persons from the laboratory, but they

are not aware which persons they are playing with. After the experiment the subjects were rewarded for their participation according to their performance depending on success, on bribes and on punishment fines.

Our research questions are:

1. What is the level of readiness to pay bribes and to accept corruption?
2. Do different detection probabilities (TRlow/TRhigh) have an influence and impact on the behavior?
3. Does leniency policy and whistle blow (WB) respectively have any consequences?

5 Results

We run the treatments with 120 participants: 80 bidders and 40 employers. All participants received substantial monetary rewards according to their performance, mainly depending on the bribes, the fines and the detection probability.

The data analysis of the contract study part clearly shows that the majority of subjects—namely 61 and 67 %—inhibits dishonest or corrupt behavior, where about 21 % show a huge dishonesty factor (Honesty factor: HF = 0,25) and therefore can be classified as very dishonest participants. They decide in the majority of the cases dishonestly and show corrupt behavior. Participants in the role of employers show even a more dishonest behavior than bidders.

The study of different detection probabilities shows a significant increase of corruption if there is no detection. The bribes offered by bidders in the treatment without detection are significantly higher than in treatments with detection. In cases of a high probability (40 %) the number of corruption decreased significantly. This significant relationship between detection probability and dishonesty confirms the result that the frequency for bribes falls with increasing detection probability.

The results show that commission-based decisions of employer in stranger design are much higher than in the partner design. In the partner design the substantial majority of contract awards were awarded to the same bidder.

The modeling of leniency allows examining the willingness to whistle blow the corrupt partner. The data analysis shows that the vast majority of detected negotiators - namely 87,5 % - makes use of the leniency. Furthermore the data obtained indicates that the leniency clearly leads to a significant reduction of dishonest offers. The analysis of bribes, regarding offer and acceptance of bribes shows that the median of the bribes is higher in the treatment with leniency than in the treatment without leniency.

This suggests that although the leniency leads to more honesty the bribe offers are higher. Thus it can be concluded that the willingness to accept the leniency is very high. A connection between testimony and whistle blowing of the corrupt partner and criminal reduction is evident.

In summary the leniency can be confirmed based on the analyzed data as an appropriate means to combat corruption. However, it was also noted that the leniency

Table 2 Results for detection probability TRlow vs. TRhigh, without (CO)/with whistle blow WB

Phase	Treatment	Bidder		Employer		
		Honest %	Non-Honest %	honest		non-honest %
				must %	can %	
Without detection	Partner design	7.50	*92.50	0.00	20.00 %	80.00
		HF≤0.25	76.25			60.00
	CO-TRlow	25.00	**75.00**	0.00	12.50	**87.50**
		HF≤0.25	25.00			62.50
	WB-TRlow	43.33	**56.67**	26.67	0.00	**73.33**
		HF≤0.25	3.33			73.33
With detection	WB-TRhigh	50.00	50.00	41.18	0.00	58.82
		HF≤0.25	0.00			58.82
	Total	42.50	*57.50	27.50	2.50	70.00
		HF≤0.25	6.25			27.50

*highly significant

program due to the subsequent possibility of a penalty reduction leads to higher amounts in bribes at contract awards.

Table 2 shows the results for all treatments. HF means honesty factor measured as follows:

$$HF = 1 - \frac{\text{offered bribes (bridder) resp. accepted bribes (employer)}}{\text{possible bribes}}$$

In the treatment with detection there are only very few bidders with a low honesty factor ($HF \leq 0.25$) especially when the detection probability is high (TRhigh). With high detection probability the number of non-honest employers also reduces from 73.33 % to 58.82 %. Obviously dishonesty is the highest in the treatment without detection.

References

1. Abbink, K., Fair Salaries and the Moral Costs of Corruption. http://www.nottingham.ac.uk/cedex/documents/papers/2002-05.pdf (2002)
2. Abbink, K., Irlenbusch, B., Renner, E.: The moonlighting game. An experimental study on reciprocity and retribution. J. Econ. Behav. Organ. **42**, 265–277 (2000)
3. Abbink, K., Irlenbusch, B., Renner, E.: An experimental bribery game. J. Law, Econ. Organ. **18**, 428–454 (2002)
4. Apesteguia, J., Dufwenberg, M., Selten, R.: Blowing the whistle. Econ. Theor. **31**, 143–166 (2007)
5. Berentsen, A., Bruegger, E., Loertscher, S.: On cheating, doping and whistleblowing. Eur J. Polit. Econ. **24**, 415–436 (2008)
6. Brenner, G.: Im Schatten der Wirtschaft. öffentliche Sicherheit. Das Magazin des Innenministeriums, 7–8, 39–41 (2010)

7. Büchner, S., Freytag, A., Gonzáles, L.G., Güth, W.: Bribery and public procurement: An experimental study. Public Choice **137**, 103–117 (2008)
8. Christoefl, A., Leopold-Wildburger, U., Mietek, A.: An experimental study on corruption and leniency policy. WP of the Faculty of Economics and, Social Sciences (2012)
9. Falk, A., Fischbacher, U.: A theory of reciprocity. Games Econ. Behav. **54**(2), 293–315 (2006)
10. Fehr, E., Gächter, S., Kirchsteiger, G.: Reciprocity as a contract enforcement device: Experimental evidence. Econometria **65**, 833–860 (1997)
11. Fehr, E., Schmidt, K.: A theory of fairness, competition and cooperation. Q. J. Econ. **114**, 817–851 (1999)
12. Fischbacher, U.: z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Exp. Econ. **10**, 171–178 (2007)
13. Kastlunger, B., Kirchler, E., Mittone, L., Pitters, J.: Sequence of audits, tax compliance, and taxpaying strategies. J. Econ. Psychol. **30**, 405–418 (2009)
14. Vehrkamp, R.B.: ökonomische Konsequenzen der Korruption. Wirtschaftsdienst **12**, 776–783 (2005)

Two-Stage Market with a Random Factor

Ekaterina Daylova and Alexander Vasin

1 Introduction

Large electricity producers have the ability to affect the market price and they take this ability into account choosing their strategies. This causes a deviation from the competitive equilibrium and leads to higher market prices, a decrease of a total output, a reduction of social welfare and its redistribution to the benefit of producers and to the detriment of consumers. We investigate the introduction of the forward market as a way to reduce market power of producers. Production concentration is typically high in electricity markets while consumers have no market power ([1]). We consider a symmetric oligopoly with constant marginal cost as a market structure.

Equilibrium strategies, relationship between spot and forward prices and prices of the competitive equilibrium and the Cournot equilibrium are examined in [2–5]. Paper [6] shows that the Nash equilibrium corresponding to the Cournot outcome is stable for a model of a uniform price auction. James Bushnell ([2]) considered a two-stage market with Cournot competition in the spot market and constant marginal cost. The results show that the introduction of the forward market has the same impact on market power as does an increase of the number of firms in the market from n to n^2. However, there are several restrictive assumptions in the paper and random factors that often affect the outcome in the spot market are not taken into account.

We construct and examine a two-stage model of the market taking into account the presence of arbitrageurs and a random factor that affects the outcome in the spot market. The present paper describes the strategic interaction between producers, consumers, and arbitrageurs. We determine optimal strategies that correspond to

E. Daylova (✉)
Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskiye Gory 1-52,119991Moscow, Russia
e-mail: e.daylova@gmail.com

A. Vasin
e-mail: vasin@cs.msu.su

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_32,

the subgame perfect equilibrium and examine the properties of the equilibrium, in particular, the deviation of the equilibrium price from the Walrasian price.

2 Strategic Model of Agents' Interaction

There is a finite set of producers A. All firms produce a homogeneous product, i.e. there is no product differentiation. $C_a(q)$ is a cost function of firm a. There are many small consumers. Each consumer b aims to buy one unit of the product and is described with reserve price r_b (the highest price a buyer is willing to pay for the product). The demand function has the following form: $D(p) = \int_p^{p_m} \rho(r)dr$, where $\rho(r)$ is the density of reserve price distribution among consumers. Consumer b is also characterised by a parameter of risk attitude λ_b: $\lambda_b \in [\lambda_{min}, \lambda_{max}]$, $\lambda_{min} < 0 < \lambda_{max}$. His utility function $U_b(\Delta, \lambda_b)$ monotonically increases in $\Delta = r_b - p$ if $\Delta > 0$. If $\Delta \leq 0$, then $U_b(\Delta, \lambda_b)$ equals zero for any λ_b, since a consumer never buys in this case. For risk-averse consumers, λ_b is positive. Their utility function is concave. For risk-neutral consumers, λ_b equals zero and the utility function is linear. For risk-seeking consumers, λ_b takes negative values and the utility function is convex. Risk-aversion increases with the growth of λ_b, and $\lambda = \lambda_{max}$ corresponds to the utility function $U(\Delta, \lambda_{max}) \equiv U_{max}\ \forall \Delta > 0$.

Arbitrageurs also act in the market. They can first sell forward contracts and then buy goods in the spot market, or they can carry out a return operation. All arbitrageurs are risk-neutral.

At the first stage, the firms choose volumes q_a^f, $a \in A$ for the forward market. Let q^f denote the volume sold by producers in the forward market: $q^f = \sum_{a \in A} q_a^f$. Arbitrageurs can also offer the product. If arbitrageurs first sell forward contracts and then buy goods in the spot market, $q_{arb} > 0$ shows the volume sold by arbitrageurs in the forward market. If arbitrageurs first buy forward contracts and then sell goods in the spot market, then $q_{arb} < 0$ and $|q_{arb}|$ shows the volume bought by arbitrageurs in the forward market. Let q_t^f denote the volume bought by consumers in the forward market.

Each consumer decides whether he will participate in the auction in the forward market. Each participant of the auction makes a bid with his reserve price. $D^f(p)$ denotes consumers' demand function for the forward market. The forward price p^f is determined by the equality

$$D^f(p^f) = q_t^f = q^f + q_{arb}.$$

At the second stage, there is a Cournot auction in the spot market. The residual demand function for the spot market $D^s(p)$ depends on the strategies of the agents: $D^s(p) = D(p) - q_t^f$ if $p < p^f$; $D^s(p) = D(p) - D^f(p)$ if $p \geq p^f$. The firms choose volumes for the spot market q_a^s. The spot price p^s is determined by the equality

$$D^s(p^s) + q_{arb} = \sum_{a \in A} q^s.$$

We take into account a random factor that affects the outcome in the spot market. The random factor takes values $i = 1, 2, \ldots, k$ with probabilities w_i; $\sum_{i=1}^{k} w_i = 1$, $w_i > 0, i \in \overline{1,k}$. Producers choose output volumes for the spot market depending on the value of the random factor. A strategy of each firm is a profile $(q_a^f, q_a^s(i); i \in \overline{1,k})$. It determines a volume for the forward market q_a^f and a volume for the spot market $q_a^s(i)$ depending on the value of the random factor. The spot price p^s is a random value that is determined by the equality $D^s(p^s) + q_{arb} = \sum_{a \in A} q^s(i)$. The spot price is equal to p_i with probability $w_i, i \in \overline{1,k}$: $p_{min} = p_1 \leq p_2 \leq \ldots \leq p_k = p_{max}$. The distribution of the considered random value depends on the strategies of producers.

3 Optimal Strategies

Rational arbitrageurs first sell forward contracts and then buy goods in the spot market if the forward price is higher than the spot price. Otherwise, they perform a return operation. We assume perfect competition among arbitrageurs and their risk-neutrality. Then arbitrageurs' activity results in equality of the forward price and the expected value of the spot price: $p^f = \mathbb{E}(p^s)$.

Further we determine the optimal behaviour for consumers under the fixed forward and spot prices.

Theorem 1 *The optimal behaviour of consumers is determined as follows: consumers with reserve prices $r^b < p^f$ buy goods only in the spot market if their reserve price r_b is higher than the market price p^s. Behaviour of risk-neutral and risk-seeking consumers with reserve prices $p^f < r_b < p_k$ is the same. For risk-averse consumers there exists such a value $\lambda(r)$ that consumers with $\lambda_b > \lambda(r)$ buy in the forward market; consumers with $\lambda_b < \lambda(r)$ buy goods only in the spot market if their reserve price r_b is higher than the market price p^s. Risk-seeking consumers with reserve prices $r_b > p_k$ buy in the spot market. Risk-averse consumers with $r_b > p_k$ buy in the forward market.*

Let $\alpha(p)$ denote the fraction of risk-seeking consumers. The next step is to determine the residual demand function in the spot market depending on the forward strategies.

Lemma 1 *The residual demand function for the spot market $D^s(p)$ that corresponds to the equilibrium is determined as follows:*

$$D^s(p) = \begin{cases} D(p) - q_t^f & \text{if } p < p^f; \\ \int_p^{r_{max}} \int_{\lambda_{min}}^{\lambda(r)} \rho(r,\lambda) d\lambda dr & \text{if } p^f < p < p_{max}; \\ \alpha(p) D(p) & \text{if } p > p_{max}. \end{cases}$$

The next step is to find the subgame perfect equilibrium. First we consider the second stage. The optimal strategies depend on the forward positions. The Cournot first-order condition is used to find the optimal output volumes. The optimal volumes q_a^* and the equilibrium price p^* for a Cournot auction with demand function $D(p)$ and n firms with marginal cost c satisfy the equation

$$q_a^* = (p^* - c)|D'(p^*)| = \frac{D(p^*)}{n}, \quad a \in \overline{1,n}.$$

Thus we obtain

Lemma 2 *Let the function $D(p)$ be concave. Then in the subgame perfect equilibrium price p_1 is determined by the equation*

$$(p_1 - c)|D'(p_1)| = \frac{D(p_1) - q_t^f + q_{arb}}{n}$$

and satisfies the inequality $p_1 \le p^f$.
For $i \in \overline{2,k-1}$, $p_i > p^f$ and satisfies the condition

$$(p_i - c)|D^{s\prime}(p_i)| = \frac{D^s(p_i) + q_{arb}}{n},$$

where

$$D^s(p) = \int_p^{r_{max}} \int_{\lambda_{min}}^{\lambda(r)} \rho(r,\lambda) d\lambda dr,$$

p_k meets the equation

$$(p_k - c)|(\alpha(p_k)D(p_k))'| = \frac{\alpha(p_k)D(p_k) + q_{arb}}{n}.$$

Each producer a aims to maximize the total profit π_a, $a \in A$. The total profit of a firm is a sum of the profit in the forward market and the average value of the profit in the spot market. The optimal volumes for the first stage are determined as solutions to the profit maximization problem.

Further we consider a special case and assume that the demand function is linear: $D(p) = max\{d(r_{max} - p), 0\}$. The fraction of risk-seeking consumers is constant: $\alpha(p) = \alpha$. The random factor takes two values. Thus the spot price equals p_1 with the probability w and it equals p_2 with the probability $1 - w$. Let q_a^{s1} denote the

volume sold by consumer a if the spot price equals p_1; q_a^{s2} denotes the volume sold by consumer a if the spot price equals p_2.

Theorem 2 *If there exists a subgame perfect equilibrium, the prices and the volumes meet the equations:*

$$p_1 = p^* - \frac{q^f}{d(n+1)}, \quad p_2 = p^* + \frac{q_{arb}}{\alpha d(n+1)}, \quad p^f = p^* - \frac{wq^f}{d(n+1)} + \frac{(1-w)q_{arb}}{\alpha d(n+1)},$$

$$q_a^{s1} = d(\Delta^* - \frac{q^f}{d(n+1)}), \quad q_a^{s2} = \alpha d(\Delta^* + \frac{q_{arb}}{\alpha d(n+1)}),$$

where p^* is the price in the Nash equilibrium of the classic Cournot oligopoly model, and $\Delta^* = p^* - c$.

The volumes q_a^f are determined by the equations

$$\frac{\partial \pi_a(q_a^f, q_i^f, i \in A \setminus \{a\})}{\partial q_a^f} = 0, \quad a \in A,$$

where $\pi_a(q_a^f, q_i^f, i \in A\setminus\{a\}) = q_a^f(\Delta^* - \frac{w(\sum\limits_{i\in A\setminus\{a\}} q_i^f + q_a^f)}{d(n+1)} + \frac{(1-w)q_{arb}(q_a^f, q_i^f, i\in A\setminus\{a\})}{\alpha d(n+1)})$
$+ wd(\Delta^* - \frac{(\sum\limits_{i\in A\setminus\{a\}} q_i^f + q_a^f)}{d(n+1)})^2 + (1-w)\alpha d(\Delta^* + \frac{q_{arb}(q_a^f, q_i^f, i\in A\setminus\{a\})}{\alpha d(n+1)})^2.$

4 Stability

The first-order conditions determine only a local SPE, since they are necessary and sufficient for the true equilibrium only for a concave demand function. In this case it is not profitable for producers to deviate from their strategies. However, the residual demand function is convex in the spot market, so the given strategies may correspond to a local equilibrium. It might be profitable for a producer to deviate from his strategy. In particular, when the spot price equals p_1 it might be profitable for a producer to reduce his output in order to increase the market price. When the spot price equals p_2 it might be profitable for a producer to increase his output in order to reduce the market price. The profit in the true equilibrium should be higher than the profit under deviation.

Theorem 3 *The subgame perfect equilibrium exists only if the following conditions hold:*

$$
\begin{aligned}
(n-1+2\sqrt{\alpha})\frac{p_1-c}{\Delta^*} &\geq (n+1)\alpha+\frac{q_{arb}}{d\Delta^*};\\
\frac{p_2-c}{\Delta^*}(2\sqrt{\alpha}+(n-1)\alpha) &\geq n+1-\frac{q_a^f}{d\Delta^*};\\
\frac{q_{arb}}{d} &\leq (n-1)(p_2-c)\alpha.
\end{aligned}
\tag{1}
$$

5 Special Cases

Let the rate of risk-aversion be so low for the consumers with $\lambda > 0$ and reserve prices $p^f < r_b < p_2$ that all of them buy in the spot market. The analysis shows that it is always profitable to deviate in this case. Therefore, the true SPE does not exist.

Let the rate or risk-aversion be so high for the consumers with $\lambda > 0$ and reserve prices $p^f < r_b < p_2$ that all of them buy in the forward market. The computation results show that the introduction of the forward market reduces market power of producers. We compare the subgame perfect equilibrium with the Nash equilibrium of a one-stage model. Let γ denote the ratio of forward price deviation from the marginal cost to the similar deviation in the classic Cournot oligopoly model: $\gamma = \frac{p^f-c}{\Delta^*}$. For instance, if $n = 6$, this ratio equals 0.18 for James Bushnell's model. In our model γ equals 0.304 for the set $(n, a, w) = (6, 0.4, 0.422)$, and γ equals 0.229 for $(n, a, w) = (6, 0.4, 0.818)$.

6 Conclusion

The results indicate that the introduction of forward contracts substantially reduces market power of producers. Market power decreases in the probability of the outcome with a low spot price and also decreases in the fraction of risk-seeking consumers. The values of the probability w for which the equilibrium exists are restricted depending on the number of firms and the fraction of risk-seeking consumers. The results show that it is not optimal for risk-seeking consumers to buy in the forward market. Risk-seeking consumers with low reserve prices buy in the spot marker if their reserve price is higher than the market price. The same behaviour is optimal for risk-averse consumers with low reserve prices. The fraction of consumers who buy in the spot market decreases with the growth of the reserve price.

References

1. Botterud A., Bhattacharyya A.K., Ilic M., Futures and spot prices: an analysis of the Scandinavian electricity market. In: Proceedings of North American Power, Symposium (2002)
2. Bushnell J.: Oligopoly equilibria in electricity contract markets. CSEM Working Paper WP-148 (2005)
3. Green, R.R.: The electricity contract market in England and Wales. J. Ind. Econ. **47**(2), 107–124 (1999)
4. Newbery D. M.: Competition, contracts, and entry in the electricity spot market. RAND J. Econo. **29**(4), 726–749 (1998)
5. Vasin A.A., Kartunova P.A., Sharikova A.A., Dolmatova M., Comparative analysis of one-stage and two-stage markets. Contributed paper for the Conference on, Economic Design (2009)
6. Vasin, A.A., Vasina, P.A.: On organization of markets of homogeneous goods. Journal of Computer and Systems Sciences International. **46**(1), 93–106 (2001)

A New Allocation Method for Simple Bargaining Problems: The Shapley Rule

Francesc Carreras and Guillermo Owen

1 Introduction

The proportional rule has a long tradition in collective problems where utility is to be shared among the agents. However, while its (apparent) simplicity might seem a reason for applying it in simple bargaining affairs, where only the whole and the individual utilities matter, its behavior is, in fact, very questionable. A consistent alternative, the *Shapley rule*, will be suggested as a much better solution for this kind of problems. Utilities will be assumed to be completely transferable.

The organization of the paper is as follows. In Sect. 2, the notions of *simple bargaining problem* (*SBP*, for short) and *sharing rule* are stated. In Sect. 3 we attach to any SBP a quasi-additive game (*closure*). By using this idea, in Sect. 4 we introduce the *Shapley rule* for SBPs. Section 5 provides axiomatic characterizations of this rule. Section 6 is devoted to criticize the proportional rule. Section 7 collects the conclusions of the work. Proofs and more details can be found in [3].

2 Simple Bargaining Problems and Sharing Rules

Let $N = \{1, 2, \ldots, n\}$ ($n \geq 2$) be a set of agents and assume that there are given: (a) a set of utilities $u_1, u_2, \ldots, u_n$ available to the agents individually, and (b) a total utility u_N that, alternatively, the agents can jointly get if all of them agree. A vector $u = (u_1, u_2, \ldots, u_n | u_N)$ collects all this information and we say that it represents a

F. Carreras (✉)
Technical University of Catalonia, Barcelona, Spain
e-mail: francesc.carreras@upc.edu

G. Owen
Naval Postgraduate School, Monterey, California, U.S.A.
e-mail: gowen@nps.edu

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_33,

simple bargaining problem (*SBP*, in the sequel) on N. The *surplus* of u is defined as

$$\Delta(u) = u_N - \sum_{j \in N} u_j.$$

The problem consists in dividing u_N among the agents in a rational way, i.e., in such a manner that all of them should agree and feel (more or less) satisfied with the outcome. Of course, the individual utilities $u_1, u_2, \ldots, u_n$ should be taken into account. The transferable utility assumption will mean that any vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $x_1 + x_2 + \cdots + x_n = u_N$ is feasible if the n agents agree.

Let $E_{n+1} = \mathbb{R}^n \times \mathbb{R}$ denote the $(n+1)$-dimensional vector space formed by all SBPs on N. In order to deal with, and solve, all possible SBPs on N, one should look for a *sharing rule*, i.e., a function $f : E_{n+1} \longrightarrow \mathbb{R}^n$. Given $u \in E_{n+1}$, for each $i \in N$ the i-coordinate $f_i[u]$ will provide the share of u_N that corresponds to agent i according to f. The classical *proportional rule*, denoted here by π and given by

$$\pi_i[u] = \frac{u_i}{u_1 + u_2 + \cdots + u_n} u_N \quad \text{for each } i \in N, \tag{1}$$

is a sharing rule defined only in the subdomain

$$E^{\pi}_{n+1} = \{u \in E_{n+1} : u_1 + u_2 + \cdots + u_n \neq 0\}. \tag{2}$$

3 Closures and Quasi-additive Games

The Shapley value [4], denoted here by φ, cannot be directly applied to SBPs as a sharing rule. We will therefore associate a TU cooperative game with each SBP in a natural way. Let $\mathcal{G}_N$ be the vector space of all cooperative TU games with N as set of players and let us define a map $\sigma : E_{n+1} \longrightarrow \mathcal{G}_N$ as follows. If $u = (u_1, u_2, \ldots, u_n | u_N)$ then $\overline{u} = \sigma(u)$ is given by

$$\overline{u}(S) = \begin{cases} \sum_{i \in S} u_i & \text{if } S \neq N, \\ u_N & \text{if } S = N. \end{cases}$$

The idea behind this definition is simple. Since, given a SBP u, nothing is known about the utility available to each intermediate coalition $S \subset N$ with $|S| > 1$, a reasonable assumption is that such a coalition can get the sum of the individual utilities of its members. Game $\overline{u}$ will be called the *closure* of u. It is not difficult to verify that σ is a one-to-one linear map.

Let us recall that a cooperative game v is *additive* iff $v(S) = \sum_{i \in S} v(\{i\})$ for all $S \subseteq N$. If we drop this condition just for $S = N$ and give the name *quasi-additive*

to the games that fulfill it for all $S \subset N$, it follows that these games precisely form the image subspace of σ, and hence a game is quasi-additive iff it is the closure of a SBP, which is unique. Using σ, notions for games can be translated to SBPs.

4 The Shapley Rule

Definition 1 By setting

$$\overline{\varphi}[u] = \varphi[\overline{u}] \quad \text{for all } u \in E_{n+1}$$

we obtain a function $\overline{\varphi} : E_{n+1} \longrightarrow \mathbb{R}^n$. Function $\overline{\varphi}$ will be called the *Shapley rule* (for SBPs). It is given by

$$\overline{\varphi}_i[u] = u_i + \frac{\Delta(u)}{n} \quad \text{for each } i \in N \text{ and each } u \in E_{n+1}, \tag{3}$$

and hence it solves each SBP in the following way: (a) first, each agent is allocated his individual utility; (b) once this has been done, the remaining utility is equally shared among all agents. Thus, the Shapley rule shows an "egalitarian flavor" in the sense of [1]. Indeed, from Eq. (3) it follows that the Shapley rule is a mixture consisting of a "competitive" component, which rewards each agent according to the individual utility, and a "solidarity" component that treats all agents equally when sharing the surplus.

5 Axiomatic Characterizations of the Shapley Rule

When looking for a function $f : E_{n+1} \longrightarrow \mathbb{R}^n$, some reasonable properties should be imposed. Let us consider the following:

- *Efficiency*: $\sum_{i \in N} f_i[u] = u_N$ for every u.
- *Inessentiality*: if $\Delta(u) = 0$ then $f_i[u] = u_i$ for each $i \in N$.
- *Symmetry*: if $u_i = u_j$ then $f_i[u] = f_j[u]$.
- *Additivity*: $f[u + v] = f[u] + f[v]$ for all u, v.

These four properties deserve to be called "axioms" because of their elegant simplicity. It is hard to claim that they are not compelling. The question is the following: is there some function satisfying them? If so, is it unique? The positive answers are given in the following result.

Theorem 1 *(First main axiomatic characterization of the Shapley rule) There is one and only one function $f : E_{n+1} \longrightarrow \mathbb{R}^n$ that satisfies efficiency, inessentiality, symmetry and additivity. It is the Shapley rule $\overline{\varphi}$.*

Following [2, 5], we adapt strong monotonicity to the SBP setup, i.e., for sharing rules. First, for each $i \in N$ we define a preorder $\succeq_i$ for SBPs as follows:

$$u \succeq_i v \quad \text{iff} \quad u_i \geq v_i \ \text{ and } \ \Delta(u) + u_i \geq \Delta(v) + v_i .$$

Then, we say that a sharing rule f satisfies *strong monotonicity* iff, for each $i \in N$, $u \succeq_i v$ implies $f_i[u] \geq f_i[v]$.

Theorem 2 *(Second main axiomatic characterization of the Shapley rule) There is one and only one function $f : E_{n+1} \longrightarrow \mathbb{R}^n$ that satisfies efficiency, symmetry and strong monotonicity. It is the Shapley rule $\overline{\varphi}$.*

It is worthy of mention that characterizations analogous to Theorems 1 and 2 hold for (the restriction of) the Shapley rule to a series of interesting subsets of E_{n+1}: e.g., the domain of the proportional rule, the open orthant of positive (resp., negative) SBPs, the open cone of superadditive SBPs ($\Delta(u) > 0$), and the intersection of this cone with the closed orthant of nonnegative (resp., nonpositive) SBPs.

6 Criticism on the Proportional Rule

We discuss here several aspects of the proportional rule, most of which are far from being satisfactory, and contrast them with the behavior of the Shapley rule.

- **Restricted domain**. As was already mentioned, the domain of the proportional rule π is not the entire space E_{n+1} but the subset defined by Eq. (2):

$$E^{\pi}_{n+1} = \{u \in E_{n+1} : u_1 + u_2 + \cdots + u_n \neq 0\}.$$

  By contrast, the Shapley rule $\overline{\varphi}$ applies to all SBPs without restriction.
- **Doubly discriminatory level**. Within its domain, the proportional rule coincides with the Shapley rule just on additive or symmetric SBPs. However, these are very particular cases and, in general, the two rules differ. As a matter of comparison, note that the expression of $\pi_i[u]$ given in Eq. (1) is equivalent to

$$\pi_i[u] = u_i + \frac{u_i}{\sum_{j \in N} u_j} \Delta(u),$$

  which shows that the proportional rule (a) allocates to each agent his individual utility (as the Shapley rule does) but (b) it shares the remaining utility proportionally to the individual utilities. In other words, no solidarity component exists in the proportional rule, as both components are of a competitive nature. Instead, in this second step the Shapley rule acts equitably (notice that the calculus for the Shapley rule is, therefore, easier than for the proportional rule). Then the proportional rule

is, conceptually, more complicated than the Shapley rule and may include a *doubly discriminatory level* since, when comparing any two agents, it rewards twice the agent that individually can get the highest utility on his own. It is hard to find a reasonable justification for this. This discriminatory level arises, for example, in the case of nonnegative and superadditive SBPs in E^{π}_{n+1}.

- **The axiomatic framework**. On its restricted domain E^{π}_{n+1}, where the Shapley rule has been axiomatically characterized by the analogues of Theorems 1 and 2, the proportional rule satisfies the properties of efficiency, inessentiality and symmetry. It fails to satisfy additivity and strong monotonicity. Now, in spite of its simplicity and mathematical tradition, maybe additivity is, in principle, the least appealing property and might seem to practitioners only a "mathematical delicatessen": the reason is that one does not easily capture the meaning of the sum of SBPs in practice. This will be illustrated in the next items. And, while the Shapley rule is linear, the proportional rule is only *homogeneous*, i.e., satisfies $\pi[\lambda u] = \lambda\pi[u]$ for every real number $\lambda \neq 0$ and every $u \in E^{\pi}_{n+1}$.
- **Inconsistency: cost-saving problems**. Whenever a cost SBP u^c and the related saving SBP u^s are considered, the proportional rule can never be applied to the saving SBP and no kind of consistency can then be discussed. Instead, the consistency of the Shapley rule is clear since, for each $i \in N$,

$$\overline{\varphi}_i[u^c] = u_i + \overline{\varphi}_i[u^s].$$

  The conclusion is that, using the Shapley rule, all agents are indifferent between sharing costs and sharing savings (as it should be). Moreover, in e.g., any purchasing pool, when using the Shapley rule all members of the pool are indifferent between sharing real costs (i.e., with discount) and sharing savings. Instead, almost any numerical instance shows that this is not the case if the proportional rule is applied: some firms prefer sharing costs whereas the remaining ones prefer sharing savings. Thus, the inconsistency (or lack of fairness) of the proportional procedure is obvious. The difference lies in additivity, satisfied by the Shapley rule but not by the proportional rule.
- **Inconsistency: added costs problems**. The same happens when considering added costs problems. By using the proportional rule some agents will prefer sharing the joint bill, whereas the remaining ones will prefer sharing the bills separately. This is a new sample of inconsistency of this rule. Instead, using the Shapley rule all agents are indifferent between sharing separate bills and sharing a joint bill. The reason is, again, the distinct behavior of the rules as to additivity.

7 Conclusions

The axiomatic viewpoints established by Shapley when defining the value notion for cooperative games, and by Young when replacing the dummy/null player and additivity properties with strong monotonicity, have been adapted to SBPs.

This allows us to evaluate any sharing rule and, in particular, to compare the proportional rule and the Shapley rule.

After the comparison, the conclusion is clear: the Shapley rule satisfies all properties and is better than the proportional rule in all aspects, included practical cases where the proportional rule shows inconsistency because it fails to satisfy additivity.

We therefore contend that the Shapley rule should replace in practice the proportional rule in SBPs, that is, in cooperative affairs where the coalitions of intermediate size ($1 < |S| < n$) do not matter.

Of course, the advantages of the Shapley value over the proportional rule are even greater when considering general cooperative games and not only quasi-additive games (i.e., PBPs). The proportional rule does not take into account most of the coalitional utilities: precisely, all those corresponding to the intermediate coalitions. This becomes more and more critical as the number of players increases, and it gives rise to a very low sensitivity. On the contrary, the Shapley value is always concerned with all marginal contributions without exception and enjoys therefore a nice sensitivity with regard to the data defining any given problem.

Acknowledgments Research partially supported by Grants SGR 2009-01029 of the Catalonian Government (*Generalitat de Catalunya*) and MTM 2009-08037 of the Science and Innovation Spanish Ministry and the European Regional Development Fund.

References

1. van den Brink, R., Funaki, Y.: Axiomatizations of a class of equal surplus sharing solutions for TU-games. Theory Decis. **67**, 303–340 (2009)
2. Carreras, F., Freixas, J.: A note on regular semivalues. Int. Game Theory Rev. **2**, 345–352 (2000)
3. Carreras, F., Owen, G.: Pure bargaining problems and the Shapley rule. Homo Oeconomicus **28**, 379–404 (2011)
4. Shapley, L.S.: A value for n-person games. Ann. Math. Stud. **28**, 307–317 (1953)
5. Young, H.P.: Monotonic solutions of cooperative games. Int. J. Game Theory **14**, 65–72 (1985)

Does SFE Correspond to Expected Behavior in the Uniform Price Auction?

Alexander Vasin and Marina Dolmatova

1 Introduction

Electricity markets have been developing in several countries for about 20 years. An important issue for such markets is a limitation of large producers' market power. An alternative to antitrust regulation way to solve the problem is to choose such kind of an auction that would minimize producers market power. A number of papers is devoted to modeling electricity auctions [1–3, 5, 6]. In these papers authors describe auctions as a normal form game and consider Nash equilibrium and its modifications as a model of behavior.

Another important feature of electricity markets is demand uncertainty. In this context, paper [3] proposes a promising auction model and theoretical results. The authors assume a bid of a producer to be a monotone smooth function and a demand function to depend on a random parameter. For a symmetric oligopoly, the authors derive a necessary condition for an equilibrium bid as a differential equation and describe the set of SFE. They show that SFE price is always less than the Cournot oligopoly price. For some cases, the price reduction is considerable [2, 6], so the supply function auction model leads to market power limitation.

However, computation of SFE bids is rather a sophisticated mathematical problem and requires full information on demand and costs of all competitors. Why should one expect that the actual behavior at the auction corresponds to this concept?

The answer to a similar question for Nash equilibria of normal form games is given in the framework of adaptive and learning mechanisms' investigation [4]. The study shows that for some classes of games rather simple adaptive-imitative

A. Vasin
Lomonosov Moscow State University, Moscow, Russia
e-mail: vasin@cs.msu.su

M. Dolmatova (✉)
Lomonosov Moscow State University, Moscow, Russia
e-mail: ms.marina.dolmatova@gmail.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_34,

mechanisms provide convergence of strategy profiles to stable NE for players with bounded rationality and incomplete information.

The present paper aims to consider best response dynamics for two variants of a symmetric oligopoly with a linear demand function: (a) with linear marginal cost, (b) with fixed marginal cost and capacity constraint. Our purpose is to find out whether the dynamics converges to any SFE and describe the set of SFE for each case.

2 A Model of an Auction and SFE Concept

In [3] the authors consider the following model of the market: $N = \{1, n\}$ is the set of players (producers). For each player, $C(q)$ is the total cost function depending on a production volume q, with standard properties: $C'(q) > 0, C''(q) \geq 0 \, \forall \, q \geq 0$. Consumers are characterized by demand function $D(p, t)$ depending on price $p \geq 0$ and random factor t with strictly positive density everywhere on the support $[\underline{t}, \bar{t}]$. For any p, t, the demand function meets conditions: $D_p < 0$, $\mathrm{D}_{pp} \leq 0$, $\mathrm{D}_{pt} = 0$, $\mathrm{D}_t > 0$, (so that D_p does not depend on t). A strategy of a player i is a twice continuously differentiable supply function $S^i(p)$ that determines the supplied amount of the good depending on the market price p. The players set $S^i(p)$ simultaneously without information about random factor t. After realization of t, strategy profile $\overrightarrow{S} = (S^1(p), S^2(p), .., S^n(p))$ determines the price $p(\overrightarrow{S}, t)$ from the aggregate supply and demand balance condition: $D(p(t)) = \sum_{i=1}^{n} S^i(p(t))$ provided a unique price $p(\overrightarrow{S}, t)$ exists. If the market clearing price does not exist or is not unique, then no production takes place and firms' profits are zero. Each producer i aims to maximize his profit $\pi_i(\overrightarrow{S}, t) = p(\overrightarrow{S}, t) * S^i(p(\overrightarrow{S}, t)) - C_i(S^i(p(\overrightarrow{S}, t)), i \in N$. Strategy profile $\overrightarrow{S}^* = (S^{*i}, i \in N)$ is called an SFE if $\forall t S^{*i} \in \mathrm{Argmax}_{S_i}(\pi_i(S^i, S^{*-i}, t))$.

Proposition 1 *(see Proposition 1.2 in [3]). If* $\sup_t D(0, t) = \infty$, *then* $\overrightarrow{S}^* = (S^{*1}, S^{*2}, ..., S^{*n})$ *is SFE if and only if* $S^i(p) \equiv S(p), \forall i \in N$, $S(p)$ *monotonously increases in* p *and meets equation:*

$$S'(p) = \frac{S(p)}{p - C'(S(p))} + D_p(p) \tag{1}$$

Below we assume that a bid is feasible only if it is a monotonously non-decreasing function. This corresponds to the real-life auction rules.

3 The Model with Linear Marginal Cost Function

SFE description Consider a symmetric duopoly with cost function $C(q) = (c_0 + 0,5c_1q)q$, where $c_0 > 0, c_1 > 0$, and demand function $D(p,t) = \overline{D}(t) - dp$, where $d > 0$ and $\overline{D}(t)$ is a maximal demand value depending on random parameter t. According to Proposition 1, an equilibrium supply function for this case should meet the differential equation: $S'(p) = \frac{S(p)}{p-c_0-c_1S(p)} - d$. There exists an infinite set of non-linear solutions. However, if $sup_t D(t) = \infty$ then there exists a unique SFE and the bid function is linear (See Proposition 4 in [3]):

$$S^*(p) = 0,5(p - c_0)d(-1 + \sqrt{\frac{4}{dc_1} + 1}). \tag{2}$$

Best response dynamics At every step $\tau = 1, 2, ...$ each firm sets bid $S(p,\tau)$ that is the best response to its competitor's bid $S(p,\tau-1)$ at the previous step. (We assume $S(p,0) = 0$).

Formally $S(p,\tau)$ is the best response to $S(p,\tau-1)$ if for each $t \in [\underline{t},\overline{t}]$ solution $p(\tau,t)$ to $S(p,\tau) + S(p,\tau-1) = D(p,t)$ provides the maximum profit:

$$p(\tau,t) \longrightarrow \max_p[(D(p,t) - S(p,\tau-1))p - C(D(p(t) - S(p,\tau-1))]$$

Note that the solution of profit optimization problem $\max_p(D(p)p - C(D(p)))$ with concave demand function and convex cost function meets the first order condition determining the optimal output volume:

$$q^*(p) = (p - C'(D(p)))|D'(p)| = D(p) \tag{3}$$

The function $q^*(p)$, determined according to (3), is a Cournot supply schedule (see [4]). For the first step the supply function maximizing the profit coincides with Cournot supply schedule : $S^1(p,1) = (p-c_0)\frac{d}{1+c_1d}$.Though the demand depends on the random parameter t, at every step there exists a bid that maximizes the profit under any value of this parameter. Indeed, let the competitor's bid be $S^2 = \max(0, k(p - c_0))$. Then the solution $\overline{p}(t)$ to the firm's profit maximization problem

$$\max_p[(D(p,t) - S^2(p))p - C^1(D(p,t) - S^2(p))] \tag{4}$$

determines optimal market clearing price of this auction. In order to realize this price the firm should bid so that $S^1(\overline{p}(t)) = D(\overline{p}(t),t) - S^2(\overline{p}(t))$.

Proposition 2 *The bid* $S^1(p) = \frac{(p-c_0)(d+k)}{1+c_1(d+k)}$ *is the best response to bid* $S^2(p) = k(p-c_0)$ *under any* $\overline{D}(t) > dc_0$. *Thus, the best response bid at the step* τ *is* $S(p,\tau) = k_\tau(p-c_0)$, *where* $k_\tau = \frac{d+k_{\tau-1}}{1+c_1(d+k_{\tau-1})}$. *The unique fixed point* $k^* = \frac{d}{2}(\sqrt{\frac{4}{dc_1}+1}-1)$ *for this best response corresponds to the SFE (2) of the static model of the auction.*

Proposition 3 *In the model with linear marginal cost, the best response dynamics converges to the SFE of the static model. Moreover,* $|\frac{k_\tau}{k^*} - 1| \le |\frac{k_1}{k^*} - 1|(1 + c_1 d)^{\tau - 1}$.

4 The Model with Fixed Marginal Cost and a Capacity Constraint

SFE description In this case the cost function is $C(q) = cq, c > 0$ and $q \le Q$. SFE bid for this case is a continuous monotone function that meets equation : $S'(p) = \frac{S(p)}{(p-c)} - d$ until $S(p)$ equals Q for some p or reaches its maximum in p, then staying constant. General solution of the equation $S(p, A) = (p - c)(A - d\ln(p - c))$ depends on the integration constant A (see [5]). This function reaches maximum value $q(A)$ under $p = p(A) \stackrel{def}{=} c + e^{\frac{A}{d} - 1}$ in the point of intersection of its graph and the Cournot supply schedule $d(p-c)$. The inverse function is $A(q) = d(ln(q) - ln(d) + 1)$. If the capacity constraint is not binding then the equilibrium price $\tilde{p}(A)$ is determined from equation $2S(p, A) = \overline{D} - dc$. We characterize the set of SFE depending on the maximum value of demand for $p = c$. Denote $D^* = \sup_t \overline{D}(t) - dc$. Define a function $D^*(Q)$ from the equation $D^* - d(p - c) = 2Q = 2d(p - c)$. Then $D^*(Q) = 3Q$. The value $D^*(Q)$ determines the maximum demand function parameter for which the capacity constraint is not binding in the Cournot auction.

Theorem 1 *If* $D^* \ge 3Q$ *then there exists a unique SFE in the model. The equilibrium bid is* $S^*(p) = \begin{cases} S(p, A(Q)) \text{ for } c \le p \le p(A) \\ Q, \text{ for } p \ge p(A) \end{cases}$

If $Q < D^* < 3Q$, *then for any* $A \in (A(D^*/3), A(Q))$ *bid* $\overline{S}(p, A) = \begin{cases} S(p, A), p \le p(A) \\ S(p(A), A), p \ge p(A) \end{cases}$ *determines SFE, and for any* $A \in (A(Q), \overline{A}(D^*))$, *where* $\overline{A}(D^*)$ *meets equation* $S(\tilde{p}(A, D^*))(\tilde{p}(A, D^*)) = \frac{(D^* - Q)^2}{4d}$, *bid* $\overline{\overline{S}}(p, A) = \min\{\overline{S}(p, A), Q)$ *also determines SFE.*

If $D^* < Q$, *then the capacity constraint is not binding for any feasible bids. In this case* $\forall A > A(D^*/3)$ *the bid* $S(p, A)$ *determines SFE. For* $A \to \infty$ *SFE tends to Walrasian equilibrium.*

Best response dynamics The best response dynamics for the model with fixed marginal cost and no capacity constraint ($C(q) = cq, c > 0$) is a particular case of the model in the Sect. 3, so the BR function converges to the Walrasian equilibrium.

Now consider the adaptive mechanism for the model with capacity constraint $q_i \le Q, i = 1, 2$. Let $c = 0$ (as non-zero marginal cost will only replace p with $p - c$ in further equations).

For a model with concave residual demand function, the bid, maximizing firm's profit for a given demand value, can be found as a solution of the problem (4).

Proposition 4 *The best response dynamics for* $D^* \le Q$ *is*

$$S(p,\tau) = \begin{cases} \tau d(p-c),\ p < \frac{Q}{\tau d} + c \\ Q,\ p \ge \frac{Q}{\tau d} + c. \end{cases}$$

It converges to the SFE that corresponds to the Walrasian equilibrium for this case.

Consider the case where $D^* \ge 3Q$. Let $c = 0, d = 1$. The third step best response function is $S(p,3) = \begin{matrix} min\{3p, Q\}, 0 \le p \le Q/2, & \texttt{for } \overline{D} \le \frac{7}{3}Q, \\ min\{p, Q\}, \frac{2}{3}Q \le p, & \texttt{for } \overline{D} > \frac{7}{3}Q. \end{matrix}$ (see Fig. 1)

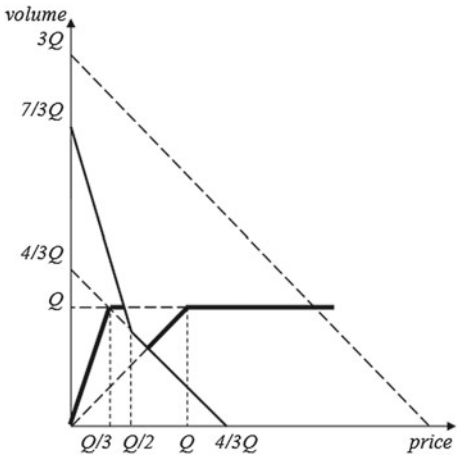


Fig. 1 It converges to the SFE that corresponds to the Walrasian equilibrium Case $D^* \ge 3Q$

Thus, there is no monotonous best response function for the third step and the best response dynamics does not converge to SFE.

Proceeding from this result, we consider another approach. Below we fix the value of $\overline{D}$ and define the best response dynamics depending on the ratio between $\overline{D}$ and Q. Our aim is to define ranges of parameters with similar best response dynamics.

Hereinafter we consider demand function $D(p) = max\{0, \overline{D} - dp\} = S(p,0)$, $c = 0$. A bid of each player is $S(p,\tau) = min(k(\tau)p, Q)$, so the problem of BR determination is to find the optimal slope $k(\tau)$.

Proposition 5 *The best response dynamics depends on the parameter values as follows:*

For $\overline{D} \ge 3Q$ $S(p,\tau) = min\{Q, dp\}, \forall\tau$. *For any* τ *the SFE coincides with the Cournot supply schedule. The outcome corresponds to the Walrasian equilibrium that is equal to the Cournot equilibrium in this case.*

For $Q < \overline{D} < 3Q$ $S(p,\tau) = min\{Q, d\tau p\}$ *at step* $\tau = 1, \ldots, T(\overline{D})$, *then the best response functions repeat in cycle. The length of the best response dynamics*

$$cycle\ T(\overline{D}) = \begin{cases} 2, & for\ \frac{7}{3}Q < \overline{D} < 3Q, \\ 3, & for\ 2Q < \overline{D} \le \frac{7}{3}Q, \\ |\frac{\overline{D}}{\overline{D}-Q}^2|, & for\ Q < \overline{D} \le 2Q. \end{cases}$$

For $\overline{D} \leq Q\ S(p,\tau) = min\{Q, d\tau p\}, \forall\tau$. *For* $\tau \to \infty$ *the best response dynamics converges to the SFE with the outcome corresponding to the Walrasian equilibrium.*

Generalization to the oligopoly The BRD for the model with a constant marginal cost and unconstrained capacity is similar to the duopoly case.The best response dynamics can be described as $S(p,\tau) = k(\tau)p, k(\tau) = d\sum_{s=0}^{\tau-1}(n-1)^s$. For $\tau \to \infty$ the BRD converges to the SFE which corresponds to the Walrasian equilibrium.

Consider the BRD for the model with the capacity constraint.

Proposition 6 *The BRD for the model of n-firm oligopoly with constant marginal cost and capacity constraint depends on the ratio between parameters* $\overline{D}$ *and* Q *as follows:*

$\mathbf{\overline{D} \leq (n-1)Q}$ *:* $S(p,\tau) = min\{Q, d\tau p\}, \forall\tau \geq 1$. *For* $\tau \to \infty$ *the BRD converges to the SFE that is equal to the Walrasian equilibrium;*

$\mathbf{(n-1)Q < \overline{D} < (n+1)Q}$ *: At the step* $\tau = 1,\ldots,T$ *the best response is* $S(p,\tau) = min\{Q, d\tau p\}$, *then the best response functions repeat cyclic. Length of the cycle* T *is the minimum integer number satisfying inequality* $\sum_{s=0}^{T}(n-1)^s > \frac{4q(\overline{D}-Q)}{(\overline{D}-(n-1)Q)^2}$;

$\mathbf{(n+1)Q \leq \overline{D}}$ *:* $S(p,\tau) = min\{Q, dp\}, \forall\tau \geq 1$. *The BRD converges to the SFE coinciding with the Cournot supply schedule (which is equal to the Walrasian equilibrium in this case).*

5 Conclusion

Discontinuous marginal cost is a typical feature of real-life power markets. The discontinuity is caused by generators capacity constraints and difference in marginal costs for different generators types. The obtained results imply that, being implemented in practice, the SFE auction proposed in [3] does not provide convergence of suppliers bidding strategies to the SFE.

References

1. Green, R., Newbery, D.: Competition in the British Electricity Spot Market. J. Polit. Econ. **100**, 929–953 (1992)
2. Green, R.: The electricity contract market in England and Wales. J. Ind. Econ. **47**, 107–124 (1999). doi:10.1111/1467-6451.00092
3. Klemperer, P.D., Meyer, M.A.: Supply function equilibria in oligopoly under uncertainty. Econometrica **57**(6), 1243–1277 (1989)
4. Milgrom, P., Roberts, J.: Adaptive and sophisticated learning in normal form games. Games Econ. Behav. **3**(1), 82–100 (1991)
5. Newbery, D.: Analytic solutions for supply functions equilibria: uniqueness and stability. Cambridge Working Papers in Economics 0848, Faculty of Economics, University of Cambridge (2008) http://www.eprg.group.cam.ac.uk/category/publications/
6. Newbery, D.: Competitions, contracts and entry in the electricity spot market. RAND J. Econ. **29**(4), 726–749 (1998)

Part IX
Health Care Management

A Three-Objective Optimization Approach to Cost Effectiveness Analysis Under Uncertainty

Walter J. Gutjahr

1 Introduction

One of the most important quantitative decision analysis tools for healthcare resource allocation is cost-effectiveness analysis (CEA). A known drawback of CEA lies in the fact that usually, neither cost nor effect of a health programme (medical intervention, treatment, screening programme etc.) are known with certainty in advance. Thus, a stochastic portfolio optimization problem for health programmes with the two different objectives "cost" and "health effect" arises. Several articles have studied healthcare resource allocation under stochastically modelled uncertainty [1, 3, 9, 11, 13]. However, up to now, no broadly accepted method for CEA under uncertainty seems to exist. Multiobjective stochastic optimization problems are notoriously hard (cf. [2, 7]). The possible dependence between the random objectives increases the difficulty, especially in the case where the decision maker is risk-averse. A broadly accepted analytical approach to decisions under risk aversion is *expected utility theory*, which is closely related to stochastic dominance concepts. This approach has recently also been extended to the multi-objective case [4, 8]. Nevertheless, the models in the last-mentioned articles assume that the objectives can be traded against each other, and this stipulates a perfect market which is usually not present in the health sector.

In this paper, we propose an alternative approach that also relies on the sound axiomatic foundations of expected utility theory. It will be argued that frequently, a healthcare decision maker (DM) may be risk-averse toward cost, but risk-neutral toward health effect. On this assumption, a recent theoretical result [6] will be used to show that stochastic dependence between the two objective functions is not detrimental: The random objectives in the bivariate stochastic optimization problem can be suitably decoupled, which leads to a three-objective optimization model.

W. J. Gutjahr (✉)
Department of Statistics and Operations Research, University of Vienna, Vienna, Austria
e-mail: walter.gutjahr@univie.ac.at

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_35,

2 A Mean-Mean-Risk Model for Healthcare Resource Allocation

We agree with [1] and other authors that in practice, healthcare DMs can rarely be risk-neutral toward cost, because they almost always face budget limits. No matter whether such limits are hard or soft, they imply that the utility function of the DM is not linear anymore in the term "total cost". As a consequence, uncertainty in the actual cost must lead to risk-averse behavior.

For health effects, measured e.g. in quality-adjusted life years (QALYs), the situation is more intricate. A usual argument for recommending risk aversion of the DM toward health effect is that *individuals* are risk-averse w.r.t. the outcome of their treatments, so if the public DM aims at taking account of the risk attitude of the individuals, she should be risk-averse w.r.t. the aggregated outcome (the total effect). However, this argument—although intuitive—is misled, as can be shown by simple examples distinguishing between *uncertainty* (parameters are not known with certainty) and *variability* (patients react differently to treatment).

If, e.g., a choice has to be made between two programmes A and B where A is subject to large uncertainty, but small variability, and for B, the opposite holds, then risk-averse individuals can prefer A, whereas a risk-averse public DM can prefer B.

From the perspective of expected utility theory, risk aversion has its source in a concave utility function. An individual is risk-averse w.r.t. QALYs, if her utility is concave in the amount of (additional) QALYs. If the public DM wishes to take account of the individual preferences, an utilitarian stance would require to aggregate the individual utilities to a total utility, the expected value of which is then to be maximized. The proper way to take individual risk aversion into account would be to re-scale the effect measure (if necessary) in such a way that it represents individual utilities. With respect to the aggregated effect, the DM should be risk-neutral.

Mathematically, we have a bivariate stochastic optimization problem of the form

$$\max\ (f_1(x, \boldsymbol{\omega}),\ f_2(x, \boldsymbol{\omega})) \quad \text{s.t.} \quad x \in S,$$

where f_1 represents negative cost, f_2 represents effect, x is the solution (decision), $\boldsymbol{\omega}$ is the influence of randomness and S is the set of feasible solutions. The goal is to determine the set of stochastically nondominated solutions of this problem or a good approximation to it. This set can then be presented to the DM for final selection of her most preferred solution. To make the problem well-defined, we have to fix a notion of bivariate stochastic dominance. Let $\mathcal{U}_2$ denote the set of nondecreasing, concave utility functions $\mathbf{u} : \mathbb{R}^2 \to \mathbb{R}$. This is a customary class of utility functions representing the preferences of a DM who is rational and not risk-seeking (cf. [10]). Furthermore, let $\mathcal{N}_2$ denote the set of utility functions $\mathbf{u} : \mathbb{R}^2 \to \mathbb{R}$, $\mathbf{u} = \mathbf{u}(f_1, f_2)$, that are linear in the second argument f_2 for each fixed first argument f_1 and thus risk-neutral w.r.t. f_2. We say that solution $x \in S$ *dominates* solution $x' \in S$, if

$$\mathbf{E}[\mathbf{u}(f_1(x, \omega),\ f_2(x, \omega))] \ \geq\ \mathbf{E}[\mathbf{u}(f_1(x', \omega),\ f_2(x', \omega))] \tag{1}$$

for all $\mathbf{u} \in \mathcal{U}_2 \cup \mathcal{N}_2$, with at least one $\mathbf{u}$ making the inequality strict. A solution $x \in S$ is called *efficient* if there is no $x' \in S$ that dominates x. Theorem 2 in [6] implies then that x dominates x' in the above-mentioned sense if and only if

$$\mathbf{E}[u(f_1(x,\omega))] \geq \mathbf{E}[u(f_1(x',\omega))] \text{ for all nondecreasing and concave } u : \mathbb{R} \to \mathbb{R}, \tag{2}$$

and

$$\mathbf{E}[f_2(x,\omega)] \geq \mathbf{E}[f_2(x',\omega)], \tag{3}$$

with at least one strict inequality. Condition (2) is the (ordinary) second-order stochastic dominance, it is usually written $f_1(x,\omega) \succeq_{(2)} f_1(x',\omega)$.

The determination of the set of efficient solutions x w.r.t. conditions (2)–(3) gives a multi-objective optimization problem with a large (possibly infinite) number of objectives. This difficulty already arises in financial portfolio optimization where objective f_2 and hence condition (3) is absent. In this area, a frequently applied approach is to restrict the dimensionality by considering only two "aspects" of f_1, the mean and a risk measure. We obtain a so-called *mean-risk* model. Let us employ the term *Pareto-optimal* solutions instead of efficient solutions in the case of a restriction of the objectives to mean and risk measures. The set of Pareto-optimal solutions of the mean-risk model is typically much smaller than the set of efficient solutions of the original problem.

As an approximation to the set of nondominated solutions in the sense of (1), we take the set of Pareto-optimal solutions of the following *mean-mean-risk* model:

$$\max\ (-\mathbf{E}[\text{cost}(x)],\ \mathbf{E}[\text{effect}(x)],\ -\text{risk}(x)) \quad \text{s.t. } x \in S \tag{4}$$

Therein, as a risk measure, we start by using either the absolute semideviation of cost, i.e. $\text{risk}(x) = \bar{\delta}(\text{cost})$, or the sum of expected cost and absolute semideviation of cost, i.e. $\text{risk}(x) = \mathbf{E}[\text{cost}] + \bar{\delta}(\text{cost})$. The absolute semideviation of a random variable X is given as $\bar{\delta}(X) = \mathbf{E}[\max(\mathbf{E}[X] - X, 0)]$. In the case where cost is the only relevant objective function, the second alternative has the advantage that, as shown in [12], each Pareto-optimal solution of the mean-risk model is an efficient solution of the original problem. By Theorem 2 of [6], this guarantee carries over to the case of both cost and effect as objectives. It is easy to see that the set of Pareto-optimal solutions under the second risk definition is a subset of those under the first risk definition.

Let us consider now the specific healthcare portfolio model investigated in [1]. It assumes that there are n perfectly divisible health programmes $i = 1, \ldots, n$. Each programme i can be implemented to some degree $x_i \in [0,1]$. Thus, the decision is given by the vector $x = (x_1, \ldots, x_n)$. Furthermore, let us consider the *discrete probability* case where the stochastic model is given by N possible scenarios with probabilities p_ν $(\nu = 1, \ldots, N)$. Under scenario ν, programme i, if completely implemented, incurs cost $c_i^{(\nu)}$ and provides health effect $b_i^{(\nu)}$. In this specific situation, problem (4) under the first choice for the risk measure can be given the following mathematical programming formulation:

$$\max(f_1, f_2, f_3) \quad \text{s.t.}$$

$$f_1 = -\sum_{i=1}^{n} c_i\, x_i, \quad f_2 = \sum_{i=1}^{n} b_i\, x_i, \quad f_3 = -\sum_{v=1}^{N} p_v\, \eta_v$$

$$\eta_v \;\geq\; f_1 + \sum_{i=1}^{n} c_i^{(v)}\, x_i \quad (v = 1, \ldots, N),$$

$$\eta_v \;\geq\; 0 \quad (v = 1, \ldots, N), \quad 0 \leq x_i \leq 1 \quad (i = 1, \ldots, n)$$

where $c_i = \sum_{v=1}^{N} p_v\, c_i^{(v)}$ and $b_i = \sum_{v=1}^{N} p_v\, b_i^{(v)}$ for $i = 1, \ldots, n$. In the case of the second choice for the risk measure, the term f_1 has to be added on the r.h.s. of the equation defining f_3. Both versions are multi-objective linear programs (MOLPs) with $n + N$ variables and can be solved by one of the MOLP algorithms proposed in the literature, see [5]. For some probability models, also the *continuous probability* case can be treated, as will be shown below.

3 Example

In this section, the three-programmes illustration example presented in [1] is slightly extended to an example with $n = 5$ health programmes, see Table 1. As in [1], it is assumed now that costs and effects are independent and normally distributed. For this model, the objective functions in (4) can be calculated analytically (we omit the formulas). Figure 1 shows the Pareto frontier (i.e., the image of the set of Pareto-optimal solutions in the objective space) for the test case. Risk has been defined as absolute semideviation. The frontier has been projected to the plane given by the two axes "expected cost" and "expected effect". The values of the third objective (risk) are only indicated in a rough classification.

The second alternative for the risk measure (expected cost plus absolute semideviation) yields in this example only the upper boundary curve in Fig. 1 as the Pareto frontier. This means that in the considered test case, it is actually not necessary to take account of the risk measure. Note that the risk measure "expected cost plus absolute

Table 1 Expected cost c_i, expected effect b_i and standard deviation of cost σ_i of five health programmes

i	c_i	b_i	σ_i
1	500	1,500	0
2	450	1,500	200
3	400	2,500	240
4	350	600	110
5	180	1,300	0

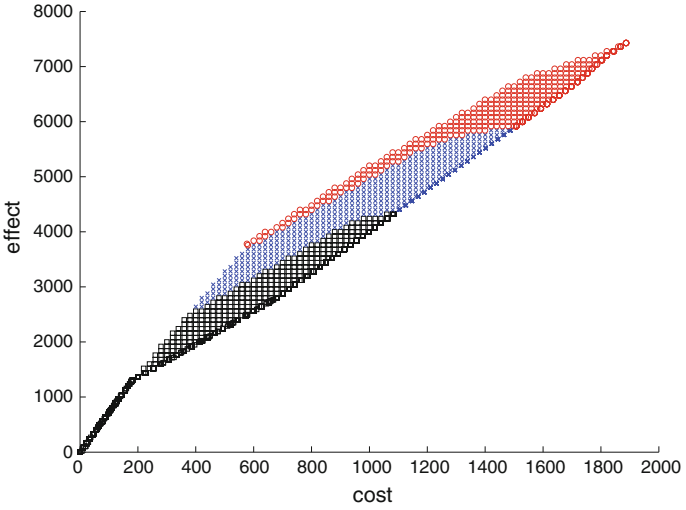


Fig. 1 Projection of the Pareto frontier for the test case, with absolute semideviation as the chosen risk measure, to the cost-effectiveness plane. The risk values are classified into low (*black squares*), medium (*blue crosses*) and high (*red circles*)

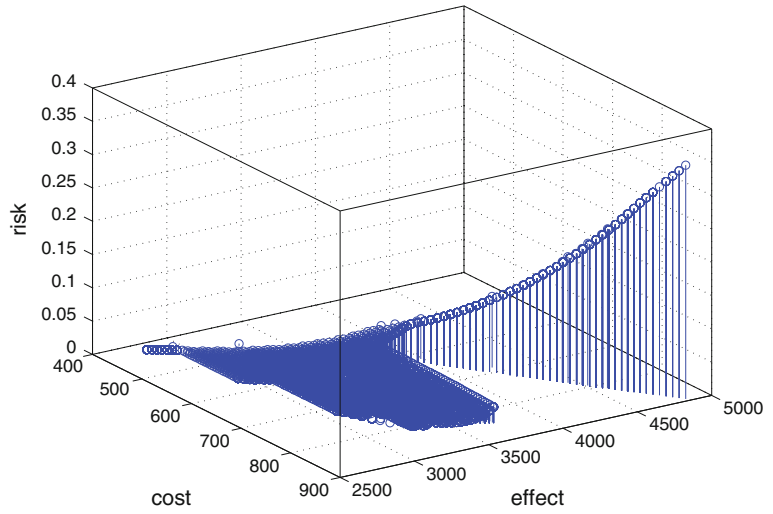


Fig. 2 Stem plot of the relevant part of the Pareto frontier for the test case, with budget overrun probability as the chosen risk measure

semideviation" does not cover utility functions making the DM *very* risk-averse. Therefore, the measure might be suitable for the situation of soft budget constraints. In the presence of hard budget constraints, a higher degree of risk aversion will prevail. We propose that for the latter case, the (more conservative) budget overrun

probability is taken as the suitable risk measure in the mean-mean-risk model. Figure 2 shows a stem plot of the Pareto frontier of the example test case if the third objective (risk) is defined as the probability that the budget threshold $\beta = 1000$ is exceeded. We see that the Pareto frontier degenerates to a curve in a range of higher expected effect values, but remains an area in a medium range.

References

1. Al, M.J., Feenstra, T.L., van Hout, B.A.: Optimal allocation of resources over health care programmes: dealing with decrasing marginal utility and uncertainty. Health Eco. **14**, 655–667 (2005)
2. Abdelaziz, Ben: F. Solution approaches for the multiobjective stochastic programming. Eur. J. Oper. Res. (2011). doi:10.1016/j.ejor.2011.03.033
3. Bridges, J.F.P., Stewart, M., King, M.T., van Gool, K.: Adapting portfolio theory for the evaluation of multiple investments in health with a multiplicative extension for treatment strategies. Eur. J. Health Econ. **3**, 47–53 (2002)
4. Dentcheva, D., Ruszczyński, A.: Optimization with multivariate stochastic dominance constraints. Math. Program. **117**(Ser. B), 111–127 (2009)
5. Ehrgott, M.: Multicriteria Optimization. Springer, Berling (2005)
6. Gutjahr, W.J.: Multi-objective stochastic optimization under partial risk neutrality. University of Vienna,Technical Report(2012)
7. Gutjahr, W.J., Reiter, P.: Bi-objective project portfolio selection and staff assignment under uncertainty. Optim. Methods Softw. **59**, 417–445 (2010)
8. Hu, J., Homem-de-Mello, T., Mehrotra, S.: Risk-adjusted budget allocation models with application in homeland security. IIE Trans. **43**, 819–839 (2011)
9. McKenna, C., Claxton, K.: Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. Med. Decis. Making **31**(6), 853–865 (2011)
10. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. Wiley, Chichester (2002)
11. O'Brian, B.J., Sculpher, M.J.: Building uncertainty into cost-effectiveness rankings - portfolio risk-return tradeoffs and implications for decision rules. Med. Care **15**, 460–468 (2000)
12. Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and mean-semideviation models. Math. Program. **89**, 217–232 (2001)
13. Sendi, P., Al, M.J., Gafni, A., Birch, S.: Optimizing a protfolio of health care prgrams in the presence of uncertainty and constrained resources. Soc. Sci. Med. **57**, 2207–2215 (2003)

Part X
Information Systems, Neural Nets and Fuzzy Systems

Spot and Freight Rate Futures in the Tanker Shipping Market: Short-Term Forecasting with Linear and Non-Linear Methods

Christian von Spreckelsen, Hans-Joerg von Mettenheim and Michael H. Breitner

1 Introduction and Methodology

In this paper we investigate the forecasting and trading performance of non-linear forecasting models, to generate short-term forecasts of spot rates and corresponding freight forwards respectively Forward Freight Agreements (FFA) in the dirty tanker shipping market. In recent time freight derivatives become interesting in the maritime market, due to the fact, that freight rates are very volatile. Derivative markets provide a way in which these risks may be transferred to other individuals who are willing to bear them, through hedging. Therefore, actors in the shipping market are forced to use forecasting techniques for the purpose of risk management.

Most studies on forecasting freight rates—see Kavussanos and Visvikis [3] for a brief survey about forecasting freight derivatives—use traditional time series models and focus on statistical forecast performance measures. Batchelor et al. [2] compared a range of time series models in forecasting spot freight rates and FFA prices. They concluded, that freight forward contracts are suitable to detect the tendency of future spot freight rate, but FFA rates do not seem to be helpful in predicting the direction of future spot rates. However, the use of linear time series models for freight rates is sometimes criticized, due to the fact that most financial time series show non-linear patterns (see Adland and Culliane [1]). Li and Parsons [4] and Lyridis et al. [5] attempted to investigate the advantages of neural networks in predicting spot freight rates. They pointed out that neural networks can significantly outperform simple time

C. von Spreckelsen (✉) · H.-J. von Mettenheim · M. H. Breitner
Leibniz Universitaet Hannover, Institut fuer Wirtschaftsinformatik, Koenigsworther Platz 1, 30167Hannover, Germany
e-mail: spreckelsen@iwi.uni-hannover.de

H.-J. von Mettenheim
e-mail: mettenheim@iwi.uni-hannover.de

M. H. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_36,

series models for longer-term forecasting of monthly tanker spot freight rates. A more recent study of Mettenheim and Breitner [6] showed that neural networks achieve encouraging forecasting and trading results in predicting the Baltic Dry Index (BDI).

Nevertheless, we find a lack of jointly spot and forward forecasting and trading investigations with neural networks. We build on the mentioned investigations, but we extend our study on freight derivatives and a wider range of time series models. The main objective of this paper is to investigate neural networks' prediction ability for maritime business short-term forecasting and provide a practical framework for actual trading applications of neural networks. We therefore implement a simple trading simulation to investigate the economic evaluation of the predicted spot freight rates and FFA prices.

Our methodology comprises of a feedforward NN, which also appear to have potential application in time series modelling and forecasting. In addition, we apply traditional linear time series models like the univariate Auto-Regressive Integrated Moving Average (ARIMA) model. Correspondingly to univariate models, statistical multivariate time series methods include the Vector Auto-Regressive process (VAR) and the Vector Error Correction model (VECM) which take into account the simultaneous information content in the spot and futures price movement. We separate all models in univariate and multivariate model classes: The univariate models consist of an ARIMA and a NN model, where we include only the relevant single spot or FFA time series. For the multivariate models VAR, VECM and a multivariate neural network, namely NN+, we include both spot and all FFA rates of each route. All models seem to be well specified, as indicated by relevant diagnostic tests.

The paper is organized as follows. Section 2 describes the data and data preparation. Section 3 gives a brief introduction about our forecasting and trading strategy and shows the performance measures of the neural network and alternative models. We evaluate the statistical forecasting performance via a simple trading simulation. Finally, Sect 4 summarizes our conclusions.

2 Description of Data and Data Preparation

Tanker routes are centralized around the biggest physical routes for shipments of crude oil, known as trade dirty (TD) or trade clean (TC) followed by a numeral to designate the vessel size and cargo. We sample daily prices of the most liquid International Maritime Exchange (Imarex) TD3 and TD5 freight futures (FFA) contracts. These contracts are written on daily spot rates for TD3 and TD5 published by the Baltic Exchange. The spot and FFA data is available from 5 April 2004 to 1 April 2011 (1,748 observations). To avoid expiry effects, we calculate "perpetual" futures contract for one month (22 trading days; FFA 1M) and two month (44 trading days; FFA 2M) as a weighted average of a near and distant futures contracts, weighted according to their respective number of days from maturity.

All prices are transformed to natural logarithms. Summary statistics of logarithmic first-differences ("log-returns") of daily spot and FFA prices are available for the whole period in the two dirty tanker routes. The result's excess kurtosis in all series,

and the skewness does not necessarily imply a symmetric distribution. The Jarque-Bera tests indicate departures from normality for both spot and FFA prices in all routes. This seems to be more acute for the spot freight rates. Augmented Dickey Fuller (ADF) and Phillips-Peron (PP) unit root tests indicate that all variables are log first-difference stationary, but the levels indicate, that most price series follow unit root processes.

3 Forecasting and Trading Performance Test

For purpose of forecasting and trading, each data set is divided into two subsets: The first subset runs from 5 April 2004 to 16 February 2010 (in-sample), the second subset from 17 February 2010 to 1 April 2011 (out-of-sample). This implies that we get a sample of 1,466 daily observations for the estimation period and a sample of 282 daily observations for the forecasting and trading period—a ratio of 5.25 to 1.

3.1 Statistical Forecasting Performance Results

We generate one-step ahead out-of-sample forecasts of each model, estimated over the initial estimation period. The forecasting performance of each model is presented in matrix form in Table 1 for all contracts. Forecasts made using first-differences will be transformed back to levels to ensure that the measures presented above are comparable for all models. The forecast performance of each model is assessed using the conventional root mean square error metric (RMSE).

All models outperform their naïve benchmark, except the ARIMA model in predicting the TD5 FFA 2M. Some regularities stand out from the table. First, the FFA rates are much harder to forecast than the spot rates. This phenomenon is not unusual for freight rates and confirms prior studies in the tanker market. Second, in most cases the multivariate models are superior against the univariate representives. We can find

Table 1 One-step ahead forecast performance for Route TD3 and TD5

				Univariate		Multivariate		
Route	Contract	Measure	RW1[a]	ARIMA	NN	VAR	VECM	NN+
TD3	Spot	RMSE	0.0468	0.0425	0.0406	0.0408	0.0397	0.0403
	FFA 1M	RMSE	0.0307	0.0293	0.0289	0.0295	0.0295	0.0291
	FFA 2M	RMSE	0.0284	0.0279	0.0277	0.0279	0.0279	0.0276
TD5	Spot	RMSE	0.0578	0.0540	0.0516	0.0485	0.0471	0.0496
	FFA 1M	RMSE	0.0298	0.0287	0.0287	0.0290	0.0288	0.0287
	FFA 2M	RMSE	0.0222	0.0223	0.0219	0.0216	0.0216	0.0220

[a] RW1 means a no-change random walk

this advantage especially for spot freight rates. But this error difference or advantage declines in the FFA contracts. Furthermore, the VECM, which has an equilibrium correction feature, perform better than VAR models for forecasts of spot rates, but not for forecasts of FFA rates. The neural network results are comparable to those of the other models. It is interesting that the univariate NN achieve relatively good results, but the multivariate NN+ is not able to reinforce this advantage significantly. It seems, that the neural network as a non-linear approximator is already able to extract sufficient information out of the univariate time-series. The additional information contained in other time-series is therefore not needed.

3.2 Trading Simulation and Performance Results

Statistical performance measures are often inappropriate for financial applications. Trading strategies guided by forecasts on the direction of price change may be more effective and generate higher profits. Based on the generated results we provide a simple trading simulation to evaluate our forecasting results in this section. We can generate trading signals now by the following rule:

$$\begin{cases} \text{long, if } \hat{p}_{t+1} > p_t \\ \text{short, if } \hat{p}_{t+1} < p_t, \end{cases}$$

where p_t denotes the log price and $\hat{p}_{t+1}$ is the one-step ahead forecast. A long signal is to buy contracts at the current price, while a short signal is to sell contracts at the current price. This approach has been widely used in the literature. In the trading experiment, it is assumed that during the initiation period, an investor will invest 1 monetary unit at the beginning of each contract period (282 trades). So far, our results have been presented without accounting for transaction costs during the trading simulation.

Some selected trading performance measures are shown in Table 2 for TD3 and TD5. We see some implications: First, all models earn a positive trading result in case of no transaction costs. Furthermore, it is obvious, that the trading results in spot rates are more profitable than those for FFA contracts. This is also valid for the directional measure "winning trades per %". In most cases all models outperform the naïve RW2 model, except some time series models in predicting FFA prices. The multivariate NN+ undermatch the RW2 benchmark trading results for TD3 spot freight rates.

The results generated by NN are encouraging in comparison to the other models—for every predicted asset the univariate NN shows the best performance across all models (univariate and multivariate) with respect to the important measures of net gain and risk-adjusted return as measured by the Sharpe ratio. ARIMA and VAR results show no unambiguous picture. Both models outperform the RW2 in case of spot rates. But ARIMA does not perform for TD5 FFA 2M contracts and VAR get worse results for TD3 FFA contracts. The multivariate VECM shows relatively

Table 2 Trading performance for route TD3 and TD5

Route	Contract	Measure (%)	RW2[a]	Univariate		Multivariate		
				ARIMA	NN	VAR	VECM	NN+
TD3	Spot	Net gain	4.76	5.52	6.07	5.72	5.48	4.64
		Log-returns	1.69	1.96	2.15	2.03	1.94	1.65
		Sharpe ratio	0.39	0.46	0.52	0.48	0.46	0.38
		Winning trades	0.73	0.73	0.71	0.73	0.72	0.71
	FFA 1M	Net gain	2.01	2.12	2.35	1.79	1.81	2.05
		Log-returns	0.71	0.75	0.83	0.64	0.64	0.73
		Sharpe ratio	0.24	0.25	0.28	0.21	0.21	0.24
		Winning trades	0.55	0.57	0.58	0.56	0.58	0.59
	FFA 2M	Net gain	1.01	1.12	1.23	0.61	1.14	1.72
		Log-returns	0.36	0.40	0.44	0.22	0.40	0.61
		Sharpe ratio	0.13	0.14	0.16	0.08	0.14	0.22
		Winning trades	0.51	0.54	0.53	0.50	0.57	0.54
TD5	Spot	Net gain	5.73	6.45	6.95	7.13	7.15	7.32
		Log-returns	2.03	2.29	2.47	2.53	2.54	2.60
		Sharpe ratio	0.38	0.43	0.51	0.49	0.49	0.50
		Winning trades	0.73	0.74	0.74	0.73	0.72	0.74
	FFA 1M	Net gain	2.01	2.12	2.35	1.81	1.81	2.24
		Log-returns	0.71	0.75	0.83	0.64	0.64	0.79
		Sharpe ratio	0.20	0.22	0.27	0.22	0.23	0.28
		Winning trades	0.55	0.59	0.62	0.59	0.59	0.60
	FFA 2M	Net gain	1.01	0.58	1.28	1.40	1.52	1.08
		Log-returns	0.36	0.21	0.45	0.49	0.54	0.38
		Sharpe ratio	0.16	0.09	0.21	0.23	0.25	0.17
		Winning trades	0.51	0.52	0.56	0.57	0.59	0.56

[a] For trading purposes, we switch from the no-change RW1 to a constant (last) change random walk (RW2), where the actual rate of return is the forecast of the next period

good and stable results, except for the FFA 1M contracts. The multivariate NN+ achieves only in some cases preeminent trading results, e.g. for the TD5 spot freight rates. As mentioned above, additional time series do not improve the neural network performance. We conclude, that both VECM and univariate NN may generate more robust trading results for this time series and perform better than the other forecasting models.

4 Conclusions and Recommendations

In this paper, we have examined the forecasting and trading performance of various standard linear time series models and a non-linear neural network to jointly predict spot and forward freight rates (FFA prices). We have focused on short-term forecasting. To our knowledge there is a lack in the literature of joint predictions of freight

rates and derivatives with neural networks and traditional time series models. We conclude, that neural networks are suitable for short-term forecasting and trading of tanker freight rates and derivatives.

For the two most liquid tanker routes TD3 and TD5 we implicate that short-term forecasting with neural networks leads to better results than other traditional time series models. Our forecasting results confirm prior studies concerning time series models. However, out-of-sample forecasting with multivariate forecasting models show that spot freight rates are not helpful in predicting FFA prices, but FFA prices do help predict spot freight rates. The results of neural networks are in line with these findings. We have also implemented a simple trading simulation to evaluate the forecasting performance with economical criteria. In our evidence, both VECM and univariate neural networks may generate more robust trading results for the analyzed time series than the other forecasting models.

Nevertheless, we think that further research with freight rates and corresponding derivatives is developable for decision and trading applications with enhanced forecasting models. Several extensions for further research are also thinkable, e.g. longer investment horizons and inclusion of further exogenous input variables in multivariate models like crude oil prices, maritime data or any other variables.

References

1. Adland, R., Cullinane, K.: The non-linear dynamics of spot freight rates in tanker markets. Trans. Res. Part E **42**, 211–224 (2006)
2. Batchelor, R., Alizadeh, A.H., Visvikis, I.D.: Forecasting spot and forward prices in the international freight market. Int. J. Forecast **23**(1), 101–114 (2007)
3. Kavussanos, M.G., Visvikis, I.D.: Shipping freight derivatives: A survey of recent evidence. Marit. Policy Manage. **33**(3), 233–255 (2006)
4. Li, J., Parsons, M.: Forecasting tanker freight rate using neural networks. Marit. Policy Manage. **24**(1), 9–30 (1997)
5. Lyridis, D.V., Zacharioudakis, P., Mitrou, P., Mylonas, A.: Forecasting tanker market using artificial neural networks. Marit. Econ. Logistics **6**, 93–108 (2004)
6. Mettenheim, H.-J., Breitner, M. H.: Robust decision support systems with matrix forecasts and shared layer perceptrons for finance and other applications. ICIS 2010 Proceedings 83 (2010).

Forecasting Daily Highs and Lows of Liquid Assets with Neural Networks

Hans-Jörg von Mettenheim and Michael H. Breitner

1 Motivation and Introduction

For a long time open-high-low-close (OHLC) data has been available cheaply or even for free on financial websites. While there is a host of forecasting studies using OHLC data as inputs for close-to-close analysis the (academic) literature on intraday systems trading highs and lows is scarce. At the other end of the spectrum are studies using high-frequency data. Curiously, a literature search reveals more high-frequency studies than OHLC studies.

However, studies which even just use close-open or open-close data show attractive results. For example [1] analyze returns on different S&P indices. In the present study we additionally use high-low data to trade on. We derive our methodology from [2, 3]. They analyze the Brazilian stock market and trade the high-low range resulting in outstanding economic performance.

The question arises what the reason for this performance is. Is it due to the novel method, to the pecularities of the Brazilian stock market or to a combination of both? The authors themselves acknowledge that there is not much literature available on the Brazilian stock market. We can therefore not put the authors' results in comparison to other studies in Brazil. We can, however, transfer and adapt the methodology to the US stock market and qualitatively compare the results. As it turns out, results on the US stock market are generally very promising, too. The Brazilian stock market studies by Martinez et al. [2], Gomide and Milidiu [3] use a 3-layer perceptron. In contrast we use a Historically Consistent Neural Network (HCNN), introduced by Zimmermann [4]. HCNNs provide remarkably robust forecasts, see [5, 6].

H.-J. Mettenheim (✉) · M. H. Breitner
Leibniz Universität Hannover, Institut für Wirtschaftsinformatik, Königsworther Platz 1, 30167Hannover, Germany
e-mail: mettenheim@iwi.uni-hannover.de

M. H. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_37,

We structure the remainder of this paper as follows. Section 2 briefly recalls the HCNN and presents general characteristics of the learning data. Section 3 deals with the peculiarities of building a testable trading system using only OHLC data. Section 4 presents performance results. Section 5 concludes, discusses limitations of our approach and outlines further work.

2 Neural Network Modeling

The goal of our study is to forecast the daily high and low of a security. To achieve this we select HCNN. This advanced neural network models several time series at once and facilitates multi-step forecasts. The following state equation computes new states in the network:

$$\mathbf{s}_{t+1} = \tanh(W \cdot \mathbf{s}_t) \tag{1}$$

W is the weight matrix which we optimize. As usual in the context of neural networks we use the abbreviated notation $\tanh(\ldots)$ to mean a component-wise computation of tanh on the resulting vector $W \cdot \mathbf{s}_t$. The upper part of $\mathbf{s}$ contains our observable time-series. These are the inputs and outputs although we do not make this distinction here, because we consider every time-series equal. The lower part of $\mathbf{s}$ contains hidden states. We can think of these hidden states as representing the unknown dynamics of the system under investigation. All raw time series are preprocessed to simple-returns:

- returns of the lows
- returns of the highs
- five-period exponential moving average of the lows
- five-period exponential moving average of the highs
- five-period lower bollinger band of the closes
- five-period upper bollinger band of the closes
- returns of the the open
- same-day return open to low (to be forecast)
- same-day return open to high (to be forecast).

We have to keep in mind that the first six time-series are only available with a lag of one day when we consider the open of the current day our starting point. Our out-of-sample dataset ranges from 2010-12-09 until 2012-03-02 (310 trading days). As is common in HCNN application we use a rolling window forecast. In this work we stick to the parameters of the original study and use 128 trading days of training data to forecast the next ten days. We then move the training data forward by ten days, train and forecast again. The first training interval for the final model therefore begins 128 trading days before 2010-12-09. We also used the trading year before 2010-12-09 as combined training and validation set for preliminary experiments.

3 Decision Support System for Trading

Each trading day, at the open, our system delivers a forecast for low and high of the day. We want to exploit this information in a viable trading system using only OHLC historical data for reasons of practicability. This imposes the limitation that we cannot tell in which order high and low will occur. We also do not know whether prices near high and low will be hit only once or several times. In other words: we can check if the range forecast is correct, but we cannot see, if we would have been able to exploit several swings within that range. On the other hand the times of open and close are well defined, obviously.

In the following we explain a possible trading system using only OHLC values. Let o, h, l, c represent the realized open, high, low, close values, and l_f, h_f the corresponding forecasts for l and h.

First, every forecast should follow the following consistency rules:

- only trade if $l_f < o < h_f$ (forecast of low and high should be consistent with the open price)
- no trade is possible when $h < l_f$ or $h_f < l$ (forecasts are entirely outside the range)

Then we can trade into the direction of the more promising extreme (low or high) at the open. When the price hits the presumed extreme we reverse the position and exit at the close. We also exit at the close if we do not reach the forecast extreme. For this scheme we first have to determine the forecast returns from the open:

$$l_r = \left| \frac{l_f}{o} - 1 \right| \tag{2}$$

and

$$h_r = \left| \frac{h_f}{o} - 1 \right|. \tag{3}$$

We define the more promising direction as the greater of l_r and h_r. For the purpose of illustration let's say that h_r is more promising. We would therefore go long at the open and eventually reverse our position if we reach h_r. For the case $h_r > l_r$ the following rules apply:

- if $h_f < h$ then book $h_f - o$ and $h_f - c$ to the PnL (long at the open, reverse at h_f, buy to cover at c)
- if $h < h_f$ then book $c - o$ to the PnL (long at the open, h_f not hit, exit at c)

The reverse rules apply for $l_r > h_r$. We would short at the open, eventually reverse at l_r and exit at c in any case.

Like in the original study we enhance the system by introducing a parameter α to encourage or discourage trading. If we narrow the forecast range we encourage trading, because we make it more likely that price will hit our range. If we make the forecast range wider we discourage trading, because we make it more unlikely

that price will hit our range. Assuming that the forecasts are consistent ($l_f < h_f$) we denote the (positive) range by $r = h_f - l_f$. We can then modify our forecasts according to

$$\tilde{l}_f = l_f + \alpha \cdot r \tag{4}$$

and

$$\tilde{h}_f = h_f - \alpha \cdot r. \tag{5}$$

$\alpha > 0$ corresponds to narrowing the range and makes trading more likely. We set α heuristically by preliminary experiments on the in-sample data set. We do not perform a strict optimization for α.

4 Economical Results

The core of our trading system is trading the high-low intraday range. For this to work, we need some amount of intraday volatility. We expect that volatile stocks will yield better results than comparatively "duller" stocks. At the same time the stocks should be liquid in the sense that our transaction does not affect the price significantly. For the purpose of our analysis we will consider S&P500 stocks as liquid. We also expect that we will have to encourage trading ($\alpha > 0$) on low volatility stocks and discourage trading on high volatility stocks. Table 1 shows the stocks in our sample.

We include these stocks because they represent different types of behavior: XOM is a big volatile stock, while MSFT is also big but not very volatile. DMND, FSLR, and SHLD on the other hand are among the most volatile stocks in the in-sample period.

Our system trades quite often: at least twice a day (open to close) or three times a day (open to extreme, extreme to close). We could expect that transaction costs severely impact our returns. As a conservative estimation we take USD 0.005 per share traded. This is a generally available flat rate at Interactive Brokers. As we do not use market orders we do not consider slippage for the remainder of the analysis.

Table 2 shows the out-of-sample results. The first row shows the number of trades. With decreasing α the number of trades also tends to decrease, because the system reaches the forecast extreme less often. Decreasing α therefore lets the system act more like a trend rider than like a swing trader. The second row shows the realized

Table 1 Analyzed stock tickers

Ticker	Company name	Exchange
XOM	Exxon mobil corporation	NYSE
MSFT	Microsoft corporation	NASDAQ
DMND	Diamond foods, inc.	NASDAQ
FSLR	First solar, inc.	NASDAQ
SHLD	Sears holdings corporation	NASDAQ

Table 2 Out-of-sample results

	XOM	MSFT	DMND	FSLR	SHLD
trades	815	790	748	647	672
return	0.644	0.139	0.912	0.934	2.359
ann. ret.	0.498	0.111	0.694	0.709	1.678
max dd	0.119	0.138	0.514	0.331	1.46
PnL / max DD	5.436	1.005	1.774	2.817	1.614
ann. ret. after tc	0.455	−0.005	0.623	0.692	1.644
α	0.2	0.2	0.0	−1.0	−0.5

return over the entire trading period of 310 days. For example, XOM realized a return of 64.4 %. The third row describes the annualized return assuming 252 trading days per year. The next row, max dd, refers to the maximum (relative) encountered drawdown and the row after that is an important risk measure. It can be interpreted as a reward to risk ratio. A higher ratio is better because it means that for the same amount of risk (as measured by drawdown) a higher reward is reached. Finally we compute the annualized return after transaction costs. The last column shows the value of α. As expected the more volatile stocks needed to be traded less often but on larger ranges.

We see from the result table that the volatile stocks generally perform well, with good annualized returns and also mostly attractive reward to risk ratios. On the other hand, MSFT included as a large cap stock with low volatility does not perform well. A closer look at the trade history of DMND, FSLR, and SHLD reveals that the loss producing trades are mostly clustered at the beginning of the out-of-sample period. It is obvious in hindsight that a simple no-trade filter that prevents trading when a certain monthly loss has been surpassed would have greatly reduced the seriousness of the drawdowns.

5 Conclusion and Outlook

Our analysis shows that it is possible to successfully model the intraday dynamics of liquid US securities with only a few (technical) indicators as inputs to a neural network. The trading strategy always goes flat at the close of the day. This eliminates the overnight gap risk present in all daily strategies. This paper confirms the good results that [2] obtained for the Brazilian stock market.

This is also a limitation of our work. It is quite probable that we could enhance the risk-reward ratio of the strategy with tick data. For example, we could imagine trading the range more than once a day if the occasion presents itself. With OHLC data we we don't know the order of possibly recurring daily highs and lows.

Another limitation of our work is the lack of diversified input factors. One could argue that we use four different price series (open, high, low, close). But these price series are strongly correlated. It would be interesting to see if the results improve if

we add different time series, for example from securities in the same industry group. Also general market indices could prove a useful addition.

Investigating other markets could prove enlightening. How does the same methodology fare when forecasting foreign exchange, commodities, or other assets?

Our study only analyzes a specific kind of neural network, the Historically Consistent Neural Network. Other types of neural networks could be investigated as well. On a related note the choice of meta-parameters for learning and network design are heuristics. Although preliminary experiments show that our results are robust with respect to the choice of meta-perameters a more detailed and systematic analysis seems justified.

References

1. Dunis, C., Laws, J., Rudy, J.: Profitable mean reversion after large price drops: A story of day and night in the s&p 500, 400 midcap and 600 smallcap indices. J. Asset Manage. **12**(3), 185–202 (2011)
2. Martinez, L.C., da Hora, D.N., de M. Palotti, J.R., Jr., W.M., Pappa, G.L.: From an artificial neural network to a stock market day-trading system: A case study on the bmf bovespa. In: Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, 14–19 June 2009
3. Gomide, P., Milidiu, R.: Assessing stock market time series predictors quality through a pairs trading system. In: Neural Networks (SBRN), 2010 Eleventh Brazilian Symposium on (Oct 2010) pp. 133–139 (2010)
4. Zimmermann, H.G.: Forecasting the Dow Jones with historical consistent neural networks. In: Dunis, C., Dempster, M., Terraza, V. (eds.) Proceedings of the 16th international conference on forecasting financial markets, Luxembourg, 27–29 May 2009
5. Zimmermann, H.G.: Advanced forecasting with neural networks. In: Dunis, C., Dempster, M., Breitner, M.H., Rösch, D., von Mettenheim, H.J. (eds.) Proceedings of the 17th international conference on forecasting financial markets, Hannover, 26–28 May 2010
6. von Mettenheim, H.J., Breitner, M.H.: Neural network model building: A practical approach. In: Dunis, C., Dempster, M., Girardin, E., Péguin-Feissolle, A. (eds.) Proceedings of the 18th international conference on forecasting financial markets, Marseille, 25–27 May 2011

Part XI
Managerial Accounting

Modeling Aggregation and Measurement Errors in ABC Systems

Jan-Gerrit Heidgen, Stephan Lengsfeld and Arndt Rüdlin

1 Introduction

The seminal papers by Labro/Vanhoucke (see [3] and [4]) made strong contributions to the analysis of errors in ABC systems by using simulation-based methods. We pick up their two-stage approach in modeling ABC systems as well as the implementation of measurement and aggregation errors and discuss the importance of sequence by which these errors are implemented on the magnitude of the overall cost distortion. As an extension to Labro/Vanhoucke we analyze different types of aggregation errors and discuss aspects of simultaneous implementation of aggregation and measurement errors. The existing literature identifies several cases with compensatory interaction effects. Addressing these effects our results show that some of theses effects are not the outcome of interaction among errors but caused by technical assumptions and restrictions. Furthermore, we give an overview of different ways of modeling errors and analyze the interaction between aggregation and measurement errors and their impact on cost distortion.

2 Product Costs, Aggregation and Measurement Errors in ABC Systems

After a short description of cost allocation in ABC systems we characterize aggregation and measurement errors and their implementation in our simulation study:

J.-G. Heidgen (✉) · S. Lengsfeld · A. Rüdlin
Finance and Accounting, University of Freiburg, Bertoldstr. 17, 79085Freiburg, Germany
e-mail: heidgen@vwl.uni-freiburg.de

S. Lengsfeld
e-mail: lengsfeld@vwl.uni-freiburg.de

A. Rüdlin
e-mail: arndt.ruedlin@vwl.uni-freiburg.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_38,

Calculation of product costs in ABC systems. In our model the usual two-stage cost allocation in ABC systems is implemented[1]: First, after assigning costs to resource cost pools, at stage I costs are allocated to activity cost pools (ACP) based on resource usage by activities. At stage II the demand for activities by a cost object (e.g. a product) is the basis for allocating costs of activity cost pools to the cost objects.[2]
Implementation of aggregation and measurement errors in simulation-based studies. In this framework aggregation and measurement errors can arise on each stage.[3] These errors distort cost information of activity cost pools or cost objects respectively, even though the sum of total costs on each stage is equal to the sum of the total resource costs. Aggregation errors can emerge if costs of aggregated pools do not vary proportional to each other. In our paper we also analyze different types of aggregation errors which differ in the way they treat outgoing cost driver links of the aggregated pools.[4] A measurement error can arise in two different ways: First, a resource cost pool may be over- and another one undercharged. Second, a measurement error can occur if incorrect estimates are used as a basis for cost allocation.
Features of the simulation model. In line with Labro/Vanhoucke we model an ABC system with 50 resource cost pools, 40 activity cost pools and 50 cost objects (later in our analysis we run simulations with an expanded model with 125, 100 and 125 pools). For generating an ABC system the total amount of costs (1,000,000 in the regular, 2,500,000 in the expanded model) is randomly assigned to the resource cost pools. Cost drivers between resource cost pools and activity cost pools as well as drivers between activity cost pools and cost objects are randomly generated. In this manner a benchmark system is generated and product costs for the cost objects are calculated representing the true cost information. In total we model 500 different benchmark systems and for each of them 100 noisy systems (in total 50,000 observations).

These noisy systems are determined by combining an aggregation and a measurement error. Each error varies in intensity: 0.0, 0.1, 0.2, ... 0.9. This represents the percentage of cost pools that are subject to the error. For each noisy system product costs are calculated and the bias between the noisy system and the corresponding benchmark system is determined. As a measure for cost system accuracy the euclidean distance (EUCD) over all cost objects between noisy and true system is calculated.[5]

---

[1] For a more detailed description and discussion of cost allocation in ABC see e.g. [1], pp. 7–13.

[2] In line with Labro/Vanhoucke we focus on overhead costs and assume that direct costs are measured without errors (cf. [3], p. 941).

[3] In the following *AE-ACP* denotes an aggregation error at activity cost pools and *ME-AD* denotes a measurement error at activity drivers.

[4] For further details see [2].

[5] Cf. [3], p. 943 and the cited literature for using the euclidean distance as a measure of overall error in the cost system.

3 Results

On the one hand our analysis confirms the main results of Labro/Vanhoucke. On the other hand we get the following additional results:

Explanation of assumed compensatory interaction effects. For combinations of aggregation and measurement errors occurring on the same stage of the cost system, we find (using our basic design parameters) initially alleged compensatory effects for high intensity of aggregation errors (see Fig. 1, Panel A and [3], Panel B, p. 949, respectively). However, these effects can be explained by technical restrictions or rather assumptions of the simulation. Precisely, they can be traced back to rounding the number of affected cost pools (which is necessary, since they can only take integer values). Furthermore, we show a way to avoid these effects by using sufficiently large cost systems. Both is explained briefly in the following:

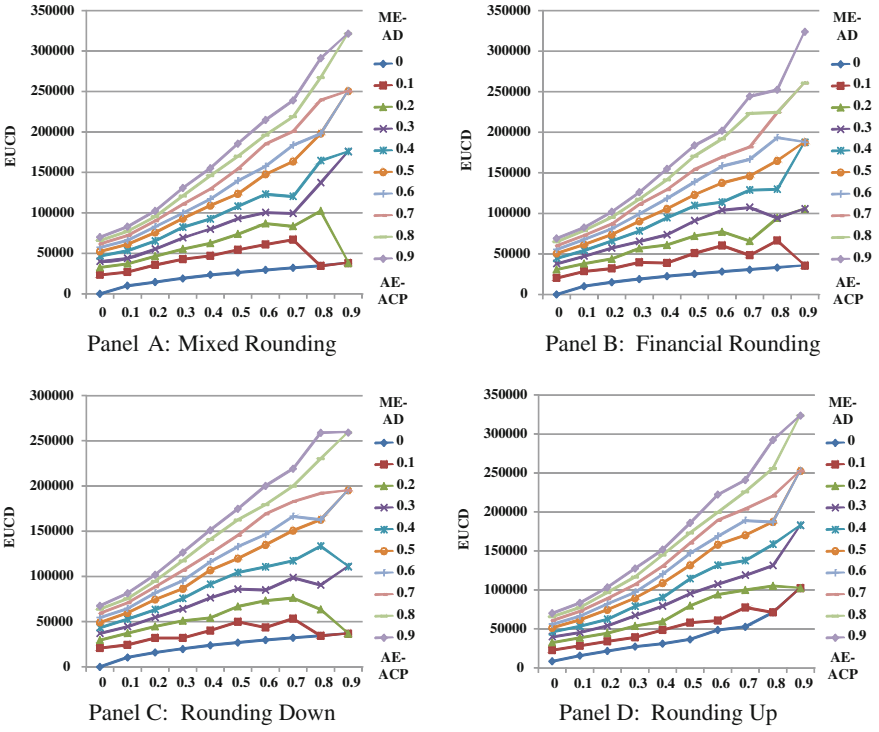
For example, if you consider the highest possible aggregation error (90 %) on activity cost pools and take into account that there are only 40 activity cost pools, then (1–0.9)*40 = 4 activity cost pools remain after aggregation. Only these cost pools are possible candidates for a measurement error. A 20 %–measurement error leads to 0.8 affected cost pools. Since the number of cost pools can only take integer values, these numbers have to be rounded up or down. Depending on the type of rounding, different effects between the two errors seem to occur.[6] Figure 1 shows the results for four different types of rounding: Panel A shows a case of "mixed rounding" where numbers less than 1 are rounded down and numbers larger than 1 are rounded up, i.e., 0.9 becomes 0 and 1.2 becomes 2. Financial rounding is shown in Panel B, strictly rounding down and strictly rounding up are shown in Panel C and Panel D, respectively.

Avoiding effects caused by rounding. The problems arising with alleged interaction effects described before can be avoided by making sure, that rounding is not necessary. In our analysis we model an adequate expanded cost system by multiplying the parameters of the basic setting with 2.5.[7] Doing so, it is obvious that reinforcing effects of aggregation and measurement errors on each other are not only observable for small up to middle intensities of the aggregation error but also continues for high intensities of the aggregation error (see Fig. 2, Panel A).

Relevance of the sequence of aggregation and measurement errors. One focus of our simulation is to analyze impacts of the sequence of the error implementation on the interaction between two errors in a cost system. The economic relevance to this question is the procedure of implementing and adjusting the cost system: Does a company first determine the organizational structure (i.e., cost pools) or does it

[6] When strictly rounding down, 0 cost pools are subject to measurement error and hence a measurement error up to 20 % seems to have no impact at all. When strictly rounding up, measurement errors of 10 and 20 % have the same impact.

[7] In our setting the highest aggregation error has to leave at least 10 cost pools, so a variation of the measurement error from 0.0 to 0.9 with steps of 0.1 strictly increases the number of cost pools subject to measurement errors.



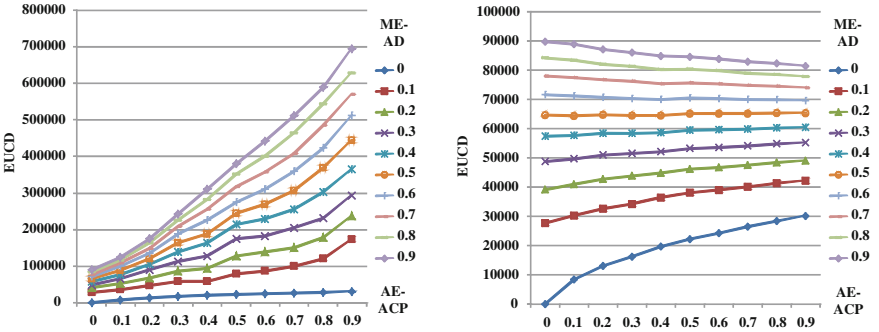
AE-ACP: Aggregation Error at Activitiy Cost Pools
ME-AD: Measurement Error at Activity Drivers
EUCD: Euclidean Distance

Fig. 1 Different types of rounding with primarily implemented aggregation error

change the organizational structure after deciding on the quality of the measurement process (e.g. after deciding on employees' skills)?

Our simulation identifies cases where the sequence has no impact on the interaction between errors but also cases where the sequence leads to different and even contrary effects. Figure 2, Panel A shows mutual enforcing effects on cost distortion, when the aggregation error is primarily implemented. A primarily implemented measurement error leads to compensating effects, i.e., the impact of the error combination on the overall error is negative and decreasing in error intensity of the measurement error (see Fig. 2, Panel B). Thus, an increasing aggregation error may even increase overall accuracy when the measurement error is sufficiently high (60 % and higher).

Different types of aggregation errors. So far, modeling aggregation assumes to keep all cost driver links of the aggregated pools. Additionally, we discuss in our paper a second type of aggregation error (type 2). In order to reduce the complexity and henceforth the maintaining expenses of the cost system, only selected cost drivers

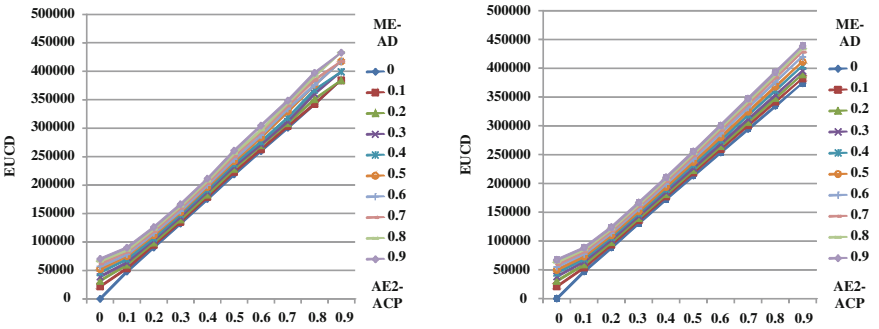


Panel A: AE-ACP primarily implemented

Panel B: ME-AD primarily implemented

AE-ACP: Aggregation Error at Aggregation Cost Pools
ME-AD: Measurement Error at Activity Drivers
EUCD: Euclidean Distance

Fig. 2 Aggregation and measurement error in the extended cost systems with two different implementation sequences



Panel A: AE2-ACP primarily implemented

Panel B: ME-AD primarily implemented

AE2-ACP: Aggregation Error (type 2) at Activitiy Cost Pools
ME-AD: Measurement Error at Activity Drivers
EUCD: Euclidean Distance

Fig. 3 Aggregation error with selected cost drivers (type 2) and measurement error with two different implementation sequences

are used to allocate total costs of the aggregated pools to the following stage. As Fig. 3 shows dominance of aggregation errors occurs and is independent of the sequence of the errors, i.e., the overall error is mainly driven by the intensity of the aggregation error.

4 Conclusion

We analyze the impact of interacting aggregation and measurement errors on reported product costs in ABC systems and present first results. First, we provide an explanation for alleged compensating effects which are described in previous simulation studies. Second, our simulation studies show that the sequence of occurrence of errors is crucial for both: the type of interaction effects and the quality of the cost information system (measured by the cost distortion of reported product costs). Third, we discuss different types of aggregation errors. Our simulation shows that measurement errors can be dominated by aggregation errors. This is the case, if only selected cost drivers are chosen for the allocation of costs.

References

1. Balakrishnan, R., Labro, E., Sivaramakrishnan, K.: Product costs as decision aids: an analysis of alternative approaches (part 1 + 2). Acc. Horiz. **26**(1), 1–41 (2012)
2. Heidgen, J.-G., Lengsfeld, S., Rüdlin, A.: Aggregations- und Messfehler in ABC-Systemen - Eine simulationsgestützte Analyse. University of Freiburg, Workingpaper (2012)
3. Labro, E., Vanhoucke, M.: A simulation analysis of interactions among errors in costing systems. Account. Rev. **82**(4), 939–962 (2007)
4. Labro, E., Vanhoucke, M.: Diversity in resource consumption patterns and robustness of costing systems to errors. Manage. Sci. **54**(10), 1715–1730 (2008)

Part XII
Production and Operations Management

Condition-Based Release of Maintenance Jobs in a Decentralized Multi-Stage Production/Maintenance System

Ralf Gössinger and Michael Kaluzny

1 Problem

Condition-based maintenance is analyzed for multi-stage production with a separate maintenance unit. Deterioration depends on multiple production parameters. After reaching a critical deterioration level a machine failure occurs and reactive maintenance is required. Preventive maintenance reduces the failure probability. In both cases the machine status is set up to "as good as new" (AGAN). The task of maintenance is to keep up agreed availabilities α_i of m production stages ($i = 1, \ldots, m$), defined as the probability of no machine failure between two consecutive AGAN states. In this context the problem of determining optimal triggering conditions α_i' to release maintenance jobs at the right situation arises. Since maintenance jobs compete for limited maintenance capacity, scheduling is required. Due to division of labour, some problem specifics are relevant: (a) Information asymmetry between production stages and maintenance unit, (b) delays W_i between releasing and carrying out maintenance jobs, (c) production during W_i with modifiable production parameters, and (d) influences of scheduling and triggering onto production performance. Previous approaches of decentralized production/maintenance-coordination[1] assume instantaneous information exchange and immediate availability of maintenance resources. Since this is not always fulfilled in decentralized systems, we present a problem-specific continuous condition monitoring (*CM*). Thereby, we operationalize the components of waiting time, identify decision relevant effects

[1] Hierarchical: cf. [1, p. 155]; [2, p. 173]; heterarchical: cf. [3, p. 1584]; [4, p. 4611]; [5, p. 168]; [6, p. 77].

R. Gössinger · M. Kaluzny (✉)
Dept of Business Administration, Production Management and Logistics, University of Dortmund, Martin-Schmeißer-Weg 12, 44227 Dortmund, Germany
e-mail: ralf.goessinger@udo.edu

M. Kaluzny
e-mail: michael.kaluzny@tu-dortmund.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_39,

of setting α_i', develop a maintenance priority rule and a stochastic approach to determine α_i' based on chance constraints and analyze it by simulations.

2 Solution Approach

The coordination approach has to consider the interaction among information flows and goods flows. One suitable way is the application of the Transfer Authorization Card (TAC) approach[2]. It is a PAC system[3] with maintenance-specific extensions. Figure 1 gives an overview on a TAC system with two production stages P_i, P_k and one maintenance unit M. The production stages are directly interconnected via their output stores and the queues $QP_{i,2}$, $QP_{k,2}$ for production orders and maintenance jobs. The interconnection between production stages and maintenance unit is considered in different ways: (1) Triggers TR_i, TR_k send out preventive maintenance orders to the maintenance order queue QM_2 if the triggering condition emerges. (2) Maintenance preparedness is signalized by pushing a message to the order queue $QP_{i,2}$, $QP_{k,2}$ of the corresponding production stage. (3) Requests of service submitted from a production stage to the maintenance requisition queue QM_3 in case of a machine failure (reactive maintenance) or machine preparedness for a preventive maintenance job. (4) The transfer of maintenance resources from the maintenance department to the production stages and vice versa.

In the process of preventive maintenance order fulfilment, the queues QM_2, QP_2 and QM_3 are accompanied by stochastic delays W_{OT_M}, W_{AT_M} and W_{RT_M}. Further delays are induced by the transfer of maintenance resources W_{Tr_M} and the residual time for finishing the last production job W_{ST_M} before the execution of the maintenance job can start. For reactive maintenance only the delays W_{RT_M} and W_{Tr_M} are relevant. From a production order's perspective the delays W_{OT_P} at the order queue QP_2, the processing times O at the production stages and the transfer times W_{Tr_P} from an output store to the next production stage have to be considered. To keep up α_i, TR_i has to release preventive maintenance jobs in such a way that the probability of reaching the critical deterioration level during W_i is at most $1 - \alpha_i$. From the maintenance point of view, an objective can be expressed with the ratio π of preventive maintenance jobs per production stage and the maximum norm $\|\cdot\|_\infty$:

$$min \|e(\alpha')\|_\infty := max\left(\frac{\alpha_i - \pi(\alpha_i')}{1 - \alpha_i}\right)$$

CM allows for considering influencing factors $v_i^p \in \mathbb{R}^n$ on deterioration by assuming of v_i^p-conditioned failure time distributions: The family $(F_{T_i|v_i^p})_{v_i^p}$ forms a hypersurface and the availability constraints an Iso-hypersurface to the value $1 - \alpha_i$. The failure probability developement $F_{T_i}(t, v_i^p(t))$ as a conditioned deterioration process

[2] Cf. [7, p. 191].

[3] Cf. [8, p. 34].

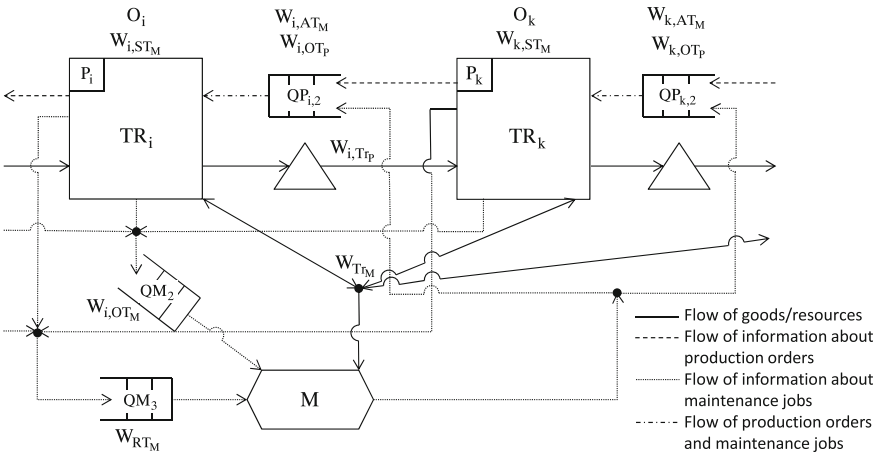


Fig. 1 Aggregated structure of the TAC system

is an utilization path on $(F_{T_i|v_i^p})_{v_i^p}$. To determine α_i', on this basis TR_i has to calculate two machine conditions at time t: (a) failure probability $\rho_i(t) := F_{T_i}(t, v_i^p(t))$ on the basis of the chosen production parameters after an AGAN status and (b) failure probability $\rho_i(t+W) := F_{T_i|\hat{v}_i^p}(\tau + \hat{W})$ that would be achieved after an estimated W_i with estimated v_i^p, if a maintenance job would be triggered instantly. Thereby, $\tau_i := F^{-1}_{T_i|\hat{v}_i^p}(\rho_i(t))$ is the time equivalent of reaching $\rho_i(t)$ with constant $\hat{v}_i^p$. To meet α_i under the assumption of $\hat{W}_i$ and $\hat{v}_i^p$, α_i' has to be defined as follows:

$$\alpha_i'(\hat{W}_i, \hat{v}_i^p) = F_{T_i|\hat{v}_i^p}(F^{-1}_{T_i|\hat{v}_i^p}(1-\alpha_i) - \hat{W}_i)$$

Numerical analyses have confirmed the suitability of the following chance constrained estimations:

$$\hat{W}_i := F^{-1}_{W_i}(p_{W_i})$$

$$\hat{v}_i^p \in \{v* \in \mathbb{R}^n | P(F_{T_i}(t, v_i^p(t)) \leq F_{T_i}(t, v*(t))) = p_{v_i^p} \forall t \in [0; \infty]\}$$

To configure TR_i, the probabilities of overestimation p_{W_i} and $p_{v_i^p}$ can be determined by a search algorithm that successively adjusts the parameters in the ordering of their influence on deterioration.[4] Because M serves all P_i, a scheduling problem at *QM3* arises, which can be solved by a maintenance-specific priority rule *PR*. Assuming a fixed inventory policy of production, the objective of the overall system is to minimize the mean lead time $\bar{D}$ of production jobs. On this basis the decision about the triggering conditions α_i' and the priority rule *PR* can be modelled as follows:

[4] Cf. [7], pp. 191, for a detailed description.

$$min \bar{D}, \text{ s.t.}$$

$$D(PR, \alpha', PAC) \tag{1}$$

$$\|e(\alpha')\|_{\infty} \leq \xi \tag{2}$$

$$\alpha_i'(\hat{W}_i, \hat{v}_i^p) = F_{T_i|\hat{v}_i^p}(F^{-1}_{T_i|\hat{v}_i^p}(1-\alpha_i) - \hat{W}_i)\ \forall i \tag{3}$$

$$W_i(PR) = W_{i,OT_M} + W_{i,AT_M} + W_{i,RT_M} + W_{i,Tr_M} + W_{i,ST_M}\ \forall i \tag{4}$$

$$\forall i, k = 1, \ldots, m; i \neq k : W_i, W_k \text{ stochastically dependent} \tag{5}$$

$$\hat{W}_i := F^{-1}_{W_i}(p_{W_i})\ \forall i \tag{6}$$

$$\hat{v}_i^p \in \{v* \in \mathbb{R}^n | P(F_{T_i}(t, v_i^p(t)) \leq F_{T_i}(t, v*(t))) = p_{v_i^p} \forall t \in [0; \infty]\}\ \forall i \tag{7}$$

$$p_W, p_{v_i^p} \in [0; 1]\ \forall i \tag{8}$$

$$0 \leq \alpha_i' < 1 - \alpha_i \leq 1\ \forall i \tag{9}$$

$$W_{i,l} \geq 0\ \forall i, k, l \tag{10}$$

(1) takes into account that lead times depend on PR, α_i' and other PAC-parameters PAC.[5] (2) and (3) consider the maintenance objective. (4) captures the delay components. (5) reflects the fact that waiting times at different production stages are interdependent due to limited maintenance resources. (6) and (7) are the chance constrained estimations. (8), (9) and (10) specify the domains. In particular (1) and (5) point out the problem complexity and the need for a heuristic solution approach supported by simulations.

For scheduling the maintenance requests waiting at QM_3, priority rules can be applied, which express the following preferences: Reactive maintenance first (*RM*) in case of equal α_i. Highest availability first (*HA*) in case of same-natured jobs (either preventive or reactive). Longest elapsed waiting time first (*LOW*) in case of same-natured jobs and equal α_i. In case of different-natured jobs and unequal α_i the differences in nature χ_i, elapsed waiting time t_i and α_i of the jobs have to be regarded simutaneously (*LOWHARM*). With $c > 1$, as the factor that a preventive maintenance job has to wait longer than a reactive to be prioritized in case of equal α_i, all these preferences can be considered by advising a prioritization with

$$LOWHARM(t_i, \alpha_i, \chi_i, c) := ((c-1) \cdot \chi + 1) \cdot \frac{(1+\alpha_i) \cdot t_i}{\sum_{i=1}^{n} \alpha_i};\ c > 1;\ \chi_i = \begin{cases} 1, & \text{if reac.} \\ 0, & \text{if prev.} \end{cases}$$

[5] Cf. [8], pp. 34.

Scenarios / Priority rules	β=1				β=1,5				β=2				β=5			
	ξ=0.4	ξ=0.3	ξ=0.25	ξ=0.2	ξ=0.4	ξ=0.3	ξ=0.25	ξ=0.2	ξ=0.4	ξ=0.3	ξ=0.25	ξ=0.2	ξ=0.4	ξ=0.3	ξ=0.25	ξ=0.2
FCFS	4984	4987	4983	Infbl.	Infbl.	Infbl.	Infbl.	Infbl.	4996	Infbl.	Infbl.	Infbl.	5072	5095	5095	5095
FCFS/HA	5005	5001	5006	5001	5001	4997	4998	Infbl.	5022	Infbl.	Infbl.	Infbl.	5065	Infbl.	Infbl.	Infbl.
LOWHARM	4985	4975	4975	4982	4984	4976	Infbl.	Infbl.	4985	4995	4996	Infbl.	5054	Infbl.	Infbl.	Infbl.

Fig. 2 Mean lead times per scenario and applied priority rule

In case of $c = 1$ the differentiation between reactive and preventive jobs has no influence on the schedule. Therefore, this rule is called *FCFS/HA*.

3 Numerical Analysis

The intention of the numerical analysis is to compare the performance of *LOWHARM*, *FCFS/HA* and *FCFS* in *QM3* by simulating a 5-stage Kanban-controlled production line. System characteristics are:

- Production: Demand for final goods is modelled as a Poisson process, maintenance-first/production-FCFS-rule in *QP1* and one sequentially working machine per stage with v_i^p-conditioned Weibull-distributed failure times $T_i|v_i^p$ (shape parameters $k_i = 3.6$, scale parameters $\lambda_i = 47,085 \cdot (v_i^p)$), whereby $v_i^p \in \{0,25; 0,5; \ldots; 2,5\}$ is a accelleration/deceleration factor chosen as a function of the *QP1*-length L_i (balancing).
- Maintenance: Weibull-distributed durations of maintenance jobs (shape parameter $k = 3.6$) with longer durations ($\mu_p \leq \mu_r$) and higher volatility ($\sigma_p \leq \sigma_r$) of reactive actions, deterministic transfer times and *HA* in QM_2.

We vary the relation β between the mean duration of reactive and preventive maintenance, the upper bound ξ of the maintenance objective and the *LOWHARM*-parameter $c \in [1.05; 5]$ systematically. Per scenario 30 simulations are carried out.[6] The results (Fig. 2) show that the proposed coordination approach is suitable for decentralized production/maintenance-systems. Its performance depends on the applied priority rule, the ratio between the durations of reactive and preventive maintenance actions as well as on the upper bound for the maintenance objective. There is no positive correlation between the production and maintenance objective values, i.e. dropping the lower bound ξ by tendency induces higher mean lead times. For lower upper bounds ξ no feasible solutions (Infbl.) were found by applying any priority rule. Comparing the tested priority rules for all β, *LOWHARM* performs best regarding the mean lead time. But for $\beta = 1,5$ and $\beta = 5$, *FCFS/HA* and *FCFS* respectively allow for lower upper bounds ξ. Depending on c there were no differences in the results of *LOWHARM*.

[6] The simulation length is chosen according to a failure probability $\varepsilon < 5\%$ and a tolerance range $\delta < 1\,\%$ for the measured objectives.

4 Concluding Remarks

Besides defining triggering conditions for preventive maintenance, in multi-stage production systems *CM* induces a maintenance scheduling problem. Starting point of the solution approach was a TAC system to ensure the required information exchange among the production stages and the maintenance department. A chance constraint model that can be solved by an iterative search considers the triggering subproblem. For the scheduling subproblem, we developed a maintenance-specific priority rule that takes different preferences of maintenance and production into account. The numerical analysis shows that the coordination approach performs well. Its performance depends on the applied priority rule whereby *LOWHARM* is a promising candidate. Since *LOWHARM* simultaneously applies different relevant criteria and hierarchically structured priority rules perform well in the area of production scheduling, further research is directed to to develope such rules and to compare them to *LOWHARM* by a multi-criteria analysis. Thereby, the optimal paramterization of *LOWHARM* has to be analyzed in more detail.

References

1. Cho, D.I., Abad, P.L., Parlar, M.: Optimal production and maintenance decisions when a system experience (sic!) age-dependent deterioration. Optimal Control Appl. Methods **14**, 153–167 (1993)
2. Nodem, Dehayem: F.I., Kenne, J.P., Gharbi, A.: Hierarchical decision making in production and repair/replacement planning with imperfect repairs under uncertainty. Eur. J. Oper. Res. **198**, 173–189 (2009)
3. Berrichi, A., et al.: Bi-Objective ant colony optimization approach to optimize production and maintenance scheduling. Comput. Oper. Res. **37**, 1584–1596 (2010)
4. Coudert, T., Grabot, B., Archiméde, B.: Production/maintenance cooperative scheduling using multi-agents and fuzzy logic. Int. J. Prod. Res. **40**, 4611–4632 (2002)
5. Feichtinger, G.: The Nash solution of a maintenance-production differential game. Eur. J. Oper. Res. **10**, 165–172 (1982)
6. Jørgensen, S.: A pareto-optimal solution of a maintenance-production differential game. Eur. J. Oper. Res. **18**, 76–80 (1984)
7. Gössinger, R., Kaluzny, M.: Condition-based release of maintenance jobs in a decentralised production/maintenance-system: an analysis of alternative stochastic approaches. In: Grubbström, R.W., Hinterhuber, H.H. (eds.) Pre-Prints of the Seventeenth International Working Seminar on Production Economics, vol. 1, pp. 191–201. Innsbruck (2012)
8. Buzacott, J.A., Shanthikumar, J.G.: A general approach for coordinating production in multiple-cell manufacturing systems. Prod. Oper. Manag. **1**, 34–52 (1992)

An Exact Approach for the Combined Cell Layout Problem

Philipp Hungerländer and Miguel F. Anjos

1 Introduction

In a cellular manufacturing system, the parts that are similar in their processing requirements are grouped into part families, and the machines needed to process the parts are grouped into machine cells. Ideally, cells should be designed in such a way that the part families are fully processed in a single machine-cell so that the machine-cells are mutually independent with no inter-cell movement. In a real-world situation however it may be impractical and/or uneconomical to require mutually independent cells. The consequence is that some parts will require processing in more than one cell.

This paper is concerned with finding the optimal layout of each cell in the presence of parts that require processing in more than one cell. Cell layout usually takes place after the machine cells are determined, see e.g. [1]. While much research has been done on the cell formation problem, the layout of machines within the cells has received less attention. We propose an exact solution method that simultaneously minimizes the inter-cell and intra-cell material handling costs for a given machine-cell assignment. To the best of our knowledge, this is the first exact method proposed for this problem. The method is based on the application of semidefinite optimization models and algorithms. The machines of each cell can be arranged in a row or on a circle. We denote this problem as Combined Cell Layout Problem (CCLP). The two types of cells that we consider have been previously studied in the literature but independently rather than combined.

The first type of cell layout is single-row layout.The Single-Row Facility Layout Problem (SRFLP), sometimes called the one-dimensional space allocation

P. Hungerländer (✉)
Institute of Mathematics, Alpen-Adria Universität Klagenfurt, Klagenfurt, Austria
e-mail: philipp.hungerlaender@uni-klu.ac.at

M. F. Anjos
Canada Research Chair in Discrete Nonlinear Optimization in Engineering,
GERAD and École Polytechnique de Montréal, Montreal, QCH3C 3A7, Canada
e-mail: anjos@stanfordalumni.org

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_40,

problem [2], consists of finding a permutation of the machines such that the total weighted sum of the center-to-center distances between all pairs of machines is minimized. This problem arises for example as the problem of ordering stations on a production line where the material flow is handled by an automated guided vehicle (AGV) travelling in both directions on a straight-line path [3]. Other applications are the arrangement of rooms along a corridor in hospitals, supermarkets, or offices [4] and the assignment of airplanes to gates in an airport terminal [5].

The second type is circular layout. The Directed Circular Facility Layout Problem (DCFLP) seeks to arrange the machines on a circle so as to minimize the total weighted sum of the center-to-center distances measured in the clockwise direction. Notice that three facility layout problems that are extensively discussed in the literature, namely the Equidistant Unidirectional Cyclic Layout Problem, the Balanced Unidirectional Cyclic Layout Problem and the Directed Circular Arrangement Problem are special cases of the (DCFLP) Problem (for details see [6]).

2 Problem Description and Matrix-Based Formulation

We assume that only one part is produced at a time and that we want to produce more than one part using the same layout. Hence we want to find the arrangement of the machines that minimizes the overall material flow. Let $\mathscr{P} = \{1, \ldots, P\}$ be the set of part types. Each part type is associated to a different process plan S_p that gives the sequence in which part type p visits the machines. (For simplicity we will assume that the process plan of each part type is unique although in general this may not be the case.) So for each part, the cost function is different and dependent on n_{ij}^p, the number of moves part type p makes from machine i to machine j per period of time. n_{ij}^p is easily determined from S_p. Furthermore n_p denotes the number of parts of type p that are processed per period of time. Hence we assume that each part forms a percentage of the total production, this gives us weights to aggregate the various cost functions into a single cost function that optimizes the layout for all the parts at once. Now the total number of parts of type $p \in \mathscr{P}$ that flow from machine i to machine j per period of time is obtained as $f_{ij}^p = n_p n_{ij}^p, i, j \in \mathscr{M},\ i \neq j,\ p \in \mathscr{P}$, where $\mathscr{M} = \{1, \ldots, M\}$ is the set of machines. The total part flow from machine i to machine j per time period can be expressed as $f_{ij} = \sum_{p \in \mathscr{P}} f_{ij}^p,\ i, j \in \mathscr{M},\ i \neq j$.

Each machine has a given integer length and is pre-assigned by the function $c : \mathscr{M} \to \mathscr{C}$ to one of the cells in the set $\mathscr{C} := \{1, \ldots, C\}$. Furthermore each cell is associated to one of two different layout types: $\ell : \mathscr{C} \to \{$ (SRFLP), (DCFLP)$\}$. The parts enter and exit each cell at a pre-specified machine of that cell, specified by $s : \mathscr{C} \to \mathscr{M}$, where we assume w.l.o.g. that $s(i) < j,\ i \in \mathscr{C}, c(s(i)) = c(j)$, $s(i) \neq j$. Finally we have given an integer distance $e_{ij},\ i, j \in \mathscr{C},\ i < j$ for each pair of cells.

To model the arrangement of the machines in a cell we introduce ordering variables $y_{ij},\ i, j \in [n],\ i < j,\ c(i) = c(j)$,

$$y_{ij} = \begin{cases} 1, & \text{if machine } i \text{ lies left of machine } j, \\ -1 & \text{otherwise,} \end{cases} \tag{1}$$

Any feasible ordering of the machines has to fulfill the 3-cycle inequalities

$$-1 \le y_{ij} + y_{jk} - y_{ik} \le 1, \quad i, j, k \in [n], \quad i < j < k, \; c(i) = c(j) = c(k), \tag{2}$$

It is well-known that the 3-cycle inequalities together with integrality conditions on the ordering variables suffice to describe feasible orderings, see e.g. [7, 8].

For the (SRFLP) the center-to-center distances between machines d_{ij} can be encoded using products of ordering variables [9]:

$$\begin{aligned} d_{ij} = \frac{1}{2}(l_i + l_j) &- \sum_{\substack{k \in [n] \\ k<i, \; c(k)=c(i)}} l_k y_{ki} y_{kj} + \sum_{\substack{k \in [n] \\ i<k<j, \; c(k)=c(i)}} l_k y_{ik} y_{kj} \\ &- \sum_{\substack{k \in [n] \\ k>j, \; c(k)=c(i)}} l_k y_{ik} y_{jk}, \\ i, j \in [n], \; i < j, \; c(i) = c(j), \; \ell(c(i)) &= \text{“(SRFLP)”}. \end{aligned}$$

As one machine is fixed to be “first” in any cell (the one where the parts enter and exit the cell), we set $y_{ij} = 1, \; j \in [n], \; s(c(j)) = i, \; i < j$.

Similar reasonings lead to a linear expression in terms of the ordering variablesfor the center-to-center distances between machines in a circular layout and between machines in different cells. The details are omitted due to space limitations and will be provided in a forthcoming paper.

Hence we can model all distances between machines in the (CCLP) as linear-quadratic expressions in ordering variables. To reformulate the (CCLP) we define an appropriate cost matrix C, collect the ordering variables in a vector y and introduce a matrix $Z := \begin{pmatrix} 1 \\ y \end{pmatrix} \begin{pmatrix} 1 \\ y \end{pmatrix}^{\mathrm{T}}$ containing all products of ordering variables. Letting t denote the total number of ordering variables, we can give a matrix-based formulation of the (CCLP).

Theorem 1 *Minimizing* $\langle C, Z \rangle$ *over* $y \in \{-1, 1\}^t$ *fulfilling* (2) *solves the* (CCLP).

Proof The inequalities (2) together with the integrality conditions on y suffice to induce a feasible layout for both single-row and circular layout and the definition of C ensures that the distances between machines are computed correctly. □

3 Semidefinite Relaxations and Computational Experience

We apply standard techniques to the matrix-based formulation of the (CCLP) proposed in Theorem 1 to construct (SDP) relaxations over the multi-level quadratic ordering polytope

$$\mathscr{P}_{MQO} := \text{conv}\left\{\begin{pmatrix}1\\ y\end{pmatrix}\begin{pmatrix}1\\ y\end{pmatrix}^{\mathrm{T}} : y \in \{-1, 1\}^t \text{ , } y \text{ satisfies (2)}\right\}.$$

Similar relaxations have already been successfully applied to combinatorial optimization problems arising in the area of graph drawing [10–14].

The core of our semidefinite approach is to solve our semidefinite relaxation ($\text{SDP}_{\texttt{full}}$) by using the bundle method in conjunction with interior point methods. The resulting fractional solutions constitute lower bounds. By the use of a rounding strategy, we can exploit such fractional solutions to obtain upper bounds, i.e. integer solutions that describe a feasible layout of the machines. Hence, in the end we have some feasible solution, together with a proof how far this solution could possibly be from the true optimum. We will discuss these two steps in more detail in a forthcoming paper.

We report the results for different computational experiments with our semidefinite relaxation ($\text{SDP}_{\texttt{full}}$). All benchmark instances used can be downloaded together with the best layouts found from http://anjos.mgi.polymtl.ca/flplib. The (SDP) computations were conducted on an Intel Xeon E5160 (Dual-Core) with 2 GB RAM, running Debian 5.0 in 64-bit mode. The algorithm was implemented in Matlab 7.7.

We set the number of parts P to 50. For each part type we generated the process plans S_p as follows: We took a random integer number r_p between 1 and M from a uniform distribution to determine the number of machines to be visited. Then we computed r_p additional integer numbers from the uniform distribution $U(1, M)$ that represented the handling order of type p. Notice that each machine is allowed to occur more than once but not consecutively. Finally n_P is chosen from uniform distributions $U(1, 10)$, $U(1, 50)$ and $U(1, 100)$ respectively to model low, medium and high variations in the part flows (for details see Tansel and Bilen [15]). We marked the instances with the letters L, M and H to identify their type of variation.

We generated 10 instances for each $\mathscr{M} \in \{20, 30, 40, 50, 60\}$ and for each variation type. The integer length of the machines are taken from $U(1, 10)$. We haven given two cells, one has a single-row and the other one a circular layout. The machines are randomly assigned to the cells. We consider here only the case that both cells have the same number of machines but the method can handle any number of machines per cell. We randomly pick the machine at which the parts enter and exit the cell (for the (SRFLP) this machine is fixed to be the first in the ordering). Finally the integer distance between the two cells e_{12} is chosen from $U(20, 40)$.

We summarize our computational results in Table 1. For the instances under consideration, we obtained globally optimal solutions in reasonable time.

Table 1 Results for instances with up to 60 machines that are assigned to 2 cells, each containing 30 machines

Instance	Time	Instance	Time	Instance	Time
CR20_L1	4.7	CR20_M1	4.2	CR20_H1	13.3
CR20_L2	3.7	CR20_M2	4.9	CR20_H2	3.2
CR20_L3	3.7	CR20_M3	6.6	CR20_H3	42.1
CR20_L4	3.6	CR20_M4	3.5	CR20_H4	3.4
CR20_L5	4.0	CR20_M5	27.0	CR20_H5	44.5
CR20_L6	3.4	CR20_M6	4.5	CR20_H6	3.1
CR20_L7	3.6	CR20_M7	9.8	CR20_H7	3.4
CR20_L8	3.3	CR20_M8	2.8	CR20_H8	2.4
CR20_L9	6.9	CR20_M9	1.9	CR20_H9	4.2
CR20_L10	4.3	CR20_M10	3.5	CR20_H10	4.2
CR30_L1	28.3	CR30_M1	1:09	CR30_H1	3:03
CR30_L2	59.8	CR30_M2	7:12	CR30_H2	16.8
CR30_L3	55.0	CR30_M3	2:55	CR30_H3	27.7
CR30_L4	22.3	CR30_M4	17.6	CR30_H4	44.9
CR30_L5	14.8	CR30_M5	26.1	CR30_H5	20.8
CR30_L6	19.2	CR30_M6	27.9	CR30_H6	29.5
CR30_L7	19.1	CR30_M7	54.7	CR30_H7	43.3
CR30_L8	41.0	CR30_M8	1:45	CR30_H8	23.4
CR30_L9	20.8	CR30_M9	24.0	CR30_H9	56.6
CR30_L10	20.4	CR30_M10	1:08	CR30_H10	51.7
CR40_L1	1:37	CR40_M1	3:43	CR40_H1	3:13
CR40_L2	3:01	CR40_M2	10:38	CR40_H2	5:38
CR40_L3	3:03	CR40_M3	4:40	CR40_H3	3:40
CR40_L4	4:39	CR40_M4	7:00	CR40_H4	4:34
CR40_L5	7:29	CR40_M5	1:18:45	CR40_H5	4:58
CR40_L6	13:00	CR40_M6	10:44	CR40_H6	2:56
CR40_L7	3:29	CR40_M7	3:00	CR40_H7	5:09
CR40_L8	2:06	CR40_M8	4:30	CR40_H8	4:22
CR40_L9	8:06	CR40_M9	5:54	CR40_H9	6:29
CR40_L10	5:16	CR40_M10	2:38	CR40_H10	2:42
CR50_L1	19:12	CR50_M1	11:49	CR50_H1	14:00
CR50_L2	9:06	CR50_M2	18:02	CR50_H2	22:53
CR50_L3	16:08	CR50_M3	12:26	CR50_H3	30:53
CR50_L4	14:52	CR50_M4	34:32	CR50_H4	17:28
CR50_L5	13:20	CR50_M5	14:50	CR50_H5	14:14
CR50_L6	27:08	CR50_M6	11:50	CR50_H6	33:06
CR50_L7	11:48	CR50_M7	19:02	CR50_H7	17:11
CR50_L8	31:42	CR50_M8	24:53	CR50_H8	24:09
CR50_L9	12:20	CR50_M9	25:14	CR50_H9	15:52
CR50_L10	27:39	CR50_M10	16:38	CR50_H10	37:56
CR60_L1	2:52:26	CR60_M1	55:56	CR60_H1	1:08:02
CR60_L2	1:20:11	CR60_M2	2:01:31	CR60_H2	1:42:34
CR60_L3	2:28:11	CR60_M3	1:10:07	CR60_H3	1:46:35

(continued)

Table 1 (continued)

Instance	Time	Instance	Time	Instance	Time
CR60_L4	1:08:07	CR60_M4	45:24	CR60_H4	49:55
CR60_L5	1:55:20	CR60_M5	1:04:58	CR60_H5	59:58
CR60_L6	1:31:03	CR60_M6	1:31:18	CR60_H6	1:10:23
CR60_L7	1:19:30	CR60_M7	46:23	CR60_H7	1:12:44
CR60_L8	1:24:40	CR60_M8	1:02:08	CR60_H8	2:38:21
CR60_L9	49:01	CR60_M9	1:14:03	CR60_H9	1:26:53
CR60_L10	1:12:44	CR60_M10	1:42:33	CR60_H10	2:21:55

One cell has a single-row layout and the other one a circular layout. All instances could be solved to global optimality. The running times are given in sec or min:sec or in h:min:sec, respectively

4 Conclusion

We proposed an exact solution method based on semidefinite optimization for optimizing the layout of multiple cells in a cellular manufacturing system in the presence of parts that require processing in more than one cell. Our preliminary computational results suggest that optimal solutions can be obtained for instances with 2 cells and up to 60 machines. While we only considered single-row and directed circular cell layouts, the method can in principle be extended to other layout types and to a larger number of cells.

References

1. Chu, C-H.: Recent advances in mathematical programming for cell formation. In: Planning, Design, and Analysis of Cellular Manufacturing Systems, number 24 in Manufacturing Research and Technology, pp. 3–46. Elsevier Science B.V. (1995)
2. Picard, J.-C., Queyranne, M.: On the one-dimensional space allocation problem. Oper. Res. **29**(2), 371–391 (1981)
3. Heragu, S.S., Kusiak, A.: Machine layout problem in flexible manufacturing systems. Oper. Res. **36**(2), 258–268 (1988)
4. Simmons, D.M.: One-dimensional space allocation: an ordering algorithm. Oper. Res. **17**, 812–826 (1969)
5. Suryanarayanan, J., Golden, B., Wang, Q.: A new heuristic for the linear placement problem. Comput. Oper. Res. **18**(3), 255–262 (1991)
6. Hungerländer, P.: A semidefinite optimization approach to the directed circular facility layout problem. Technical report, submitted (2012)
7. Tucker, A.W.: On directed graphs and integer programs. Technical report,IBM Mathematical Research Project (1960)
8. Younger, D.H.: Minimum feedback arc sets for a directed graph. IEEE Trans. Circuit Theory **10**(2), 238–245 (1963)
9. Anjos, M.F., Kennings, A., Vannelli, A.: A semidefinite optimization approach for the single-row layout problem with unequal dimensions. Discrete Optim. **2**(2), 113–122 (2005)
10. Buchheim, C., Wiegele, A., Zheng, L.: Exact algorithms for the quadratic linear ordering problem. INFORMS J. Comput. **22**(1), 168–177 (2010)

11. M. Chimani and P. Hungerländer. Exact approaches to multi-level vertical orderings. INFORMS Journal on Computing. accepted, preprint available at www.ae.uni-jena.de/Research_Pubs/MLVO.html (2012)
12. Chimani, M., Hungerländer, P.: Multi-level verticality optimization: Concept, strategies, and drawing scheme. J. Graph Algorithms Appl. Accepted, preprint available at www.ae.uni-jena.de/Research_Pubs/MLVO.html (2012)
13. Chimani, M., Hungerländer, P., Jünger, M., Mutzel, P.: An SDP approach to multi-level crossing minimization. In: Proceedings of Algorithm Engineering and Experiments [ALENEX'2011] (2011)
14. Hungerländer, P.: Semidefinite approaches to ordering problems. PhD thesis, Alpen-Adria Universität Klagenfurt (2012)
15. Tansel, B.C., Bilen, C.: Move based heuristics for the unidirectional loop network layout problem. Eur. J. Oper. Res. **108**(1), 36–48 (1998)

A Cutting Stock Problem with Alternatives: The Tubes Rolling Process

Markus Siepermann, Richard Lackes and Torsten Noll

1 Overview

Cutting stock problems are a well known class of optimisation problems [2] [15] since its first mention in 1939 [8]. Even one-dimensional cutting stock problems are already NP-hard [10] and the optimal solution cannot be found in acceptable computational time [9]. There are many known variants, e.g. where orders may have to be handled until a given period of time, the capacity of cutting machines is limited and a single cut needs a certain period of time [2, 5, 7, 12]. Several algorithms lead to very good but not optimal results like the Delayed Column Generation approach (DCG) [3, 4]. Despite the large number of literature that addresses the cutting stock problem it seems that there have not been investigations where customer orders can be fulfilled with alternative cutting lengths.

In this paper, we are introducing the problem of alternative cutting lengths. This problem occurs in the tubes rolling process where tubes are made out of solid steel bars. For this problem, we are developing an adapted optimisation model before the problem solving method can also be adapted. The paper concludes with a short analysis of the benefit for the tubes manufacturer.

M. Siepermann (✉) · R. Lackes
Department of Business Information Management, Technische Universität Dortmund, 44221 Dortmund, Germany
e-mail: Markus.Siepermann@tu-dortmund.de

R. Lackes
e-mail: Richard.Lackes@tu-dortmund.de

T. Noll
Valloirec and Mannesmann Tubes V and M Deutschland GmbH, Plant Mülheim, Schützenstr. 124, 45476 Mülheim an der Ruhr, Germany
e-mail: Torsten.Noll@vmtubes.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_41,

2 The Tubes Rolling Process Problem

Depending on the length and the diameter of ordered tubes the needed lengths of input material can be calculated that have to be cut out of solid steel bars. This input material is heated in a furnace where there is a fixed number of places. Each place can be filled with one bar irrespective of its length. The furnace is rotating permanently. A steel bar is heated within one rotation. After the heating, a mandrel is put through the steel bars which are rolled out to tubes in several rolling steps. Then, the tubes are cooled and the final cutting is done where the tubes are cut into the desired customer lengths. Because the end sections of the tubes cannot be used due to the heating and rolling they are cut off. Due to several degrees of freedom within the rolling process, it is possible to produce the same output with different cutting lengths that can be used alternatively. For example, it can be possible to fulfil a customer order with 10 pieces of input length 12 or with 14 pieces of input length 10. The input parameters for the cutting planning are the numbers of different alternatives that can be used for a customer order which is called rolling lot consisting of several alternative cutting lengths. The longest cuting length is called main variant. It is optimal for the rolling process because each place in the furnace can only be used to heat one bar at the same time irrespective of its length. Furthermore, there is a production waste in the tubes rolling process that mainly depends on the number of bars to be rolled. This waste consists in the end sections of the rolled tubes that cannot be used and have to be cut off.

3 Optimisation Model

In addition to the costs for the input material, we have to minimise the costs for the production waste and the production inefficiency we get when we use alternative cutting lengths. Let J be the number of possible cutting patterns and C_j the costs of cutting pattern j. C_j consists of the costs CWC_j for the waste of pattern j, of the production waste PW_j and of the penalty costs PC_j of pattern j that arise because of production inefficiencies when cutting pattern j is not the main variant. Then, we get the following objective function that minimises the total production costs C with x_j is the number of times cutting pattern j is used to fulfil the customer orders:

$$\text{Min } C = \sum_{j=1}^{J} C_j \cdot x_j \tag{1}$$

$$\text{subject to } C_j = CWC_j + PW_j + PC_j.$$

A rolling lot consists of several variants, each one comprising the length and the number of times that the variant has to be produced. Thus, we have to consider W

different rolling lots each one having up to I_w alternative lengths. Rolling lot w, $w = 1, \ldots, W$, consists of an ordered set of tuples (di_w, li_w), $i = 1, \ldots, I_w$, where li_w is the length of alternative i of rolling lot w and di_w is the number of times length li_w has to be produced if no other alternative length of rolling lot w is used. The first cutting length is the longest one and is called main variant. The others are called side variants. Let a_{iwj} be the number of times that alternative length i of rolling lot w appears in cutting pattern j. Then, concerning the length L of the input material the following restriction must hold:

$$\sum_{w=1}^{W} \sum_{i=1}^{I_w} l_{iw} \cdot a_{iwj} \leq L \; \forall j = 1, \ldots, J \tag{2}$$

$$\text{subject to } a_{iwj} \in \mathbb{N}_0.$$

The cutting waste of pattern j determines its costs CWC_j. These costs are proportional to the length of waste. Let c be the costs of waste proportional to its length. Then, the cutting waste costs CWC_j of cutting pattern j can be computed as follows:

$$CWC_j = \left(L - \sum_{w=1}^{W} \sum_{i=1}^{I_w} l_{iw} \cdot a_{iwj} \right) \cdot c. \tag{3}$$

All orders must be fulfilled and we can use the different alternative lengths simultaneously. The relation between the input and the output of the tubes rolling process is not known when planning the cutting. Thus, we only know the number of input material to fulfil a rolling lot but we do not know the number of output material we get with a certain cutting length. The correct substitution slope between the different alternative lengths is unknown. Therefore, we work with the percentage with which a cutting pattern j fulfils a rolling lot w:

$$\sum_{j=1}^{J} \left(\sum_{i=1}^{I_w} \frac{a_{iwj}}{d_{iw}} \right) \cdot x_j \geq 1 \; \forall w = 1, \ldots, W. \tag{4}$$

In addition to the cutting waste of input material there is also waste PW_j of the production process. The main length is optimal concerning the production process. We cannot reach a better waste. Therefore, we need at least $l_{1w} - d_{1w}$ of input material for the optimal production process. Each alternative length needs $l_{iw} \cdot d_{iw} \geq l_{1w} \cdot d_{1w}$ of input material. Thus, the production waste of alternative i is $l_{iw} \cdot d_{iw} - l_{1w} \cdot d_{1w}$. This waste can now percentally be distributed among the cutting patterns. Again, we can use the costs c that are proportional to the waste length:

$$PW_j = c \cdot \sum_{w=1}^{W} \sum_{i=2}^{I_w} (l_{iw} \cdot d_{iw} - l_{1w} \cdot d_{1w}) \cdot \frac{a_{iwj}}{d_{iw}}. \tag{5}$$

Additionally, when using alternative lengths instead of the first alternative there is also a waste of time: The alternative lengths are shorter than the first length so that we need more pieces of length 2 to I_w than of length 1 and therefore more time is consumed. Let V be the monetary value of production time per second and T be the time (in seconds) that is needed to roll one length. Due to the production process this time is fixed and independent of the length. In order to consider the time waste in the model we can calculate a penalty for each alternative cutting length. This penalty is based on the main length and the number of times d_{1w} it has to be cut within lot w. The alternative lengths in general have more cuts d_{iw}. Therefore, the penalty costs pc_{iw} of alternative i of lot w are determined as follows:

$$pc_{iw} = (d_{iw} - d_{1w}) \cdot T \cdot V. \tag{6}$$

Then, the penalty PC_j of one pattern j can be computed as follows:

$$PC_j = \sum_{w=1}^{W} \sum_{i=2}^{I_w} \frac{a_{iwj}}{d_{iw}} \cdot pc_{iw}. \tag{7}$$

In the general problem, a cutting pattern a_j is represented by a tuple with the number of times the cutting lengths l_m appear in the pattern $a_j = (a_{1j}, \ldots, a_{Mj})$. The transposed representation is used together with the vector $d = (d_1, \ldots, d_m)^T$ within the DCG which is based on a revised simplex. Because of the alternative lengths, this representation of cutting patterns cannot be used anymore. The alternative lengths cannot be handled independently from each other. Therefore, we have to adapt the representation of a cutting pattern. Again, the representation is a tuple of W values. Instead of using the number of times a cutting length appears in a pattern we use the percental share of a cutting pattern to fulfil rolling lot w. Now, each value represents the degree cutting pattern j fulfils rolling lot w:

$$a_j = \left(\sum_{i=1}^{I_w} \frac{a_{i1j}}{d_{i1}}, \ldots, \sum_{i=1}^{I_w} \frac{a_{iWj}}{d_{iW}} \right). \tag{8}$$

For example: Given length $L = 10$ of input material and two orders O_1 and O_2. Order O_1 requires 5 pieces of length $l_1 = 5$, $O_{11} = (5, 5)$, or 8 pieces of length $l_2 = 4$, $O_{12} = (8, 4)$, and O_2 requires 7 pieces of length $l_3 = 3$, $O_2 = (7, 3)$. With the classic representation (see Fig. 1) the first two rows affect order 1 and the third row affects order 2.

Fig. 1 Not dominated cutting patterns—classic representation

$$\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

Fig. 2 Not dominated cutting patterns—percental representation

$$\begin{pmatrix} 2/5 \\ 0 \end{pmatrix}, \begin{pmatrix} 13/40 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/5 \\ 1/7 \end{pmatrix}, \begin{pmatrix} 1/4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/8 \\ 2/7 \end{pmatrix}, \begin{pmatrix} 0 \\ 3/7 \end{pmatrix}$$

In order to get only one row per customer order we now use the adapted cutting pattern representation with its percental notation, see Fig. 2. Then, the rows of a cutting pattern do not represent the cutting lengths anymore. Instead, they represent the degree the cutting pattern fulfils an order O_w.

4 Problem Solving Method

The optimisation model we described above is NP-hard and cannot be solved exactly in acceptable computational time [9]. That is because the number of cutting patterns is growing exponentially with the number of cutting lengths [11]. There are several approaches that solve the standard problem [6]. Many of them are linear programming solution procedures e.g. [14] based on the Delayed Column Generation approach (DCG) of Gilmore and Gomory [3, 4]. Others are so-called sequential heuristic procedures, hybrid solution procedures [6] or artificial intelligence solutions e.g. [6]. Despite his age the DCG is still one of the most preferred solution procedures because of its good computational time, the quite simple implementation and the quite good results. For one-dimensional cutting stock problems new approaches often do not operate better [1] or faster [13] than the DCG. Furthermore, sequential heuristic procedures for example may lead to a solution with much higher waste [6]. For these reasons and because we are facing a new problem variant of the one-dimensional cutting stock problem we have chosen the DCG as solution procedure and adapted it. Future work can be done by analysing if other solution procedures would operate better than the standard DCG.

5 Results

We analysed the algorithm and its benefit varying the problem instances. Taking typical numbers of rolling lots $W \in (5, 10, 20, 30)$ and average numbers of ordered main cutting lengths $\bar{d} \in (50, 100, 150)$ we computed the percental divergence of the solution to the lower bound of the theoretically possible optimal solution. This lower bound is calculated within the DCG. The results are not surprising: The more lots we are taking into account and the bigger the rolling lots are at an average the better the results are. The difference between the lower bound of the optimal solution and the computed solution decreases from over 8 % to less than 1 %. If we assume, that the manual planning has about the same solution quality as the first problem instance of $W = 5$ and $\bar{d} = 50$ (the most typically instance) then we are facing a decrease of waste of up to 8 %. In fact, an analogue decrease could be observed.

Much more lots could be taken into account so that the planning period for the cutting could be widened. As a side effect the factory in deed could operate at full capacity. If we widen the analysis from the exclusive use of the main variant to the use of multiple alternatives the results improve only little. The reason is the penalty for the use of side variants that leads to the fact that they are used more seldom than the main length.

For ten examples also the optimal solution was computed. The algorithm found five from ten optimal solutions. Three times the divergence was less than 2 % and only two times more than 2 % (maximum was 4.2 %). Thus, the algorithm really improves the situation of the manufacturer. It is less time consuming and in most cases it reduces the waste at a very high level. A typical customer order ranges between 5 and 3,000 t of steel. At a normal price level for steel of about 300US$, a saving of only 1 % of input material leads to savings of approximately 5000 US$ and overall savings greater than 1 M US$. Concerning the high steel prices of 2008 and saving of more than 4 % averaged the manufacturer's benefit is much greater.

References

1. Belov, G., Scheithauer, G.: A branch-and-cut-and-price algorithm for one-dimensional stock cutting and two-dimensional two-stage cutting. Eur. J. Oper. Res. **171**(1), 85–106 (2006)
2. Chenge, C.H., Feiring, B.R., Cheng, T.C.: The cutting stock problem: A survey. Int. J. Prod. Econ. **36**, 291–305 (1994)
3. Gilmore, P.C., Gomory, R.E.: A linear programming approach to the cutting stock problem. Oper. Res. **9**, 849–859 (196)
4. Gilmore, P.C., Gomory, R.E.: A linear programming approach to the cutting stock problem: Part II. Oper. Res. **11**, 863–888 (1963)
5. Golden, B.L.: Approaches to the cutting stock problem. AIIE Trans. **8**, 265–274 (1976)
6. Haessler, R.W., Sweeney, P.E.: Cutting stock problems and solution procedures. Eur. J. Oper. Res. **54**, 141–150 (1991)
7. Johnston, R.E.: A direct combinatorial algorithm for cutting stock problems. In: Anderssen, R.S., De Hoog, F.R. (eds.) Application of Mathematics in Industry. Addison-Wesley Publishing, Massachusetts (1982)
8. Kantorovich, L.V.: Mathematical methods of organizing and planning production. Manage. Sci. **6**(4), 366–422 (1960)
9. Korte, B., Vygen, J.: Combinatorial Optimization: Theory and Algorithms, 2nd edn. Springer, Berlin (2002)
10. Lai, K.K., Chan, W.M.: An evolutionary algorithm for the rectangular cutting stock problem. Int. J. Ind. Eng. **4**(2), 130–139 (1997)
11. Pierce, J.F.: Some Large Scale Production Problems in the Paper Industry. Prentice Hall, Engelwood Cliffs (1964)
12. Reinertsen, H., Vossen, W.M.: The one-dimensional cutting stock problem with due dates. Eur. J. Oper. Res. **201**(3), 701–711 (2010)
13. Stadtler, H.: A comparison of two optimization procedures for 1 and 1.5-dimensional cutting stock problems. OR Spektrum **10**, 97–111 (1988)
14. Stadtler, H.: A one-dimensional cutting stock problem in the aluminium industry and its solution. Eur. J. Oper. Res. **44**, 209–223 (1990)
15. Wäscher, G., Haußner, H., Schumann, H.: An improved typology of cutting and packing problems. Eur. J. Oper. Res. **183**, 1109–1130 (2007)

Part XIII
Renewable Energy and New Mobility

E-mobility: Assessing the Market Potential for Public and Private Charging Infrastructure

Robert Rieg, Stefan Ferber and Sandrina Finger

1 Introduction

Except for hybrid drives all electric cars need charging in some or another way. Charging stations are prerequisite components of e-mobility infrastructure. It is plausible to assume that a wide net of charging stations will ease decisions to buy an e-vehicles. Yet, these decisions are further complicated by extreme range expectations of potential buyers. Potential buyers compare ranges of e-vehicles to traditional combustion-based ones although the average travel distance of consumers per day is clearly lower than the average range of current e-vehicles and probably even lesser relevant given future improvements in battery technology. But such a "range anxiety" can hinder the diffusion of e-vehicles which is still low.

The general public and governments assume that private investors will build an adequate charging infrastructure. In the light of complex consumer preferences and a slow diffusion, it seems reasonable for investors to wait for investments until a certain number of e-vehicles is demanding for charging services. However, a lack of infrastructure would hinder the diffusion of e-vehicles. This interaction between infrastructure and demand produces a "chicken-egg" problem which could paralyse relevant decisions to invest in infrastructure and to buy e-vehicles respectively.

A decision to invest in infrastructure depends on an assessment of potential markets for charging infrastructure and services. A promising way is to understand causal

R. Rieg (✉)
Institute of Applied System Dynamics, Aalen University, Beethovenstrasse 1,D-73430 Aalen, Germany
e-mail: robert.rieg@htw-aalen.de

S. Ferber
Bosch Software Innovations GmbH, Stuttgarter Strasse 130,D-71332 Waiblingen, Germany
e-mail: stefan.ferber@bosch-si.com

S. Finger
Aalen University, Beethovenstrasse 1,D-73430 Aalen, Germany
e-mail: sandrina.finger@gmail. com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_42,

forces driving or dampening diffusion of such innovations and to integrate that in a dynamic business model. The outcome of such a model is not a point prediction but gives insight in main variables and policies that could lead to a successful e-mobility market as well as degrees of magnitudes of expected market size.

2 Capturing Dynamics of Business Models for E-mobility Markets with System Dynamics

2.1 Dynamics of Business Models and System Dynamics

A dynamic perspective is important to understand the long-term success or failure of a firm, a new venture or new product. Success or failure of e-mobility will be determined by the interplay of many factors, so understanding it needs appropriate methods to capture the dynamic interaction of causal factors. We suggest to use System Dynamics for two reasons: First, the effects under scrutiny develop over time so a method to understand dynamic processes is appropriate and a key feature of System Dynamics [1]. Second, the perspective of System Dynamics is to explain behavior from the structure of systems, i.e. endogenously [2]. The latter seems especially important, since it directs the analysis toward finding policies to enable e-mobility in the first place.

2.2 Diffusion of Innovations as Theoretical Underpinning

E-mobility and its infrastructure can be modelled as diffusion of innovations [3]. Diffusion and adoption are general processes which can be applied to physical goods, services as well as ideas [4]. Cumulative adoption seem to form an S-shaped curve over time which is a product of at least two processes: one process reinforcing adoption and one process dampening or reversing it [1]. Factors advancing adoption and diffusion can be grouped in two broad categories, however factors may work in an intertwined fashion [4–6]: (a) Social factors like normative and regulatory pressures to conform to common behaviour and practices or imitation and herding; (b) Economic or efficiency factors like expected profits or cost-savings through innovation. Adoption decision is a result of cost-benefit calculations of potential adopters as well as of decisions to stay with innovations.

The increase of diffusion over time can be explained by a process similar to "infection of a disease": potential adopters get in contact, learn about the innovation through for example word of mouth and decide to adopt it. High adoption rates lead to fast diffusion. Limits to diffusion are exerted by the limited number of potential adopters and may be additional factors like incompatible preferences held by the majority (e.g. [7]). Adopters can be divided in several groups [4], according to their time of adoption. Typically, a small group of so-called innovators (2.5 %) will be the

first to adopt. They are followed by early adopters (13.5 %), the majority (68 %) and finally by the laggards (16 %). It seems plausible to assume the same will happen with e-mobility so all efforts should focus on the first two groups to lead the way to broad diffusion.

3 Modelling Charging Infrastructure and Services

3.1 Focus on Alpha Cities and Charging Services

The amount of innovators and early adopters depends in reality on attributes of innovations as well as characteristics of adopters [4]. Successful diffusion of e-mobility should focus on market segments whith high amounts of innovators and early adopters. That seems true for so-called "Alpha Cities" like New York and Singapore: larger cities with a high degree of connectivity, openness to innovations and change and a need for new mobility solutions.

We consider conductive charging as main charging method, since it requires only modest investments in infrastructure at home and elsewhere. It is reasonable to assume that energy and charging infrastructure provider will agree on some form of "roaming" procedure to allow for charging in any order and subsequent clearing of claims and payments. Roaming will be important to convince potential customers of EVs.

Given the costs of charging stations compared to prices of electricity, revenues from selling electricity will be not sufficient to make investments profitable. Additional sources of revenues are needed, especially from services like [8]: (1) Services for private customers: leasing of charging station installed at home (if not bought), installation, maintanance as well as updates, (2) services for fleet operators (taxi, mail or parcel delivery services): EVs are especially attractive for inner cities with short distances, frequent stops and lesser pollution. Fleet operators with large fleets of EVs can exploit economies of scale which make EVs even more attractive. (3) Services for other corporate customers: lease rates (optional), installation, and maintenance as well as operation of charging stations for charging private or corporate EVs at work including billing services. (4) Services for Utility firms: billing for electricity supply to EVs and smart load management (SLM) [9]. (5) Services for general businesses: income from offering advertising space on charging stations. The costs of advertising space in larger cities is significant. A large number of charging stations should then lead to substantial revenues.

3.2 Modeling Demand and Competition

Private and corporate customers differ in making purchasing decisions, in needs and preferences. Hence, it is plausible to distinguish between services for private

(called basic charging service 1, BCS 1) and corporate customers (BCS 2). Further revenue streams will come from advertising and smart load management. Demand will be influenced by word of mouth (WOM) and advertising. WOM is seen as more important for innovators and early adopters than advertising which affects later adopter groups. The main driver of WOM is the visibility of the innovation to potential adopters as well as an ongoing accessability of it in the minds of these people [10]. Marginal rate of net benefits of WOM will decline over time when innovators and early adopters already adopted. Later adopters react more on advertising. An additional effect on demand is the network effect [11] that stems from benefits by an increasing number of charging stations and services. Finally, demand and supply for services and demand and supply for e-vehicles will interact in a non-linear feedback process as mentioned in the beginning (chicken-egg-problem).

At best the market for charging services will grow due to diffusion of e-vehicles and parallel contracting of services. Firms offering services will benefit not only directly from providing services with better cost-benefit-relations than competitiors. They will also benefit indirectly from a growing market and other competitors because of increased awareness and attractiveness of charging services.

4 Simulation and Scenarios

Figure 1 shows the core structure of the system dynamics model with two main main feedback processes: (a) network effect and (b) market share saturation. The network effect depends on number of installed services and by the sensitivity to the product's network effect. The attractiveness of a service is derived by a network effect on attractiveness as well as other factors like availability, convenient handling or green image for the customers' lifestyle. The market share of each charging service provider is derived from the own attractiveness compared to total attractiveness of charging services in general. The revenue stream of each service results from the number of installed services and their prices, which decrease with a rising number of services. All installed charging stations are assumed to be integrated into the smart load management service for utilities. The total revenue stream compounds the revenue streams of all services.

The model simulates a base case and a best case scenario. The best case scenario differs from the base case by more optimistic assumptions like faster WOM and stronger network effects. The results of base case and best case in terms of revenues indicate that basic charging services for private and corporate customers will be the main sources of revenue, only in the best case advertising and smart load management produce noteworthy earnings.

As a result, firms willing to invest in charging infrastructure should then concentrate on offering charging services and try to foster word of mouth in the 1st years, later on accompanied by advertising. However, that does not solve the "chicken-egg problem", someone has to make the first move.

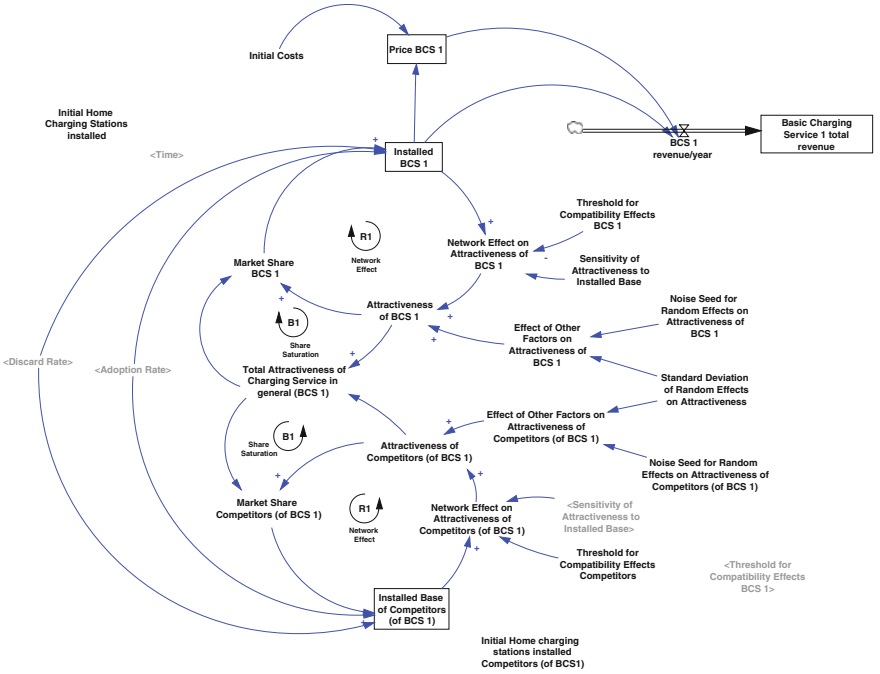


Fig. 1 Core structure of the system dynamics model

5 Conclusion

Widespread diffusion of e-mobility will materialize only if there are adequate investments in infrastructure esp. charging stations. Until the near future, selling electricity alone will not produce enough net earnings to make such investments profitable. Additional services should be offered. Our system dynamics model suggests successful business models for charging stations should focus on a combination of services for private and corporate customers including smart load management and advertising. The main purpose and benefit of this and other system dynamics models is not to predict the future but to help firms to prepare for it: through highlighting the factors and policies necessary to succeed.

References

1. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill, Boston u.a (2000)
2. Richardson, G.P.: Reflections on the foundations of system dynamics. Syst. Dyn. Rev. **27**(3), 219–243 (2011)
3. Bass, F.M.: A new product for growth model consumer durables. Manag. Sci. **15**(5), 215–227 (1969)

4. Rogers, E.M.: Diffusion of Innovations. The Free Press, New York (2003)
5. Westphal, J.D., Gulati, R., Shortell, S.M.: Customization or conformity? An institutional and network perspective on the content and consequences of tqm adoption. Adm. Sci. Q. **42**(2), 366–394 (1997)
6. Kennedy, M.T., Fiss, P.C.: Institutionalization, framing, and diffusion: The logic of tqm adoption and implementation decisions among U. S. hospitals. Acad. Manag. J. **52**(5), 897–918 (2009)
7. Fiss, P.C., Zajac, E.J.: The diffusion of ideas over contested terrain: The (non) adoption of a shareholder value orientation among german firms. Adm. Sci. Q. **49**(4), 501–534 (2004)
8. Kley, F., Lerch, C., Dallinger, D.: New business models for electric cars-a holistic approach. Energy Policy **39**(6), 3392–3403 (2011)
9. Brown, S., Pyke, D., Steenhof, P.: Electric vehicles: The role and importance of standards in an emerging market. Energy Policy **38**(7), 3797–3806 (2010)
10. Berger, J., Schwartz, E.M.: What drives immediate and ongoing word of mouth? J. Mark. Res. **48**(5), 869–880 (2011)
11. Katona, Z., Zubcsek, P.P., Sarvary, M.: Network effects and personal influences: The diffusion of an online social network. J. Mark. Res. **XLVIII**(June):425–443 (2011)

A Decentralized Heuristic for Multiple-Choice Combinatorial Optimization Problems

Christian Hinrichs, Sebastian Lehnhoff and Michael Sonnenschein

1 Introduction

In 0-1 multiple-choice combinatorial optimization problems, multiple sets or classes of elements are given, from which each exactly one element has to be chosen to form a solution. The goal is to find a solution that minimizes (or maximizes) a given objective function. Such problems are typically NP-hard in the weak sense, so that they can be solved in polynomial time using dynamic programming [4]. By exploiting the specific structure of a given problem, even linear time can be achieved. For example, the well-known multiple-choice knapsack problem can easily be solved using the *core* concept [6]. But such methods cannot be applied to all problems of this class for two reasons: First, a core may be difficult to find, as it is the case in some (multiple-choice) subset-sum problems [7]. Second, and more importantly, the problem may have to be solved in a decentralized way, so that global knowledge is not available. This is especially the case in distributed systems with autonomous actors. For example, such a system may not support gathering global knowledge due to its openness, or the cost for doing this may be very high. Also, the collection of global knowledge at certain stages of an algorithm (i.e. at startup) leads to synchronization points, which are undesirable in many cases. Finally, such actions may violate privacy considerations. Typical application fields for these kinds of problems are distributed resource allocation, wireless sensor networks, media streaming in bandwith-contrained environments, logistics and decentralized energy management.

C. Hinrichs (✉) · M. Sonnenschein
University of Oldenburg, Oldenburg, Germany
e-mail: hinrichs@informatik.uni-oldenburg.de

M. Sonnenschein
e-mail: sonnenschein@informatik.uni-oldenburg.de

S. Lehnhoff
OFFIS, Institute for Information Technology, Oldenburg, Germany
e-mail: lehnhoff@offis.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_43,

In this contribution, we present a **C**ombinatorial **O**ptimization **H**euristic for **D**istributed **A**gents (COHDA). Unlike other population-based heuristics (i.e. particle swarm optimization), the individuals ("agents") in COHDA do not describe candidate solutions. Instead, each agent represents a class of elements of the considered multiple-choice combinatorial opimization problem, so that solutions comprise individual decisions from all participating agents. The heuristic solves the given problem in a completely decentralized manner, relying on individual knowledge and utilizing cooperative behaviour. The individual knowledge bases are formed through local perception plus beliefs about (possibly incomplete or outdated) global knowledge. Convergence of the search process is achieved by distributing the beliefs to other agents. This kind of heuristic corresponds to the *Cooperative Algorithmic-Level Parallel Model* defined in [9].

2 Problem Definition and Model

As stated in the introduction, multiple-choice subset-sum problems (MC-SSP) belong to the rather difficult types among combinatorial optimization problems, because their structure is less exploitable. Hence, in this contribution, we focus on this problem type. We are given m classes with each class i containing n_i elements. The jth element of class i has weight w_{ij}. From each class, exactly one element has to be chosen for a feasible solution. In MC-SSP, there exists a capacity c which defines an upper bound that should be approximated, but not exceeded by the sum of the weights of the chosen elements. We generalize from this formulation by removing the upper bound constraint. The goal is therefore to approximate c as close as possible from any side, which yields a problem with increased solution space. We call this generalization *multiple-choice combinatorial optimization problem* (MC-COP). Formally, it can be expressed with an integer programming model:

$$\text{(MC-COP)} \qquad \min \; d\left(c, \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(w_{ij}\cdot x_{ij}\right)\right) \tag{1}$$

$$\text{subject to} \quad \sum_{j=1}^{n_i} x_{ij} = 1, \; i = 1\ldots m,$$

$$x_{ij} \in \{0,1\}, \; i = 1\ldots m, \; j = 1\ldots n.$$

The objective function d can be defined arbitrarily, i.e. the 1-norm might be used:

$$d(u,v) = \|u - v\|_1 \tag{2}$$

Since we are targeting distributed systems, we mapped this model to a multi-agent system (MAS). In this MAS, each agent represents an element class. Hence, each agent a_i has to select one of its elements to be included in the solution by assigning $x_{ij} = 1$, where j is the index of the selected element. The difficulty of the problem arises from the lack of global knowledge at agent level, i.e. there is no global decision maker, and the solution has to be determined in a decentralized fashion. In order to accomplish that, for each agent a_i, a neighborhood set $\mathcal{N}_i$ of other agents is defined with whom a_i is able to communicate. The communication network may be expressed with an undirected graph. An important property of the communication layer are message delays: We assume that sent messages take an arbitrary time until they are received by the target agent, as it is the case in most real communication networks like the internet. However, we simplify from reality by limiting this delay to a known upper bound $t_{msg,max}$, and we assume that no messages are lost during transmission. Finally, we presume a central operator, who is able to broadcast the target value c to all agents and thus can initiate the heuristic.

3 Heuristic

The task of each agent is to select one of its own controlled elements, so that the sum of the weights of all selected elements in the population minimizes the objective function d. To accomplish that, each agent communicates with its neighbors and shares knowledge that helps in selecting optimal elements. In our approach, an agent a_i needs to exchange only two items with its neighbors: The best combination of weights $\hat{W}_{i,best}$ of selected elements that the agent has seen during the process so far and the currently selected weights $\hat{W}_{i,current}$ that the agent is aware of. Each of these communicated weights is labelled, where a label $l\left(w_{ij}\right) = (i,\ s_i)$ contains the unique ID i of the agent the selected weight belongs to, and a counter s_i that reflects the "age" of the selection. Note that, however $\hat{W}_{i,best}$ and $\hat{W}_{i,current}$ must each satisfy the constraints of MC-COP (i.e. they must not contain more than one chosen weight per agent, see (1)), these sets are allowed to be incomplete. At the beginning of the process, $\hat{W}_{i,best}$ and $\hat{W}_{i,current}$ will each contain only the weight $w_{i,init}$ of the initially selected element of the agent a_i itself, labelled $l\left(w_{i,init}\right) = (i,\ 0)$. However, as the agent begins to communicate with its neighbors, the sets will be updated with each information exchange. The agent is allowed to change its element selection with the help of the knowledge contained in $\hat{W}_{i,best}$ and $\hat{W}_{i,current}$ at any time under one condition: It has to make this selection public to its neighbors, while labelling it with its age counter increased by one. Hence, the age counter s_i of an agent a_i increases each time the agent changes its selection, thus making it possible to decide if received information concerning an agent is newer than an already stored value. The complete heuristic executed by each agent comprises three steps and can be described as follows.

1. **(update)** Every time an agent a_i receives a message, its local knowledge base is updated with the information received. This can be either the target value c, which starts the process and is broadcasted by the central operator, or information from a neighbor agent a_h, containing the sets $\hat{W}_{h,best}$ and $\hat{W}_{h,current}$. In the first case, the value c is stored locally, and step 3 is executed without further conditions. Otherwise, the local knowledge base $\hat{W}_{i,current}$ is updated with the items in $\hat{W}_{h,current}$ by adding selected weights from agents not known so far, and replacing outdated selections according to the age counters in the associated labels. The local set $\hat{W}_{i,best}$ is replaced by $\hat{W}_{h,best}$ if the latter contains more elements, or if it yields a better value using the objective function. If either $\hat{W}_{i,best}$ or $\hat{W}_{i,current}$ has been changed during this process, step 3 is executed afterwards.
2. **(choose**) In this step, the agent iterates through its own weights and calculates the value of the objective function for each of these weights when combined with the items in $\hat{W}_{i,current}$, thus testing which selection $w_{i,new}$ fits the best into the currently existing configuration. Note that this local view on the system defined by $\hat{W}_{i,current}$ will most likely already be outdated due to the asynchronous execution of the heuristic in each agent. Yet it describes an approximation to the global system state that helps in finding optimal selections. The best resulting objective value found in the iteration of weights is then compared to the objective value of $\hat{W}_{i,best}$. If the found objective value of $\hat{W}_{i,current} \cup \{w_{i,new}\}$ is better than the value of $\hat{W}_{i,best}$ and the size of $\hat{W}_{i,current} \cup \{w_{i,new}\}$ is at least as large as the size of $\hat{W}_{i,best}$, the weight $w_{i,new}$ is marked as selected and the new best known configuration is stored as $\hat{W}_{i,best} := \hat{W}_{i,current} \cup \{w_{i,new}\}$. Otherwise, the agent reverts its selection to the one stored in $\hat{W}_{i,best}$.
   If the selected weight of agent a_i has changed during this process, step 3 is executed afterwards.
3. **(ublish)** In this step, the agent sends its currently stored sets $\hat{W}_{i,best}$ and $\hat{W}_{i,current}$ to all of its neighbors.

Utilizing these actions, the agents will update each other iteratively with the newest information they are each aware of, while simultaneously trying to find the best global combination of weights. This leads to some important properties of COHDA:

Convergence. The conditions in step 3 will first cause the individual sets $\hat{W}_{i,best}$ to increase in size until completeness, hence spreading knowledge in the system, even between agents that are not directly connected. Afterwards, the heuristic will start to converge, because only better rated configurations survive the selection process in each iteration. When no agent is able to find a better configuration, all agents will eventually stick to a common best known configuration.

Cooperation. The agents need to cooperate with each other, i.e., they need to know the global objective function d, and have to be trustworthy in their exchanged messages.

Termination. After convergence, no agent will change its selected element any more. Hence, termination of the heuristic can be determined by observing inactivity of all participating agents. From the point of view of the global observer, this is the

case when no messages have been sent for a duration $t_{msg,max}$ by any agent (thus eliminating the uncertainty introduced by arbitrary message delays).

Completeness. Due to the lack of global knowledge, the algorithm is not complete; it cannot be guaranteed that the optimum is found.

4 Evaluation

We implemented COHDA in a discrete-event simulation system that simulates a communication network using integer-valued message delay time steps. Each simulation step represents one simulated time step. The evaluation example corresponds to an application from the domain of decentralized energy management. In the electric grid, the supply and demand of energy has to be balanced at every point in time. In the context of operations research, this problem has been formulated as the electrical generation unit commitment problem (UC, see [5] for a comprehensive survey). In our evaluation, we do not only consider generators, but include flexible loads as well, which is known as demand side management (DSM). We mapped this problem to MC-COP by replacing the single-valued integer variables in (1) by q-dimensional vectors and allowed negative values, thus being able to model a time series of electrical load (with k discrete time steps, $k = 1 \ldots q$), which should be approximated by the sum of the agent's individual load curves; see [2] for a more detailed formalization. The considered problem instances were generated using fixed parameters $m = 30$ (number of element classes), 100 different elements in every class with $q = 96$ dimensions each, and strongly correlated weight values w_{ij}, which have been identified as hard instances for SSP-like problems [3, 6, 7]. Five different target time series c_h were generated as described in [1], thus $1 \leq h \leq 5$. Each of these five configurations was simulated 100 times with a set of 100 seeds for the random number generator, so that within each configuration the same 100 seeds were used for the 100 simulation runs. The communication network was modelled after the small world paradigm [8] with a rewiring propability $\Phi = 2.0$ and message delays uniformly distributed in $[1,\ 10]$. The output of the objective function was normalized to $[0,\ 1]$ using an optimal-to-worst-interval according to [9].

Table 1 COHDA evaluation results: final objective value d, number of messages per agent per time step msg, number of time steps until termination t for different targets c_h

	$h = 1$		$h = 2$		$h = 3$		$h = 4$		$h = 5$	
	mean	std	mean	std	mean	std	mean	std	mean	std
d	0.0044	0.0122	0.0024	0.0084	0.0031	0.0142	0.0031	0.0096	0.0027	0.0088
msg	0.5495	0.0201	0.5400	0.0222	0.5354	0.0234	0.5456	0.0279	0.5477	0.0255
t	373.59	171.66	465.41	216.07	526.43	236.98	443.58	226.83	408.16	210.93

Summarized in Table 1 are the means and standard deviations of the final objective value d, the number of messages sent per agent per time step (msg), and the number of time steps t until termination for each of the five configurations.

5 Conclusion

The evaluation results show that COHDA is able to find near-optimal solutions within a range of 0.0031 ± 0.0106 (mean over all configurations) of the theoretical optimal value $d_{opt} = 0$ for the given multiple-choice combinatorial optimization problem. While the run-times vary rather large (443.434 ± 212.493 simulated time steps), the exchanged messages remain very constant at 0.5436 ± 0.0238 messages per agent per time step on average over all simulation runs. We will publish a more thorough evaluation in a subsequent paper. There, the approach will be compared to related work, and a multi-objective variant of COHDA will be presented.

Future work will mainly focus on adaptivity. In this context, we will extend MCCOP with a dynamic objective function, and we will incorporate the ability to handle non-constant agent populations as well as adaptive communication networks into COHDA. An interesting research topic resulting from these extensions is the stability of the heuristic against destructive/harmful agents.

References

1. Han, B., Leblet, J., Simon, G.: Hard multidimensional multiple choice knapsack problems, an empirical study. Computers & Operations Research **37**(1), 172–181 (2010). doi:10.1016/j.cor.2009.04.006
2. Hinrichs, C., Vogel, U., Sonnenschein, M.: Approaching Decentralized Demand Side Management via Self-Organizing Agents. In: Yolum, Tumer, Stone, Sonenberg (eds.) ATES Workshop, Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011). Taipei, Taiwan (2011).
3. Lust, T., Teghem, J.: The multiobjective multidimensional knapsack problem: a survey and a new approach. International Transactions in Operational Research 19(4), 495–520 (2012). doi:10.1111/j.1475-3995.2011.00840.x
4. Martello, S., Toth, P.: Knapsack problems, 1 edn. John Wiley & Sons (1990).
5. Padhy, N.: Unit Commitment-A Bibliographical Survey. IEEE Transactions on Power Systems **19**(2), 1196–1205 (2004). doi:10.1109/TPWRS.2003.821611
6. Pisinger, D.: A minimal algorithm for the multiple-choice knapsack problem. European Journal of Operational Research 83(2), 394–410 (1995).doi:10.1016/0377-2217(95)00015-I
7. Pisinger, D.: Linear Time Algorithms for Knapsack Problems with Bounded Weights. Journal of Algorithms **33**(1), 1–14 (1999). doi:10.1006/jagm.1999.1034
8. Strogatz, S.H.: Exploring complex networks. Nature **410**(6825), 268–276 (2001). doi:10.1038/35065725
9. Talbi, E.G.: Metaheuristics. John Wiley & Sons, Inc., Hoboken, NJ, USA (2009). 10.1002/9780470496916.

An E-Clearinghouse for Energy and Infrastructure Services in E-Mobility

Andreas Pfeiffer and Markus Bach

1 Introduction

Electrical transportation are globally perceived as the technology of the future for sustainable private transport. Supplied with renewable energies, transportation based on electrified powertrains contribute in an important way to climate protection and reduce local emissions including noise and particulate matter in the cities. Due to ambitious environmental aims in Europe, a change in mobility behaviour and steadily rising gas prices the market breakthrough seems to be only a matter of time.

Among the availability of extensively usable electric vehicles, a comprehensive charging infrastructure (CI) is discussed as a critical factor for the success of e-mobility. For battery electric vehicles (BEV) a large scale network of charging stations has to be implemented in public areas, especially since in the inner city area there are few garages or private parking spaces with a power supply. Following the current discussion a non-discriminated customer access to CI is essential at this time, to strengthen the trust of the customers in the subject of e-mobility [5]. Various methods can be used to realise these efforts.

In addition to the testing of customer neutral methods of direct payment, systems for the support of comprehensive business models derived from roaming processes are developed and tested. On the basis of "e-Roaming agreements" relevant Triple-A-Data is already exchanged [1].

By an "Clearinghouse for e-mobility" or "e-Clearinghouse" an information system is understood, which allows end customers access and usage of CI of different providers in the e-Roaming market model. It manages roaming

A. Pfeiffer (✉) · M. Bach
Wirtschaftsinformatik und Operations Research RWTH Aachen, Templergraben 64,
52052 Aachen, Germany
e-mail: pfeiffer@rwth-aachen.de

M. Bach
e-mail: markus.friedrich.bach@rwth-aachen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_44,

agreements between numerous providers and handles the exchange of Triple-A-Data. Data
Clearing, Nettoclearing for B2B-Accounting and the Settlement form the economic basis for comprehensive business models in e-mobility. Open interfaces and Data Mining applications serve as the essential component for future provider-specific and general value-added services as well as a cross-linking of transport and energy systems.

2 Fundamentals of a Public Charging Infrastructure

It can be assumed that for the success of e-mobility an extensive network of freely accessible charging stations is a necessity. Thereby result, due to the free access in the public area, higher access requirements than with private stations [5]. To prevent the abuse of non-paying customers and violations, it is important to control the access to the stations and to authenticate customer relationships.

A benchmark of different approaches showed, that at this point no consistent means of authentication could prevail. Nevertheless, chips with radio frequency identification (RFID) in the form of customer cards are the most common. In future mobile devices or a direct communication between e-cars and infrastructure as recommended in ISO/IEC 15118 could gain in importance. Using of ISO/IEC 15118 compatible devices will open an efficient way for new system services like V2G-applications in the near future.

The online capability of CI is proposed and gives advantages like status and error informations for the charging station operators (CSO). This optimises operations as well as contribute to the grid quality and the inclusion of renewable energies. Apart from these mere practical reasons, customer friendly offers such as an intelligent tariff control or real-time information in traffic and navigation systems can be realised. In that way the high demand on an ecological and economic reasonable contribution of e-mobility in current energy systems and mobility concepts can only be realised by an interlinked infrastructure [4].

To allow the billing of individual customers, a distinct identification and allocation of every charging to a contractual relationship has to take place. With the electricity output at public charging stations the usual customer-provider relationships in the energy sector, which assign one power supply (house connection) to one customer, do not necessarily apply anymore. At one charging point (grid connection) any number of customers can obtain electricity and customers can charge at any number of charging points [3]. As a market model e-Roaming can facilitate a discrimination-free end user access to the public CI, which is realizable more pragmatically and without complex alterations to energy-management processes [6]. Hereby e-Roaming means a customer can authenticate himself via one single authentication device at all charging stations of different e-mobility providers. The end customer contractual relationship remains with the own provider, i.e. at charging stations no change of supplier has to take place during the day. This is allowed by a contractual relation-

ship between the charging station operators respectively energy suppliers based on the roaming agreements similar to mobile communications.

3 Conception of an Open Architecture for E-Roaming a Clearinghouse for E-Mobility

For the possibility of e-Roaming into foreign station networks apart from the mentioned contractual relationship between the CSOs a data exchange between the independent CSMS is necessary. At the moment this is tested between individual research projects. However, for a nationwide e-Roaming in Germany in a mass market strived for until 2020 numerous bilateral agreements and interfaces to other CSMS become necessary, since the majority of the 1,050 German energy providers like municipal utilities operate only locally. In addition there would have to be agreements on a European level. Bilateral connections between all partners would result in a complicated network of agreements and data links.

For this reason the contractual administrative expenses and the data exchange have to be steadily simplified respectively standardised by appropriate, productive systems and efficient structures. This could be realised for example with an information system as an automated interface between CSO/CSMS by enabling an efficient real-time data exchange. A central system could link the spread CSMS and at the same time clearly display and manage hundreds of Roaming agreements. Furthermore by such a system additional value-added services for e-mobility can be provided, such as the central provision of geo- and status information for geo information services.

Similar challenges had to be faced by branches with comparable structures, whose solutions (e.g. GSM-Roaming and credit card authorizations) are partly transferable to e-mobility. Some of these solutions were examined in more detail and on this basis in the following a concept for an e-Clearinghouse (CH) in e-mobility is presented, which supports e-Roaming and enables structures suitable to a mass market. According to the presented definition a CH supports provider-individual business models in e-mobility, such as the use of batteries as grid buffers (reserve capacities) and at the same time increases the customer comfort by a standardised and easy use of CI.

The development of Roaming in mobile communications additionally showed that hundreds of bilateral roaming agreements with an increasing diversification of technologies and services become impractical, which is the reason for designing the present CH as an open Hub-model. Thereby bilateral structures are avoided and replaced by one single roaming agreement with an open Hub. This minimises the interfaces on the contractual and data level.

The Use-Case diagram in Fig. 1 shows the relevant participants, systems and business processes in an overview.

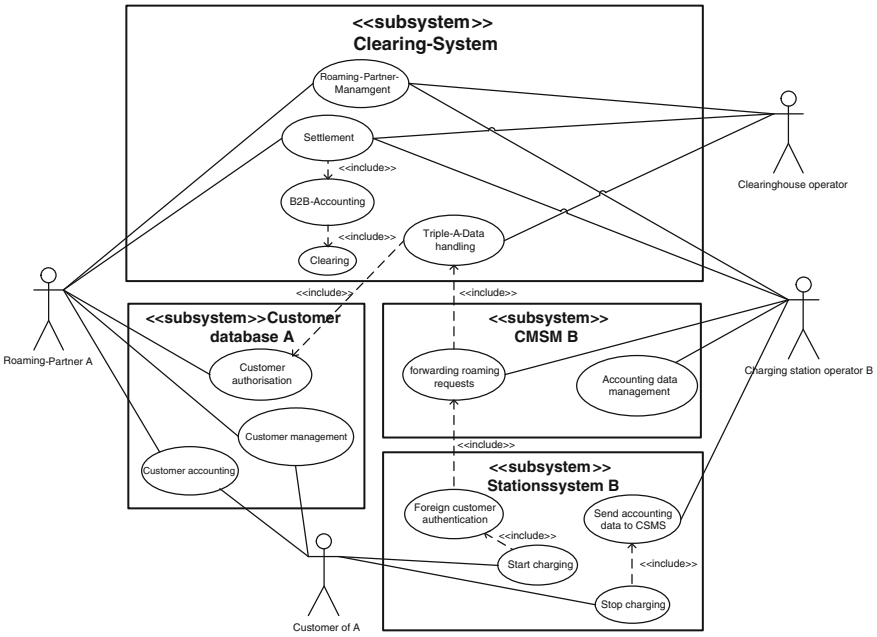


Fig. 1 Use-case diagram with relevant participants, systems and business processes

4 An E-Clearinghouse for Energy and Infrastructure Services in E-Mobility

The new structures of e-mobility necessitate intelligent Triple-A-Systems using modern ICT. The e-Roaming model was proposed as an sustainable and efficient method for a discrimination-free customer access to e-mobility infrastructre. For a comprehensive access for customers the responsible charging station management systems have to be linked and a B2B-Accounting system has to be established. The generic system architecture for a Clearinghouse for e-mobility "e-Clearinghouse" can be seen as an approach to describe an enabling technology setup, see Fig. 2.

The generic architecture described below was developed as a technology neutral recommendation, which can serve as a basis for future system developments. An adjustment on the typically pre-competitive market development is strongly recommended.

Correspondingly the architecture is designed open and modularly. In this way the necessary range of functions in each case can be constantly adjusted to the demand. As long as no economically viable business models exist and the billing process is more cost-intensive than the overall billing amount, initially no elaborate financial-Clearing and Settlement systems will be operated. Nevertheless they basically have to be developed in view of future business models and be provided already today [2].

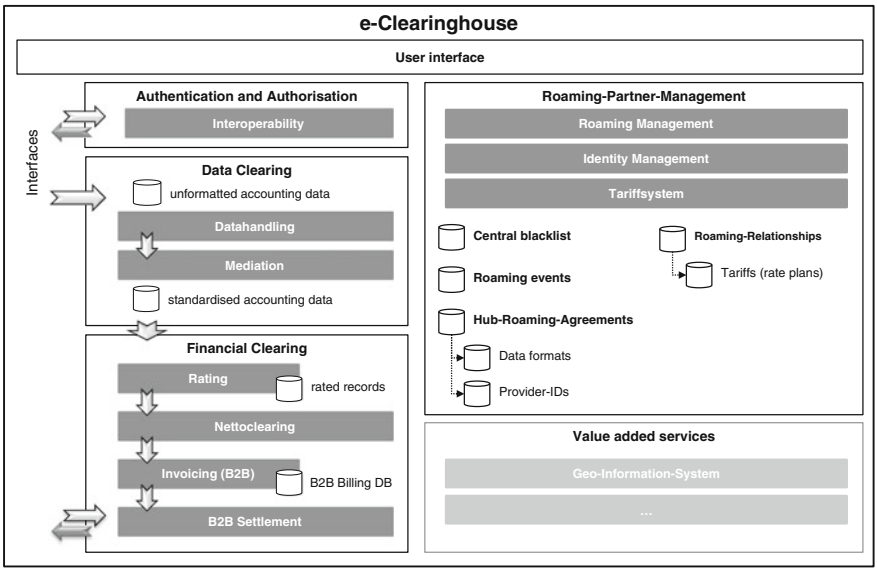


Fig. 2 System architecture of the E-clearinghouse

The CSMS will make a final authorisation decision with the knowledge of further (unknown to the CH) parameters such as the current load on the grid. A direct contact from the CH-System to charging stations or end customers will usually not occur; therefore it can be seen as a B2B platform.

With experiences from the examined industry solutions and the research work (i.e. econnect Germany) a structuring of the CH into the following five subsystems is proposed:

- Authentication and authorisation module
- Data Clearinghouse for relevant billing data
- Financial Clearinghouse (including Settlement system)
- Roaming-Partner-Management system
- User Interface

By means of the predicted numbers of the market growth and the structure especially of the German market it became clear that a European-wide comprehensive as well as discrimination-free access for customers to public charging stations via bilateral roaming agreements does not seem practical. The greatest problem consists of the contractual and administrative expenditure. For a thoroughly efficient arrangement of the simplest possible roaming relationships especially the Hub concept of the mobile communications was identified as a practical solution with Proof-of-Concept.

Already in the current market development phase an authentication and authorisation module should be implemented. Based on conversion engine the needed roaming data exchange between the attached countries and charging station operators could

be realised quickly. Thereby cross-border interoperability could be enabled based on different charging station management systems.

The subsystem to be implemented subsequently, the Data Clearinghouse, collects and harmonises relevant billing data sets of the charging stations operators. These bundled and uniform data sets are evaluated in the next step by a Financial Clearinghouse via bilateral tariffs. This results in mutual assets and liabilities, which are subjected to a nettoclearing, a balancing, so that merely the balances between the individual partners have to be accounted.

Thus European-wide roaming could be efficiently processed in a mass market with numerous charging stations, charging station management system and providers. In addition market entry barriers for small market participants would be kept low, since even with a limited effort a comprehensive Roaming can be offered. Uncertain general conditions can represent therefore opportunity as well as risk of an e-Clearinghouse.

References

1. BMWi: Deutsch/Niederländische Kooperation zur grenzüberschreitenden Elektromobilität in Europa (2010). http://www.ikt-em.de/de/1058.php. Cited 10 Oct 2011
2. Fest, C., Franz, O.: Szenario 7: Elektromobilität mit Roaming, pp. 114. (2011). http://www.modellstadt-mannheim.de/moma/web/media/pdf/FGR11_E-Energy_DATENSCHUTZ_Public_Onl.pdf. Cited on 17 Nov 2011
3. Fest, C., Franz, O., Gaul, A.: Verträge und Datenströme beim E-Roaming, p. 3. (2010). http://www.e-energy.de/documents/Gaul.pdf. Cited on 09 Jun 2011
4. Günther, M., Pfeiffer, A.: Anforderungen an eine Software für den Betrieb von Ladeinfrastruktur aus Sicht eines Stadtwerkes. In: Mattfeld, D.C.; Robra-Bissantz, S. (Eds.): Multikonferenz Wirtschaftsinformatik 2012. Tagungsband der MKWI 2012 (MKWI 2012), 29 Feb 02 Mar, Braunschweig (Deutschland), Berlin: GITO, 2012, S. 1421–1432 (2012)
5. NPE (Nationale Plattform Elektromobilität): Zweiter Bericht der Nationalen Plattform Elektromobilität, p. 37–42 (2011)
6. Zerres, A.: Elektromobilität - Stand der Diskussion, Regulierungsbedarf und Weiterungen. netconomica. Bundesnetzagentur, p. 6. Bonn (2011). http://www.wik.org/fileadmin/Konferenzbeitraege/netconomica/2011/Zerres.pdf. Cited on 07 Nov 2011

Decision Support Tool for Offshore Wind Parks in the Context of Project Financing

André Koukal and Michael H. Breitner

1 Introduction

The development of renewable energy technologies has been increasingly furthered by governments in various countries in order to reduce greenhouse gas emissions. In this context the offshore wind energy sector represents one major part as in some regions there is enormous potential for electricity production [7]. The importance is increasing due to the usage of ever larger dimensions of the wind energy turbines and its engines along with respective economies of scale. Due to high total costs of the wind parks project financing gains an increasingly important role. The profitability of the wind parks, however, depends on the regional feed-in tariffs. Behind this background it is important to evaluate the profitability of offshore wind parks for every individual project. This includes the identification and analysis of individual risk factors in the context of the risk management of a project in order to be able to make a differentiated statement regarding the influence of specific risk factors on the project success.

2 Decision Support Tool

The provided DST for evaluating offshore wind projects applies different methods and considers specific characteristics of project financing.

Project financing Project financing is characterized by three important characteristics [2]: For the *cash-flow related lending* the future project cash-flow represents the

A. Koukal (✉) · M. H. Breitner
Leibniz Universität Hannover, Hannover, Germany
e-mail: koukal@iwi.uni-hannover.de

M. H. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_45,

relevant criterion for lending. *Risk sharing* is the allocation of the inherent risks of a project among the most important participants. The objective of the *off-balance sheet financing* is to keep the balance sheets of the project developers largely unchanged in the course of the implementation of the project.

Risk management In order to be able to measure the effects of all risks on the profitability of the project, all relevant risk factors have to be identified, classified and evaluated individually. One key figure which considers the project risks is the value-at-risk (VaR). It indicates the maximum loss in monetary units that is not exceeded within a specific time frame and a specific confidence level. The cash-flow-at-risk (CFaR) applies this method on cash-flows. In order to determine the VaR or the CFaR a Monte Carlo simulation is conducted.

Financial key figures The most important key figure in the context of project financing for lenders is the debt service cover ratio (DSCR) [2]. It is the quotient of the cash flow available for debt service (CFADS) and the debt service (DS) and represents the coverage of debt service for every period of a project.

$$DSCR_t = \frac{CFADS_t}{DS_t} \tag{1}$$

Additional key figures are the loan life cover ratio (LLCR) and the project life cover ratio (PLCR) which are only useful in combination with the DSCR. They are the quotient of future CFADS discounted by the cost of debt (r_D) and the amount of debt (D_{t-1}). The LLCR refers to the CFADS of the remaining loan life, whereas the PLCR refers to all outstanding CFADS of a project.

$$LLCR_t = \frac{\sum_{l=t}^{L} \frac{CFADS_l}{(1+r_{D,l})^l}}{D_{t-1}} \tag{2}$$

$$PLCR_t = \frac{\sum_{p=t}^{P} \frac{CFADS_p}{(1+r_{D,p})^p}}{D_{t-1}} \tag{3}$$

Discounted cash-flow (DCF) method The calculation of the project value and the key figures is performed by discounting the future cash-flows. The adjusted present value (APV) method is a good choice in case the debt-equity ratio is not constant in the course of time [5] which is the case for project financing. The first part of the project value calculation in Eq. 4 represents the present value of all free cash-flows (FCF) and the second part the tax-shield of the project [8].

$$Projectvalue = \sum_{t=1}^{T} \frac{FCF_t}{(1+r_U)^t} + \sum_{t=1}^{T} \frac{\tau * [r_{D,t} * D_{t-1}]}{(1+r_{D,t})^t} \tag{4}$$

The discount factor r_U is the average of the return on equity (r_E) and the costs of debt weighted with the share of equity and debt on the company value (Eq. 5).

$$r_U = r_E * \frac{E}{V} + r_D * \frac{D}{V} \tag{5}$$

The costs of debt are determined by loan agreements, whereas the return on equity has to be determined with the capital asset pricing model (CAPM) (Eq. 6).

$$r_E = r_f + (r_M - r_f) * \beta \tag{6}$$

It is based on the risk free interest rate r_f, the market yield $(r_M - r_f)$ which includes the market interest rate r_M and the beta. The latter implies the systematic risk of the investment compared to risks on general markets.

3 Case Study: Offshore Wind Park in Germany

The case study is based on a fictitious offshore wind park in the German North Sea. The assumptions about the characteristics of the wind park are based on projects which are in operation or currently being planned.

Characteristics of the wind park The offshore wind park consists of 80 wind energy plants with a nominal power output of 5 MW each. It is built with a distance of 90 km to the coastline at a water depth of 40 m. By achieving 3,850 full load hours after technical unavailability the annual energy output results in 1,540 GWh [7]. The compensation for the produced energy is initially based on the compression model imposed by the German Renewable Energy Sources Act 2012. The total investment costs amount to 1,440 M € and the initial annual operating cost to 46.2 M €. Both costs are divided into multiple cost components. The breakdown and the amount of the costs are based on analyses of the recent past [1, 3, 4, 7]. An equity ratio of 40 % is assumed for the investment costs so that the amount of debt is 864 M €.

Monte Carlo-simulation For every cost component and other influence factors like the inflation rate or the annual full load hours individual probability distributions are set up to take the risks of the project into consideration. Due to an insufficient database about risks in the offshore wind sector, BetaPERT distributions are used because they only need a minimum, a maximum and a most likely value in order to be completely discribed [6].

Discounted factor The project value is determined by applying the APV method (Eq. 5). The required discount factor results in 10.56 %. It is based on costs of debt of 6.66 % and a return on equity of 16.4 %. The latter is the result of Eq. 6 with the assumptions of a risk free interest rate of 2 %, a market yield of 10 % and a beta of 1.8.

Table 1 Calculation of the expected project value

	$\sum$	2013	2014	2015	2016	2017	...	2035
1. Investment cash-flow		145.0	−575.0	−720.0			...	
2. Operating cash-flow				123.2	245.5	244.5	...	−89.1
2.1 Earnings				146.3	292.6	292.6	...	59.5
2.2 Expenditures				−23.1	−47.1	−48.1	...	−148.7
3. Taxes						5.6	...	
Free cash-flow		−145.0	−575.0	−596.8	245.5	239.0	...	−89.1
Tax-shield			6.7	20.6	21.1	19.8	...	
Discounted free cash-flow	−23.35	−131.2	−470.4	−441.7	164.3	144.7	...	−8.9
Discounted tax-shield	121.9		6.3	18.1	17.4	15.3	...	
Project value	98.5	−131.2	−464.2	−423.6	181.7	160.0	...	−8.9
Cumulative project value		−131.2	−595.3	−1, 018.9	−837.2	−677.2	...	98.5

3.1 Results

All assumptions about the fictitious OWP are combined in a cash-flow model. At first, the free cash-flow is calculated by adding up the operating cash-flow and the investment-cash-flow with simultaneous consideration of taxes, depreciation and amortization. The discounted free cash-flow is the result of applying the discount rate on that cash-flow. The product of tax rate and interest payments is discounted with the costs of debt in order to obtain the discounted tax shield. As shown in Eq. 4, the adding up of both discounted values results in the project value.

The expected results of the cash-flow model are presented in Table 1. They allow some statements about the OWP:

1. The project value of 98.5 M € is positive and therefore investors achieve a return on equity which is bigger than the required 16.4 % of the discount factor.
2. The calculated internal rate of return (IRR) which represents the discount factor that results in a project value of zero is 13.29 %.
3. The cumulative project value points out at which time the project turns into a positive investment. This is the case after 13 years in the year 2026.

The Monte Carlo-simulation is applied on the cash-flow model. The consequences on the project value are presented in Fig. 1 and allow different statements:

1. At a confidence level of 95 % the project value is −34.9 M € ($CFaR_{95\ \%} = -34.9$). Therefore, with a certainty of 95 %, the value of the project is at least at this level.
2. The calculation of the IRR for the $CFaR_{95\ \%}$ can be made analogous to point 2 of the considerations about the expected project value. In this case the IRR is at least 9.81 % with a certainty of 95 %.
3. With a probability of 87.8 % the project has a minimum value of 0. Investors get an interest yield that is at the same level as other investments with similar risk.

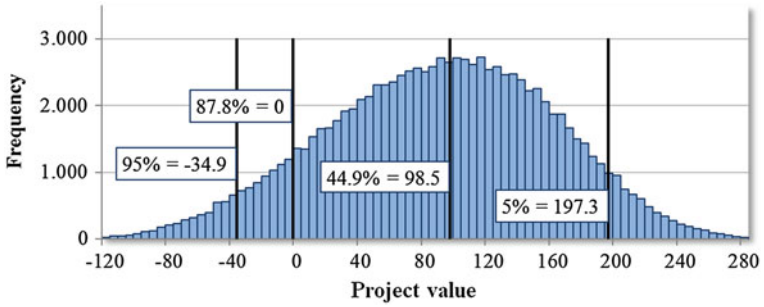


Fig. 1 Distribution of the project value (millions of e , 100,000 simulations)

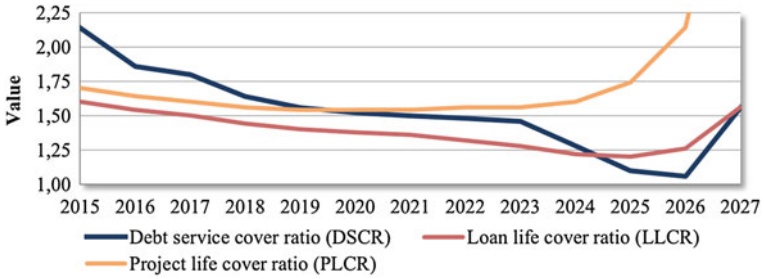


Fig. 2 Key figures at a confidence level (VaR) of 95 %

4. The probability to reach a project value of at least 98.5 M € which corresponds with the expected project value is only 44.9 %.

The Monte Carlo-simulation is also used to take the consequences of all risks on the financial key figures into consideration. The VaR principle is applied to these key figures (Fig. 2). Lenders normally demand a value between 1.35 and 1.45 [4]. Most of the time the DSCR has an uncritical value greater than 1.5. In the years between 2023 and 2027, in which the compensation for the produced energy guaranteed by law will have ended, the DSCR is only over a value of 1. Due to the consideration of the project risks at a 95 % confidence level and values of the LLCR and especially the PLCR which are significantly higher in the previously mentioned critical periods the project cash-flows offer adequate debt service coverage for the entire loan life.

The fictitious OWP represents a typical offshore-project in Germany. Therefore a general answer on the economic perspective of the German offshore wind energy can be given. The investment opportunity of the OWP provides sufficient returns to investors as well as adequate debt service coverage. Therefore, the current feed-in tariff is suitable for supporting the expansion of the offshore wind power in Germany. Due to the special geographical and economic conditions in Germany a transfer of the results to other countries is only possible to a limited extent.

4 Conclusion

The DST gives an answer about the economic efficiency of specific offshore wind projects which are constructed and operated within the context of project financing. The tool is used to perform a case study of an offshore wind park in the German North Sea. The results of the case study clearly indicate that investments into similar projects in Germany are profitable for project developers as well as for lenders.

References

1. BMU: Vorbereitung und Begleitung der Erstellung des Erfahrungsberichtes 2011 gem §65 EEG. Vorhabe IIe Windenergie. 2. Wissenschaftlicher Bericht (2011)
2. Böttcher, J.: Finanzierung von Erneuerbare-Energien-Vorhaben. München, Oldenbourg (2009)
3. EWEA: The Economics of Wind Energy (2009)
4. KPMG: Offshore Wind in Europe. 2010 Market Report (2010)
5. Luehrman, R.: Using APV: a better tool for valuing operations. Harvard Bus. Rev. **75**, 145–154 (1997)
6. Pleguezuelo, R.H., Prez, J.G., Rambaud, S.C.: A note on the reasonableness of PERT hypotheses. Oper. Res. Lett. **31**(1), 60–62 (2003)
7. Prässler, T., Schaechtele, J.: Comparison of the financial attractiveness among prospective offshore wind parks in selected European countries. Energy Policy **45**, 86–101 (2012)
8. Ruback, R.S.: Capital cash flows: a simple approach to valuing risky cash flows. Financ. Manage. **31**(2), 85–103 (2002)

Optimization Strategies for Combined Heat and Power Range Extended Electric Vehicles

H. Juraschka, K. K. T. Thanapalan, L. O. Gusig and G. C. Premier

1 Introduction

Currently available battery electric vehicle (BEV) concepts suffer from reduced energy density compared to internal combustion engine (ICE) vehicles. BEV either have to be designed to limited ranges or are much more heavy and expensive due to size of battery-packs. To attract more customers and thus increase the number of electrical powered vehicles the usability has to be improved. One possibility is to integrate a small internal combustion engine as a range-extender into the concept. These range-extended electrical vehicles (REEV) recharge high-voltage batteries when driving long distances and are a means to reduce the size of the battery. The range of a REEV—like other BEV—is strongly dependent on climate conditions and applied driving cycles. To ensure temperature conditions in passenger compartment and battery-packs a significant amount of energy is needed for thermal management. Depending on vehicle concept, temperature conditions and drive cycle 28–35 % [1] of fuel is used for heating ventilation and air conditioning (HVAC). Thus it is sensible to take improved techniques to generate thermal energy inside the car into account. Solutions like range-extenders using exhaust gases for heating or combined-heat-and-power generation (CHP) range-extender as proposed in [5] can improve efficiency and reduce emissions. To find optimal balance of mechanical, electrical and thermal power, dynamic simulation of concepts at different temperatures is necessary.

H. Juraschka (✉) · L. O. Gusig
Department of Mechanical Engineering,University of Applied Sciences and Arts Hannover, Ricklinger Stadtweg 120, D-30173 Hannover, Germany
e-mail: lars.gusig@fh-hannover.de

K. K. T. Thanapalan · G. C. Premier
Sustainable Environment Research Centre (SERC), University of Glamorgan, Wales, UK

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_46,

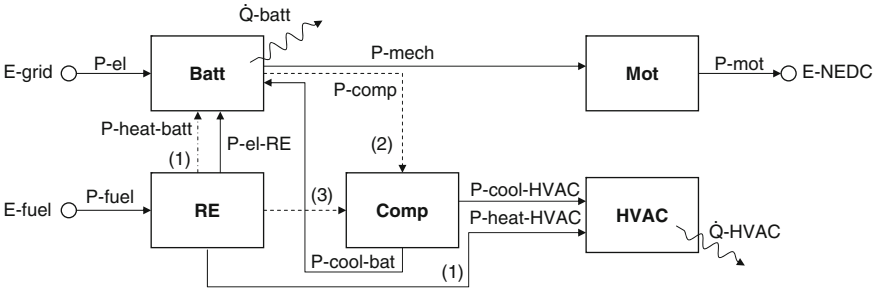


Fig. 1 Integration of alternative range-extender concepts into vehicle power management

2 Vehicle Concepts

Three alternative REEV-concepts have been examined in comparison to conventional BEV and ICE vehicles. A pure range-extender (REEV), an exhaust heat coupled range-extender (EXH-RE) and a combined-heat-and-power range-extender (CHP-RE) with directly linked climate compressor are analysed. Figure 1 describes the energy flow from grid and fuel towards the applied driving cycles and the thermal losses $\dot{Q}_{batt}$ and $\dot{Q}_{HAVC}$. The pure REEV uses only its electrical output power $P_{el,RE}$ to charge the battery, all thermal losses are then powered from the battery. In the EXH-RE concept the exhaust gases of the engine are used to heat battery and passenger compartment (1) in low temperature conditions. For cooling situation the compressor is still powered directly by the battery (2). Only CHP-RE uses a mechanical link from combustion engine to power the climate compressor (3).

For the BEV it is assumed, that all power for propulsion and conditioning is taken form the battery. The ICE which is used for comparison takes all power directly from the engine, exhaust gases for heating are existing in great quantity, the climate compressor is coupled directly to the engine like in the CHP-RE. To generate sensible results for a comparison of these very different concepts masses of components have to be assumed. Table 1 lists differences resulting from alternative sizes of battery packs etc. The data presented here are the calculated values, however it may represent a future vehicle in the Golf-class.

Apart from the weights different component efficiencies have to be taken into account for the three RE-concepts. Efficiency for charging and discharging battery packs is assumed to be independed of its size. The η-ICE for the REEV and EXH-RE are expected to be relatively high according to the constant speed these units can be operated. The CHP-RE is expected to be a smaller unit which possibly could be consisting a rotary engine, thus giving a lower efficiency of 25 %. The heating for REEV is done by electrical heating elements with high efficiency, the EXH-RE and CHP-RE need to take heat transfer from the exhaust gases into account. The efficiency of the compressor, the coefficient of performance, is assumed to be less for REEV and EXH-RE due to electrical energy conversion losses.

Table 1 Masses and efficiencies of components for the analyzed vehicle concepts

	BEV	REEV	EXH-RE	CHP-RE	ICE
Usable traction battery size (kWh)	30	16	16	16	–
Weight traction battery cells (kg)	290	170	150	200	–
Weight bat.packaging/cond. syst. (kg)	90	60	50	70	–
Weight range-extender/engine (kg)	–	50	50	30	100
Weight chassis, motor, compartment etc (kg)	1,100	1,100	1,100	1,100	1,000
Total weight incl 120kg load (driver, luggage) (kg)	1,600	1,500	1,470	1,520	1,220
Efficiencies η-battery / η-ICE (%)	–	95/40	95/40	95/25	–
Efficiencies η-heating / η-cooling (%)	–	98/200	80/200	80/230	–

3 Environmental Conditions and Driving Scenarios

For the comparison of the different vehicle concepts, the environmental temperature and the applied driving cycles have to be defined, as these have significant impact on the performance.

Environment Temperature Passenger compartment and battery-packs shall be operated at defined temperature ranges. In a simplified model all thermal losses to the environment due to higher or lower outside temperature can be calculated from the temperature difference. Figure 2 shows the energy demand for a BEV at different outside temperatures. It can be seen, that the demand nearly doubles when the temperature drops from an ideal value of 20 °C to an extreme cold conditions at −40 °C. Although many data about temperature, humidity and solar radiation are published [3], to reach realistic results a specific temperature distribution has to be taken into account. Figure 2 shows the number of days within a certain temperature band in Hannover based on measurements for the year 2010. For each day in that year the minimum, maximum and average temperature were used for further calculations.

Required Range and Drive Cycle As with the temperatures an averaging of drive cycle data would produce mis-leading results. Therefore two different scenarios were calculated separately, these are short and long distance scenarios. In a short distance trip scenario it is assumed, that on every day of the year one ride is conducted. Each

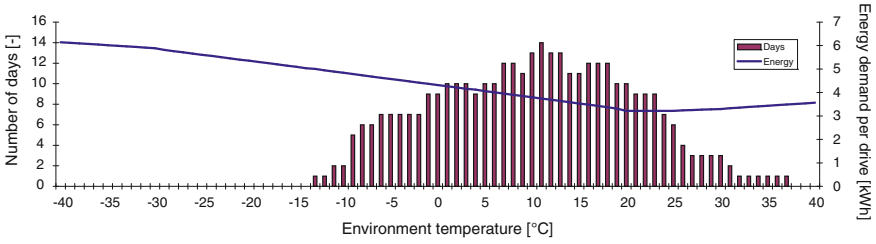


Fig. 2 Temperature distribution and energy demand for a BEV vehicle

consists of a number of New European Driving Cycles (NEDC), whereas each adds four urban driving cycles and one exurban cycle giving 11.02 km. When assuming an overall of 10,000 km per anno each day 2.48 NEDCs are conducted. Thus it is possible to calculated each ride with its corresponding ambient temperature for a whole year. In a long distance trip scenario it is assumed that three rides are conducted every month. These rides consist of a number of 39.94 exurban driving cylces, which gives a trip of 277.8 km. Thus it is possible to calculate one ride at the minimum, maximum and average temperature of every month. In both cases the annual overall kilometers travelled are the same.

4 Dynamic Modeling of Vehicle Concepts

A great number of vehicle models is under investigation within several institutions (e.g. [7, 8]). To enable direct access to modelling parameters and to integrate temperature distribution and vehicle data a simplified model has been programmed in a spreadsheet calculation tool [2]. Basic data for cooling and heating power are presented in [6]. Using macros in spreadsheet-model [2] changes in temperatures, powers and state-of-charge (SOC) as time series for an array of settings has been calculated. In this paper for an illustrative purpose one example is presented in Fig. 3. In this example at SOC of 5 % the RE starts to recharge the HV-batteries. Integrating power values over the drive cycle durations result in one point of energy consumption for that specific value of ambient temperature. Adding up all energy values for the number of days within a year with regards to the temperature distribution at a specific location gives the energy consumption and thus the CO_2-emissions for the short and long driving scenarios for a whole year.

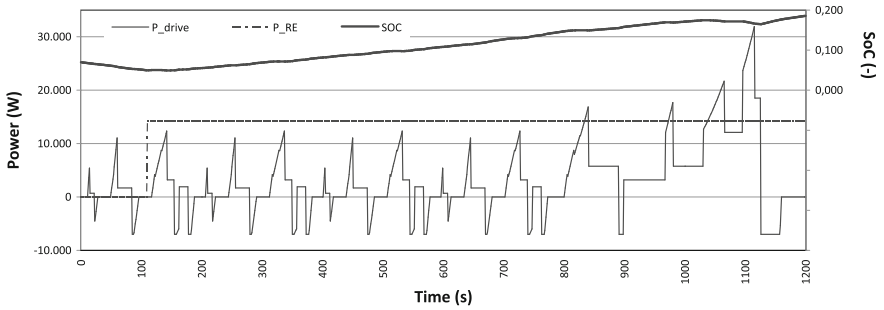


Fig. 3 Power and SOC-development for REEV concept at 0 °C

5 Results

Results for the three RE-concepts under investigation in comparison with the BEV are shown in Fig. 4. It can be seen, that emissions for all RE-concepts are above the BEV-values for short distances. As the BEV is not capable of enduring the long distance scenario only the three RE-concepts can be compared. It can be seen that the EXH-RE can reduce the emissions from 8,774 to 8,190 kg through use of exhaust heat at the cold days of the year. The CHP-RE gives a further reduction to 5,406 kg due to direct coupling of the climate compressor.

The power of the CHP-RE has a significant effect on the emissions. Figure 5 shows the results for different power levels at short and long driving scenarios. It can be seen that an increase of power generally leads to higher emissions due to efficiency losses in the RE itself. But two levels can be identified as optimal for short (s) and long (l) scenarios. For short rides a small power of just 1 kW can be sufficient of delivering just enough energy between the short intervals of recharging at the grid. For long rides a power of 4 kW delivers enough energy for all heating and cooling demands in cold and warm periods over the year.

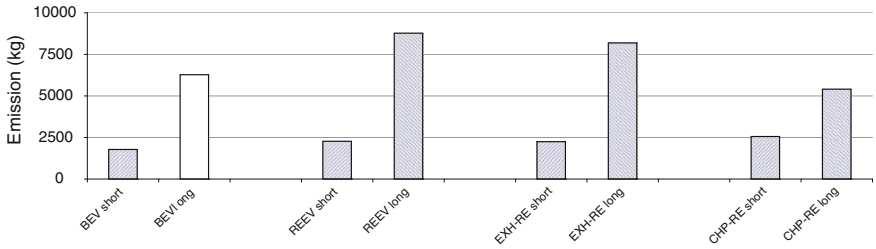


Fig. 4 CO2 emissions of different vehicle concepts at short and long distance scenarios

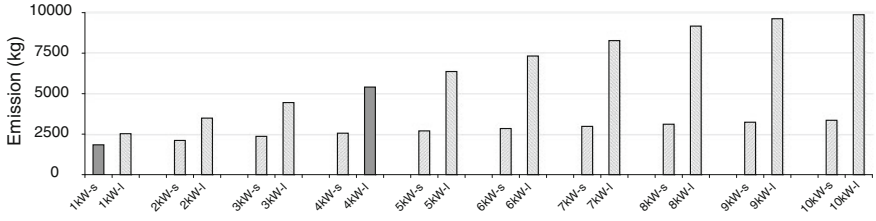


Fig. 5 CO2 emissions of different CHP-power concepts for short and longs distance scenarios

6 Conclusion

Integrating REs in vehicle concepts can reduce battery size, weight and cost while making them capable for typical long distance rides. As the outside temperature has a significant effect on energy demand the temperature distribution of a whole year for a specific location has to be taken into account. Comparing different RE-concepts it can be seen, that a CHP-RE offers a potential of reducing emissions. To maximize this reduction the optimal RE-power has to be identified. It was found that 4 kW can be sufficient for long trips, bigger REs would result in increasing emissions. Thus a CHP-RE with two operating points could be sensible. For further improvement of results in-depth analysis of component sizes and efficiencies shall be conducted theoretically and in experiments.

References

1. Farrington, R., Rugh, J.: Impact of Vehicle Air-Conditioning on Fuel Economy, Tailpipe Emissions, and Electric Vehicle Range. National Renewable Energy Laboratory, in: Earth Technologies Forum Washington (2000), NREL/CP-540-28960
2. Juraschka, H.: Evaluation of Battery-Electric- and Serial-Hybrid-Powertrains for Road Vehicles. Bachelor Thesis, Hochschule Hannover (2012)
3. Koehler, J., Strupp, C., Lemke, N.: Klimatische Daten und Pkw-Nutzung Klima-daten und Nutzungsverhalten zur Auslegung, Versuch und Simulation an Kraftfahrzeug-Kaelte-/ Heizanlagen in Europa, USA, China und Indien. FAT - Schriftenreihe 224, Forschungsvereinigung Automobiltechnik e.V., Berlin (2009)
4. Konz, M., Lemke, N., Foersterling, S., Eghtessad, M.: Spezifische Anforderungen an das Heiz-Klimasystem elektromotorisch angetriebener Fahrzeuge. FAT - Schriftenreihe 233, Forschungsvereinigung Automobiltechnik e.V., Berlin (2011)
5. Laudien, M., Reimann, W.: Hilfsaggregat fuer elektromotorisch betriebene Fahrzeuge. IAV GmbH Ingenieurgesellschaft Auto und Verkehr, Patent DE102009048719A1 (2009)
6. Prokop, G., Lewerenz, P.: Thermomanagement - Loesungen fuer neue und alte Herausforderungen. ATZ **113**(11):812–817 (2011), Jahrgang
7. Roscher, M., Leidholdt, W., Trepte, J.: High efficiency energy management in BEV applications. Int. J. El. Power Energy Syst. **37**(1), 126–130 (2012)
8. Thanapalan, K., Zhang, F., Juraschka, H., Gusig, L., Premier, G., Guwy, A.: Optimal power management of hydrogen fuel cell vehicles. World Renew. Energy Forum May 13–17, 2012 Denver (2012)

Towards a Decision Support System for Real-Time Pricing of Electricity Rates: Design and Application

Cornelius Köpp, Hans-Jörg von Mettenheim and Michael H. Breitner

1 Introduction

Due to the development of renewable energy sources like wind turbines and photovoltaic systems, there is a steadily increasing proportion of uncontrollable variations [2] on the supply side. In contrast to conventional power plants, these energy sources do not produce precisely defined quantities. Electrical energy can be saved only with great losses, unlike, for example gas. Known memory options such as pumped-storage power plants have been built in the past at appropriate places. A further expansion is hardly possible. As the unpredictable flows [10] will continue to increase in future, alternative solutions are required. Not only on producer but also on consumer side. Variable electricity rates can provide customers with an incentive to shift their electricity consumption to times in which, for example, there is a surplus of wind energy.

In the past—and in general still now—electricity rates are constant all over the day. Well known exceptions are night storage heaters with a two-rate electricity meter. This will change in the future: In Germany, for example, since 2011 the energy suppliers have to offer special electricity rates to encourage people to save energy or to shift their power consumption to times of lower consumption. This can be realized by high prices for peak-load times (typically during the day) and low prices for other times (typically during the night). That concept (similar to night storage heaters in the past) is still very static—unlike the electricity production of renewable energies.

C. Köpp (✉) · H.-J. von Metthenheim · M. H. Breitner
Institut für Wirtschaftsinformatik, Leibniz Universität Hannover, Hannover, Germany
e-mail: koepp@iwi.uni-hannover.de

H.-J. von Mettenheim
e-mail: mettenheim@iwi.uni-hannover.de

M. H. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_47,

Short term, variable, dynamic electricity prices allow the reduction of the variance between demand and supply [3]. Real time pricing is considered the most direct and efficient price based approach for demand side management [1]. Basic idea of dynamic rates is to reflect the actual costs of electricity generation [1]. Following [7] the electricity rates will be higher when there is high demand, and lower when there is low demand. Mohsenian-Rad et al. [8] assume that the load is proportional to the costs. This is suitable for a constant generation, but not for an integration of fluctuating renewable energy sources. For proper load shifting, Mohsenian-Rad et al. [8] propose a rate model independent of actual energy costs.

Typical devices listed as suitable for load management are dishwashers, washing machines, dryers, refrigerators, freezers and heaters with storage. Other devices like night storage heaters, air conditioners, heat pumps and batteries, especially in plug-in hybrid electrical vehicles are also mentioned; see for example [1, 3, 5, 7–9]. Many concepts include still a manual operation of the devices by the user. This leads to a significant loss of comfort and requires the presence of the user. In addition, the user must actively inform about the current and future prices.

There is as potential for controlling more than 10 % of electricity [6] in the long run. This can be used to (partly) compensate the fluctuations of renewable generation. To create the desired consumption curve there is the need of an pricing signal. This is a challanging task, as the load not only depends on the price signal for the same time. A decision support system (DSS) should support the energy provider in setting the appropriate pricing signal.

2 Real-Time Pricing Model for Consumers and Reaction

A pricing model must be a balance between the interests of suppliers (maximal controllability) and customers (understandable and affordable). Only a fair contract will convince customers to participate. Our pricing model is straightforward. It is the result of discussions with experts, and will also be used in a field experiment to examine the consumers reaction to RTP. We use price levels of {14, 17, 20, 23, 26, 29, 32} Ct/kWh with a fixed gap of 3 Ct/kWh. The minimum rate of 14 Ct/kWh (much lower than current prices in Germany) offers a saving potential as incentive for consumers. Using seven price levels makes control more precise than is available with rate models with only three steps. The rate is fixed six hours in advance. The preview period of six hours provides planning reliability to the consumer. It enables predictive scheduling of smart appliances, as well as manual device control by the consumer. The electricity supplier can use a more exact short time forecast of wind and solar energy production for price setting. The rate is defined for each hour, as hourly prices are easy to understand. A constant price for at least two hours prevents price changes from becoming too frequent and leaves the option of manual device control by the user open.

As there are presently no smart devices available (see for example [4]), we choose a simulation-based approach for expected behavior. Based on the assumption that

participants in price-incentive programs are interested in minimizing the cost of electricity without losing comfort, we expect the following scheduling strategy as a reaction to price signals:

1. Schedule the execution of device to the latest cheapest start time, within the preview period and before the deadline for the latest start.
2. Wait until the next price signal is received.
3. If the execution has not started yet, check a reschedule by starting again with 1.

This strategy satisfies the following properties: An optimal solution is guaranteed when the time window is shorter than the preview period. There is a guarantee that execution will occur before the deadline. It prevents users from missing the cheapest time at the beginning, but as a drawback it also results in the risk of missing a better price in the time after the preview period. By always using the latest cheapest start time, we reduce this risk. The strategy could be used by integrated device controllers, external control boxes, and humans controlling the devices manually.

We implemented the rate model and the device control algorithm as a software prototype in Java. Based on the price model, the simulation represents the period of 90 full days using a resolution of one minute. The price curves according to the model are randomly generated. We simulate an average of 10,000 activated devices per day. Each device is defined by a time window for scheduling of four to twelve hours, a runtime of 15–150 min, and a constant power consumption of 1,000 W. The device parameters are set randomly. The simulation returns the aggregated loads for uninfluenced (devices will start at once), realized (devices use the presented strategy) and optimal (devices know the full price curve in advance) scheduling.

3 Getting the Next Price Signal Using Neural Networks

The realized load at a certain hour depends not only on the price at the same hour, but also on the prices from the previous and following hours. So setting the price for a certain hour also affects the load in the previous and following hours. Our pricing model induces on very simple case: the next price signal must be set to be equal to the last price signal, if the last price signal differs from the second to last price signal.

Figure 1 show the data available at the time of decision. At a certain hour, the DSS knows all prices of the past hours, the price of the current hour, and the price of the following six hours of the preview period. Historical load data is available only with a certain delay. This delay is caused by data measurement, transmission, and processing. We assume a fast provisioning of consumption data. The consumption data for the current hour is available as a linear extrapolation from the first 45 min of current hour. The decision will be made within the last 15 min of the current hour. As the price is set for full hours only, all data used for the decision will be aggregated to an hourly base.

The price signal depends on the desired load, so this is required as an external input. Minimizing the need of balancing energy determines considers the short-term

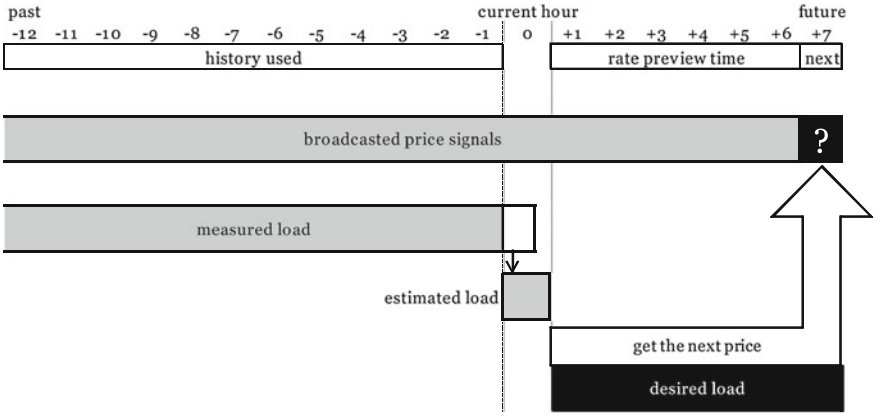


Fig. 1 Available data for getting the next price. Fixed values are hightlighted with *gray background*, variable inputs and the next price as output are hightlighted with *black background*

forecasts of volatile renewable energy sources. The short forecast horizon of slightly more than seven hours is better suited than a one-day ahead forecast. The desired load profile starting at the second hour is not fixed permanently. In the following hours, the DSS will replace the desired load with a refined version based on newer forecasts.

We approximate the desired relationship using a three-layer neural network. We select approximately 30 % of available data for validation and control network quality with early stopping. The following time-series defines 58 inputs to the network: Price of past 12 h, current hour, and future six hours (price signal already sent) coded as absolute price and difference to the previous hour. Load history for 12 h past. Load extrapolation for current hour. Desired load for seven hours future. The network has a single output: Absolute price of seventh hour in the future (after price preview period).

We select the best five networks from neural network training, according to the errors on the training set. We aggregate the result by computing the mean of the five ensemble members. This further helps to smooth out possible outliers.

We calculate the prices from evaluation on a separate out-of-sample data set of three weeks. Price follows the target signal satisfactorily. The ensemble generates a hitrate of up to 93 % if we allow a fluctuation of the price around the target price. More important than the absolute value of the price signal is the local rank in comparison to neighboring prices, and hitting the point in time for price rise. As the neural networks return continuous results, we have to map these raw values to valid price levels and ensure a constant price for at least two hours:

$$price\,Valid\;\;Step_t = \text{round}((\max(14, \min(rawValue_t, 32)) - 14)/3) * 3 + 14$$

$$price_t = \begin{cases} price\,Valid\;\;Step_{t-1} & \text{if} \quad price\,Valid\;\;Step_{t-1} \neq price\,Valid\;\;Step_{t-2} \\ price\,Valid\;\;Step_t & \text{else} \end{cases}$$

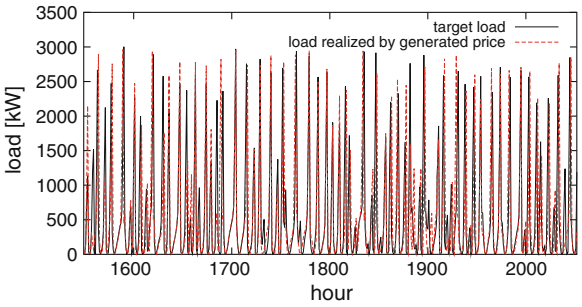


Fig. 2 Load realized by generated price compated to target load

Adjusting raw prices to the pricing model reduces the above hitrate to 85 %. The DSS produces an acceptable price for the presented rate model. The shape fitting of price curve is reduced by the post processing, but the result is still a good representation of the local price ranking and changes. As explained before, we are less interested in an exact match of prices. Rather, price ranks (lower or higher) at specific point in times drive device behavior. As next step in evaluation, we combine the DSS with the simulation. Prices are generated to fit the hourly load defined by the evaluation data set. The result is shown in Fig. 2. There is good matching between target load and simulated realization by generated price.

4 Conclusions and Outlook

We present first steps toward a DSS for real-time pricing of electricity rates. In the context of the current simulation our neural network ensemble is capable of producing adequate price signals to fit a given load profile. The DSS use aggregate load information only and does not need individual upstream information from any device. This reduces the technical infrastructure that we need and protect the users privacy. A prototypical realisation will be evaluated in a field experiment. The neural network approach is flexible: By continuous retraining on price signals and measured data, we can gradually replace the simulation with real data. Splitting the aggregate into sub aggregates could extend the concept with the option of localized control.

We focused on demand side management of private consumers only. Several concepts on consumer side, including the management of industrial consumption are ignored. Our DSS can be one of many solutions towards an increase in share of renewable electricity. We argue that individual solutions can only have a limited effect. Rather, a combination of different methods is necessary to exploit the full potential of renewable energies.

References

1. Albadi, M.H., El-Saadany, E.F.: Demand response in electricity markets: An overview. In: Proceedings of Power Engineering Society General Meeting (2007). doi: 10.1109/PES.2007.385728
2. Cardell, J.B., Anderson, C.L.: Analysis of the system costs of wind variability through Monte Carlo simulation. In: Proceedings of the 43rd Hawaii International Conference on System Sciences (2010)
3. Chen, C., Kishore, S., Snyder, L.V.: An innovative RTP-based residential power scheduling scheme for smart grids. In: Proceedings of the IEEE Conference on Acoustics, Speech, and, Signal Processing (2011)
4. Hauttekeete, L., Stragier, J., Haerick, W., and De Marez, L.: Smart, smarter, smartest... the consumer meets the smart electrical grid. In: Proceedings of the 9th Conference on Telecommunications Internet and Media Techno Economics (CTTE), Ghent, Belgium (2010)
5. Köpp, C., von Mettenheim, H.-J., Klages, M., Breitner, M.H.: Analysis of electrical load balancing by simulation and neural network forecast. In: Operations Research Proceedings 2010, pp. 519–524 (2010). doi:10.1007/978-3-642-20009-0_82
6. Köpp, C., von Mettenheim, H.-J., Breitner, M.H.: Price-induced load-balancing at consumer households for smart devices. In: Operations Research Proceedings 2011 (in print)
7. Lee, J., Jung, D.K., Kim, Y., Lee, Y.W., Kim, Y.M.: Smart grid solutions, services, and business models focused on telco. In: Proceedings of Network Operations and Management Symposium Workshops (2010). doi: 10.1109/NOMSW.2010.5486554
8. Mohsenian-Rad, A.H., Wong, V.W.S., Jatskevich, J., Schober, R.: Optimal and autonomous incentive-based energy consumption scheduling algorithm for smart grid. In: Proceedings of Innovative Smart Grid Technologies (2010). doi: 10.1109/ISGT.2010.5434752
9. Molderink, A., Bakker, V., Bosman, M. G.C., Hurink, J., Smit, G. J.M.: A three-step methodology to improve domestic energy efficiency. In: Proceedings of the Innovative Smart Grid Technologies (2010). doi: 10.1109/ISGT.2010.5434731
10. Sioshansi, R.: Evaluating the impacts of real-time pricing on the cost and value of wind generation. In. IEEE Transactions on Power Systems (2010). doi:10.1109/TPWRS.2009.2032552

100 % Renewable Fuel in Germany 2050: A Quantitative Scenario Analysis

Maria-Isabella Eickenjäger and Michael H. Breitner

1 Introduction

The use of fossil or nuclear energy results in climate change, radioactive pollution and the rise of energy prices. These outcomes result from the finiteness of fossil energy, the ever-increasing energy demand and the resulting increase in energy production effort. To ensure survival and high quality of life, also of future generations, the energy-provision should take the following criteria into account: energy has to be available and in a environmentally conscious manner, appropriately used and converted efficiently [1]. One of the largest energy consumers in Germany is the transport sector, which accounts for around 30 % [2] of the country's total consumption. An energy source change does not only occur because of declining public acceptance, but is also supported by politics. That leads to the question, which renewable energy sources meet the required energy demand, so that a sustainable change can take place? The question of the impact of different climate policies of the GHG emission has been addressed in [3]. STEEDS is a DSS that supports multi-criteria decision aid for transport energy problems, able to assist global policy-making until 2020 [4]. Which policies should be used in Germany to obtain 100 % renewable fuel use, while ensuring availability of energy for the transport sector as well as economic and environmental change?

The remainder of this paper is structured as follows: Section 2 formulates the model and describes the application of the model. Some limitations are shown in the discussion in Sect. 3. Finally, Sect. 4 concludes, and further suggestions are offered.

M.-I. Eickenjäger (✉) · M. H. Breitner
Institut für Wirtschaftsinformatik, Leibniz Universität Hannover, Hannover, Germany
e-mail: eickenjaeger@iwi.uni-hannover.de

M. H. Breitner
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_48,

2 Model

Model Formulation Our Objective is to determine the single net amounts of fuel of the respective means of transport. The amount of the different fuel alternatives corresponds to either the surface capacity, production capacity or capacity of the vehicle. In this section a mathematical formula for the German fuel market has been developed: The $costs_{t,net}$ describes the cost of fuel i in year t. Furthermore $quantity^{i,j,k}_{t,net}$ represents the amount of fuel i, usable form engine type k and transport sector j in year t. For this purpose an engine k is limited to certain fuels i:

$$\text{quantity}^{i,j,k}_{t,\text{net}} = \begin{cases} \text{quantity}^{i,j,k}_{t,\text{net}}, & \text{if fuel } i \text{ can be used by engine } k, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The following model is derived:

$$\min \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \text{cost}^{i}_{t,\text{net}} \cdot \text{quantity}^{i,j,k}_{t,\text{net}} \tag{2}$$

$$s.t.: \sum_{i=1}^{I} \sum_{k=1}^{K} \text{quantity}^{i,j,k}_{t,\text{net}} \geq \text{total consumption}^{j}_{t,\text{net}} \; \forall j \tag{3}$$

$$\text{quantity}^{i,j,k}_{t,\text{net}} \geq 0 \; \forall i, \, j, \, k \tag{4}$$

$$\sum_{i=m+1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\text{quantity}^{i,j,k}_{t,\text{net}}}{\text{efficiency}^{k}_{t} \cdot \text{net energy yield}^{i}_{t}} \leq \text{field-capacity}_{t} \tag{5}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\text{quantity}^{i,j,k}_{t,\text{net}}}{\text{efficiency}^{k}_{t} \cdot \text{net energy yield}^{i}_{t}} \leq \text{oil-capacity}_{t} \tag{6}$$

$$\sum_{j=1}^{J} \sum_{k=1}^{K} \text{quantity}^{i,j,k}_{t,\text{net}} \leq \text{production capacity}^{i}_{t,\text{net}} \; \forall i \tag{7}$$

$$\sum_{j=1}^{J} \sum_{k=1}^{K} \text{quantity}^{i,j,k}_{t,\text{net}} \leq \text{import share}^{i}_{t,\text{net}} \; \forall i \tag{8}$$

$$\sum_{i=1}^{I} \text{quantity}^{i,j,k}_{t,\text{net}} \leq \text{vehicle-capacity}^{j,k}_{t,\text{net}} \; \forall j, k \tag{9}$$

As the fuel demand will be met as cheap as possible, the goal for each year is the cost-optimized distribution of the fuel supply to fuel demand. The objective function for optimization of a simulation year is represented by (2). The constraints (3) guarantee that the sum of all $quantities_{t,net}^{i,j,k}$, which are used in a year t by the transport sector j must comply with the $total\ consumption_{t,net}^{j}$ of the transport sector j in year t. Since the used fuel quantities cannot be negative the amounts of fuel is limited by the constraint (4) to positive values. To perform the optimization, it is also necessary to calculate the optimization-determined factors first. These are firstly, the various capacity constraints and the fuel price. The determination of capacity restrictions is, as described below: To represent the dependencies of the fossil fuel quantities, these are without loss of generality (WLOG) the fuels $i = 1$ to m, the oil resources and the dependencies of the renewable fuel quantities, these are WLOG the fuel $i = m + 1$ to I of the field capacity, it is necessary to separate the direct connection between the production capacity. (5), (6), (7), (8) and (9) reflect the capacity constraints.Eq. (5) is based on renewable fuels and it indicates that the total renewable fuel volumes, each of which is adjusted for the conversion losses of the provision and use of all transport sectors and fuels must not be larger than the field of capacity. On this occasion $efficiency_t^k$, is the efficiency of an engine k in in year t, $net\ energy\ yield_t^i$ is the net energy output of the fuel i in t and the $field\ capacity_t$, field in year t available for cultivation of renewable fuels. Constrain (6) restrict the fossil fuel quantities: The sum of fossil fuel, each of which is adjusted for the conversion loss, over all transport sectors and fuels must not be greater than $oil\ capacity_t$, oil capacity of the available oil reserves in year t. Furthermore inequality (7) require that the sum of fuel of the transport sectors is not greater than the capacity of the fuel i in year t, $production\ capacity_{t,net}^{i}$. The user-dependent specification of the import share of a fuel i in year t $import\ share_{t,net}^{i}$ is used as a restriction. The imported fuel is thus not restricted in the model by the size of the land nor by the number of production facilities. Therefore inequality (8) states that the capacity of the imported fuel corresponds to the scenario dependent import share. Inequality (9) states that the quantity of fuels of a transportation sector, which can be used by the number of vehicles with a suitable motor k must not be greater than their capacity, thus the vehicle-capacity $vehicle\ capacity_{t,net}^{j,k}$.

Model Application The effects of parameters energy tax exemptions, penalties on the emission of carbon-dioxide-equivalents and major restrictions for renewable fuels are examined for the following: gasoline, diesel, kerosene, electricity (fossil), bio-ethanol from wheat, bio-ethanol from sugar beet, bio-ethanol from lignocellulosic biomass, bio-ethanol from sugar cane, biodiesel, vegetable oil, dimethlyether, biomass-to-liquid, biogas, bio-kerosene, renewable electricity. The model was iterative solved by 2050. The implementation of the model is done in Excel and the optimization is automated using VBA. An example of the output result of a scenario in the user interface is shown in Fig. 1.

Considering the different scenario results it can be concluded that the political measure of the energy tax exemption have been used as a cornerstone for sustainable change. However, it is important to insure that this goal will continually followed.

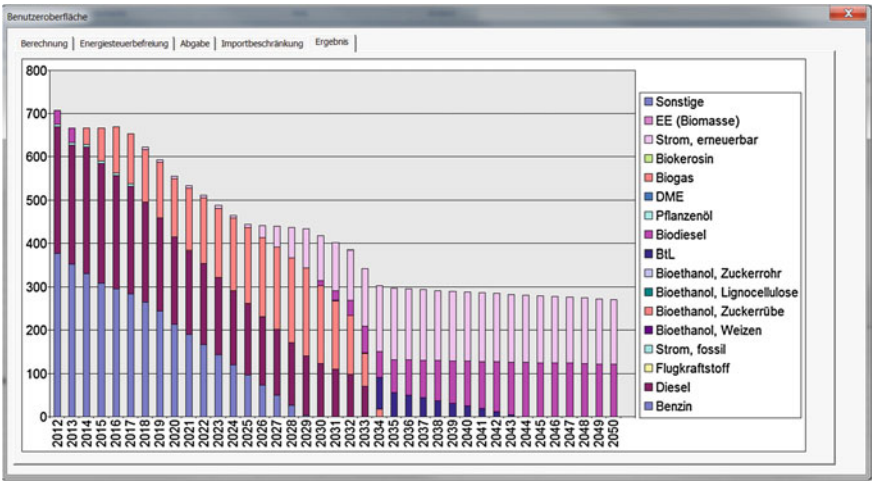


Fig. 1 Example for the representation of the result of the user interface.

Scenarios in which the energy tax exemption will continue until 2050 show the best results. Furthermore, the scenarios show that the government also introduced penalties help to reduce emissions. This supports both the provision and use of more efficient and lower-emission renewable fuel alternatives. A lesser penalty should be selected to reduce the pressure on the consumer. If energy tax exemption were to be reduced more rapidly, then a penalty should also be introduced. The rapid reduction of energy tax exemption up to the year 2039 showed poor results, which almost correspond with the results, which would be no further intervened by the state. In this case it is necessary to introduce higher penalty so that the results are satisfactory. Otherwise, this will always lead to a substitution of the same requirements for supplies of fossil fuel and thus to a lower success in terms of sustainability. In the event that energy taxes are risen in 2016 for all fuels to 100 %, then renewably produced electricity for the transport sector should be additionally supported. Only then will it be possible to achieve a basis for a sustainable change in the mid-term. However, policies should be broader and relate not only to one single renewable alternative in order to avoid supply uncertainty. Since alternatives not have to be obtained i.e. cheaper through import or in cases of a poor harvest resulting in insufficient supplies to meet the demand. The results of the scenarios with imports, which are initially low in price and then become very expensive are not realistic, because consumers take more time for the decision to use this fuel However, these results also show that it is necessary to correctly communicated to consumers, in order to avoid uncertainty and to simultaneously increase the acceptance.

3 Discussion

The input of parameters allows a simulation of the effects of government measures. The results of the simulation show the distribution and use of fuels in different transport sectors as well as the price of fuel and the consumption capacity for each year. These results allow statements for various input parameters, which relate to the change in fuel prices, the required capacity and the possibility of replacing fossil fuels with renewable fuels. Please note; the statements only refer to the specified reference fuel composition. It should also be kept in mind that the model calculates the total fuel quantities of each year. Some results of the fuel quantities for vehicles equipped with the same engine type are not realistic, i.e. diesel and biodiesel. The reasons for this are the existing blending of biodiesel with fossil diesel and that in the model, the total requirement of one year should be covered and not the daily requirement. Therefore, these results do not correspond to reality, because i.e. the entire capacity of biodiesel used to meet the demand of road transport and thus the demand of other transport sectors cannot be covered. In addition, the model only considers the respectively monetary impact on the fuel market. Other aspects, such as corporate strategies are largely neglected. Therefore, the following assumptions are considered: The model does not consider the human factor. It is possible that the public are more influenced by their routine or through ignorance of a fuel such as hydrogen, continue to purchase vehicles which run on fuels with which they have experience. It is also possible that, i.e. by rapidly rising fuel prices in the future there could be a change of mobility preferences, people may use more public transport, resulting in a decrease in fuel consumption and a shift of the modal sectors in total consumption. It should also be noted that in the model the refueling cycles and the infrastructure are not observed. Substitute fuels will not be considered by customers, if they find a lack of nationwide retail network. It is also possible that vehicle efficiency improvements and measures for energy recovery systems will lead to lower consumption rates. In addition to the purchase price of vehicles and personal transportation energy demand of the user, a possible new registration of passenger cars, which use different fuels, must be considered. This is necessary because people who have high mileage may choose different vehicles then those who drive less. Governments should note that a price difference of fuels must be avoided if necessary through subsidization Also the life expectancy of vehicles could increase and thus the replacement of one fuel by another may be questionable.

4 Conclusions and Directions for Further Work

For the creation of scenarios various parameters were modified: energy tax exemptions, penalties on the emission of carbon-dioxide-equivalents and major restrictions for renewable fuels. Parameter dependent results show optimized fuel combinations for every year up to 2050. Results from scenario analyses show further that it is

possible to increase the share of renewable fuel consumption in total up to 100 % by 2050. This can be achieved by constant mobility preferences and by more efficient vehicles. Furthermore, it is necessary to support the substitution of fossil fuels with bio-fuels, hydrogen and renewable electricity by using policies for a sustainable change. Simulations show that long-term energy tax exemptions with long-term taxes on emissions of CDE lead to significantly better results.

Although it is possible to transfer the results of the scenarios on fuels with similar properties and requirements, it should be considered that these refer to the German fuel market. Different factors, such as population or space availability, prevent utilizing the scenarios without approximation for other countries. Also it is necessary to maintain the data for further simulations to ensure valid results. At the time of the simulation various parameters could not be documented as being valid. An example currently there is no valid data on the production capacity of renewable fuels, because there are few standardized production methods. Many are currently under development and have to be tested. In order to extend the model and the Excel tool, these conclusions together which Sect. 3, can be used as a basis for further work. In Future new fuels, competing factors or changes in production processes may lead to deviating results.

References

1. Nitsch J.: Ökologisch optimierter Ausbau der Nutzung erneuerbarer Energien in Deutschland. (2004) Available from: http://www.ifeu.de/landwirtschaft/pdf/Oekologisch_optimierter_Ausbau_Langfassung.pdf Cited 26 Mar 2012
2. Radke, S.: Verkehr in Zahlen 2010/2011. DVV-Media Group GmbH, Hamburg (2011)
3. Creutzig, F., McGlynn, E., Minx, J., Edenhofer, O.: Climate policies for road transport revisited (I): Evaluation of the current frameworks. Energy Policy **39**, 2396–2406 (2011)
4. Brand, C., Mattarelli, M., Moon, D.: Wolfer Calvo, R.: STEEDS: A strategic transport-energy-environmet decision support. Eur. J. Oper. Res. **139**, 416–435 (2002)

Multi-objective Planning of Large-Scale Photovoltaic Power Plants

M. Bischoff, H. Ewe, K. Plociennik and I. Schüle

1 Introduction

Photovoltaic (PV) power plants play a decisive role in switching the global energy supply from fossil to renewable energies [1]. Compared to typical roof-top PV installations, it is a complex task to design the layout of a large-scale power plant due to a variety of free optimization parameters, many interdependent goals, and rather complex design principles [2]. Without further support, it is very hard for the responsible engineers to estimate the impact of their design decisions on the plant's overall technical and economic efficiency. While this is already the case with a single objective function to be optimized, the situation gets even more difficult when trying to balance the different objectives, of which many are contradicting on top of that.

In this paper, we present the problem of designing a large-scale PV power plant and describe our solution approach: We provide the engineer with a multitude of reasonable plant layouts, each having its own benefits and drawbacks for the different objectives. For each layout, we determine a set of key performance indicators (KPIs) using detailed financial calculations and a simulation of the plant's performance. Here, the topography and weather data for the area on which the plant is to be built are considered [3]. Based on these KPIs, the engineer can choose the best layout in a multicriterial decision support system. We illustrate the superiority of our multi-objective approach compared to single-objective methods with the help of some examples and report on the practical usage of our software tool at Siemens Energy.

M. Bischoff
Siemens AG, Energy Sector, Hugo-Junkers-Str. 15-17, Nürnberg 90411, Germany
e-mail: martin.bischoff@siemens.com

H. Ewe · K. Plociennik · I. Schüle(✉)
Fraunhofer ITWM, Fraunhofer Platz 1, Kaiserslautern 67663, Germany
e-mail: ingmar.schuele@itwm.fraunhofer.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_49,

2 Problem Description

In a large-scale PV plant, PV *modules* convert the sun's irradiation into continuous electrical current which is converted into alternating current by inverters. The voltage is then increased by transformers, and finally the electrical power is fed into the grid. Given the outline and topography of an area on which a customer wants to build a plant, the engineer has to find the best possible layout. This includes choosing the right module and inverter types and defining the number and positions of the components like the service and cable ways, the so called *tables* carrying the PV modules, and the inverters. Also, the tables have to be assigned to the inverters. Finally, the cables' routing has to be computed and cross sections have to be chosen to meet certain safety standards and to limit cable losses. What "the best" layout is depends on different aspects like the customer's preferences regarding the technical and financial KPIs, the latitude, topography, and typical weather conditions of the area, legal restrictions, feed-in tariffs, and subsidies. Here, several objectives have to be balanced, some of which are contradicting. For example, the engineers want to minimize construction costs and maximize energy yield, but often a higher yield demands for higher costs.

2.1 Degrees of Freedom and Optimization Goals

There are many degrees of freedom in the design of a plant layout, and the best choice for a parameter strongly depends on the values chosen for the other parameters and, as mentioned, on environmental and financial conditions and the customer's preferences. We now give some examples.

Increasing the number of service ways may decrease maintenance costs due to better accessability but may decrease the energy yield. Since more space in the area is covered by ways, the available area for locating tables is reduced. Interestingly, due to the complex design principles, increasing the number of ways for some areas may even increase the number of placeable tables. Another optimization parameter is the so called *row distance* of the plant. Figure 1 shows two layouts for the same area with different distances of the table rows.

In the left layout there is a larger amount of mutual table shading, due to the smaller row distance. Consequently, a simulation for the typical weather conditions of the area over one year shows that this layout has a smaller energy yield per table than the right layout. But the larger number of placed tables compensates for this: The left layout has a yield of 7.6 GWh while the right layout's yield is only 6.6 GWh. However, due to the increased component counts, the left layout's construction costs are higher than for the right one. It now depends on whether the customer is willing to spend more money for building the plant to achieve a higher energy yield. Further examples of degrees of freedom are the *tilt angle* by which the tables are inclined, the module and inverter types, the number and placement of the tables and inverters, the assignment of the tables to the inverters, and the cable routing and cross sections.

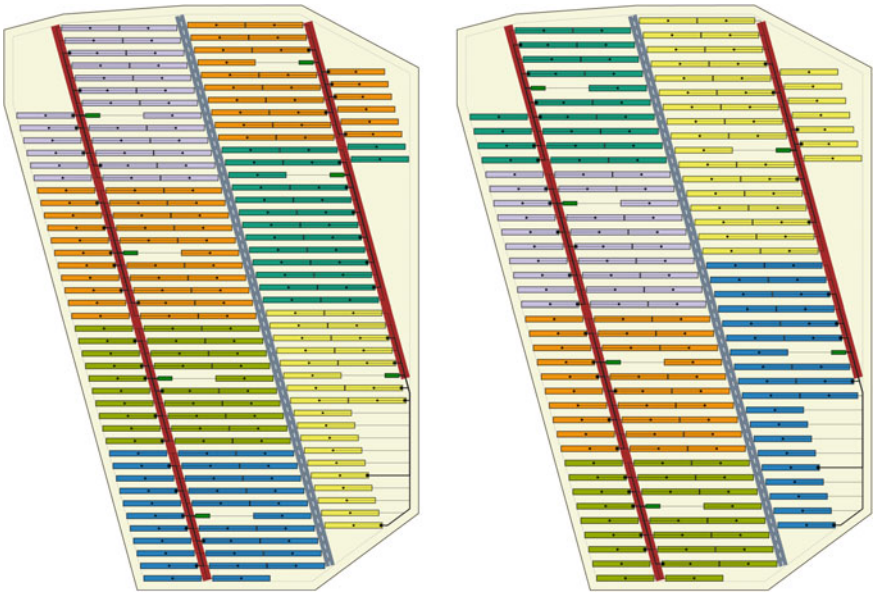


Fig. 1 Two plant layouts with small (*left*) and large (*right*) row distance. The tables are shown as differently *colored rectangles*, tables belonging to the same inverter having the same color. Inverters are drawn as *smaller rectangles* close to the cable ways. Cables are shown as *black lines*

As mentioned, the definition of a "good" plant layout strongly depends on the customer's preferences and several other environmental and financial conditions. Some typical KPIs [4, 5] with their respective optimization goals (maximize or minimize) are: Number of tables (max for high peak power), number of inverters (min for low costs), peak power (max), energy yield (max), yield per installed table (max), construction costs (min), operation and maintenance costs (min) and levelized cost of energy (LCOE, the costs for producing one kWh of energy over the plant's lifetime) (min). The engineer now has to find a layout which balances out the KPIs regarding the customer's preferences. Additional complexity is added by the fact that sometimes one does not want to maximize or minimize a certain KPI but wants to achieve a certain value, e.g., the customer might want to exactly meet a specific peak power due to legal contracts. Furthermore, factors which are not easily expressible as KPIs have to be considered. For example, a layout with well-balanced KPIs might still not be the best choice in the end since the plant is hard to build for some reason, e.g., it might have a too high wiring complexity. Also, esthetical aspects might play a role for the customer.

3 Solution Approach

We now describe our solution approach to handle the complex tasks described in the previous section. Due to the large number of degrees of freedom and different, partially contradicting optimization goals, one cannot expect to find the best possible plant layout by formulating and solving one single optimization problem, e.g., by formulating one large mixed integer program. Such an approach would not be computationally manageable. Also, it would be very complex to incorporate changes in design principles or new module and inverter technologies. Furthermore, approaches computing a single solution will not succeed in finding the best layout, since the engineer cannot a priori know the "right balance" of the KPIs. The one single solution computed by such an approach will not have the best balance of the KPIs, and it will not be clear by how far the individual KPIs can be improved and which dependencies exist between the KPIs.

Consequently, we rather generate a multitude of reasonable plant layouts for the engineer, from which he or she can determine the most suitable layout using an interactive multicriterial decision support system. We decompose the problem into several subproblems which are solved with specialized fast heuristics. For each subproblem, we determine not only one single solution, but consider a multitude of reasonable solutions instead. This way, we generate a tree of partial solutions, the leafs of which being complete solutions, i.e., layouts. The set of all generated solutions is fed into our multicriterial decision support tool. A screenshot of the graphical user interface (GUI) is shown in Fig. 2.

The GUI consists of two components. The first one is a two-dimensional plot, in which the axes can be chosen from any of the computed KPIs. The second component is a set of sliders that can be used to enable or disable solutions by restricting some of the KPIs. For example, in Fig. 2 the construction costs (CAPEX) have been restricted to at most 7.5 million Euro, and the number of ways has been restricted to be exactly 3. Plants which do not comply with these restrictions are grayed out (inactive) in the plot. Now, iteratively using the sliders and choosing different axes for the plot, the engineer can navigate on the set of all generated solutions, determining a small number of interesting solutions with well-balanced KPIs and optimization parameters. For each of these solutions, the engineer can show a preview of the layout as in Fig. 1, with the opportunity to edit the layout in a graphical way, e.g., remove or add new inverters and change the assignment of tables to the inverters. The KPIs and optimization parameters shown in the multicriterial decision support system (Fig. 2) are then updated accordingly.

We now give some details on our decomposition of the layout problem into subproblems. We use three subproblems as follows. Firstly, we generate placements for the ways, which can vary in the number of ways, orientation, way width, the minimum and maximum number of table columns between consecutive ways (set to two columns in the example in Fig. 1), and the actual positions of the ways in the area. Secondly, we place the tables in the area between the already placed ways. This placement can, among other parameters, vary in the tilt angle of the tables and

the individual row distances, which are chosen topography-dependent (on hillsides facing the equator, one can place table rows closer because of reduced mutual table shading). Thirdly, inverters are placed and tables are assigned to the inverters, varying the number of inverters. Due to the smaller impact of the choices for the wiring on the KPIs, computing a wiring is not included in our tree-like exploration of partial solutions, but instead we compute a single wiring for a generated layout in a single-objective manner, trying to minimize cable cross sections such that specified losses are not exceeded. Notice that our approach to decompose the problem into subproblems which are then explored in a tree-like manner is very flexible: For example, we could change the decomposition into subproblems in general, e.g., increase the number of subproblems if demanded or formulate other problems to create the partial solutions. This could be necessary if the design principles change or if one wants to create layouts in a conceptually different way, e.g., by using predefined standard building blocks consisting of one inverter and some tables arranged in a certain way. Due to its flexibility, our approach can easily be adapted to such changes in the future.

4 Results and Conclusion

We implemented our solution approach as a single-threaded application in the C# language using .NET 4.0. On a machine with an Intel Core i5 M560 CPU at 2.67 GHz , having 4 GB of RAM and running Windows 7, we were able to calculate

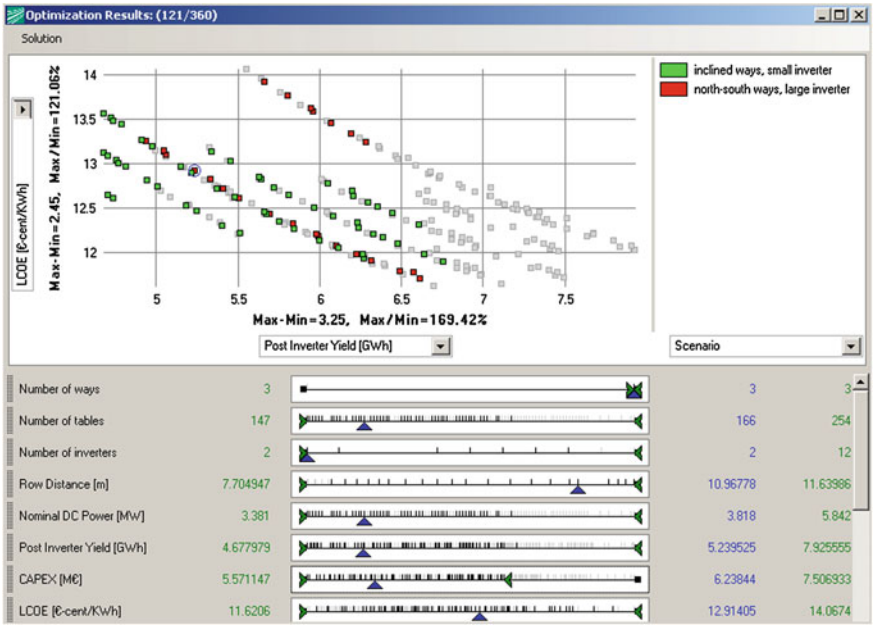


Fig. 2 The graphical user interface of our multicriterial decision support tool. Each small square in the plot area corresponds to a calculated PV layout.

the 360 layouts shown in Fig. 2 in 1.48 s for a real-world area similar to the one in Fig. 1. In each scenario, two placements for the ways (one with 1 and one with 3 ways) were generated. For the tilt angle, the six values 25°, 27°, 29°, 31°, 33°, and 35° were used. The table row distances were varied using the so called *shadow angle* placement with shadow angles[1] of 20°, 25°, 30°, 35°, and 40°, i.e., 5 values. Finally, to vary the number of tables assigned to each inverter, three different values (100 %, 110 %, 120 %) for the so called DC/AC ratio were explored. So, in total $2 \cdot 6 \cdot 5 \cdot 3 = 180$ solutions per scenario were generated.

Looking again at Fig. 2, it is clear that a multicriterial approach is superior to optimizing any single objective function: For example, the numbers next to the sliders show that the best solution in terms of LCOE (y-axis) has an objective function value of 11.62 Euro-Cent/kWh, and the best solution in terms of total yield (x-axis) has value 7.93 GWh (grayed out in the figure). But both solutions show quite weak performance in the objective they were not optimized for (6.70 GWh and 12.02 Euro-Cent/kWh). Our multicriterial approach offers many intermediate solutions, among them one with values 7.50 GWh and 11.71 Euro-Cent/kWh, for example, which might be a good compromise.

The presented approach is currently being integrated in the planning processes at Siemens Energy. Here, the engineers not only save a lot of time while designing new layouts, but moreover obtain an overview on the whole solution space and the interdependencies of the different layout parameters. Based on this approach, the best plant configuration can be selected which would possibly not have been found in an iterative trial-and-error process.

References

1. European Photovoltaic Industry Association: Solar Generation 6: Solar Photovoltaic Electricity Empowering the World, 2011. http://www.epia.org/publications/epiapublications/solar-generation-6.html
2. Messenger, R.A., Ventre, J.: Photovoltaic Systems Engineering, 3rd edn, CRC Press, Boca Raton, FL (2010).
3. Quaschning, V.: Simulation der Abschattungsverluste bei solarelektrischen Systemen. Verlag Dr. Köster, Berlin (1996)
4. Fraunhofer-Institut für Solare Energiesysteme ISE: Studie Stromgestehungskosten Erneuerbare Energien, 05/2012
5. Short, W., Packey, D.J., Holt, T.: A Manual for the Economic Evaluation of Energy Efficiency and Renewable Energy Technologies, NREL (National Renewable Energy Laboratory), Technical, Report NREL/TP-462-5173, 03/1995

[1] Roughly speaking, the shadow angle placement works as follows. On the northern hemisphere, a table t_2 in the next row to the north of a table t_1 is placed such that a ray of light with the given shadow angle from the upper edge of t_1 hits the lower edge of t_2.

Market Modeling in an Integrated Grid and Power Market Simulation

Torsten Rendel and Lutz Hofmann

1 Introduction

The liberalization of the European power market causes increased cross boarder transmission of electrical energy. The transmission capacity of the power lines was originally designed to support grids in neighbor countries in case of power shortages to supply the consumers and to assist in case of power plant malfunctions. Therefore, the power lines are not designed for an almost unlimited transport of energy. In fact, their transmission capacity is limited. It is not possible to handle an unlimited trade with the use of the actual available power lines. Due to the increased construction of decentralized generation units, especially renewable power generation, the amount of fluctuating generation and feedback from lower voltage levels will increase. This will also affect the power market and though the power flows within the European transmission grid.

To foresee and analyse the upcoming changes in the European power system, the Institute of Electric Power Systems in Hannover is developing a grid and power market simulator which generates a power plant schedule based on a marginal cost model. This schedule is optimized by an evolutionary algorithm in order to organize the exchange of energy between the interconnected countries. The goal is to minimize the electricity price in the whole market area. Based on the results of the dispatch, various load flow and short circuit analyses are possible, to make sure that the system runs in stable conditions. Open databases and a graphical user interface allow an easy design and modification of the scenarios in order to create possible future power systems [1].

T. Rendel (✉) · L. Hofmann
Institute of Electric Power Systems, Appelstrasse 9a, 30167 Hannover, Germany
e-mail: rendel@iee.uni-hannover.de

L. Hofmann
e-mail: hofmann@iee.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_50,

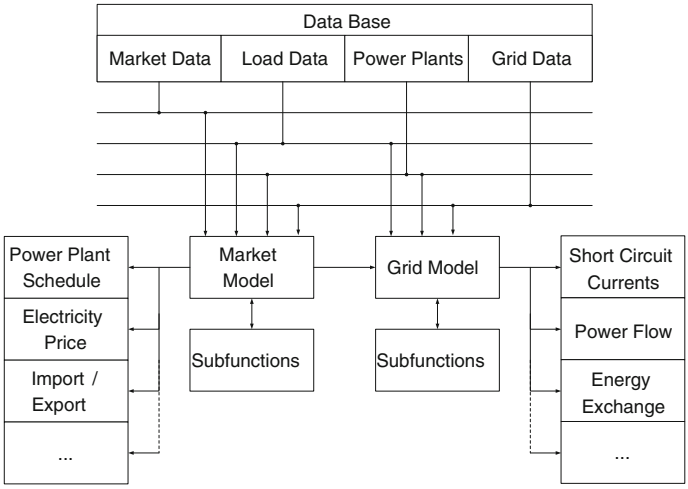


Fig. 1 Schematic view of modular structure of the grid and power market simulator

2 Integrated Grid and Power Market Simulation

The core part of the simulation is the analysis of the electric power system. Due to the strong correlation between the power market and the European transmission grid, both are combined in the integrated grid and power market simulation (see Fig. 1). The first part of the simulation is the handling of the input data, which consists of load, grid and generation data. These datasets are assigned to regions as described in [2]. Based on these input data the market model, which is explained in detail below, is creating a power plant schedule. With the help of this schedule the load of the power grid can be calculated and detailed grid analysis is possible.

2.1 Market Model

The market model consists of several core parts. One part is a power derivative market, where options and futures are traded to hedge the electricity price. In addition, the energy exchange provides a spot market on which day ahead and intraday products are traded [3]. Besides the standard products on the derivative and spot market OTC-trades are possible. In this market model, a derivative market and a spot market simulation are feasible. At first the market area is subdivided in two steps. The first step is a segmentation in bidding zones, which are connected by power lines with a limited amount of transmittable energy. Everey bidding zone operates its own derivative and spot market. The economical connection between the bidding zones markets is provided by an evolutionary algorithm which is optimising the

cross boarder trades [4]. Within these bidding zones it is assumed that an unlimited exchange of energy is possible in every possible system state. The bidding zones themselves are subdivided in several regions in order to model decentralized generation units and the energy consumer distribution. In general, the boarder of bidding zones are the frontier lines of the countries and the regional boarders are given by administrative region boarders.

2.1.1 Marginal Cost Model

The cost function of a power plant is depending on its fixed and variable costs. Participants of the European Energy Exchange have several possibilities to buy and sell energy. Therefore the European Energy Exchange is operating a plattform on which the markets are held Fig. 2. As mentioned above, several products are traded on these markets. In this thesis a derivative market and a spot market model are implemented. Therefore, the marginal costs of the power plants are calculated by Eq. (1) in which the marginal costs C_{marg} are calculated via the fuel costs C_{fuel}, the efficiency of the unit η_{el}, the start-up and shut-down costs C_{su}, the emission costs C_{emission} and the unit specific emission factor EF. The marginal costs are the costs which have to be payed to produce one more unit of energy. The power plants are sortet according to their marginal costs. Thus the merit order of the power plant park is received.

$$C_{\text{marg}} = \frac{C_{\text{fuel}}}{\eta_{\text{el}}} + C_{\text{su}} + \frac{C_{\text{emission}}}{\eta_{\text{el}}} EF \tag{1}$$

Equation (1) can be summarized with the use of (2).

$$C_{\text{marg}} = C_{\text{Run}} + C_{\text{su}} \tag{2}$$

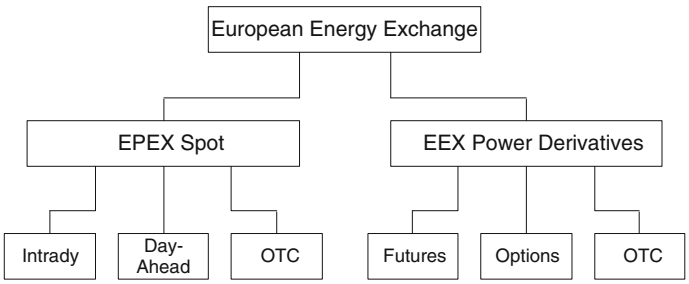


Fig. 2 EEX overview of possible trades on the spot and power derivatives market online-eex

2.1.2 Splitting the Markets

For the splitting of the markets, the load as well as the generation must be distributed on both market areas. Therefore a load dividing factor p_L is set to exclude the derivative market load P_{DM} from the residual load P_R. The residual load is the difference between the complete load P_L and the generation of renewable sources P_{RG} (see Eq. (3)). This derivative load is used to create a long term schedule for the power plants. The short term schedule is organized by the spot market. The spot market load P_{SM} is calculed in Eq. (4) [5].

$$P_{DM} = p_L \left(P_L - P_{RG}\right) \tag{3}$$

$$P_{SM} = (1 - p_L)\, P_L + p_L P_{RG} \tag{4}$$

2.1.3 Derivative Market

In the simulation the derivate market is based on a load forecast which is generated by the evaluation and reproduction of former demand timelines. This load is segmented in sveral blocks (see Fig. 3). Every power plant operator can place a bet for every block, according to the marginal cost model. The derivative market model will assign for every step the cheapest power plants to a certain block. Based on this model, a

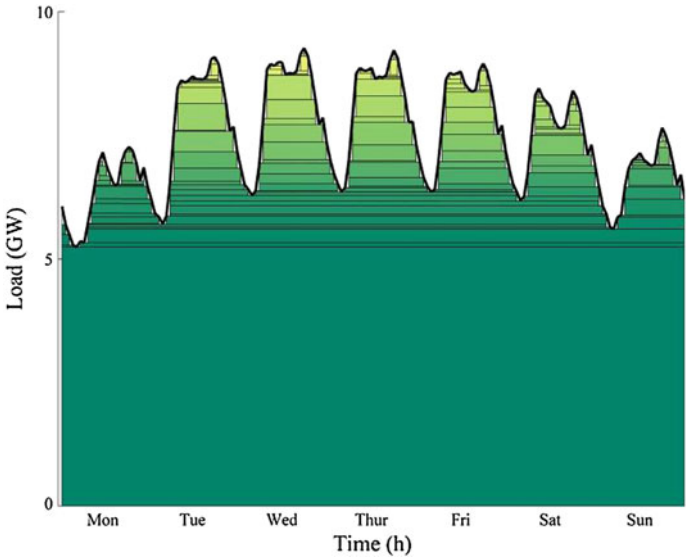


Fig. 3 Segmentation of the load in blocks for the power plant dispatch

long term power plants dispatch is organized. Aditionally it is assured that every power plant will fulfill its individual minimum load requierements.

2.1.4 Spot Market

The spot market is splitted in two submodules which are based on a 48 h forecast and a 24 h forecast. The first step is a power plant dispatch 48 h ahead. The input data for this step is the spot market load P_{SM} and the power of all units which is not scheduled yet as well as the power of the renewable energy sources P_{RG}. The second part is a short-term schedule which is acting 24 h ahead. In this module power plant operators are able to lower their market price below their marginal costs (see Eq. (5)). If it is possible to avoid a shut-down of the plant and thereby a start-up and start-up costs. Affected by this action spot market prices which are below the marginal costs of the most expensive unit on line are possible. This effect can cause negative electricity prices.

Based on the 48 h forecast, the renwable energy sources are dispatched. The amount of spot market load which is not served yet is also dispatched based on the 48 h forecast. The market model which is used here is based on the algorithm which is used on the derivative market. On this market also so called must run units can trade their energy. A unit can be assigned the state must run unit by the transmission system operator, if it is needed to provide system stability. In this case the power plant operator tries to benefit from the must run situation and places an unlimited bet.

The 24 h forecast is used to handle load and generation forecast inaccuracies. If the energy demand of the consumers is lower than assumed for a certain amount of hours $T = \{t_1, \ldots, t_n\}$, or the generation of renewable sources is higher than assumed, several power plants have to shut down during this time, because their marginal costs are higher than the market clearing price MCP.

$$MCP\,(t) < C_{\mathrm{marg}}\,(t)\ \forall\, t \in T \tag{5}$$

If the sum given in Eq. (6) is smaller than the start-up costs for the next point in time (t_{n+1}), where the market clearing price is higher than the marginal costs, the operator of the power plant can decide to run his plant below its marginal costs.

$$\sum_{t=t_0}^{t_n} \left(C_{\mathrm{marg}}\,(t) - MCP\,(t)\right) < C_{\mathrm{su}}\,(t_{n+1}) \tag{6}$$

If the consumption of energie is higher than assumed in the forecast or the wind generation is lower than assumed, additional energy must be provided by the conventional power plants.

3 Summary

The shown market model is able to model negative electricity prices. This is simulated by two assumpmtions. The first is the possibility to set certain power plants as must run power plants which have to be on line to provide system stability and place unlimited bets on the spot market. The other assumtion is the splitting of the market in a derivative market for long-term descisions and a 48 and 24 h spot market for mid- and short term-descisions.

References

1. Rendel, T., Rathke, C., Hofman, L.: Integrierter Netz- und Strommarktsimulator, ew - Das Magazin fr die Energiewirtschaft, Jg. 110 (2011), Heft 20, pp. 20–23, VWEW Energieverlag (2011)
2. Rendel, T., Rathke, C., Hofman, L.: Kraftwerkseinsatzplanung in einem integrierten Netz- und Strommarktsimulator, 12. Symposium Energieinnovation, Graz (2012)
3. Panos, K.: Praxisbuch Energiewirtschaft. Springer, Berlin (2009)
4. Rathke, C., Rendel, T., Hofmann, L.: Entwicklung eines europischen Strommarktsimulators, Elektrie Deutsche Elektrotechnik, Jg. 65, 85–87/2011, pp. 14–22, Verlag Dr. Heide und Partner GmbH (2011)
5. Doerbaum, M.: Erweiterung eines Strommarktsimulators um ein Modell zur Berücksichtigung negativer Preisimpulse in der Kraftwerkseinsatzplanung, Hannover (2012)
6. European Energy Exchange http://www.eex.com. Cited 15 Jun 2012

Customer-Oriented Delay Management in Public Transportation Networks Offering Navigation Services

Lucienne Günster and Michael Schröder

1 Introduction

Today car navigation is standard. Contrary, in public transport (PT) the ease and "never get lost"-feeling of car navigation has not yet been reached. However there are several projects that try to close the gap from the available mobile trip planners[1] towards the comfort of true navigation. The European research project SMART-WAY[2]. aims at bringing PT navigation a step further. In the cities of Dresden and Torino SMART-WAY is currently available in a piloting phase. It enables several new use cases like spontaneous entry into a vehicle, assistance for staying on the planned route, guidance if connections are at risk, etc.

It is worthwile to ask for the impact of PT navigation services on PT network performance. In this paper we investigate the situation when a PT network offers navigation services and a substantial group of passengers uses it. We have created a microscopic simulation model that allows to study navigation-enabled PT networks. Will passengers, modeled as individual entities in the simulation, reach their destinations with less delay on average? Can such a reduction in delay be even improved by an appropriate delay management of the PT operator?

The latter question is based on the fact that for navigated passengers the PT operator knows their planned routes. The operator can thus decide at which connections it makes sense that a vehicle waits for a delayed feeder vehicle so that passengers can

[1] In Germany a very well known mobile trip planner is the "DB Navigator", available for most smartphones.
[2] www.smart-way.mobi

L. Günster · M. Schröder (✉)
Fraunhofer ITWM, D-67663 Fraunhofer, Kaiserslautern
e-mail: lucienne.guenster@itwm.fraunhofer.de

M. Schröder
e-mail: schroeder@itwm.fraunhofer.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_51,

transfer. These wait-or-depart decisions are interdependent and must be considered simultaneously for all pending transfer options.

We embedded a specific mixed integer programming model for optimized customer-oriented delay management, taken from the literature [4], into the dynamic approach of our simulation. The MIP model is solved periodically on the basis of the currently known routes of navigated passengers. Wait-decisions are pushed into the timetable information system, thus navigated passengers will know about actual transfer possibilities. Particularly this can be transfers that emerge dynamically and are not planned in the published timetable.

Since navigated passengers are always informed about the best routes to their destinations, taking the dynamic transfer options into account, it can be expected that they reach their destinations with less delay than non-navigated passengers. But what happens to this group of passengers? Do they also profit from the wait-or-depart decisions, even if their routes have not been considered during the opimization? Our simulation experiments aim at getting insight into these questions.

The paper is organized as follows. In Sect. 2 we briefly describe the simulation model. Section 3 addresses dynamic customer-oriented delay management. Section 4 reports some results of the simulation experiments with regard to the average delay of passengers at their destinations.

2 The Simulation Model

The simulation model consists of six building blocks, which are—as well as their primary responsibilities—depicted in Fig. 1. The simulation component, the travel demand model and the delay generator essentially model the transport network with a discrete event simulation approach. The route information system and the dynamic schedule synchronization model implement the services of the PT operator.

The latter model decides which connections are to be maintained and which ones are to be dropped if delays occur. The main task of the route information system is to keep the navigated passengers updated about their current route options. The evaluation component collects data necessary to analyze the simulation results.

The most important entities in the simulation model are passengers and vehicles.

Passengers We simulate passengers as individual objects, equipped with the essential behavior. This resembles a multi-agent approach. Based on an origin-destination matrix modeling the demand, we create a large number of passenger objects as well as their destination and arrival locations and desired start times.

As said above, the model distinguishes between two different groups of passengers: those who use the navigation service and those who are non-navigated. Navigated passengers are frequently informed about their further route options and continue their journeys according to the currently best routes. Contrary, non-navigated passengers only update their routes in case they miss a connection. Further, they have only access to the published timetable, not to real-time schedule data.

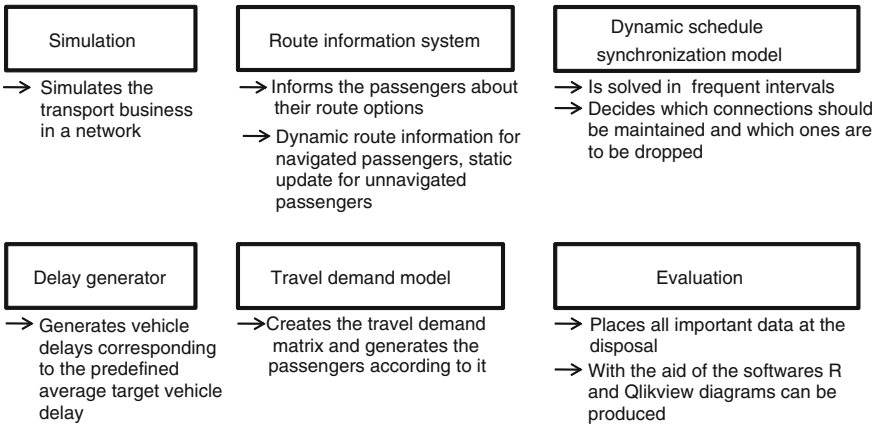


Fig. 1 Structure of the SMART-WAY simulation model: primary components and their main tasks.

Vehicles Buses and trains are also modeled as individual objects. Basically, they move according to their schedules. For simulating the delay of a vehicle we introduce two parameters, the *current delay*, stating the vehicle's actual tardiness, and the *target delay*, indicating the delay the vehicle will have reached when arriving at its final destination. The simulation adapts the vehicle's current delay towards the target delay by slowing it down. The target delays are exponentially distributed with a predetermined expectancy value representing the vehicles' mean target delay. This way, different amounts of perturbation can be imposed on the PT network.

Simulation scenarios Currently, the model is based on an artificial PT network ("SmartCity") and a suitable schedule. Simulation scenarios are defined by varying data as, for example, the travel demand and the mean vehicle delay. Further, scenarios distinguish if delay management measures are applied or not. Finally, significant parameters like the percentage of navigated passengers, are varied.

3 Delay Management as a Dynamic Planning Problem

The simulation model integrates a module for customer-oriented dynamic schedule synchronization. The decision whether to maintain a certain connection or not, is taken in such a manner that most passengers will profit from it. Consequently, we want to maximize the passengers' conveniences which can be realized by minimizing their target delays as well as the waiting times they have to wait in case of missed connections.

We implemented a dynamic planning model based on a mixed integer program, developed by Schöbel [2, 3].
This MIP for the static case is embedded in a dynamic planning approach where the

relevant data is only revealed over time. Contrary to static planning where a plan is computed only once, in a dynamic setting the plan has to be updated and adapted regularly to the current situation.

When shifting from the static case to a dynamic setting, different topics arise. For instance it turned out that it is necessary to define a so-called *frozen zone*. This is a predetermined time interval in which all decisions concerning the near future should be kept stable.

The route information system is informed after each run of the MIP model about the solution, i.e. which connections should be maintained and which ones are to be dropped. Only connections lying within the timeframe of the frozen zone are disclosed to the system, for a simple reason: Once we have notified the information system about a preserved connection, this connection should definitely be maintained (unless the delay of the delayed feeder vehicle becomes too large). Consequently, the corresponding wait-or-depart decision must not be overturned in the next planning updates since the PT operator has promised the concerned passengers to keep this connection.

4 Simulation Study and Results

We examined the impact of customer-oriented delay management by a series of simulation experiments. Under the variation of important parameters of the model we ran each scenario twice, once with applying delay management measures and a second time without. Afterwards, the average passenger delays in both cases are compared. As discussed above, non-navigated passengers are not considered in the objective function of the MIP.

For each scenario we ran the simulation ten times with different initializations of the random numbers generator. In each of these 10 experiments around 10,000 passengers are created. Each passenger is either navigated or non-navigated and travels within the simulated PT network to reach his destination.

We set the mean target vehicle delay to 4.5 min and consider a frozen zone time span of 10 min. The time interval between two runs of the MIP amounts to 30 s of simulated time. Moreover, we assume a timetable periodicity of 15 min.

We vary the percentage of navigated passengers to investigate the profit of dynamic schedule synchronization for these passengers opposed to the group of all passengers. We consider four different values, namely 10, 25, 50 and 75 % of navigated passengers. In all cases, we measure the average delay of the passengers at their destinations.

For each experiment, we compare the outcome with delay management actions to the situation without delay management measures by taking the difference of the average passenger delay in the two settings. Based on the 10 experiments for each scenario, we perform statistical tests, to find out the gain of customer-oriented delay management.

Table 1 p-values and 95 % confidence intervals for different percentages of navigated passengers and different groups of passengers

Percentage of navigated passengers	Navigated passengers profit		All passengers profit	
	p-value	95-PCI	*p*-value	95-PCI
10	2.303e-06	13.21986	0.2117	−1.186115
25	3.253e-07	12.58336	0.04874	0.02171689
50	2.895e-08	11.52843	6.548e-06	4.476398
75	2.62e-08	11.27876	5.066e-07	7.580077

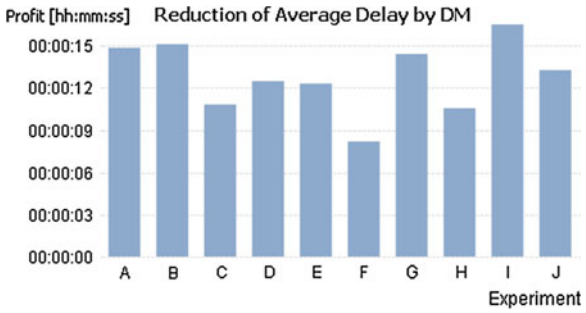


Fig. 2 Reduction of average passenger delays, depicted for navigated passengers only

We test the hypothesis that over all 10 experiments for a scenario the average delay of passengers under customer-oriented delay management is strictly smaller than the average delay of passengers in a network without delay management measures by the PT operator.

Table 1 summarizes the results of the *t*-tests. In each case, the subsequent hypotheses are examined by depicting the corresponding *p*-values and 95 % confidence intervals (95-PCI; average delay at destination in seconds):

1. "All *navigated* passengers on average benefit from customer-oriented dynamic schedule synchronization."
2. "*All* passengers, navigated and non-navigated, will, on average, profit from delay management measures."

Figures 2 and 3 exemplarily show the underlying reductions in average delay over the 10 different experiments for 50 % of navigated passengers.

Small *p*-values close to zero are obtained for navigated passengers. Consequently, the first hypothesis has a strong support: navigated passengers profit on average from customer-oriented delay management. By contrast, the second hypothesis cannot be approved in the event that only 10 % of the passengers use navigation. A significance level of $\alpha = 0.2117$ results in this case; the wait-or-depart decisions are based on a small group of passengers and show a negative effect on the whole group of passengers. However, if enough passengers are navigated, all passengers will indeed profit from the conducted delay management measures.

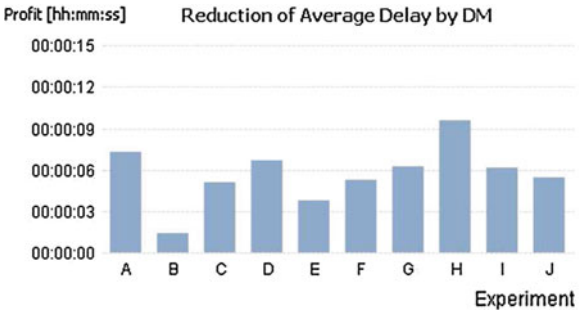


Fig. 3 Reduction of average passenger delays, all passengers evaluated

5 Conclusions

We have carried out extensive simulation experiments to research the benefit of customer-oriented delay management in PT networks that offer navigation services to the passengers. A few examples of the results obtained have been shown. Many more outcomes are discussed in [1].

Optimal wait-or-depart decisions that take the routes of navigated passengers into account reduce their average delays at destination significantly. This has to be contrasted to the fact that the decision to wait for a delayed feeder vehicle always increases the total delay of vehicles in the PT network. However, to make sure that measures for securing connections in case of delays do not negatively affect the whole group of passengers, the group of navigated passengers has to be large enough. Otherwise only an "elite" group of passengers who are using navigation will have a gain.

Acknowledgments The research in this paper was partially supported by the European Commission under grant FP7-248251 SMART-WAY.

References

1. Günster, L.: Models and algorithms for customer-oriented dynamic schedule synchronization in regional public transport networks. Diploma Thesis, Kaiserslautern (2012)
2. Schöbel, A.: Integer Programming Approaches for Solving the Delay Management Problem. In: Geraets, F., Schöbel, A., Wagner, D., Zaoliagis, C. (eds.) Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, vol. 4359, pp. 145–170. Springer, Berlin (2007)
3. Schöbel, A.: A model for the delay management problem based on mixed-integer-programming. Electronic Notes Theoret. Comput. Sci. **50**(1), 3–10 (2001)
4. Schöbel, A.: Capacity constraints in delay management. Public Transp. **1**(2), 135–154 (2009)

Sustainability Assessment and Relevant Indicators of Steel Support Structures for Offshore Wind Turbines

Peter Schaumann and Anne Bechtel

1 Introduction

Within the last decades cost effectiveness and constructional aspects of structures were main design drivers for new constructions within the offshore wind industry. Increasing importance of environmental friendly products and the reduction of CO_2-emissions have a growing impact on the development of holistic design concepts in all industries. Especially the wind energy section producing so-called 'green' energy is presupposed to follow sustainable concepts.

The annual installation (Fig. 1) regarding onshore and offshore wind energy in Germany from 1990 to 2030 shows a growing market especially for offshore. The upcoming onshore power installation will be decisively affected by repowering whereas the offshore sector indicates a significant grow of new electrical power for the prospective years. In combination with a mean steel amount of 150 t per Megawatt (MW) for onshore and 250 t per MW for offshore wind turbines (OWT) the annual predicted consumption of more than 1.2 Mio tons of steel in 2023 reflects the significant potential of wind energy constructions related to the steel demand. The material steel representing 90 % of the material mass used in an OWT has the biggest effect on environmental aspects as presented by Wagner et al. [8].

Growing market, great steel demand, and optimization potential lead to the motivation for the development of a life cycle assessment that takes all relevant sustainability dimensions into account.

P. Schaumann (✉) · A. Bechtel
Leibniz Universität Hannover, Institute for Steel Construction, Appelstraße 9A,
30167 Hanover, Germany
e-mail: schaumann@stahl.uni-hannover.de

A. Bechtel
e-mail: bechtel@stahl.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_52,

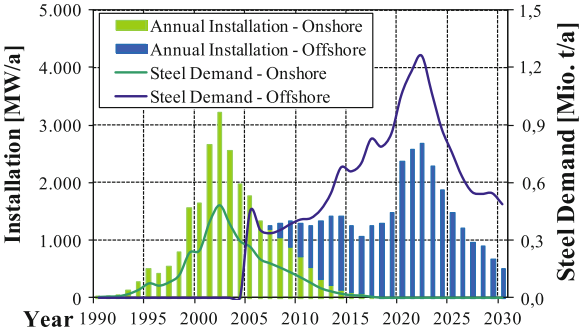


Fig. 1 Annual installation of offshore and onshore wind energy and related steel demand for Germany

2 Life Cycle Sustainability Assessment

For a holistic sustainability concept additionally to environmental aspects the economical, social, technical and procedural dimension have to be taken into account. The following sections describe the basic methodological features of the modified life cycle sustainability assessment.

2.1 Theoretical Background

Existing rating systems for buildings, as e.g. DGNB [1] constitute the basis for the establishment of a life cycle sustainability assessment for steel constructions of renewable energies.

The rating system of the DGNB handbook [1] consists of the six sustainable categories: ecology, economy, socio-cultural and functional criteria, as well as technical aspects, process criteria, and local effects. Each of these categories is defined by a certain number of criteria and indicators reflecting the impact of building and materials used. The rating results from weighted categories and additionally weighted criteria and indicators. The degree of fulfilment of the criteria leads to the degree of performance. To reach a holistic rating the entire life cycle of the building including all products used has to be considered.

2.2 New Approach

Due to the basic understanding of sustainability reflecting the elements ‘ecology, economy, society, process, and technique’, the rating system for steel constructions

of renewables consist of these sustainability dimensions. Each of these categories consists of numerous criteria and indicators. Within the research project 'Sustainability of steel support structures for Renewables (NaStafEE), a total of 35 criteria were developed to evaluate the sustainability of e.g. steel support structure of OWT, see Sect. 3. In order to compare different constructional solutions regarding the sustainability the functional unit is defined as the 'steel support structure'.

To reach a holistic consideration all decisive life cycle stages need to be included. Table 1 shows the main life cycle stages which have to be taken into account within the life cycle sustainability assessment. These stages were defined referring to EN 15804 [3], EN 15978 [4], and Hauke and Siebers [5]. Especially the life cycle stages A and D play an important role in the evaluation of steel structures due to the high recyclability of steel.

3 Indicators of Steel Support Structures

Different indicators of steel support structures depending on the type of renewable energy and structure influence the sustainability in different ways. Within this section the steel support structure of OWT as well as related sustainability effects and criteria are presented.

3.1 Steel Support Structures for Offshore Wind Turbines

Steel support structures of OWT consist of the tower and the substructure which includes piles and steel structure under water representing the interface between tower and piles (Fig. 2). As shown by Wagner et al. [8] one of the decisive components of OWT regarding ecological sustainability criteria is the substructure. Depending on the type of substructure and the pile length this substructure requires up to five-times more steel than the tower. For the use of steel substructures and in relation to soil conditions and water depth it can be differentiated between monopile, jacket, tripile and tripod substructure. Due to large water depth in the German Exclusive Economic Zone (EEZ) of about 30–50 m preferably spatial structures as tripod and jacket will

Table 1 Life cycle stages according to Hauke and Siebers [5]

	Life cycle stages			
A1–A3	A4–A5	B	C	D
Before lifetime of the structure: product stage	Manufacturing and construction stage	Use stage	End of life stage	After lifetime of the structure: benefits and loads

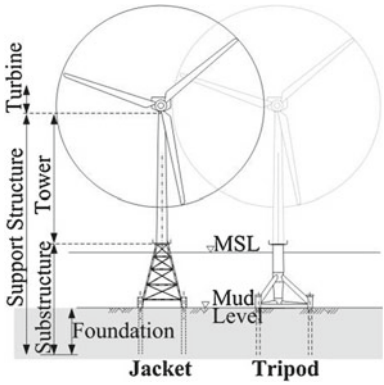


Fig. 2 Offshore wind turbines with jacket (*left*) and tripod (*right*) substructure

be used. Hence, for a first comparison of results the focus is set to the reference substructure type jacket and tripod.

The water depth for tripod and jacket is equally 30 m. With a pile length of 50 m for the tripod and between 30 and 45 m for the jacket the steel mass results to about 1,300 t for the tripod and 830 t for the jacket. Besides the material masses, welds and corrosion protection were considered. The corrosion protection for both substructures consists of a coating system in the splashing zone and anodes in the underwater zone.

3.2 Relevant Criteria and Indicators

Regarding the steel substructure for jacket and tripod relevant criteria and indicators affecting the sustainability of the structure are discussed subsequently.

Ecology. The ecological evaluation is commonly based on the standardized and approved Life Cycle Assessment (LCA) acc. to ISO 14040 [2]. Within a LCA several environmental impacts, primarily attributed to emissions of pollutants or the consumption of resources, caused by a product system considering its entire life cycle are gathered.

The LCA of the substructures tripod and jacket was analyzed regarding the mentioned life cycle stages and the main LCA criteria: global warming potential (GWP), ozone depletion potential (ODP), acidification potential (AP), eutrophication potential (EP), photochemical ozone creation potential (POCP), and cumulative energy demand (CED). Results for both substructures considering these impact criteria are included in a polardiagram (Fig. 3). The two spanned areas, grey for tripod and red for jacket, show that the tripod has a worse ecological impact than the jacket which is caused by the higher amount of steel. The AP is relatively high for the jacket due to large transport distances and deployed transport carrier. For both structures the evaluation for the individual life cycle stages showed that the construction phase

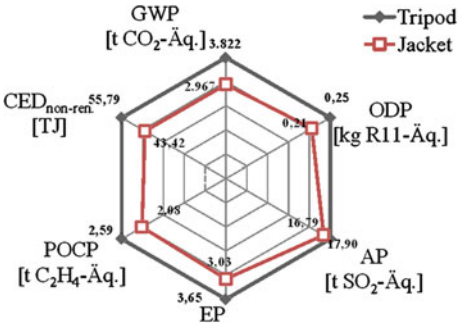


Fig. 3 Polardiagram reflecting ecological impact of tripod and jacket

influences the ecologic impact indicators significantly, whereas phase B and C are less important.

In addition to the traditional criteria covered by the LCA noise emission is one important factor concerning offshore substructures. Regarding the pile driving process noise protection has to be performed to protect the marine fauna in the North Sea.

Economy. The economical effect of steel structures can be evoked by the criteria: life cycle costs, investment for research & development, and employment effects. Regarding the life cycle costs own studies have shown that transport, manufacturing and installation of wind turbines influence the costs decisively compared to operation and removal costs.

Social Quality. The social impact encompasses the effect of the product over the life cycle to workers, users and persons concerned. Regarding the mounting and erection of steel structures the working safety plays especially offshore a significant role evaluated by the criterion 'industrial safety and accident frequency'.

In addition the two social criteria 'family friendliness' and 'social engagement' include company oriented social activities.

Technical Quality. The technical aspects reflect the technical quality of the product regarding monitoring, repair, and inspections over the lifetime of the construction. Due to harsh environmental conditions 'offshore' the technical quality regarding the criterion corrosion protection is of fundamental importance. On the one hand the lifetime of steel structures is effected directly by the corrosion protection, whereas on the other hand the applied corrosion protection affects ecological criteria.

Process Quality. The Process quality is influenced by planning and process technical elements. The criterion transport which includes transport way as well as type of transport carrier is determinative for offshore substructures. Analysed substructures have shown that large transport routes combined with shipping lead to significant influence of ecological dimensions, Schaumann et al. [7].

Additionally the productivity as well as the project realisation time is included in the evaluation of the process dimension.

4 Conclusions

Based on existing and well-established sustainability rating system for buildings a life cycle sustainability assessment has been developed for steel structures of renewable energies. Relevant impact to the sustainability dimensions of the reference substructure types jacket and tripod of Offshore Wind Turbines have been shown by decisive criteria within the categories ecology, economy, social, technical and process quality.

Acknowledgments Presented results and approaches are achieved within the research project 'NaStafEE' [6] funded by the German Federal Ministry of Economics and Technology (BMWi) via the AiF and FOSTA accompanied by more than 30 industry project partners. The support by the partners especially the research partners Prof. N. Stranghöner and scientific assistants, IML, University of Duisburg-Essen and Prof. Wagner and scientific assistants, LEE, Ruhr-University Bochum is kindly acknowledged.

References

1. Deutsche Gesellschaft für Nachhaltiges Bauen e.V.: DGNB Handbuch - Neubau Büro—und Verwaltungsgebäude (in German). DGNB, Stuttgart (2009)
2. Deutsches Institut für Normung e.V.: DIN EN ISO 14040— Umweltmanagement—Ökobilanz—Grundsätze und Rahmenbedingungen (in German). Beuth Verlag, Berlin (2009)
3. Deutsches Institut für Normung e.V.: DIN EN 15804—Nachhaltigkeit von Bauwerken—Umweltproduktdeklaration—Grundregeln für die Produktkategorie Bauprodukte (in German). Beuth Verlag, Berlin (2012)
4. Deutsches Institut für Normung e.V.: DIN EN 15978—Nachhaltigkeit von Bauwerken—Bewertung der umweltbezogenen Qualität von Gebäuden—Berechnungsmethode (in German). Beuth Verlag, Berlin (2012)
5. Hauke, B., Siebers, R.: Life cycle assessment comparison of a typical single storey building. bauforum stahl e.V., Düsseldorf (2011)
6. Schaumann, P., et al.: Zur Nachhaltigkeitsbewertung von Stahlkonstruktionen für regenerative Energien (in German). Stahlbau **80**(10), 711–719 (2011a)
7. Schaumann, P. et al.: Indicators for Environmental and Social Assessment of Steel Support Structures for Offshore Wind Turbines. In: Proceedings of the EWEA Offshore Conference 2011. Amsterdam (2011b)
8. Wagner, H.-J., et al.: Ökobilanzierung des Offshore Windparks alpha ventus (in German). LIT-Verlag, Münster (2010)

A Quantitative Car Sharing Business Model for Students

Michael H. Breitner and Judith Klein

1 Introduction

"Many people, especially younger ones, don't need to possess everyday things. It's enough for them to just be able to use them temporarily.", see [15] and additionally [2, 6, 7, 10], and [11]. This aspect is the focus of an "Electromobility Showcase" currently running in the metropolitan region of Hannover, Braunschweig, Göttingen, and Wolfsburg. The goal of one of the projects, the "Campus E-Car Sharing" project, is to introduce users to the topic of electromobility. To this end, universities in the metropolitan region plan to set up stations for Quicar, which is Volkswagen's car sharing offering. These stations will be available to people associated with the university that want to drive electric cars. The International Association of Public Transport (UITP) describes car sharing as follows: "Car-sharing is based on the simple idea of being able to use a car individually without having to own one but in a way that is just as convenient. It is an innovative local mobility service complementary to public transport.", see [14]. Typically, the costs of car sharing are based on time and kilometer usage, and these costs include insurance, tax, and sometimes fuel, too. Quicar, the subject of the study that follows, only charges for time: 0,20 Euro/min for driving time and 0,10 Euro/min for parking time.

The upward trend in the number of car sharing users, which experienced especially a sharp growth in 2011, shows that car sharing is continuing to gain acceptability as a mobility service, and provides a positive environment for the projects mentioned above, see, e.g., [3–5, 8], and [12]. This increase, especially in 2011, is also due to market entries from BMW, Daimler and Volkswagen: all of whom recognized the attraction of car sharing and all were able to acquire a large numbers of users, even

M. H. Breitner (✉) · J. Klein
Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany
e-mail: breitner@iwi.uni-hannover.de

J. Klein
e-mail: klein@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_53,

in the first year, see [9], and [13]. These numbers are the result of new providers with large fleets of vehicles joining the market and winning over customers with very attractive offers. Because the automobile manufacturers are not offering car sharing as their main line of business, as opposed to most of the other providers, and due to the size of their companies and the resulting higher levels of liquidity, they have more options both in communication and pricing. On this basis, it makes sense to actively address students, as potential affine and long-term customers, with new offers. The gap between car sharing in its current form and students, who are generally keen on innovation and enthusiastic, and who are also especially flexible and open to a wider variety of potential uses for car sharing, can be shortened with well designed offers. In this way, a new mobility service for students is beneficial for all stakeholders: it increases students' mobility, introduces them to new technologies, possibly draws them towards buying a car later, making them potential new customers, and promotes an universities' reputation and customer orientation.

2 A Quantitative Car Sharing Business Model for Students

We calculate a suitable price model for students, and this price forms the basis of further considerations in the "Campus E-Car Sharing" project. To determine the needs and students' willingness to pay, as well as possible use scenarios for students, we created an internet-based survey for more than 1,000 first year Economics and Engineering students at the Leibniz Universität Hannover. The results were then used to draw important conclusions about pricing. In order to be able to determine students' willingness to pay, we need a precise definition of demand. Those surveyed were confronted with scenarios that require a car, and were not asked about general, time-variant car sharing rates. These scenarios were evaluated by participants with distinct willingness to pay, and these were used to determine the frequency distribution and the empirical distribution function with these new car sharing rates. With the goal of introducing a large number of students to the subject, in the future, with these new rates, we hope to make it possible for at least half of students to use a car occasionally and according to his or her specific ability to pay.

In order to drive the necessary benefit increase for students, we need to find scenarios that require a much higher level of effort with public transportation than they would with a car, because all students in Hannover have a semester ticket for travel within Hannover and Lower Saxony for free. In most scenarios, we differentiate between short-term use and long-term use of a car, see Table 1. The former includes use of up to ten hours (and includes fuel costs), and this type of use forms the basis of this model. The latter is served by Quicar Plus, which works like a typical car rental service (and excludes fuel costs) and as an overall concept, is meant to complement car sharing as part of cross selling. On this basis, the following scenarios make a car suitable:

Table 1 Parameters and assumptions for scenarios 1–7

Scenario	Driving time (min)	Parking time (min)	Travel distance (km)	Covered by	Vehicle model	Assumptions
1	30	0	21	Quicar	Golf blue motion	No minimum rental du- duration, one-way drive drive between Herren- hausen and Garbsen possible
2	40	120	40	Quicar	Golf blue motion	
3	300	420	450	Quicar Plus	Up!	
4	480	2,400	550	Quicar plus	Beetle	
5	30	30	20	Quicar	Golf blue motion	
6	20	0	15	Quicar	Golf blue motion	One-way drive possi-ble with-out restrictions
7	120	480	75	Quicar Plus	Transporter	

- **Scenario 1:** A trip from the main building of the Leibniz Universität Hannover in Hannover Herrenhausen to the Produktionstechnisches Zentrum Hannover in Garbsen and back.
- **Scenario 2:** A trip from the city center in Hannover to IKEA in Laatzen and back.
- **Scenario 3:** A one-day excursion within Lower Saxony.
- **Scenario 4:** A weekend trip within Germany (Friday– Sunday).

The scenarios are clearly described using the various parameters in order to relate them to the current pricing structure of Quicar and to determine students' willingness to pay. Assumptions are also set that differ from the current Quicar concept. However, these assumptions are valid for the current business model. In addition to information about students' willingness to pay, the survey asks for proposals for other scenarios that might require a car. Some typical proposals can be summarized into three general scenarios:

- **Scenario 5:** Shopping for food and drinks within Hannover.
- **Scenario 6:** A one-way trip to the train station or to a residence within Hannover.
- **Scenario 7:** Moving within Hannover (long-term use).

3 Results and Recommendations

Over five-hundred students of the Leibniz Universität Hannover participated in the survey and for scenarios 1-6, named the amount that they would be prepared to pay for the respective scenario. The empirical distribution function ($\hat{F}$), which shows the cumulative relative likelihood for each value and assumes a linear course of relative frequency within the class boundaries, is the basis of the analysis. It is also denoted as cumulative frequency, and thus for each value of students' willingness to pay, it indicates which portion of those surveyed indicated a value that was less than or equal to the value in question. The following is thus true:

$$\hat{F}(x) = \hat{F}(x_{j-1}) + \frac{n_j}{n} * \frac{(x - x_{j-1})}{\Delta x_j} \text{with}$$

$$n_j = \text{absolute frequency of answers in class } j \text{ and}$$

$$\Delta x_j = \text{quantity of class } j.$$

The most important parameter is the median ($x_{0.5}$). It bisects those surveyed with regard to their answers: exactly half of all cases are below and above this value. Because the data for this survey are distinct, the median can be found in the class in which the empirical distribution function exceeds the value 0.5. If the median is found in class j, it can be calculated as follows:

$$x_{0.5} = x_{j-1} + \frac{0,5 - \hat{F}(x_{j-1})}{\frac{n_j}{n * \Delta n_j}}$$

Table 2 shows the results of the survey, whereby the medial students' willingness to pay can be de-rived for scenarios 5–7 from their medial willingness to pay for scenarios 1–4. The conceptual design of a new car sharing rate that is based on students' willingness to pay now focuses on car sharing and not on Quicar Plus (car rental). The four scenarios that were served by Quicar are each characterized by

Table 2 Willingness to pay for scenarios 1–7

Scenario	Students' willingness to pay (median)	Quicar revenue	Quicar plus revenue
1	5,70 Euro	6,00 Euro	–
2	10,87 Euro	20,00 Euro	–
3	47,64 Euro	-	71,73 Euro
4	80,50 Euro	-	154,45 Euro
5	7,42 Euro	9,00 Euro	–
6	4,67 Euro	4,00 Euro	–
7	27,11 Euro	-	Approx. 60,00 Euro

three parameters: driving time, parking time, and the medial students' willingness to pay. The revenue for student use was developed from those parameters under the following assumptions:

- The revenue for a scenario cannot exceed the students' willingness to pay in the median for this scenario.
- The new car sharing rate can only be a maximum of 25 % lower than the current rate.
- The total revenue for all scenarios combined realistically should be maximized.

The result is the following general optimization model. The model fulfils the objective of developing revenue that provides the highest possible sales volume for Quicar, but nevertheless has the potential to draw the highest percentage of students to car sharing:

$$\max \sum_{j=1}^{n} revenue_j$$

where

$revenue_j = driving\ time_j * a + parking\ time_j * b$ obeying the constraints
$revenue_j \leq wtp_j$ for $j = 1, 2...n,\ a \geq A * 0,75,\ b \geq B * 0,75$ with
$driving\ time_j$ = driving time in scenario j [min]
$parking\ time_j$ = parking time in scenario j [min]
a = new student car sharing rate in driving mode [Euro/min]
b = new student car sharing rate in parking mode [Euro/min]
wtp_j = students' willingness to pay (median) [Euro]
A = standard car sharing rate in driving mode [Euro/min]
B = standard car sharing rate in parking mode [Euro/min].

In the following, the model is calculated with the quantities that were determined in the demand model for the above mentioned scenarios, in order to determine the optimal car sharing rate. However, for scenario 2, this process appears to go nowhere. Since only 6 % of the surveyed students were willing to pay the current rate for a trip to IKEA, it is not possible to offer a large enough percentage of students the use of a car at the price they are willing to pay and still stay within the constraints. For this reason, we disregard scenario 2 in this model. Solving the model results in the following new revenues: *a=0,17* [Euro/min], *b=0,075* [Euro/min].

The results show that the revenue per minute a must be reduced by less than 25 % to reach 50 % of those surveyed. Quicar Plus offers will also be reduced for students. As for Quicar, the reduction cannot be more than 25 % of the standard car sharing rate. However, because students' willingness to pay for long-term use scenarios is far below current revenue in comparison to the scenarios, no rate can be determined for Quicar Plus that enables the majority of students to use the cars with their given willingness to pay. For this reason, all Quicar Plus revenues are reduced by 25 % across the board. The calculations show that the positive trends in the number of car sharing users are also reflected in students' willingness to pay. Within the framework of this analysis, the potential of university car sharing is demonstrated for the first

time and now can form the basis of further activities. The survey made a large group of people aware of car sharing and is thus the first step towards making car sharing available at the Leibniz Universität Hannover even in 2012.

References

1. Bundesverband für CarSharing e.V.:http://www.carsharing.de
2. Borchardt, M.: Car-Sharing: Eigentumslose PKW-Nutzung als Beitrag zu einer nachhaltigen Entwicklung? Grin Verlag, Norderstedt, and Diplomica Verlag, Hamburg (2012)
3. car2go: http://www.car2go.com
4. DriveNow: DriveNow by BMW/Mini/Sixt. http://www.drive-now.com
5. Flinkster by Deutsche Bahn: http://www.flinkster.de
6. IFMO: Zukunft der Mobilität Szenarien für das Jahr 2030 (zweite Fortschreibung). Study of the Institut für Mobilitätsforschung (IFMO), München (2010)
7. Loose, W., Glotz-Richter, M.: Car-Sharing und ÖPNV - Entlastungspotenziale durch vernetzte Angebote. Ksv-verlag Thomas J. Mager, Köln (2012)
8. Quicar by Volkswagen: http://www.quicar.de
9. Schmidt, S.: Trends in der Automobilindustrie: Vermarktung innovativer Mobilitätsdienstleistungen am Beispiel Car-Sharing. Grin Verlag, Norderstedt (2011)
10. Schwieger, B.: Second Generation Car-Sharing: Developing a New Mobility Services Target Groups and Service Characteristics. Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken (2012)
11. Shell: Shell PKW-Szenarien bis 2030. Study of the Shell Deutschland Oil GmbH, Hamburg (2009)
12. stadtmobil Hannover: http:// www.stadtmobil.de.
13. Stricker, K., Matthies, G., Tsang, R.: Vom Automobilbauer zum Mobilitätsdienstleister - Wie Hersteller ihr Geschäftsmodell für integrierte Mobilität richtig aufstellen. Study of the Bain & Company, Germany/Switzerland (2011)
14. UITP: Flyer of the International Association of Public Transport (UITP). http://www.uitp.org/Working-Bodies/Car-Sharing/pics/brochure_de.pdf
15. Welt-Online: VW geht mit “Quicar" unter die Autovermieter. http://www.welt.de/finanzen/verbraucher/article13694521/VW-geht-mit-Quicar-unter-die-Autovermieter.html

The Canola Oil Industry and EU Trade Integration: A Gravity Model Approach

Dirk Röttgers, Anja Faße and Ulrike Grote

1 The Production and Trade Situation in the Biodiesel Sector

In recent years, many developed countries emphasized support for the production of biofuels in their political agenda. This new interest in biofuels arose mainly from the quest for increasing national energy sovereignty to become independent from oil, but is also founded in strong fluctuations of crude oil prices and environmental concerns (see e.g. [6]). The strongest and most concrete and concerted political decision of the European Union (EU) was to set a mandatory quota for the use of biofuel of 5.75 % biofuel in 2010 (see e.g. [16]). The path to the 5.75 goal can be set differently by the member countries, leading to diversity among members. Other less widespread or clearly defined national and other supranational measures followed, like raising excise taxes or providing capital subsidies for green investments [11].

To satisfy the increased demand for biodiesel in many European countries due to the quota and other measures, the import of raw products for canola-based biodiesel increased. It is the aim of this paper to analyze the effect the EU imposes on the trade of these biofuel commodities.

It is clear that being a member of the EU makes a difference for trade patterns of a country. The EU regulates both, international trade and the bioenergy sector heavily. Thus it creates a difference among members and, more importantly, between members and non-members. But what effect exactly drives canola oil trade: Trade regulations, bioenergy regulations, both or neither? To correctly analyze this

D. Röttgers (✉) · A. Faße · U. Grote
Institute for Environmental Economics and World Trade, Leibniz Universität Hannover,
Königsworther Platz 1,
30167 Hannover, Germany
e-mail: roettgers@iuw.uni-hannover.de

A. Faße
e-mail: fasse@iuw.uni-hannover.de

U. Grote
e-mail: grote@iuw.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_54,

question, patterns of the biodiesel market have to be taken into consideration also. Therefore the model is expanded with sector specific variables.

2 Methodological Framework and Data

To analyze trade relationships for canola oil, we use the gravity model, which is particularly suited for the analysis of bilateral flows. The model can take on additional variables for other possible influential factors and expand to correct for econometrical imperfections like zero-inflation of data sets, spatial autocorrelation and omitted multilateral resistance.

The specific flow analyzed here is the import of canola oil for non-food use (TARIC: 15141110) into EU countries [4]. Since there are not yet any trade statistics for the import of biofuels as such available, this is the next closest commodity to analyze. Unlike other oils like palm oil, there is practically no other use for this type of oil than production of biodiesel. Therefore the canola data can be considered an appropriate proxy for biodiesel trade data.

The newest and at the same time most informative data stems from 2006. It spans trade of 39 different countries, 23 EU members and 16 non-EU countries, leading to 1,300 potential pairs of trade partners. However, by far not all of those 1,300 actually trade; only 107 do. This leads to what is known as a zero-inflated dependent variable. Unfortunately, simply eliminating the irrelevant cases of non-trading pairs is not possible because there is no easy way to distinguish between relevant and irrelevant cases.

However, since this zero-inflation can be treated as a selection bias problem, it can be resolved using the method of [7] as advised by Linders and de Groot [12]. Even with this correction the gravity model might still suffer from two more flaws: the already mentioned omitted multilateral resistance and spatial autocorrelation.

Omitted multilateral resistance is caused by the lack of inclusion or observability of countries' alternatives to trade with a particular partner. While the amount of actual trade between two partners can be measured, the amount of potential trade occurring if certain factors were different, is impossible to know. This is not a new concept to the gravity model: the distance term already tries to control for the resistance to trade. However, as [1] argue, this is not enough. There are other factors about possible trade partners which are not included in a standard gravity analysis. Therefore, they advice to use a term controlling for prices in potential other trade partner countries and transaction costs.

This would require vast amounts of data on prices, not only of goods, but also of transport and information services. Since these data are not available for the canola oil case, the proposed model here reverts to a method described in [2]. Instead of calculating the omitted multilateral resistance term from a plethora of data for all countries, a fixed effects dummy is introduced for every country. This dummy is assumed to hold constant for all unmeasurable circumstances this country faces concerning trade, thereby controlling for omitted factors causing resistance to trade.

By the assumption about their composition, these dummies rather serve as indicators for having trade at all than having more or less trade. Therefore, they are introduced in the first step of the gravity model, the so-called selection equation.

Unlike multilateral resistance, which deals with the availability of trade alternatives, a further possible problem, spatial autocorrelation, deals with trade similarities. This kind of autocorrelation stems from being part of a cluster of traders or, conversely, being remote from clusters.

As suggested by Porojan [14], to correct for the part of trade that is explained by being part of a cluster, spatial weights are included in the gravity model. These weights summarize the relationship of the importer to all its trade partners relative to all other trade partners.

Apart from the distance measured in kilometers according to a geographical approach developed in [13], the previously described IMR, further data was taken from the [3], [5], [9–11], [15], IEA [8].

3 Results

The first estimation shown in Table 1 represents the basic gravity model including only total GDP of the importer and the agricultural GDP of the exporter and the distance between them. Here, only the distance as a measure for transaction costs has a significant impact on trade and interestingly exhibits a positive coefficient. As opposed to the selection model result, distance does not seem to act as a barrier in terms of additional costs due to transportation and other distance-related transaction cost but rather the opposite. An economic explanation could be economics of scale in terms of quantities and production costs. Another explanation could be of econometrical nature: The IMR carried over much of the effect of distance from the selection equation, which now more than adjusts for the expected negative effect so that it tips to the positive. Beyond this effect, however, distance might still be a proxy for some of the effects that have to be looked into more closely in the following models. This would also explain why distance becomes insignificant once other factors are introduced: distance might proxy for some of these effects. Summarizing, the basic gravity model, even with further specifications, does not seem to explain trade well.

In the second model, the dummy variable for EU trade integration, ‘EU Both Dummy’ is added. Surprisingly, we see a negative significant coefficient indicating that the trade volume is higher if one of the partners is a non-EU country. This indicates that the border effects of the European Union seem not to be a trade inhibitor for trade partnership of two EU countries but rather for a non-EU/EU-partnership. That is consistent with the interpretation of the distance coefficient of model 1: it indicates that higher transaction costs due to distances and tariffs play a minor role in the trade volume. After all, if both countries are in the EU it also means that they are close neighbors, which was captured by distance before the introduction of the new dummy. Therefore, once this effect is taken up by the newly introduced EU-Both-Dummy, distance becomes insignificant.

Table 1 Determinants of canola oil import to the European union

Variables	Basic gravity model	*Trade integration effect	*Boifuel policy effect	*Value chain effect
Dependent variable	Log import value canola oil	Log import value canola oil	Log import value canola oil	Log import value Canola oil
Intercept	4.89***	9.40***	9.14***	11.15***
	(2.51)	(3.92)	(4.02)	(4.99)
Log GDP_i	0.23	0.39**	0.23	0.19
	(1.20)	(2.04)	(1.20)	(0.75)
Log agricultural GDP_j	−0.01	0.06	0.01	−0.22
	(−0.09)	(0.34)	(0.06)	(−0.19)
Log $distance_{ij}$	1.04***	0.40	0.26	−0.04
	(3.61)	(1.15)	(0.83)	(−0.12)
EU $Both_{ij}$ Dummy		−1.83***	−1.98***	−1.67***
		(−3.00)	(−3.51)	(−3.05)
Biofuel $quota_i$			0.90***	0.85***
			(2.87)	(2.79)
Subsidy $dummy_i$			0.98	1.18
			(1.22)	(1.45)
Log production costs $ratio_{ij}$				(0.89)
				(0.86)
Canola seed $production_i$				$-4.59 \cdot 10^{-07}$*
				(−1.88)
Canola seed $production_j$				$1.72 \cdot 10^{-07}$**
				(2.04)
Biofuel consumption $transport_i$				$8.65 \cdot 10^{-04}$***
				(2.64)
Biofuel consumption $transport_j$				$1.30 \cdot 10^{-04}$**
				(2.10)
Value weighted $distance_{ij}$	$4.16 \cdot 10^{-06}$***	$4.09 \cdot 10^{-06}$***	$3.79 \cdot 10^{-06}$***	$3.21 \cdot 10^{-06}$***
	(6.39)	(6.54)	(6.72)	(5.68)
Inverse mill's $ratio_{ij}$	−0.64***	−0.59**	−0.58**	−0.50**
	(−2.37)	(−2.27)	(−2.46)	(−2.20)
Adjusted R^2	0.13	0.15	0.20	0.24
AIC		429.22	408.94	402.12
Breush-Pagan test (p-value)		0.22	0.01	0.06
Global Moran's I Test			−2.08	
N	N = 98	N = 98	N = 98	N = 98

Denotation i = importer, j = exporter
Level of significance $\alpha = 0.1$*, $\alpha = 0.05$**, $\alpha = 0.01$***; t-value in parentheses

In the third model, biofuel quotas and a dummy for the existence of subsidizing the green industry are introduced to gauge the effect of political measures. Biofuel quotas have a positive and significant coefficient whereas the dummy for a subsidization of the green industry in the importer country is not significant. The result concerning the quota is expected since the quotas are clearly defined and their ultimate goal demands an increase in production and consumption of biodiesel. Naturally that would lead to increased imports of intermediate products, too.

Lastly, the fourth and best specified model controls for up- and downstreamed value chain stages of the biodiesel chain. To avoid multicollinearity between the possible value chain variables and endogeneity with the dependent variable, we introduced only the two extreme ends of the biodiesel chain instead of the whole chain: the production of raw material, represented by canola seed production, on the one hand and the consumption of the product, represented by liquid biofuel consumption for transport, on the other hand. Both parts of the value chain are assumed to affect the trade of canola oil; raw material because of its role for sector specific supply and liquid biofuel consumption for its role for sector specific demand. For the value chain stages, all coefficients for the importer and exporter countries are significant and have the expected sign, except for the biodiesel consumption of exporter countries exhibiting a positive coefficient. This indicates that the demand in biodiesel for transport of exporter countries might have an effect on a high level of canola oil production which is not only being consumed but also exported. However, the coefficient of the importer's biodiesel transportation sector is much higher, indicating that the pull is stronger on the importer side due to a higher biodiesel consumption level.

4 Conclusion

The main objective of this analysis was to identify the effect of different EU policies on the canola oil import for non-food use of the European Union. The estimation results have surprisingly shown a negative value of the coefficient for a dummy proxying for EU trade integration. This indicates that even though the EU trade integration has been set up to foster trade among members, members do rather import canola oil from outside of the EU. The negative relationship could possibly be explained by the import pull caused by exhausted input production of canola oil in the biodiesel value chain. The magnitude of a mandatory biofuel quota showed a positive influence on the import of canola oil. Though not surprising, it reinforces the interpretation that demand for raw or intermediate products for biodiesel cannot be satisfied within the EU. Therefore it has to be imported from non-EU countries. Accordingly, the answer to the original research question would be: political measures seem to have a positive influence on trade whereas the EU trade integration cannot be found to have an enabling effect, if not even a negative effect, on canola oil trade.

Apart from this result, no further statement about political measures could be yielded since the coefficient for a green investments subsidy dummy was insignif-

icant. This result warrants a closer look at the specific kinds of different political measures and their effectiveness.

In contrast to the interpretation of distance based on the outcome equation, the decision whether to import canola oil at all is significantly negatively affected by distance, as can be seen in the selection equation. Here, a closer look at economies of scale and resource scarcity in the importer country needs to be taken. The value chain structure, which also affects the trade volume of canola oil, has to be taken into account as well.

References

1. Anderson, J., Van Wincoop, E.: Gravity with Gravitas: A solution to the border puzzle. NBER Working Paper (2003)
2. Behrens, K., Ertur, C., Koch, W.: 'Dual' gravity: Using spatial econometrics to control for multilateral resistance. SSRN eLibrary (2007)
3. Earthtrends.:Earthtrends. Data retrieved 2nd June 2009, (2007)
4. EU Export Helpdesk:EU Export Helpdesk. Data retrieved 13th May (2009)
5. FAOSTAT: Statistical databases. Food and Agriculture Organization of the United Nations, Washington, DC, pp. 128–130 (2009)
6. Florin, M. V., Bunting, C: Risk governance guidelines for bioenergy policies. J. Cleaner Prod. **7**,106 (2008)
7. Heckman, J. J.: Sample selection bias as a specification error. Econometrica: J. Econometric Soc. 153–161 (1979)
8. IEA:IEA energy statistics. IEA. Data retrieved 29th May (2009)
9. IMF:World Economic Outlook Database. Data retrieved 21st May (2009)
10. Johnston, M., Holloway, T.: A global comparison of National biodiesel production potentials. Environ. Sci. Technol. **41**(23), 7967–7973 (2007)
11. Kutas, G., Lindberg, C., Steenblik, R.: Biofuels-at what cost?. Government support for ethanol and biodiesel in the European union: International institute for sustainable development (2007)
12. Linders, G.-J. M. & de Groot, H.L.F.: Estimation of the gravity equation in the presence of zero flows. Tinbergen Institute Discussion Paper (2006)
13. Mayer, T., Zignago, S.: Notes on CEPIIs distances measures. Centre dEtudes prospectives et dInformations Internationales (CEPII), Paris (2006)
14. Porojan, A.: Trade flows and spatial effects: The gravity model revisited. Open Economies Rev. **12**(3), 265–280 (2001)
15. REN21: REN21 database. Renewable energy policy network for the 21st Century. Data retrieved 25th May (2009)
16. Schnepf, R.D., Congress, L.O., Service, C.R.: European Union biofuels policy and agriculture: an overview. Congressional Research Service, Library of Congress (2006)

Modeling the Transformation of the German Energy System Until 2050:A Multi-Criteria, Long Term Optimization Problem with Many Constraints

Michael H. Breitner

1 Introduction

First discussions about the German, and also global transformation of the energy system date back to the early 1980s, see [7]. The "energy transformation tetrahedron" has four corners considering the most important aspects (the "energy demand tetrahedron" and the "energy supply tetrahedron" are discussed below):

- A replacement of fossil energy resources (oil, gas, coal, uranium etc.) with renewable energy sources (sun, wind, water, geothermal energy, biomass etc.) for a long-term, sustainable energy supply for all energy consuming sectors (buildings, traffic and electricity).
- A reduction of total energy demand by saving energy, e.g., for air conditioning and warm water by restoration of buildings or by changing people's customs.
- A reduction of total energy demand by an increase of efficiency using fossil/renewable energy (re)sources, e.g., for motors, turbines, energy transportation, transformation or storage in general or for power-heat cogeneration.
- Various economic, societal and environmental changes, e.g., the change of energy costs, taxes, fees, and subsidies, but also new jobs, decentralized/private energy prosumers and a better German trade balance.

Figure 1 gives a glance at the interdependency and conflicts for the transformation of the German energy system. These interdependency and conflicts require joint interdisciplinary research in the fields Economics, Management (including Finance and project/risk management), Engineering and Information Systems. The status quo and also scenarios for the transformation of the German energy system are described, e.g., in the references [1], [4, 6], and [8] for regional, decentralized success stories

M. H. Breitner (✉)
Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany
e-mail: breitner@iwi.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_55,

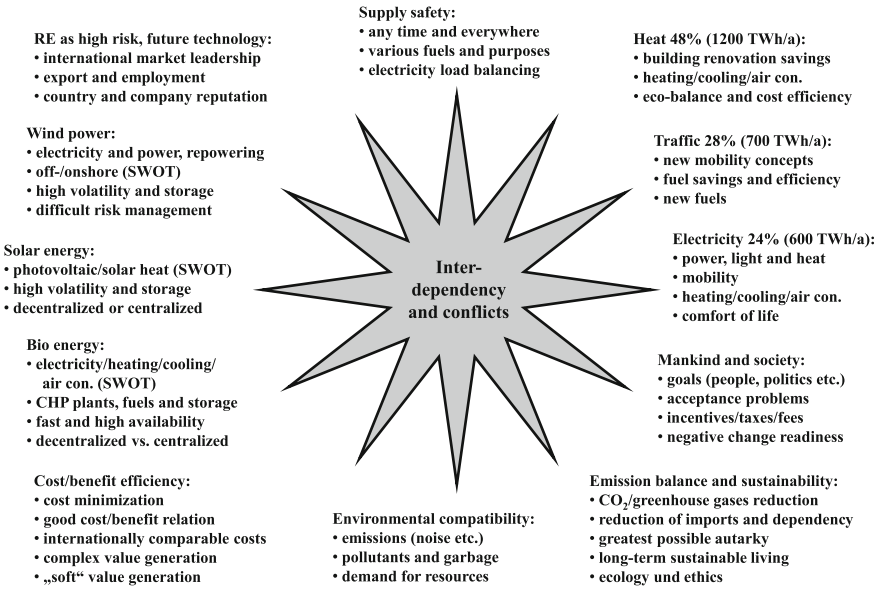


Fig. 1 Interdependency and conflicts for the transformation of the German energy system

of 100 % renewable energy supply and usage. A profound discussion of the security-political and military implications of the usage of fossil energy resources can be found in [14].

2 Towards a Simulation and Optimization Model for Germany

From the viewpoint of Operations Management and Research the transformation of the German energy system can be described as a multi-step model in time (years 2012–2050). In each time-step energy users (people, households, industry, companies etc.) minimize their energy costs for buildings, traffic and electricity. Various constraints like available air conditioners, heaters and boilers using different energy (re)sources or vehicles using different fuels must be obeyed. In each time step from year i to i+1 the energy infrastructure is optimized minimizing Germany's total energy costs, too. Again many constraints have to be obeyed. The energy infrastructure includes, e.g., various power plants and production facilities in industry, vehicles, ships, aircrafts, trains, air conditioners and warm water boilers which produce resp. consume different energy (re)sources like electricity, oil, coal, natural/bio gas, and biofuels. The optimization of Germany's energy infrastructure requires the allocation of regular investments, e.g. in new cars, but also additional investments. Control variables of the German state are, e.g., taxes, fees, and subsidies, but also other incentives, see Fig. 2. The model for Germany is based on different, interconnected submodels, e.g., for mobility and logistics, air conditioning and warm water for

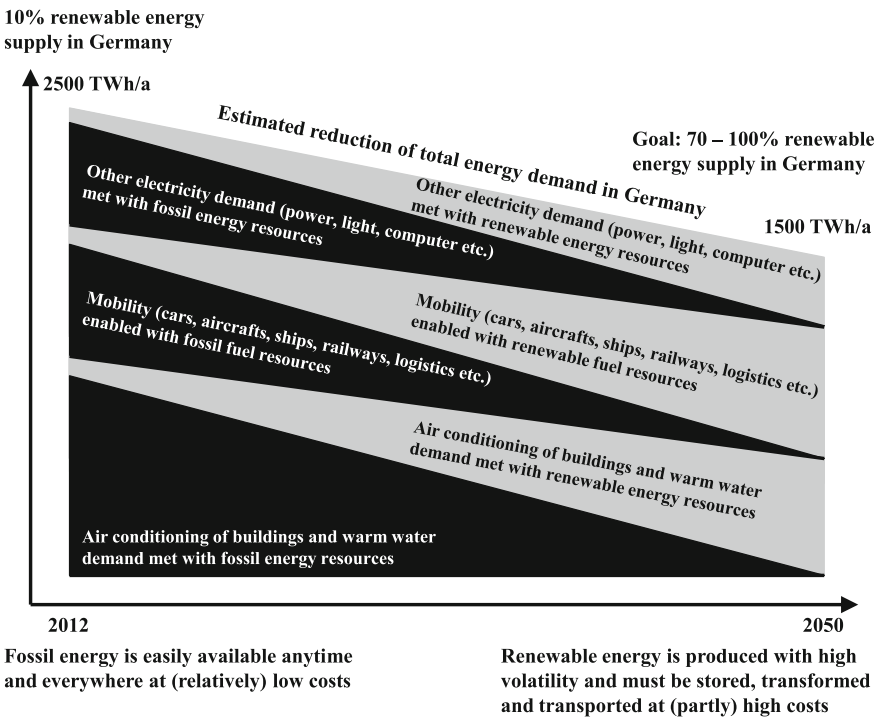


Fig. 2 Transformation of the German energy system (2012– 2050)

buildings, industrial processes and machines, and electricity. A significant reduction of greenhouse gases like carbon dioxide must be one the effects.

Following selected aspects are discussed:

1. Germany spends about 80 billion Euro p.a. for energy resources import and about 50 million Euro p.a. for domestic energy production and resources (total primary energy (re)sources are about 4,000 TWh p.a.). Imports are mainly oil (45 %; 20 % from Russia, 8 % from Great Britain, 3–4 % each from Norway, Kazakhstan, and Nigeria), natural gas (40 %; 16 % from Russia, 12 % each from Norway and The Netherlands) and black coal (15 %; 5 % from Columbia, 4 % each from the USA and Poland). Domestic energy production and resources are mainly renewable energy with about 320 TWh p.a. (13 % of the total energy demand 2,500 TWh p.a.), brown coal 140 TWh p.a. (6 %), nuclear energy 100 TWh (4 %), and natural gas 90 TWh p.a. (4 %), see [2] and [13].
2. Germany spends in total about 260 billion Euro p.a. for energy supply, i.e. 80 for fuels (traffic/mobility), 70 for electricity, 60 for gas, 30 for heating oil and 10 each for coal and long-distance heating. About 80 billion Euro p.a. are spend for mobility, 70 for heating/warm water, 60 and 50 for industrial processes and mechanical power, respectively. Total energy taxes and fees in Germany are about 70 billion Euro p.a.

3. Renewable energy production in Germany is about 320 TWh p.a. with 35 TWh biofuels (5 % of 700 TWh total for mobility), 145 TWh for heating/warm water (12 % of 1,200 TWh total used in buildings), and 140 TWh electricity (23 % of 600 TWh total electricity demand). Renewable electricity of 140 TWh p.a. comes from wind (33 %), photovoltaic (22 %), water energy (17 %), and biomass (28 %).
4. About 81 million inhabitants of Germany live in about 39 million households: for heating/warm water 50 % use natural gas, 30 % heating oil, 12 % long-distance heating, and 6 % electricity (including heat pumps). Due to isolation and replacement of old buildings today's total demand of 1,200 TWh p.a. for air conditioning/warm water decreases by 1–2 % p.a., see [5]. Today even few new buildings have a positive energy balance ("energy surplus house"), i.e. over the year an energy surplus can be used to sell electricity or—even better—to use e-cars onsite. Electric heat pumps and photovoltaic will clearly dominate in the long run.
5. The market for air conditioning and warm water energy (today primarily fossil oil and gas) will merge with the market for electricity in the long run: electricity, gas and heating transport networks must be optimized simultaneously. Additionally, the European energy market with energy imports and exports—"import cheaper renewable and also fossil energy"—will become more and more efficient, see, e.g., the Leipzig European Energy Exchange (EEX, www.eex.de) for today's successful trading of spot electricity and also electricity futures.
6. Highly efficient cogeneration and usage of electricity/power and heat must dominate energy production in the long run: both large plants with long-distance heating supply for households and industry and small, privately owned aggregates are necessary. Hundreds or thousands of these small private aggregates can be combined to a large, virtual heating power station (up to 50 MW) which can be switched on/off on demand remotely. In general Germany's energy production becomes more and more people owned. Decentralized, privately owned and operated local energy prosumers enable a democratization of the German energy system. i.e. more and more people take care themselves for their electricity and in general energy supply and demand, too. To avoid energy transport with decentralized energy prosumers (produce and consume energy at the same place), a decentralized energy storage, e.g., electricity in batteries and e-cars, hydrogen/biogas in tanks and cars and biofuels in tanks and cars, is required.
7. Smart meters which monitor and forecast private and corporate/industry electricity demand are mandatory in the next years, see, e.g., the projects Smartwatts (www.smartwatts.de) and SmartNord (www.smartnord.de). These smart meters should show time and region dependent electricity prices (demand/supply) for the next 6, 12 or 24 h. These announcements enable incentives for an automatic or a manual load balancing both for households or companies/industry.
8. Remote electricity and energy production are also possible outside Germany in favorite areas like Greece or the Sahara desert. Risky German investments are necessary in foreign countries to build up huge plants (photovoltaic or wind) and also to build up large transformation, e.g. to hydrogen or methanation, and transportation capacities. Either electricity has to be transported thousands of

kilometers in real-time, see, e.g., Desertec (www.desertec.org), or liquid/gaseous energy sources have to be transported in pipelines or by tankships.

9. The production of renewable electricity from onshore/offshore wind and sun is highly predictable today, but volatile: expensive storage or electricity plants on demand (gas) can smoothen electricity supply and e-mobility (high energy efficiency) and power-to-gas (rather low energy efficiency) can store/transform surplus electricity. Renewable energy from biomass is very flexible (mobility, electricity production on demand), but often has low energy efficiency and high energy costs unless power heat cogeneration is used. Biogas and power-to-gas processes—carbon dioxide is needed, extraction/storage from coal power plants (recovery rate 85–95 %)—can use even today's extremely high storage capacity in Germany, see [10] and [11]. Storage capacity in Germany in 2012 is only less than 1 TWh for electricity, but 200 TWh for gas and 250 TWh for oil/fuels. Electricity on demand can be supplied by stand-by gas or nuclear power plants (until 2022) which can adapt quickly. Large photovoltaic plants and wind parks, which do not run with peak performance, can be used for stand-by, too. But, investors have to be paid for stand-by instead of producing maximum energy or for local storage in batteries or local hydrogen production, see [9] and [12].
10. Available agricultural areas in Germany are insufficient for 100 % biofuel or biogas mobility in Germany: e-mobility or power-to-gas mobility or hydrogen mobility with fuel cell vehicles will be necessary for sure, see [3].
11. The transformation of the German energy system requires a bundle of many measures which have to be coordinated and combined parallely and sequentially in the next decades, see, e.g., the German EEnergy projects (www.e-energie.info). Hugh private, corporate and public investments are necessary: estimates range from 500 to 2,000 billion Euro until 2050, additionally. These huge investments and the likely increase of fossil energy prices in the next decades make an increase of electricity—and in general energy-costs likely (estimated 2–5 % increase p.a.). These increasing costs cause severe challenges both for companies/industry and private households, especially people with a lower income. A reduction of energy imports and an increase of renewable energy investments improve Germany's trade balance and provide hundred thousands of new jobs in the next years and decades. Renewable energy technology and knowhow is "high-tech" and can be exported globally, additionally.

References

1. Bundesministerium für Wirtschaft und Technologie (BMWI): Energie in Deutschland, Trends und Hintergründe zur Energieversorgung, study, www.bmwi.de/Dateien/Energieportal/PDF/energie-in-deutschland.pdf (2010)
2. Bundesverband der Energie- und Wasserwirtschaft (BDEW): Entwicklung in der deutschen Stromwirtschaft/Primärenergieverbrauch Steinkohle/Entwicklungen in der deutschen Erdgaswirtschaft im 1. Halbjahr 2012, slides, Sitzung der Arbeitsgemeinschaft Energiebilanzen, July 26–27, 2012, Berlin, download via www.ag-energiebilanzen.de (2012)

3. Eickenjäger, M.-I., Breitner, M.H.: 100 % Renewable Fuel in Germany 2050: A Quantitative Scenario Analysis. Proceedings of the international annual conference of the German OR society, Hannover, 4–7 September 2012
4. ForschungsVerbund Erneuerbare Energien (FVEE): Beiträge zur FVEE-Jahrestagung 2011: Transformationsforschung für ein nachhaltiges Energiesystem, proceedings, 12–13 Oktober, Berlin, www.fvee.de/fileadmin/publikationen/Themenhefte/th2011-2/th2011.pdf (2011)
5. Hauser, G.: Der Gebäudebereich als Motor für die Energiewende, slides, Symposium Erneuerbare Energien, Emmerthal, June 29, 2012, Institut für Solarenergieforschung (ISFH), www.isfh.de/institut_solarforschung/erneuerbare-energien-fachsymposium.php (2012)
6. Kompetenznetzwerk Dezentrale Energietechnologien (deENet): Regionale Erfolgsbeispiele auf dem Weg zu 100 % EE, Sammelband zur Posterausstellung "100 %-EE-Meile", study, download via www.100-ee.de (2010)
7. Krause, F., Bossel, H., Müsser-Reissmann, K.-F.: Energie-Wende: Wachstum und Wohlstand ohne Erdöl und Uran - Ein Alternativ-Bericht des öko-Instituts, Fischer Verlag, Stuttgart (1980)
8. Leprich, U.: Energiewende jetzt: robuste Schritte und offene Fragen, slides, Symposium Erneuerbare Energien, Emmerthal, June 29, 2012, Institut für Solarenergieforschung (ISFH), www.isfh.de/institut_solarforschung/erneuerbare-energien-fachsymposium.php (2012)
9. Solvay: Wind-Wasserstoff-Systeme - wie geht's weiter? Proceedings of the PRO H2 Technologie Forum 2011, download via www.pro-h2.de (2011)
10. Specht, M.: Power-to-Gas - Speicherung erneuerbarer Energie im Erdgasnetz, slides, Symposium Erneuerbare Energien, Emmerthal, June 29, 2012, Institut für Solarenergieforschung (ISFH), www.isfh.de/institut_solarforschung/erneuerbare-energien-fachsymposium.php (2012)
11. Sterner, M., Jentsch, M., Holzhammer, U.: Energiewirtschaftliche und ökologische Bewertung eines Windgas-Angebotes, study. Fraunhofer Institut für Windenergie und Energiesystemtechnik, IWES), Kassel (2011)
12. Verband der Elektrotechnik Elektronik Informationstechnik (VDE): Energiespeicher für die Energiewende: Speicherungsbedarf und Auswirkungen auf das übertragungsnetz für Szenarien bis 2050, study, download via www.vde.com (2012)
13. World Energy Council/Weltenergierat Deutschland: Stromerzeugung zwischen Markt und Regulierung, study, www.worldenergy.org/documents/energie_fr_deutschland_2012.pdf (2012)
14. Zentrum für Transformation der Bundeswehr: PEAK OIL: Sicherheitspolitische Implikationen knapper Ressourcen, study, Strausberg, www.peakoil.net/files/German_Peak_Oil.pdf (2010)

Part XIV
Revenue Management and Pricing

Optimization of Strategic Flight Ticket Purchase

Kathrin Armborst and Brigitte Werners

1 Introduction

In the course of contract negotiations between sales and purchase enterprises in order to support strategic and operative sourcing, it is necessary to use a structured decision support system which recommends optimal purchase decisions and offers what-if analyses. The evolution and implementation of a purchase strategy according to airline tickets is able to influence corporate profitability and perform activities due to pursue the overall business units' objectives. The developed model provides a reasonable response to revenue management [4, 5] and dynamic pricing policies of sales oriented business units.

The analysis of characteristics referring to flight ticket purchase situations and integrated types of volume discounts provides the basis for supporting in decision-making. Complex strategic purchase situations with several airlines, different routes, unit- and revenue-based volume discounts, varying tiers and targets and corresponding dependencies are considered.

Liberalization and deregulation of air transport have led to a highly competitive flight ticket market according to traditional airlines and low-cost carriers [7, 8]. New developments in the aviation sector and airline industry provide various opportunities for purchase enterprises in order to identify and implement optimal procurement strategies. Thus, traditional airlines use travel agents, contingent contracting and negotiations with sourcing business units to conclude one year contracts while low-cost carriers want to offer new possibilities without travel agents. It is important for sourcing enterprises to make optimal strategic decisions referred to their overall objectives. Current results of negotiation and supply chain research [3, 12] show the

K. Armborst · B. Werners (✉)
Faculty of Management and Economics, Ruhr-University Bochum, 44780 Bochum, Germany
e-mail: kathrin.armborst@rub.de

B. Werners
e-mail: or@rub.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_56,

importance of negotiation based decision-making. The developed decision system supports both strategic and operative purchase decisions regarding flight ticket procurement.

2 Characteristics

The proposed model supports business units' complex decision-making between traditional airlines and low-cost carriers in travel purchase. With the objective of flight cost minimization the presented decision support system conduces to assist structured decision support processes. It leads to an optimal supplier selection and allocation of airline tickets under consideration of possible strategic surplus purchasing. As a result of applied what-if-analyses, the decision support system promotes interactive model modifications and assists in (preparing for) contract negotiations between potential suppliers and the purchase enterprise considered.

Various elements of supplier selection and order quantity allocation have been treated in the literature, e.g. general literature overview with strategy focus [14] and potentials of strategic purchase [13]. Flight ticket purchase outlines a specialty in the area of supplier selection and order quantity allocation because of volume discount offers, contract negotiations and contingent contracting. Basic discount concepts [2] and literature overviews [1] have been examined, commonly relating to inventory and lot sizing models [6, 9–11].

Figure 1 shows the basic problem structure under consideration of three potential carriers and five routes. During contract negotiations other potential suppliers may

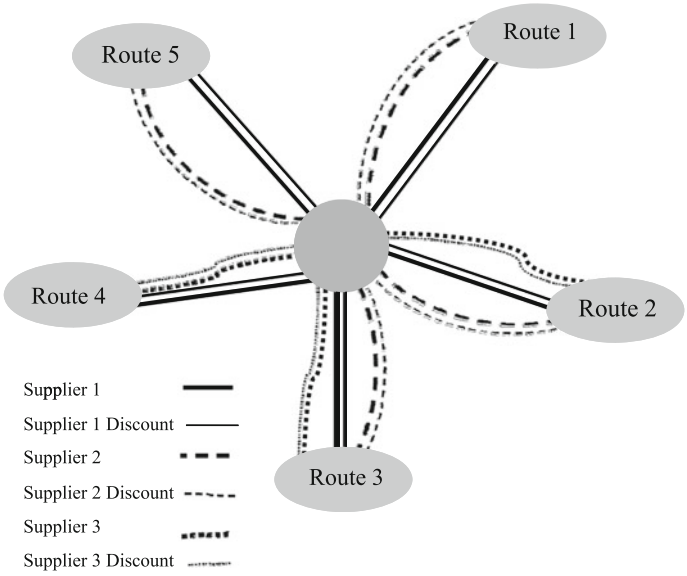


Fig. 1 Problem structure under consideration of three suppliers and five routes

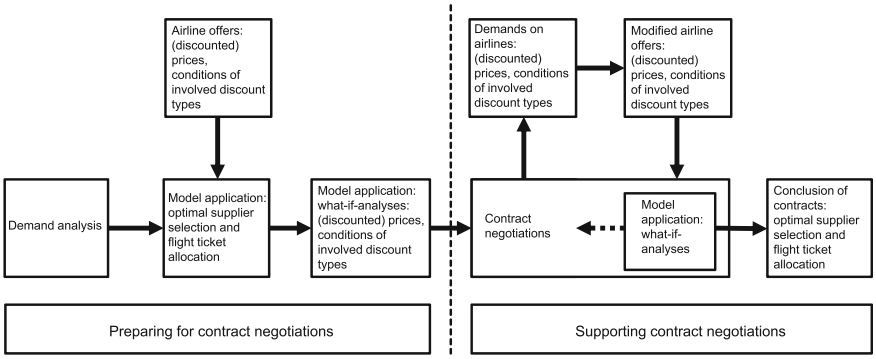


Fig. 2 Structured decision support process

be added to the portfolio. According to business units' supplier selection and product allocation it is an annual decision. In order to purchase flight tickets at discounted prices, volume discount targets have to be satisfied.

Figure 2 illustrates the structured decision process of organizing flight ticket purchase. The underlying model has got a major importance according to attend the basic decision-making process and support. Potential suppliers are able to operate on one route or more.

The current model implementation assumes that two price levels (original and discounted prices) and conditions of involved discount types (targets and tiers) as the basis for validity are known for all combinations of suppliers and routes. In addition, strategic surplus purchasing in order to minimize business units' total travel costs and segmenting order quantities for a single route to several suppliers are allowed. Because of high administration effort and low relevance, small purchase quantities are undesirable.

3 Model Structure

The decision system supports operative and strategic decision-making for a complex purchase situation by transforming the characteristics of the problem described above into a quantitative model. Different traditional airlines and low-cost carriers, several routes, unit- and revenue-based volume discounts, conditions of involved discount types, varying tiers and targets and corresponding dependencies are taken into account.

The objective function of the deterministic MILP minimizes the enterprises' total travel costs over all suppliers, routes, discount types and tiers with due regard to original and discounted prices and order quantities. The decision support system guides a decision on airline ticket allocation depending on suppliers, routes, discount types and tiers. This aspect is considered in the decision variable 'quantity'. Original and discounted prices are parameters which can be changed during contract negotia-

tions between business units and potential suppliers as a consequence of enterprises' demands on airlines after implementation and what-if-analyses.

The decision support system takes into account three different kinds of volume discounts with demands on absolute and relative quantities and total absolute revenue. For each type of volume discount there are two to three constraints which ensure that orders of discounted flight tickets are only allowed if discount targets subject to suppliers, routes and discount tiers are at least satisfied. The discount targets are given by suppliers but can be modified by suppliers after business units conduct what-if-analyses. An individual supplier is allowed to offer several absolute and relative quantity targets according to discounts but merely one discount type per route. By using a total absolute revenue target, other discount targets are impossible. Discounted prices are individually route-oriented, which means that one supplier may offer different discounts on several routes.

The discount target constraints consider this by using several binary variables depending on suppliers, routes and discount tiers with three additional constraints which summarize each corresponding binary variable to one for each supplier and route in case of quantity targets and for each supplier in case of a total revenue target.

Demand satisfaction is considered by summing up all purchased flight tickets over all suppliers, discount types and tiers for every route. It is assumed that all product requirements are known for the period considered. Strategic surplus purchasing is allowed to satisfy volume discount targets and save total travel costs. The short term decision for a certain airline product with lowest price may induce losses when volume discount margins are missed. The used volume discounts are all-unit discounts. Because of that, the unit price is uniformly applied to all airline tickets falling in the quantity discount interval for each supplier and route [1].

With a view to implement negotiation costs which vary greatly in size according to business units and decision-makers, the parameter 'relative minimum order quantity' is integrated into the developed model and two constraints ensure that small amounts of flight tickets are prohibited. Negative orders are also forbidden.

4 Results and Application

The use of the presented decision support system reveals benefits for enterprises' strategic and operative sourcing. Along with recognition profits at strategic purchase level and targeted assessments of willingness to pay and required supplier conditions, there are know-how profits which enable business units to enter into travel contract negotiations with traditional airlines and low-cost carriers with quantified demands and suggestions. In this way, it encourages an increased competition between potential suppliers to the end that cost-cutting potentials and better implementation of corporate philosophy are realized. Strategic purchasing departments are able to identify characteristic, structured sourcing relationships of different influences on strategic contract negotiations with their potential suppliers. The implementation of the decision support system in the decision-making process provides the basis for business

units' revealed willingness to pay and required purchase conditions. For this reason, enterprises which have to deal with the results of revenue management pricing policies are supported to optimize their purchase. Interactive model modifications during the business decision process enable purchase enterprises to enforce targeted analyses of offered supplier conditions according to (discounted) prices and conditions of involved discount types. Because of that reason, there is a support in contract negotiations.

The model application achieves both strategic and operative decision support. It conduces to successful analyses of strategic positioning of business units, working with business unit leaders to identify strategic options. In addition, it focuses on supporting a structured decision process and purchase optimization. Under consideration of the presented model characteristics and assumptions, the developed model reveals a wide-ranging applicability because of its transferability to different purchasing departments. The model reported in the current paper leads to additional future modeling efforts involving further aspects of strategic flight ticket purchase optimization. To support an optimal real-life decision under uncertainty and dynamics based on uncertain product demands and future interactions between current decisions and upcoming contract negotiations, the model will be extended accordingly.

References

1. Benton, W.C., Park, S.: A classification of literature on determining the lot size under quantity discounts. Eur. J. Oper. Res. **92**(2), 219–238 (1996)
2. Benton, W.C., Whybark, D.C.: Material requirements planning (MRP) and purchase discounts. J. Oper. Manage. **2**(2), 137–143 (1982)
3. Dudek, G., Stadtler, H.: Negotiation-based collaborative planning between supply chains partners. Eur. J. Oper. Res. **163**(3), 668–687 (2005)
4. Gosavi, A., Ozkaya, E., Kahraman, A.F.: Simulation optimization for revenue management of airlines with cancellations and overbooking. OR Spectr. **29**(1), 21–38 (2007)
5. Kimms, A., Klein, R.: Revenue management. OR Spectr. **29**(1), 1–3 (2007)
6. La Forge, R.L.: MRP lot sizing with multiple purchase discounts. Comput. Oper. Res. **12**(6), 579–587 (1985)
7. Mason, K.J.: The propensity of business travellers to use low cost airlines. J. Transp. Geogr. **8**(2), 107–119 (2000)
8. Nijkamp, P., Pels, E.: Developments in air transport economics: introduction. J. Air Transp. Manage. **8**(3), 137–139 (2002)
9. Reith-Ahlemeier, G.: Resource-orientated purchase planning and supplier selection—models and algorithms for supply chain optimization and e-commerce. In: Ahr, D., Fahrion, R., Oswald, M. (eds.) Operations Research Proceedings 2003, pp. 20–25. Springer, Berlin (2003)
10. Stadtler, H.: A general quantitiy discount and supplier selection mixed integer programming model. OR Spectr. **29**(4), 723–744 (2007)
11. Tempelmeier, H.: A simple heuristic for dynamic order sizing and supplier selection with time-varying data. Prod. Oper. Manage. **11**(4), 499–515 (2002)
12. Walther, G., Schmid, E., Spengler, T.S.: Negotiation-based coordination in product recovery networks. Int. J. Prod. Econ. **111**(2), 334–350 (2008)
13. Weber, C.A., Current, J.R.: A multiobjective approach to vendor selection. Eur. J. Oper. Res. **68**(2), 173–184 (1993)
14. Weber, C.A., Current, J.R., Benton, W.C.: Vendor selection criteria and methods. Eur. J. Oper. Res. **50**(1), 2–18 (1991)

Advertisement Scheduling-Revenue Management with TV Break Preference-Based Cancellations

Michael Mohaupt and Andreas Hilbert

1 Problem Statement

Broadcasting advertisements is a major (if not the predominant) source of revenues for most TV channels. In 2011, all TV stations in Germany earned 3.9bn Euro net revenues by selling ad time in commercial breaks to advertisers [14]. As these service providers are faced with limited and perishable airtime inventory (i.e. advertising time slots) that is often regulated by law on the one hand and varying demand of clients with heterogeneous preferences and willingness to pay on the other hand, a revenue management problem arise. OR techniques used to maximize total revenues are reported to generate millions of induced revenues at leading broadcasting networks in the US and France [1, 2].

Commercial breaks offered by the provider are characterized by length, broadcasting time and price. As a lot of clients only roughly specify the airtimes when advertisements should be aired, the TV channel gains some flexibility in substituting capacities (i.e. TV breaks) still satisfying customers needs [8]. Facing such multimodal or flexible products, the station has to decide simultaneously which requests to accept or to deny and when ad spots from accepted requests should be scheduled. Dealing with this multi-knapsack problem (being highly combinatorial) the provider's decisions are not only about selecting the best requests to serve but also about how best to serve a request, i.e. its assignment to the most appropriate TV break considering all requests to schedule. In addition, there may be customers that have preferences for specific ad slots (time and weekday of airing, during or between specific TV shows, short vs. long slots) with cancellation rates depending on the provider's slot assignment decisions [6, 12].

M. Mohaupt (✉) · A. Hilbert
University of Technology Dresden, 01062 Dresden, Germany
e-mail: mohaupt@wiid.wiwi.tu-dresden.de

A. Hilbert
e-mail: hilbert@wiid.wiwi.tu-dresden.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_57,

To our knowledge, there is no revenue management methodology that accounts for these preferences possibly leading to the abortion of reservations or even relationships with the service provider. Even though there are broadcasting companies that respect customers' slot preferences [10], reviewed papers only identify the chance of general cancellation [6, 8], address conflicts (e.g. elimination of competitors' ad spots in one slot) that can result in cancellations [4] or only mention the revision of rejected proposals in negotiations between client and provider [2]. Despite criticism of the less realistic assumption of no preference among the ad slots [6], neither there is a quantification of cancellation rate nor its integration in the optimization model. Even though a service provider's current availability and scheduling decisions may affect customers' (cancellation) behavior (as an endogenous factor [3, 5, 11]) and in turn can result in unused or overbooked capacity [7], the implications are ignored. The up-to-date optimization approaches do not comprise an anticipation of the customer reactions resulting from the provider's scheduling decisions albeit expressed slot preferences can be considered as an influential factor in the service valuation of the customer [9].

Additionally increasing complexity but not modeled so far, clients may differ in their worthiness and hence are grouped into different segments based on their loyalty (buying frequency), region (local vs. national), amount of ad budget, campaigns, long-term contracts or sponsoring [12]. Such as at NBC TV where about 20 % of the customers account for 80 % of the revenues [2]. Therefore, raised attention in scheduling should be paid to requests of most important clients first.

The remainder of this paper is organized as follows: In Sect. 2, we introduce an optimization model accounting for preference-based cancellations of customers from prioritized segments dependent on the provider's slot assignment decisions. In Sect. 3, the model is evaluated in a simulation and compared to traditional method. Section 4 contains closing remarks.

2 Mathematical Modeling

A TV station has $|B|$ slots $b \in B$ available (each with a capacity $\tilde{c}^b$) and offers several types of spots of varied price and length. The booking horizon is split in $T = 2$ periods. A set R of requests arrive in the first period $t = 1$. Each request $r \in R$ is typically characterized by its reward v_r and capacity c_r. Accepted requests may cancel their reservation (set C of cancellations) in the second period. At the end of each period t, the goal is to find an assignment of a subset $A \subseteq R$ of requests to the slots satisfying the problem-specific linear constraints and maximizing the objective function (1).

$$U(B, R, C, t) = max \sum_{r \in R} \sum_{b \in B} (1 - p^b_{r,t}) \cdot m_s \cdot v_r \cdot x^b_{r,t} - \sum_{r \in R} m_s \cdot v_r \cdot y_{r,t} \qquad (1)$$

$$\text{s.t.} \quad \sum_{b \in B} x_{r,t}^{b} \leq 1 \quad \forall r \in R \tag{2}$$

$$\sum_{r \in R} (1 - p_{r,t}^{b}) \cdot c_r \cdot x_{r,t}^{b} \leq \tilde{c}^{b} \quad \forall b \in B \tag{3}$$

$$x_{r,t}^{b} \in \{0, 1\}, \, y_{r,t} \in \{0, 1\} \quad \forall r \in R, \forall b \in B, t = 1, 2 \tag{4}$$

$$p_{r,t}^{b} = \begin{cases} 0 & \text{if } \mathit{preferredslot}(r) = 0 \\ & \text{or } \mathit{preferredslot}(r) = b \\ & \text{or } t = 2 \\ z & \text{otherwise} \end{cases} \quad \forall r \in R, \forall b \in B, z \in (0; 1], t = 1, 2 \tag{5}$$

$$\sum_{b \in B} x_{r,2}^{b} + y_{r,2} = \begin{cases} 0 & \text{if } \sum_{b \in B} x_{r,1}^{b} = 0 \\ 1 & \text{otherwise} \end{cases} \quad \forall r \in R \tag{6}$$

$$x_{r,2}^{b} + y_{r,2} = 1 \quad \text{iff } \mathit{preferredslot}(r) \neq 0 \text{ and } x_{r,1}^{b} = 1 \quad \forall r \in R, \forall b \in B \tag{7}$$

A binary variable $x_{r,t}^{b}$ is associated with each request r and slot b whose value is 1 when the request is allocated (at period t) to slot b and 0 otherwise (4). Each request can be assigned to at most one slot (2) noting that the capacity of each slot $\tilde{c}^{b}$ is not allowed to be exceeded (3). During the first period, requests may express a preference for a specific slot (function $\mathit{preferredslot}(r) \in B$) or not ($\mathit{preferredslot}(r) = 0$). As clients may cancel with probability $p_{r,t}^{b} = z (z \in (0; 1])$ if preferred and (in period $t = 1$) assigned slot differ from each other (5), reward and size of a request become expected values (1), (3). The final reward will further depend on the requesting customer segment $s \in \{1, \ldots, S\}$. Therefore, the reward is corrected by a multiplier m_s (1) to represent the segments' different worthiness [13].

In the second period (after receiving slot preference-based cancellations), only a scheduling of non-cancelled requests is allowed (6). Whereas requests with no preference can be re-shuffled to all available slots, requests with preference must stay in the slot assigned in $t = 1$ (7). In case the set C of cancellations is lower than expected, the provider may deny selected reservations in order to prevent overbooked slots. By introducing a nonnegative binary variable $y_{r,t}$ non-cancelled requests are either scheduled to one slot or denied (4, 6, 7). In case of denial (common preemption [12]), the provider will then loose the reward and also has to compensate the client by paying a penalty cost in the amount of the reward (1).

3 Simulation Results and Evaluation

The optimization problem described above has been implemented in MATLAB 11. General aim of the simulation is the comparison of the booking control methods *trad*, *pref* vs. $pref^2$. The former does not account for preference-based cancellations (e.g. $p^b_{r,t} = 0$). Whereas *pref* evaluates the requests based on their expected reward and size, $pref^2$ optimizes the assignment of requests based on their expected reward but real capacity consumption (i.e. without a cancellation rate). The revenue management analyst can prioritize the allocation to preferred slots by using $pref^2$ if she/he may prefer safe instead of uncertain revenues.

The TV station has 4 slots (each of 4 minutes length) available and offers 7 types of spots (numbers taken from SevenOne Media [10], a leading broadcasting network marketing German-speaking TV channels). One can see that short spots have to pay a premium (see Table 1). Over all 1,000 simulation runs, each type of spot is in demand of about 10 customers. As requests can be identified by customer ID or login name in this B2B-market [12], customized purchase history can be retrieved to estimate customer's value. Whereas half of the requests is from low-value customers ($m_3 = 1$), only 2 of 10 requests come from high-value segment ($m_1 = 2$, e.g. profitable regular clients or current high-buyers) and the rest is from customers of medium value ($m_2 = 1.5$). Half of the customers have a slot preference (equally distributed over all slots) with cancellation rate of 25 % if assigned and preferred slot differ (comparable with common modification rate (with cancellations) of up to 40 % [1]). Whereas realization of demand is deterministic at time of optimization [8], cancellations depending on assigned slots are not known a priori.

The main results of the simulation are summarized in Table 2. For evaluation of booking control's performance in isolation, total transaction-based revenues e_{trans} (with $m_s = 1$) and total segment-based revenue $e_{segment}$ (both including penalty costs

Table 1 Price and demand model

	$spot_1$	$spot_2$	$spot_3$	$spot_4$	$spot_5$	$spot_6$	$spot_7$
Reward v_r	78	120	180	206	266	300	360
Capacity c_r	6	10	16	20	26	30	36
Demand	9.8	10	9.9	9.9	10	10.1	10.1
Demand variance	3.1	3.3	3.1	3	3.1	3	3.2

Table 2 Simulation results (averaged over 1,000 runs)

Revenues	$trad_{\text{trans}}$	$pref_{\text{trans}}$	$pref^2_{\text{trans}}$	$trad_{\text{segment}}$	$pref_{\text{segment}}$	$pref^2_{\text{segment}}$
e_{trans}	9,312	10,087	10,286	9,254	10,096	10,207
e_{segment}	12,580	13,617	13,881	14,055	15,384	15,499
e_{penalty} (incl.)	0	−99	0	0	−58	0
stdev e_{trans}	446	200	114	464	162	119
stdev e_{segment}	809	670	630	1123	997	982

$e_{penalty}$) are measured. Each method can either maximize transaction or segment-based revenues during optimization (e.g. $trad_{trans}$ vs. $trad_{segment}$).

As $trad_{trans}$ does not account for preference-based cancellations, it suffers a huge loss in revenues because each slot remains with about 22 s (on average) of unused capacity. By anticipating cancellations and therefore virtually overbooking resources, $pref_{trans}$ can increase e_{trans} by 8.3 %, but it struggles with penalty costs if cancellations are lower than expected. The latter can be almost avoided when using $pref^2_{trans}$ as it foregoes uncertain customer reactions by primarily assigning requests to their preferred slots leading to an additional increase in e_{trans} of 2 % (with lower standard deviation).

By accepting customers from segment $s = 1$ with priority, transaction-based revenues may decrease slightly (if those clients request mostly long spots vs. shorter ones with price premium in some scenario runs) but relationships to customers of high value are established. Comparing $trad_{segment}$ with $pref_{segment}$ vs. $pref^2_{segment}$, the loss in $e_{segment}$ becomes even greater (9.5 vs. 10.3 %) if - in worst case - customers of high-value cancel as their preference has not been considered during assignment decisions.

Simulation results imply that preference-based revenue management is recommended when booking control decisions of the service provider are likely to have an effect on customer (cancellation) behavior [6, 12], provided that clients express slot preferences [10] and value its consideration during scheduling process [9]. In particular, the appliance of the (extended) booking control is advisable when the provider wants to establish long-term relationships to customers of high value given that requests can be classified into customer segments accordingly.

4 Conclusions

By considering effects of booking control decisions, preference-based revenue management affords both efficient capacity utilization and long-term profitable customer relationships. Simulation results emphasize the benefits of the enhanced optimization and booking control. Research remains in further estimating the cancellation rate resulting from availability and scheduling decisions of the service provider and identification of efficient optimization techniques or slot preference-based heuristics due to the huge combinatorial aspects of the problem. In addition, further simulations with other model parameters (resource, price, demand model) are required. The integration of revenue management and customer relationship management remains a research area with significant implications on improving competitiveness not only of TV stations.

References

1. Benoist, T., Gardi, F., Jeanjean, A.: Lessons learned from 15 years of operations research for French TV channel TF1 (to appear in Interfaces).
2. Bollapragada, S., Cheng, H., Phillips, M., Garbiras, M., Scholes, M., Gibbs, T., Humphreville, M.: NBC's optimization systems increase revenues and productivity. Interfaces **32**(1), 47–60 (2002)
3. Kimms, A., Müller-Bungart, M.: Revenue Management unter Berücksichtigung des Kundenwahlverhaltens [Revenue management with consideration of customer choice behavior]. Wirtschaftswissenschaftliches Studium **35**(8), 434–439 (2006).
4. Kimms, A., Müller-Bungart, M.: Revenue management for broadcasting commercials: The channel's problem of selecting and scheduling the advertisements to be aired. Int. J. Revenue Manage. **1**(1), 28–44 (2007)
5. Lindenmeier, J., Tscheulin, D.K.: The effects of inventory control and denied boarding on customer satisfaction. Tourism Manage. **29**(1), 32–43 (2008)
6. Martin, B.: Combinatorial aspects of yield management, a reinforcement learning approach. In: Proceedings of European Simulation and Modelling Conference ESM04 (2004)
7. Morales, D.R., Wang, J.: Forecasting cancellation rates for services booking revenue management using data mining. Eur. J. Oper. Res. **202**(2), 554–562 (2010)
8. Müller-Bungart, M.: Revenue Management with Flexible Products. Models and Methods for the Broadcasting Industry. Springer, Berlin (2007)
9. Ng, I.: The Pricing and Revenue Management of Services. A Strategic Approach. Routledge, London (2008)
10. SevenOne Media: Price index for advertising spots. http://www.sevenonemedia.de/preissystematik Cited 12 June 2012
11. Shen, Z.-J.M., Su, X.: Customer behavior modeling in revenue management and auctions: A review and new research opportunities. Prod. Oper. Manage. **16**(6), 713–728 (2007)
12. Talluri, K.T., van Ryzin, G.J.: The Theory and Practice of Revenue Management. Kluwer, Boston (2004)
13. Wirtz, J., Kimes, S.E., Theng, J.H.P., Patterson, P.: Revenue management: Resolving potential customer conflicts. J. Revenue Pricing Manage. **2**(3), 216–226 (2003)
14. ZAW (German Association of the Advertising Industry in Germany): Net revenues of advertising mediums in Germany. http://www.zaw.de/index.php?menuid=33 Cited 12 June 2012

An Integrated Approach to Pricing, Inventory, and Market Segmentation Decisions with Demand Leakage

Syed Asif Raza

1 Introduction

Revenue Management (RM) also known as Yield Management has been well recognized as an essential practice in many businesses. RM is loosely defined as the set of strategies adopted by a business to improve its profitability [5]. It is among the most important applications of management science and operations research [1]. Initially RM started in airline industry in 1980s and since then it has emerged as an essential practice by numerous industries that include, travel, cargo, media, utilities and retails [6]. There has been a tremendous growth in RM research, a detailed coverage of the research in RM can be found in [3] and more recently in [2]. Price differentiation is among the principal tactics of RM, in which a firm differentiates its market demand from only one segment to multiple segments. Each market segment is differentiated with price based on the willingness of customers who are attributed by the firm to that particular market segment. There are numerous examples in which a firm leads customer to different channel using differentiated prices. For example, online vs. retail store sales, in which, the firm may offer discounted prices for online sales but with less or no option of touch and feel. Whereas, retail stores sales are higher priced because the customers can interact with products and sales staff. Another commonly encountered example is that of airline ticket sales.

2 Problem Definition and Model Development

In this section, a mathematical model is proposed for RM problem for a firm offering the same product into two market segments at differentiated prices to customers and adopts a price differentiation strategy which results two market segments incurred at

S. A. Raza (✉)
College of Business and Economics, Qatar University, Doha, Qatar
e-mail: syedar@qu.edu.qa

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_58,

no additional investment. However, the segmentation is considered imperfect, and it is assumed that the customers belonging to full price market segment cannibalize to discounted market segment. The firm offers its products/services in monopoly and faces both the short sales and leftovers for unmet demand and an excess inventory respectively. The firm's problem is determine optimal integrated decision on market segmentation, pricing and order quantity. In the next sections, the models in this situation for a firm are developed assuming both the deterministic and stochastic market demands, and to the case when the demand distribution is unknown to the firm. Suppose a firm offers two product at differentiated price in two perfectly segmented markets. The market segment 1 is for the customers who are willing to pay the full price, p_1. The market segment 2 is for customers with willingness to pay the discounted price, p_2 such that $p_1 > p_2$. The product cost per unit for the firm is c. In market segment $i = \{1, 2\}$, the customers' riskless price dependent demand, $y_i(p_i) = [\alpha_i - \beta_i\, p_i]^+$, experienced by the firm in market segment i for brevity, $y_i = y_i(p_i)$, $\forall\, i = \{1, 2\}$. The riskless demand y_i follows Increasing Price Elasticity (IPE) property in non-strict sense, thus $\dfrac{\partial \eta_i}{\partial p_i} \geq 0$, where $\eta_i = -\frac{p_i \partial y_i / \partial p_i}{y_i}$, which means $\partial y_i / \partial p_i \leq 0$. In an earlier work of [5], $\beta_i, \forall\, i = \{1, 2\}$ are assumed equal, therefore, $\beta = \beta_i, \forall i = \{1, 2\}$, and also the cost of an item incurred to the firm for the distinct market segments are also assumed equal, thus $c_i = c, \quad \forall\, i = \{1, 2\}$. This research also follows this guideline. It can be inferred from riskless demand functions that maximum riskless demand the firm can experience is α_1 and it segments this demand such that $\alpha_1 - \alpha_2$ is designated to market segment 1 and α_2 is designated to discounted market segment 2. The firm's problem would be to determine the optimal market segmentation (price differentiation) that would allocate $\alpha_1 - \alpha_2$ for discounted price market segment. In order to achieve that the firm would need to set an optimal segmentation via price as differentiation strategy, υ, such that $\upsilon \leq p_1 \leq \dfrac{\alpha_1}{\beta}$, and $c \leq p_2 \leq \upsilon$. Thus, the maximum risk less demand which may be observed in the discounted market segment would be $\alpha_1 - \upsilon\, \beta$, where in this case, $\alpha_2 = \beta\, \upsilon$. Next, it is assumed that the fences (segments) observed due to this price differentiation strategy are imperfect, and therefore, there is a θ proportion of customers who belong to full price market segment and cannibalize to discounted price market segment. Although θ is independent of the prices p_1 and p_2, but, as will be identified in model, that a change in the demand incurred due to a price differentiation would impact the amount of demand that is leaked from full price market segment to discounted price which is consistent with a recent study in [5]. Now the adjusted demand for each market segment would be:

$$d_1(p_1, \upsilon, \theta) = (1 - \theta) y_1 \tag{2.1}$$

$$d_2(p_2, \upsilon, \theta) = \theta\, y_1 + y_2 \tag{2.2}$$

For brevity, $d_i = d_i(p_i, \upsilon, \theta)$, $\forall\, i = \{1, 2\}$. In addition, the firm experiences expected demand, $z_i = d_i + \mu_i$, $\forall\, i = \{1, 2\}$ where μ_i is the expected random effect [4]. The firm's constrained nonlinear optimization problem P_1 would be as

follows:

$$P_1 : \pi = \sum_{i=1}^{2} (p_i - c)\, d_i \tag{2.3}$$

subject to

$$p_2 \leq \upsilon \tag{2.4}$$

$$\upsilon \leq p_1 \tag{2.5}$$

In problem P_1, Equation 2.3 represents the deterministic revenue, $\pi = \pi(p_1, p_2, \upsilon)$. The decision variables are p_1, p_2, and υ. The optimal revenue, $\pi^*(p_1^*, p_2^*, \upsilon^*) = \underset{p_1, p_2, \upsilon}{\text{Max}}\, \pi(p_1, p_2, \upsilon)$, in which, p_1^*, and p_2^* are the optimal prices for full price and discounted price market segments. Also, υ^* is the optimal market segmentation/differentiation strategy.

Proposition 1 *For the deterministic problem P_1', the followings hold:*

1. *The optimal prices, $p_i^*, \forall\, i = \{1, 2\}$ are determined such that, $p_1^* = \frac{\mu_1}{2\beta(1-\theta)} + \frac{\alpha_1+\beta+\beta c}{2\beta}$, and $p_2^* = \frac{\alpha_1+\beta+\theta(\alpha_1-\beta(c+1))+3\beta c+\mu_1+2\mu_2}{4\beta}$*
2. *Optimal order quantities, $q_i^* = z_i^* \forall\, i = \{1, 2\}$, where $z_1^* = (1-\theta)(\alpha_1 - \beta p_1^*) + \mu_1$, and $z_2^* = \beta(p_1^* - p_2^*) + \theta(\alpha_1 - \beta p_1^*) + \mu_2$.*
3. *The optimal market segmentation is at $\upsilon^* = p_1^*$.*

`Proof`: see Appendix Following a same approach, additional models as discussed in abstract are developed, however, only the deterministic model is discussed in this paper.

3 Numerical Analysis

Adopting a numerical example from [5], $c = 5$, $\alpha_1 = 10000$, $\beta = 800$, $\mu_i = 0$, and $\sigma_i = 0, \forall i = \{1, 2\}$. For a single market segment which is also referred as an unsegmented market in this paper, optimal revenue is, $\pi = 11,250$, and the optimal price, $p^* = 8.75$, and optimal order quantity are $q^* = 3,000$ respectively. Now, as discussed in model that firm adopts a price differentiation such that out of a total, $\alpha_1 = 10,000$, it allocates, $\alpha_1 - \alpha_2$, where $\alpha_2 = \beta\, \upsilon$ for full price market segment, and thus α_2 designated to discounted price market segment with θ amount of leakage. As it is reported in Proposition 1 that an optimal segmentation would yield $p_1^* = \upsilon^*$, which means that $p_2^* < \upsilon^*$ to differentiate the market, and firm sets the price for full price market segment the same as it adopts an optimal price differentiation strategy. In Table 1, it can be noticed that as the firm exercises an integrated framework to optimize decisions on pricing, order quantity (allocation), and price differentiation it yields significant revenue gains, for example, when the firm observed perfect market segmentation (i.e., $\theta = 0$) it achieves an optimal revenue

Table 1 Impact of θ on firm's strategy for deterministic demand

θ	π^*	p_1^*	p_2^*	q_1^*	q_2^*	% Imp.
0	15762.5	10.12	7.62	1900.08	1999.96	40.11
0.1	15302.9	10.21	7.78	1648.44	2320.06	36.03
0.2	14872.1	10.29	7.93	1412.54	2621.87	32.20
0.3	14467.5	10.37	8.07	1190.95	2907.77	28.60
0.4	14087	10.45	8.20	982.37	3180.41	25.22
0.5	13728.6	10.54	8.32	785.72	3442.89	22.03
0.6	13390.6	10.63	8.44	600.00	3700.05	19.03
0.7	13072.1	10.73	8.55	424.34	3961.25	16.20
0.8	12773	10.89	8.65	257.90	4252.64	13.54
0.9	12500	11.25	8.75	100.00	4700.05	11.11

of 15762.5 which is substantially superior to the corresponding optimal revenue, 11,250, when the firm does not opt for market segmentation, thus the gain is over 40%. With an increase in the amount of proportion of demand leakage from full market segment to discounted segment, the revenue gains observed by the firm drop. When the proportion of demand leakage is $\theta = 0.9$, the optimal revenue to the firm using an optimal price differentiation is 12,500 which is still 11.11 % superior to unsegmented optimal revenue. Figure 1 shows that if the firm does not optimize its segmentation strategy and adopts an arbitrary price differentiation strategy at $\upsilon = 7$, it may fail to attain revenue gains when the leakage exceeds beyond 13 %.

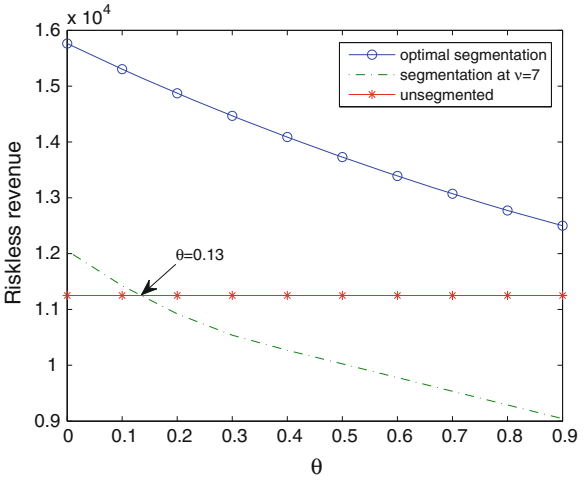


Fig. 1 Impact of optimal segmentation

4 Conclusion and Future Research Suggestions

In this paper, an integrated approach to optimal price differentiation strategy, pricing, and order quantity decisions for firm is presented. An optimal price differentiation strategy enables the firm to segment its demand. The fence that segments the market demand is considered imperfect. Due to imperfect market segmentation, the firm experiences the demand leakage from full price market segment to discounted market segment. The research develops models for RM for a firm in situations when the firm experiences price dependent deterministic demand.

Acknowledgments This publication was made possible by the support of an NPRP grant # 4-173-5-025 from the Qatar National Research Fund. The statements made herein are solely the responsibility of the author.

Appendix

Proof of Proposition 1

Using the Karush Kuhn Tucker Optimality condition we get the following Lagrange function for the problem

$$L(p_1, p_2, \upsilon, \lambda_1, \lambda_2) = (p_1 - c)z_1 + (p_2 - c)z_2 + \lambda_1(p_1 - \upsilon) + \lambda_2(\upsilon - p_2) \quad (4.6)$$

Where in Eq. 4.6, $z_1 = (1-\theta)(\alpha_1 - \beta p_1) + \mu_1$, and $z_2 = (\beta\upsilon - \beta p_2) + \theta(\alpha_1 - \beta\upsilon) + \mu_2$. Also λ_1, and λ_2 are non-negative. The First order Optimality Conditions (FOCs) are:

$$\frac{\partial L}{\partial p_1} = (p_1 - c)\frac{\partial z_1}{\partial p_1} + z_1 + \lambda_1 \leq 0,\ p_1 \geq 0,\ \left((p_1 - c)\frac{\partial z_1}{\partial p_1} + z_1 + \lambda_1\right) p_1 = 0 \quad (4.7)$$

$$\frac{\partial L}{\partial p_2} = (p_2 - c)\frac{\partial z_2}{\partial p_2} + z_2 - \lambda_2 \leq 0,\ p_2 \geq 0,\ \left((p_2 - c)\frac{\partial z_2}{\partial p_2} + z_2 - \lambda_2\right) p_2 = 0 \quad (4.8)$$

$$\frac{\partial L}{\partial \upsilon} = (1 - \theta)\beta - \lambda_1 + \lambda_2 \leq 0,\ ((1 - \theta)\beta - \lambda_1 + \lambda_2)\,\upsilon = 0 \quad (4.9)$$

$$\frac{\partial L}{\partial \lambda_1} = p_1 - \upsilon \geq 0,\ \lambda_1 \geq 0,\ (p_1 - \upsilon)\,\lambda_1 = 0 \quad (4.10)$$

$$\frac{\partial L}{\partial \lambda_2} = \upsilon - p_2 \geq 0,\ \lambda_2 \geq 0,\ (\upsilon - p_2)\,\lambda_2 = 0 \quad (4.11)$$

Consider Eq. 4.10, since $p_1 \geq \upsilon$, and $(\upsilon - p_1)\,\lambda_1 = 0$, and $\lambda_1 \geq 0$, thus the only feasible solution is, $\upsilon = p_1$, and $\lambda_1 > 0$. Now, since $\upsilon = p_1$, from Eq. 4.11, we can only have $p_2 < \upsilon$, which is possible only when $\lambda_2 = 0$. Using Eq. 4.9,

as $\upsilon > 0$, therefore,$\lambda_1 = (1-\theta)\beta$, and $\lambda_2 = 0$. To find optimal prices, p_1, and p_2, we need to solve Eqs. 4.7 and 4.8. Notice here, $\frac{\partial z_1}{\partial p_1} = -(1-\theta)\beta$, and $\frac{\partial z_2}{\partial p_2} = -\beta$. The unique optimal prices would be: $p_1 = \frac{\mu_1}{2\beta(1-\theta)} + \frac{\alpha_1 + \beta + \beta c}{2\beta}$; and $p_2 = \frac{\alpha_1 + \beta + \theta\,(\alpha_1 - \beta(c+1)) + 3\beta c + \mu_1 + 2\mu_2}{4\beta}$.

References

1. Bell, P. C.: Revenue management: That's the ticket. OR/MS Today, 1998
2. Chiang, W.C., Chen, J.C.H., Xu, X.: An overview of research on revenue management: Current issues and future research. Int. J. Revenue Manag. **1**(1), 97–128 (2007)
3. McGill, J.I., Van Ryzin, G.J.: Revenue Management: Research overview and prospects. Transp. Sci. **33**, 233–256 (1999)
4. Petruzzi, N.C., Dada, M.: Pricing and the news vendor problem: A review with extensions. Oper. Res. **47**, 183–194 (1999)
5. Philips, R. L.: Pricing and revenue optimization. Standford university press, 2005
6. Talluri, K. T., Van Ryzin, G.: The Theory and Practice of Revenue Management. Kluwer Academic Publishers, Dordrecht (2004)

Capacity Allocation and Pricing for Take-or-Pay Reservation Contracts

Mehdi Sharifyazdi and Hoda Davarzani

1 Introduction

Supply chain contracts provide the opportunity of appropriate coordination among parties by creating incentives for all the members so a decentralized supply chain acts similar to a centralized one. Among different contract types, capacity reservation contract is one of the least investigated ones related to option contracts [3]. Take-or-pay is the main category of capacity reservation contracts where buyer reserves the capacity before the demand is realized and he will be penalized for any part of the reserved capacity which has is not demanded [4]. This type of contract is mainly considered by high tech [2, 4] and transportation industry [5]. This type of contract is also related to capacity allocation literature where the contract is considered from seller's point of view. For instance, Cao et al. [1] formulated the problem of containers capacity allocation when there is stochastic demand. With the spread of contracting all over the world, there has been a growing interest in examining different contract choices. This research aims to analyze take-or-pay contracts for a two-tier supply chain with stochastic demand. The analysis is from seller perspective which is a unique approach in the literature. The proposed contract is investigated in two stages; first a non-linear optimization model is solved from buyer's perspective to decide on the amount of capacity to be reserved, in second level, the model is solved from seller's perspective considering multiple buyers and the solution of first stage. The upper-level of this model helps supplier to choose price and maximum available capacity for each buyer. The employed approach for solving the model is bi-level optimization which is suitable for game context with leader and follower.

M. Sharifyazdi (✉)
Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden
e-mail: mehdi.sharifyazdi@chalmers.se

H. Davarzani
Department of Industrial management and logistics, Lund University, Lund, Sweden
e-mail: hoda.davarzani@tlog.lth.se

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_59,

Even though the leader determines the basic parameters to make decision, they are not completely predefined in real business environment. Leader has the possibility to predict follower's decision based on the defined parameters, so she can design them in a way to maximize her profit regarding expected decision of buyer. In the following section, after a brief review on motivation of the study, the proposed structure of the take-or-pay reservation contract is formulated in the form of an optimization model. This section is followed by introducing the proposed meta-heuristic algorithm and numerical examinations. Finally the paper concludes in a succinct summary of the results and discussion potential future research.

2 The Contract Optimization Model

Assume there is a supplier who assigns a capacity of c units to a buyer with x units of demand, where x is a non-negative continuous random variable with probability density function $f(x)$. Without loss of generality assume that the normal price of the capacity is 1 per unit. According to the capacity reservation contract, if the buyer books b units of the supplier's capacity ($0 \leq b \leq c$), she only have to pay for αb units in advance, where $0 \leq \alpha \leq 1$. α is called *prepayment parameter*. Furthermore, when the buyer is paying in advance, she only pays β per unit of capacity, where $0 \leq \beta \leq 1$. β is called *discount parameter*. Therefore, the buyer only pays $\alpha\beta b$ in advance when booking b units of capacity. Based on the contract, after realization of demand, if it is not more than αb units, then the buyer do not have to pay anything else. When the demand is higher than αb but still not higher than the booking level (b), then the buyer can buy the capacity needed in addition to the prepaid αb units, for the same discounted price of β per unit. However, if she still needs more, i.e. more than b units of capacity, she must buy the additional needed units, up to the possible $c - b$ units, for the non-discounted price of 1 per unit. Now, the buyer's problem is that if she wants to cover as much demand as possible, given c, α and β by the supplier, what is the optimal b which results in the minimum purchasing cost. To solve the above stated problem, let g be the buyer's total cost. Then, the problem can be formulated as:

$$\min_b g\big(b, c, \alpha, \beta, f(.)\big) = \int_0^{\alpha b} \beta\alpha b f(x)\,\mathrm{d}x + \int_{\alpha b}^{b} [\beta(x - \alpha b) + \beta\alpha b] f(x)\,\mathrm{d}x + \int_b^c (x - b) f(x)\,\mathrm{d}x + \int_c^{\infty} (c - b) f(x)\,\mathrm{d}x \quad (1)$$

$$\text{subject to} \quad 0 \leq b \leq c \quad (2)$$

where $F(x)$ is the cumulative distribution function of x. To find the optimal b, represented by b^*, that maximizes g the first and the second derivatives of g with

respect to b have to be analysed. After performing simplifications, the results are as follows:

$$\frac{dg}{db} = \beta\alpha F(\alpha b) - (1-\beta)[1 - F(b)] \tag{3}$$

$$\frac{d^2g}{db^2} = \beta\alpha^2 f(\alpha b) + (1-\beta) f(b) \tag{4}$$

Since $\frac{d^2g}{db^2} \geq 0$, $\frac{dg}{db}$ is a non-decreasing function of b. Therefore, g is a convex function of b. Let θ be the point where $g(b)$ is minimum. Then, $\frac{dg}{db}|_{b=\theta} = 0$. Then, from (3)

$$(1-\beta)F(\theta) = 1 - \beta - \beta\alpha F(\alpha\theta) \tag{5}$$

$$\text{and since } \frac{dg}{db}|_{b=0} = \beta - 1 \leq 0, \text{ we have } \quad b^* = \min(\theta, c) \tag{6}$$

Now, consider the case where a supplier sells capacity to multiple buyers. In order to maximize her revenue, she wants to know how much capacity, discount parameter and prepayment parameter should be offered to each buyer. Assume there are n buyers. The capacity, prepayment parameter and discount parameter offered by the supplier to the buyer $i (i = 1 \ldots n)$ are represented respectively by c_i, α_i and β_i. Each buyer i reflects to the suppliers offer by reserving b_i units of capacity. Let $\mathbf{c} = (c_1, c_2, \ldots, c_n)$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$. Also, let the total capacity of the supplier be C and the whole capacity have to be shared among the buyers. Furthermore, assume that the supplier has to pay a penalty of s per unsold unit of the capacity after realization of the buyers' demand. This latter assumption makes the supplier to maximize her capacity utilization and revenue at the same time. The above stated problem can be formulated in the form of the following bi-level optimization problem.

$$\max_{\mathbf{b},\mathbf{c},\boldsymbol{\alpha},\boldsymbol{\beta}} G(\mathbf{b}, \mathbf{c}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{f}(.), s) = \sum_{i=1}^{n} \Big[g\big(b_i, c_i, \alpha_i, \beta_i, f_i(.)\big) - \int_0^{\alpha_i b_i} s(c_i - \alpha_i b_i) f_i(x_i)\, \mathrm{d}x_i - \int_{\alpha_i b_i}^{c_i} s(c_i - x_i) f_i(x_i)\, \mathrm{d}x_i \Big] \tag{7}$$

$$\text{subject to} \quad \sum_{i=1}^{n} c_i = C \tag{8}$$

$$0 \leq c_i, 0 \leq \alpha_i \leq 1, 0 \leq \beta_i \leq 1, \text{for } i = 1 \ldots n \tag{9}$$

$$b_i \in \arg \min_{0 \leq \overline{b_i} \leq c_i} g\big(\overline{b_i}, c_i, \alpha_i, \beta_i, f_i(.)\big), \text{for } i = 1 \ldots n \tag{10}$$

Note that the constraint (10) can be replaced with its equivalents according to (5) and (6). The above problem is not generally a convex optimization problem. One

simple reason is that $F(x)$ is not always a convex or concave function. Moreover, we have checked a very simple case where there are only one buyer who has a uniform probability distribution between 0 and C for demand. Even in this case where c is no longer a decision variable and b can be replaced with a function of α and β, the objective function will be non-convex. Therefore, it would be proper to have a computational algorithm for the problem.

3 The Genetic Algorithm

Here, a real-valued genetic algorithm is proposed to find at least a good solution to the above model efficiently. Furthermore, the algorithm can even handle additional constraints such as lower or upper bound limitations for α_i's and β_i's. Each chromosome has three parts of the same length respectively representing $\mathbf{c}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Both $\mathbf{b}$ and G can be calculated based on these three parts, given the demand probability distributions of all buyers and s. At first, given the population size, namely N, the first generation of the chromosomes is produced. To generate each chromosome, $3n$ uniformly distributed random numbers between 0 and 1 are generated. To produce $\mathbf{c}$ the first n numbers, namely $r_1, r_2, \ldots, r_n$ are picked. Then we let

$$c_i = \left(\frac{r_i}{\sum_{i=1}^{n} r_i}\right)C \tag{11}$$

This makes the summation of c_i's equal to C. To build the rest of the genes of the chromosome, the last $2n$ random numbers are directly assigned to α_i's and β_i's. If α_i's and β_i's have bounds, a simple linear transformation must firstly be applied to the random numbers to return new random numbers in the specified ranges. However, the population composition is different in the next generations. To generate each of the next generations, at first an intermediate generation is produced which is composed of three parts: (1) n_1 (even) chromosomes generated by crossover between chromosomes of the previous generation, (2) n_2 copies of the chromosomes of the previous generation and (3) n_3 copies of the best chromosome found so far. Aafter mutation of some chromosomes, this intermediate generation will be a part of the next generation. Note that a part of the copies of the best chromosome so far ($n_4 \leq n_3$) is kept intact (of mutation) and transferred to the next generation to make sure there are at least some copies of the best chromosome. In addition to the intermediate generation, another portion of the next generation consists of new randomly produced chromosomes. To generate the intermediate generation, at first, the optimal objective value (G), that will be called *fitness* from no on, is calculated for each chromosome according to (5) and (6). Then, roulette wheel probabilities are computed for the chromosomes relative to their fitness. Then, $n_1 + n_2$ chromosomes are selected randomly according to the roulette wheel probabilities. The first n_1 selected chromosomes are set as parents who will produce the same number of offsprings for the intermediate generation. The rest are directly transferred into the intermedi-

ate generation. We have defined two types of crossover operators. In the first type, chosen with probability p_1, genes are exchanged randomly between parents as in the regular genetic algorithms. However, there is a probability, namely p_2 that all of the genes (c_i, α_i and β_i) corresponding to each buyer to be transferred together to one of the two offsprings. So, type 1 crossover has two sub-types itself. In the second type of crossover, selected with probability $1 - p_1$, each of the parts ($\mathbf{c}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) in offspring 1, comes completely from one of the parents randomly and the remaining parts go to the other offspring. Furthermore, the offsprings will be further modified in two stages. At first, each gene is given an opportunity to randomly change within a known relatively small neighborhood radius, namely ρ_1. In case after the change a gene violates the upper or lower bound limitation, it will be rounded to the closest bound. This gives the boundary values a higher chance to appear in the genes. Then, the first part of each offspring ($\mathbf{c}$) is multiplied to $\frac{C}{\sum_{i=1}^{n} c_i}$ to keep the sum of c_i's equal to C. Finally, the offsprings are transferred to the intermediate generation as the first n_1 members. For the second n_2 members of the intermediate generation, n_2 chromosomes from the previous generation are randomly selected with respect to the roulette wheel probabilities. Thereafter, n_3 copies of the best chromosome so far are also added to the intermediate generation. Next, each member of the intermediate generation, except the last n_4 ones, is mutated with a given probability, namely μ. If a chromosome is selected for mutation, at first, one of the genes is selected randomly for change. The process of change in mutation is almost the same as that at the end of crossover, but with a different neighborhood radius, namely ρ_2 which is relatively much bigger. When mutation is completed, the $n_1 + n_2 + n_3$ chromosomes of the intermediate generation are moved to the next generation. Moreover, $N - n_1 - n_2 - n_3$ new randomly generated chromosomes are added to the next generation. The whole process of producing new generations is repeated until the termination criteria is satisfied. We have two termination criteria: (1) if the number of produced generations exceeds a given maximum, namely M, and (2) if more than m generations are produced and the change in the best fitness observed during the last m generations is not greater than ϵ, where both m and ϵ are given as parameters. To check its performance, the algorithm is coded and executed in MATLAB R2007a environment for a sample problem. As for parameter setting, $N = 800, n_1 = 0.65N, n_2 = 0.25N, n_3 = 0.05N, n_4 = 0.02N, M = 3000$, $m = 30, \epsilon = 0.002$, $p_1 = 0.65$, $p_2 = 0.5$, $\mu = 0.05$, $\rho_1 = 0.5$ and $\rho_2 = 0.02$. In the sample problem, all α_i's and β_i's belong to the interval [0, 1], $s = 0$, $C = 100$ and $n = 3$. x_1, x_2 and x_3 have continuous uniform distributions respectively in the intervals [0 , 30], [20 , 40], [10 , 20]. In this problem, the optimal objective value cannot be higher than the expected total demand which is $60 = \sum_{i=1}^{3} E(x_i)$. The algorithm managed to find the optimal solution with the objective value of 60 when terminated after just 36 generations. Although algorithm is designed to find a good solution not necessarily the optimal solution, since it lets the decision variables to get boundary values with a positive probability, it is not surprising to see sometimes it finds the optimal solution. Moreover, in this sample, the algorithm gives all of the $beta_i$'s the value of 1. This is because even if no discount is given to the buyers, if

enough capacity is assigned to them (not less than the mean demand), on average they will buy as much as the mean demand and that maximizes supplier's expected revenue.

4 Conclusion and Future Research

This paper formulates take-or-pay contract with stochastic demand to optimize seller's expected revenue. Authors contribute to capacity reservation literature by addressing a new approach to model such contracts through bi-level optimization. Moreover, almost all of the previously published works in this field have considered the context of one buyer-one seller, while the proposed model is extended to multiple buyers where the seller is the leader to make decisions. As for future research, it is recommended have multiple suppliers. This gives the buyers further options when a supplier does not offer enough discount. Accordingly, this makes the decision maker supplier to have better offers for the buyers in order to remain competitive and utilize her capacity.

Acknowledgments The research is sponsored by the Sustainable Transport Initiative (STI).

References

1. Cao, C., Gao, Z., Li, K.: Capacity allocation problem with random demands for the rail container carrier. Eur. J. Oper. Res. **217**, 214–221 (2012)
2. Erkoc, M., Wu, S.D.: Managing high-tech capacity expansion via reservation contracts. Prod. Oper. Manag. **14**, 232–251 (2005)
3. Inderfurth, K., Kelle, P.: Capacity reservation under spot market price uncertainty. Int. J. Prod. Econ. **133**, 272–279 (2011)
4. Jin, M.: and David Wu, S.: Capacity reservation contracts for high-tech industry. Eur. J. Oper. Res. **176**, 1659–1677 (2007)
5. Van Norden, L., Van De Velde, S.: Multi-product lot-sizing with a transportation capacity reservation contract. Eur. J. Oper. Res. **165**, 127–138 (2005)

Part XV
Scheduling and Project Management

Two Dedicated Parallel Machines Scheduling Problem with Precedence Relations

Evgeny R. Gafarov, Alexandre Dolgui and Frédéric Grimaud

1 Introduction

The two-dedicated-parallel-machines scheduling problem is formulated as follows:

We are given a set $N = \{1, 2, \ldots, n\} = N_1 \bigcup N_2 \bigcup N_{1or2} \bigcup N_{1and2}$ of n jobs that must be processed on two machines. Jobs from the subset N_1 have to be processed on the first machine, jobs from the subset N_2 on the second one, jobs from the subset N_{1or2} can be processed on any of them, jobs from the subset N_{1and2} use both machines simultaneously. Job preemption is not allowed. Each machine can handle only one job at a time. All the jobs are assumed to be available for processing at time 0. For each job j, $j \in N$, a processing time $p_j \geq 0$ is given. Furthermore, arbitrary finish-start precedence relations $i \rightarrow j$ are defined between the jobs according to an acyclic directed graph G. The objective is to determine the starting time S_j for each job j, $j = 1, 2, \ldots, n$, in such a way that the given precedence relations are fulfilled and the makespan $C_{max} = \max_{j=1}^{n} C_j$, where $C_j = S_j + p_j$, is minimized. Denote this problem as $P2|prec, N_1, N_2, N_{1or2}, N_{1and2}|C_{max}$.

This problem originally appeared as a sub-problem of the well-known two-sided assembly line balancing problem. To define it, firstly, we describe a simple assembly line balancing problem. A single-model paced assembly line which continuously manufactures a homogeneous product in large quantities is considered (mass production). The simple assembly line balancing problem (SALBP-1) is to find an optimal

E. R. Gafarov (✉)
Institute of Control Sciences of the Russian Academy of Sciences,
Profsoyuznaya Street 65 , Moscow, Russia 117997
e-mail: axel73@mail.ru

A. Dolgui · F. Grimaud
Ecole Nationale Superieure des Mines, FAYOL-EMSE CNRS:UMR6158, LIMOS,
42023 Saint-Etienne, France
e-mail: dolgui@emse.fr

F. Grimaud
e-mail: grimaud@emse.fr

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_60,

line balance for a given cycle time c, i.e., to find a feasible assignment of given operations to stations in such a way that the number of stations used m reaches its minimal value. The SALBP-1 is defined as follows.

Given a set $N = \{1, 2, \ldots, n\}$ of operations and K stations (machines) $1, 2, \ldots, K$. For each operation $j \in N$ a processing time $t_j \geq 0$ is defined. The cycle time $c \geq \max\{t_j,\ j \in N\}$ is given. Furthermore, finish-start precedence relations $i \to j$ are defined between the operations according to an acyclic directed graph G. The objective is to assign each operation $j,\ j = 1, 2, \ldots, n$, to a station in such a way that:

- number $m \leq M$ of stations used is minimized;
- for each station $k = 1, 2, \ldots, m$ a total load time $\sum_{j \in N_k} t_j$ does not exceed c, where N_k—a set of operations assigned to a station k;
- given precedence relations are fulfilled, i.e., if $i \to j,\ i \in N_{k_1}$ and $j \in N_{k_2}$ then $k_1 \leq k_2$.

SALBP-1 is NP-hard in the strong sense. For surveys on results for SALBP-1, see [1]. There exists a special electronic library http://www.assembly-line-balancing.de of benchmark data for this problem.

In contrast with SALBP-1, in the two-sided assembly line balancing problem of type 1 (TSALBP-1) instead of single stations, pairs of opposite stations are disposed in parallel. They work simultaneously at opposite sides of the same workpiece. Operations have to be performed on either a side of the line or can require both sides simultaneously.

While for SALBP-1 all jobs from a set N_k where $\sum_{j \in N_k} t_i \leq c$ can be processed on the single station, for TSALBP-1 the question appears: Is it possible to process all jobs from a set N_l where $c < \sum_{j \in N_l} t_i \leq 2c$ on a pair of opposite stations? So, the problem $P2|prec, N_1, N_2, N_{1or2}, N_{1and2}|C_{max}$ is obtained.

In this paper we consider the special case of the problem, where $N_{1or2} = N_{1and2} = \emptyset$, which is denoted as $P2|prec, N_1, N_2|C_{max}$. A similar problem without precedence relations was considered in [2], where jobs are assigned to the machine in advance and an incompatibility relation was defined over the tasks which forbids any two incompatible tasks to be processed at the same time.

The rest of the paper is organized as follows. In Sect. 2 some complexity results for special subcases are presented. Approximation results are discussed in Sect. 3. This paper is finishing up in Sect. 4 with the conclusion.

2 Complexity Results

Denote by $P2|chain, N_1, N_2|C_{max}$ a special subcase of the problem, where G consists only chains of jobs and by $P2|prec, p_j = 1, N_1, N_2|C_{max}$ a special subcase with equal-processing-times of jobs.

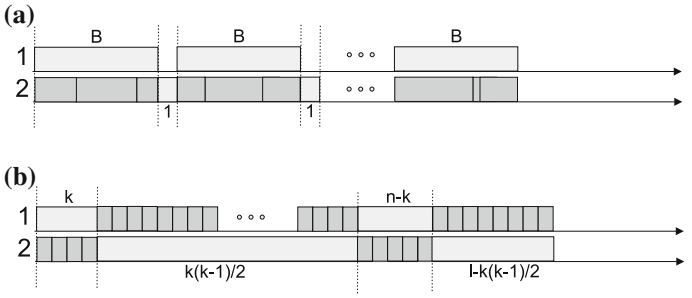


Fig. 1 Examples

3-Partition problem:

A set $N = \{b_1, b_2, \ldots, b_n\}$ of $n = 3m$ positive integers is given, where $\sum_{i=1}^{n} b_j = mB$ and $\frac{B}{4} < b_j < \frac{B}{2}$, $j = 1, 2, \ldots, n$. Does there exist a partition of N into m subsets $\overline{N}_1, \overline{N}_2, \ldots, \overline{N}_m$ such that each subset consists exactly three numbers and the sum of the numbers in each subset is equal, i.e.,

$$\sum_{b_j \in \overline{N}_1} b_j = \sum_{b_j \in \overline{N}_2} b_j = \cdots = \sum_{b_j \in \overline{N}_m} b_j = B?$$

Lemma 1 *$P2|chain, N_1, N_2|C_{max}$ is NP-hard in the strong sense.*

Proof We give a reduction from the 3-Partition problem. Given an instance of the 3-Partition problem with $3m$ numbers. Construct an instance of $P2|chain, N_1, N_2|C_{max}$ with $5m - 1$ jobs. The first $3m$ jobs are independent, $p_j = b_j$, $j = 1, 2, \ldots, 3m$, and there is a chain of jobs $3m + 1 \to 3m + 2 \to 3m + 3 \to \cdots \to 5m - 1$, where $p_j = B$, $j = 3m + 1, 3m + 3, \ldots, 5m - 1$ and $p_j = 1$, $j = 3m + 2, 3m + 4, \ldots, 5m - 2$. Furthermore, $N_1 = \{3m + 1, 3m + 3, \ldots, 5m - 1\}$ and $N_2 = \{1, 2, \ldots, 3m, 3m + 2, 3m + 4, \ldots, 5m - 2\}$. See Fig. 1a.

If and only if the instance of the 3-Partition problem has the answer "YES", there is a schedule in which a subset of jobs which corresponds to the set $\overline{N}_i$ is processed in parallel with the job $3m + (2i - 1)$, $i = 1, 2, \ldots, m$. Starting times $S_{3m+2i-1} = (B + 1)(i - 1)$, $i = 1, 2, \ldots, m$, and $S_{3m+2i} = Bi + (i - 1)$, $i = 1, 2, \ldots, m - 1$. For such the schedule $C_{max} = mB + m - 1$. □

We can present the similar reduction from a decision version of SALBP-1 to $P2|prec, N_1, N_2|C_{max}$. In the decision version, we have to answer the question, whether there is a line balance with m stations.

For this reduction there is a subset of jobs which corresponds to the set of operations from SALBP-1 with the processing times $p_j = t_j$, $j = 1, 2, \ldots, n$, and the same precedence relations. Furthermore, a chain of long and short jobs $n + 1 \to n + 2 \to n + 3 \to \cdots \to n + 2m - 1$ is given, where $p_j = c$, $j = n + 1, n + 3, \ldots, n + 2m - 1$ and $p_j = 1$, $j = n + 2, n + 4, \ldots, n + 2m - 2$.

If and only if the instance of SALBP-1 has the answer "YES", there is a schedule for which $C_{max} = mc + (m - 1)$. So, the following Lemma is proven.

Lemma 2 *Decision version of SALBP-1 can be reduced to* $P2|prec, N_1, N_2|C_{max}$ *in polynomial time.*

Since to solve TSALBP-1, solution methods for SALBP-1 can be used, where $P2|prec, N_1, N_2, N_{1or2}, N_{1and2}|C_{max}$ has to be solved as a subproblem, it seems to be interesting that SALBP-1 can be reduced to a special case of $P2|prec, N_1, N_2|C_{max}$. Furthemore, there are instances of SALBP-1 for which any known Branch and Bound algorithm with a lower bound computed in polynomial time can not solve instances with $n \geq 60$ operations in appropriate time [3]. So, it seems to be inadvisable to try to construct an effective solution algorithm for the general case of the problem under consideration.

Clique Problem:

Given a graph $G = (V, E)$ and an integer k, does G have a clique (i.e., a complete subgraph) on k vertices?

Lemma 3 $P2|prec, p_j = 1, N_1, N_2|C_{max}$ *is NP-hard in the strong sense.*

Proof We give a reduction from the Clique problem. We introduce a job J_v for every vertex $v \in V$ and a job J_e for every edge $e \in E$, with $J_v \rightarrow J_e$ whenever v is endpoint of e. Denote $\overline{n} = |V|$ and $l = |E|$. The processing times of all the jobs equal 1. Jobs $J_v \in N_2, \forall v \in V$, and $J_e \in N_1, \forall e \in E$. We also add the chain of jobs $\overline{n}+1 \rightarrow \overline{n}+2 \rightarrow \overline{n}+3 \rightarrow \overline{n}+4$, where $p_{\overline{n}+1} = k,\ p_{\overline{n}+2} = k(k-1)/2,\ p_{\overline{n}+3} = n - k,\ p_{\overline{n}+4} = l - k(k-1)/2$ and $\overline{n}+1, \overline{n}+3 \in N_1,\ \overline{n}+2, \overline{n}+4 \in N_2$. See Fig. 1b.

If and only if the instance of clique problem has the answer "YES", there is a schedule for which $C_{max} = \overline{n} + l = \sum_{i=\overline{n}+1}^{\overline{n}+4} p_i$. Denote the clique by $G'(V', E')$. Jobs $J_v,\ v \in V'$, are processed in parallel with the job $\overline{n}+1$. Jobs $J_e,\ e \in E'$, are processed in parallel with the job $\overline{n}+2$. Jobs $J_v,\ v \in V \setminus V'$ are processed in parallel with the job $\overline{n}+3$. Jobs $J_e,\ e \in E \setminus E'$ are processed in parallel with the job $\overline{n}+4$.

If there is no clique of size k, then after scheduling of k jobs $J_v,\ v \in V$, we will be able to schedule no more than $k(k-1)/2 - 1$ jobs $J_e,\ e \in E$, in parallel with the job $\overline{n}+2$.

The jobs $\overline{n}+1, \overline{n}+2, \overline{n}+3, \overline{n}+4$ can be substituted for chains of $k,\ k(k-1)/2,\ \overline{n}-k$, and $l - k(k-1)/2$ equal-processing-time jobs, respectively, i.e., the special case, where $p_j = 1$, is received.

So, the Lemma is proven. □

As a consequence from Lemma 3, for the special case $P2|prec, p_j = 1, N_1, N_2|C_{max}$ the approximation ratio of polynomial time algorithms is not less than $2/n$ and there is no FPTAS (fully-polynomial time approximation schema) for the special case.

Denote the problem, where preemptions of jobs are allowed by $P2|prec, pmtn., N_1, N_2|C_{max}$.

Corollary 1 $P2|prec, pmtn., N_1, N_2|C_{max}$ is NP-hard in the strong sense.

Denote $C^*_{max}(pmtn.)$—the minimal makespan for the problem with preemptions.

Lemma 4 *[4] For the problem* $P2|prec, N_1, N_2|C_{max}$ *an inequality* $\frac{C^*_{max}}{C^*_{max}(pmtn.)} < 2$ *holds and there is an instance for which* $\frac{C^*_{max}}{C^*_{max}(pmtn.)} \approx 2$.

3 Approximation by List Scheduling Algorithm

To solve problems with precedence relations (e.g., SALBP-1, $P2|prec|C_{max}$) enumeration schemas based on the well-known List Scheduling (LS) Algorithm are usually used. The problem $P2|prec, N_1, N_2|C_{max}$ can be solved by an algorithm based on LS as well. Different dominations rules are used in LS, which define the jobs' choice, e.g., choose a job with the maximal processing time among ready to be scheduled jobs (LPT), or choose a job which belongs to a critical path (CP) etc.

The question appears, which approximation ratios LS with different domination rules has. Let us denote the optimal objective function value by C^*_{max} and the objective function value for the solution constructed by LS with domination rules α by $C_{max}(LS_\alpha)$.

It is obvious that for the problem $P2|prec, N_1, N_2|C_{max}$, we have $\frac{C_{max}(LS_\alpha)}{C^*_{max}} < 2$, since

$$\frac{1}{2}\sum_{j\in N} p_j \le C^*_{max}, C_{max}(LS_\alpha) < \sum_{j\in N} p_j.$$

For some problems it can be useful to know the worst possible active schedule constructed by LS. Such problems with opposite optimality criteria have both theoretical and practical significance. For the problem under consideration, we can note such a problem with opposite optimality criteria, namely to maximize the makespan where only active schedules[1] are considered, by $P2|prec, N_1, N_2|C_{max} \to \max$. Unfortunately, the proposed maximization problem is strongly NP-hard, too.

Lemma 5 *[4]* $P2|chains, N_1, N_2|C_{max} \to \max$ *is NP-hard in the strong sense.*

We show that approximation ration of LS with the following domination rules is ≈ 2: CP—critical path rule (choose a job which belongs to a critical path [1]), LPT–choose a job with the maximal processing time, MS—choose a job with the maximal number of immediate successors.

Lemma 6 *[4] There are instances for which*

$$\frac{C_{max}(LS_\alpha)}{C^*_{max}} \approx 2,\ \alpha \in \{CP, LPT, MS\}.$$

[1] For which there is no job which can be shifted to an earlier starting time without violating precedence or resource constraints.

We conjecture that the same relation is true for other rules α computed in polynomial time.

4 Conclusion

In this paper, we present some complexity and approximation results for the two-dedicated-parallel-machines scheduling problem with precedence relations to minimize makespan, which is a sub-problem of two-sided assembly line balancing problem. The presented results shows that the two-machines problem is not easier than well-known SALBP-1, i.e., there is no Branch and Bound algorithm with a polynomial time computed Lower Bound that solve instances of a special case even for $n = 60$ jobs in appropriate time. For the future research a question appears: if there is a constant a, $1 < a < 2$, in which the problem is either approximable or not.

Acknowledgments The authors are grateful to Chris Yukna for his help regarding the English presentation.

References

1. Scholl, A.: Balancing and Sequencing of Assembly Lines. Physica, Springer, Berlin (1999)
2. Lushchakovaa, I.N., Strusevich, I.N., Vitaly, A.: Scheduling incompatible tasks on two machines. Eur. J. Oper. Res. **200**(2), 334–346 (2010)
3. Gafarov, E.R., Dolgui, A.: Hard special case and other complexity results for SALBP-1. Research Report LIMOS UMR CNRS, p. 6158 (2012)
4. Gafarov, E.R., Dolgui, A.: Two customized parallel machines scheduling problem with precedence relations. Research Report LIMOS UMR CNRS (2012)

MILP-Formulations for the Total Adjustment Cost Problem

Stefan Kreter, Julia Rieck and Jürgen Zimmermann

1 Introduction

Within the scope of project scheduling, subject to general temporal constraints, resource leveling problems aim at minimizing the variations in resource utilizations. A special resource leveling problem is the "total adjustment cost problem" that considers the cumulative costs arising from increasing or decreasing the requirements of resources (cf. [7]). Thus, many practical applications may be handled, where costs occur if a continuous resource utilization is not guaranteed, e.g. at heat treatment, sintering, or laser welding processes.

The paper is organized as follows: In Sect. 2, we describe an interesting application, where the resources represent different kinds of manpower and a change in the number of engaged workers induces high transport costs due to the fact that workers must be brought to their workplaces. Section 3 is devoted to describe the "total adjustment cost problem" from a theoretical point of view. In Sect. 4, we introduce two different mixed-integer linear programming (MILP) formulations for solving the problem. In a performance analysis, the two formulations are tested extensively and the computation times are compared to each other (cf. Sect. 5).

S. Kreter (✉) · J. Rieck · J. Zimmermann
Department for Operations Research, Clausthal University of Technology,
38678Clausthal-Zellerfeld, Germany
e-mail: stefan.kreter@tu-clausthal.de

J. Rieck
e-mail: julia.rieck@tu-clausthal.de

J. Zimmermann
e-mail: juergen.zimmermann@tu-clausthal.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_61,

2 Resource Leveling in Practice

A practical application for the "total adjustment cost problem" can be found in the offshore industry. Around 450 drilling rigs are positioned in the North sea and offer jobs to diverse specialists. For example, there are maintenance specialists that assess the machines and make sure that all moving parts are in a good condition, as well as offshore managers that supervise the processes carried out by workers. The activities may be split in individual tasks, e.g. maintenance specialists perform the safety-related check on engines, valves, and pipes. Between these tasks, there are several temporal constraints resulting from technological and organizational sequences. Usually, high-qualified specialists have to be brought to the platform by helicopters, and this is a very expensive transport option. For these people, a continuous employment in which successive tasks are executed without breaks has to be applied in order to minimize the transport costs. Hence, the problem should be modeled as a "total adjustment cost problem" subject to general temporal constraints.

3 Problem Description and Structural Properties

We assume that projects in question consist of n real activities that have to be carried out without interruption, and two fictitious activities, 0 and $n + 1$, which describe the start and completion of the project, respectively. In order to depict projects, we make use of activity-on-node networks $N = (V, A; \delta)$, where V represents the set of nodes (i.e. activities), A represents the set of arcs (i.e. time lags between activities) and δ describes the arc weights. We denote the start time of activity $i \in V$ by $S_i \geq 0$ and assume that every project begins at time zero, i.e. $S_0 := 0$. Then, the start time of the project completion, S_{n+1}, equals the project duration. The duration of activity $i \in V$ is indicated by $p_i \in \mathbb{Z}_{\geq 0}$, where $p_0 = p_{n+1} := 0$, and $p_i > 0$ otherwise. The time lags ensure that each activity has to be started no earlier than the project beginning and must be completed by project termination at the latest. If activity j cannot be started earlier than $d_{ij}^{min} \in \mathbb{Z}_{\geq 0}$ time units after the start of activity i, i.e. $S_j - S_i \geq d_{ij}^{min}$, an arc $\langle i, j \rangle$ with weigth $\delta_{ij} := d_{ij}^{min}$ is inserted into the network. If activity j must be started no later than $d_{ij}^{max} \in \mathbb{Z}_{\geq 0}$ time units after activity i, i.e. $S_j - S_i \leq d_{ij}^{max}$, a backward arc $\langle j, i \rangle$ with weight $\delta_{ji} := -d_{ij}^{max}$ is integrated. Hence, a prescribed project completion deadline $\overline{d} \geq 0$ may be considered by adding arc $\langle n+1, 0 \rangle$ with $\delta_{n+1,0} = -\overline{d}$ to N. With ES_i (LS_i) we denote the earliest (latest) feasible start time of activity i with respect to the minimum and maximum time lags of the project. Moreover, $\mathscr{R}$ identifies the set of renewable resources required for carrying out the project activities and r_{ik} represents the amount of resource $k \in \mathscr{R}$ used constantly during execution of activity i.

A sequence of start times $S = (S_0, S_1, \ldots, S_{n+1})$, where $S_i \geq 0$, $i \in V$, and $S_0 = 0$, is termed a "schedule". A schedule is said to be feasible if it satisfies all temporal constraints. The set of all feasible schedules is denoted by $\mathscr{S}_T$ and

represents a convex polyhedron. Given some schedule S, the set of points in time $0 \leq t \leq \overline{d}$ at which at least one activity i is started or completed, is denoted by DT (decision times concerning schedule S). Furthermore, the set of activities in progress at time t (the "active set") is given by $\mathscr{A}(S,t) := \{i \in V | S_i \leq t < S_i + p_i\}$. Consequently, $r_k(S,t) := \sum_{i \in \mathscr{A}(S,t)} r_{ik}$ represents the total amount of resource k required for those activities in progress at time t. Let $\sigma_t := \max\{\tau \in DT | \tau < t\}$ be the preceding decision time of some $t \in DT \setminus \{0\}$, in that case

$$\Delta r_{kt} := \begin{cases} r_k(S,t) - r_k(S,\sigma_t), & \text{if } t > 0 \\ r_k(S,0) & \text{otherwise} \end{cases}$$

represents the jump difference in the resource profile of resource $k \in \mathscr{R}$ at time $t \in DT$. Moreover, let $\Delta^+ r_{kt} := (\Delta r_{kt})^+$ and $\Delta^- r_{kt} := (-\Delta r_{kt})^+$ be the increase and decrease, respectively, in utilization of resource k at time t, where $(z)^+ := \max(z, 0)$. Parameters c_k^+ and c_k^- denote the costs for changing the resource utilization of k by one unit. We assume that the condition $c_k^+ = c_k^- := c_k$ holds and therefore only positive jumps in the resource profiles have to be considered. Then, the problem of finding an optimal schedule for the "total adjustment cost problem" subject to general temporal constraints may be formulated as follows:

$$\text{Minimize} \quad f(S) := 2 \sum_{k \in \mathscr{R}} c_k \sum_{t \in DT} \Delta^+ r_{kt} \tag{1}$$

$$\text{subject to} \quad S_j - S_i \geq \delta_{ij} \qquad \langle i, j \rangle \in A \tag{2}$$

$$S_0 = 0. \tag{3}$$

Constraints (2) guarantee that the minimum and maximum time lags are satisfied. Equation (3) sets the project start time to zero.

The problem at hand bears analogy to the so called "classical resource leveling problem" (cf. [2]), in particular, objective function (1) is locally concave as well. Thus, there always exists a quasistable schedule that will be optimal for the problem under consideration. As has been shown by [4], each quasistable schedule marks an extreme point of some order polytope and is integer-valued (due to the integer input parameters $p_i, \delta_{ij}, i, j \in V$). Based on this consideration, a population based local search procedure has been developed for the "classical resource leveling problem" (cf. [1]). We adapted this heuristic to the total adjustment cost problem and used the solutions as upper bounds in our computational study.

4 MILP-Formulations

In this section, we present two different MILP-formulations for solving the "total adjustment cost problem". The first formulation, MILP-1, is based on a discretization of the time horizon that was applied by [5]. Here, binary variables x_{it} allocate a

feasible start time $t \in W_i := \{ES_i, \ldots, LS_i\}$ to each activity $i \in V$, i.e.

$$x_{it} := \begin{cases} 1, \text{ if activity } i \text{ starts at time } t \\ 0, \text{ otherwise.} \end{cases}$$

Furthermore, we introduce auxiliary variables $\Delta^+ r_{kt} \geq 0$, which indicate the positive adjustment for resource $k \in \mathscr{R}$ at time $t \in T := \{0, \ldots, \overline{d} - 1\}$. With binary variables x_{it}, the requirements of resource k, at time t for some schedule S, can be specified by $r_k(S, t) = \sum_{i \in V} r_{ik} \sum_{\tau=\max\{ES_i,\, t-p_i+1\}}^{\min\{t,\, LS_i\}} x_{i\tau}$. The problem of finding an optimal schedule for the "total adjustment cost problem" may then be formulated as follows:

$$\text{Minimize} \quad 2 \sum_{k \in \mathscr{R}} c_k \sum_{t=0}^{\overline{d}-1} \Delta^+ r_{kt}$$

subject to

$$\sum_{t \in W_i} x_{it} = 1 \qquad i \in V \tag{4}$$

$$\sum_{t \in W_j} t\, x_{jt} - \sum_{t \in W_i} t\, x_{it} \geq \delta_{ij} \qquad \langle i, j \rangle \in A \tag{5}$$

$$\Delta^+ r_{kt} \geq \sum_{i \in V} r_{ik} \sum_{\tau=\max\{ES_i,\, t-p_i+1\}}^{\min\{t,\, LS_i\}} x_{i\tau} - \sum_{i \in V} r_{ik} \sum_{\tau=\max\{ES_i,\, t-p_i\}}^{\min\{t-1,\, LS_i\}} x_{i\tau} \qquad \begin{matrix} k \in \mathscr{R} \\ t \in \{0, \ldots, \overline{d} - 1\} \end{matrix} \tag{6}$$

$$x_{00} = 1 \tag{7}$$

Constraints (4) guarantee that each activity only starts at one point in time $t \in W_i$. Since $S_i = \sum_{t \in W_i} t\, x_{it}$ for all $i \in V$, inequalities (5) ensure that the temporal constraints will be satisfied. Constraints (6) measure the positive jumps in the resource profiles at each point in time. Finally, condition (7) sets the start time for the project to zero.

The second formulation, MILP-2, involves binary variables indexed by start and end times (henceforth called events) of the activities (cf. [3], where the resource-constrained project scheduling problem with only precedence constraints is considered). Therefore, the number of possible events E can be restricted to twice the number of real activities plus two for fictitious activities 0 and $n + 1$, or the project completion deadline, i.e. $E := \min\{2n + 2, \overline{d}\}$. Let $\mathscr{E} := \{0, 1, \ldots, E - 1\}$ be the set of events, then MILP-2 uses binary variables

$$XS_{ie}(XC_{ie}) := \begin{cases} 1, \text{ if activity } i \text{ starts (ends) at event } e \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

For every event $e \in \mathscr{E}$, the corresponding start time is represented by variable $t_e \in [0, \overline{d}]$. Additionally, continuous variables $\Delta_{ke}^+ \geq 0$ indicate the positive adjustment of resource k at event e. The "total adjustment cost problem" may then be specified by:

$$\text{Minimize} \quad 2 \sum_{k \in R} c_k \sum_{e \in \mathscr{E}} \Delta_{ke}^{+} \tag{9}$$

subject to

$$\sum_{e \in \mathscr{E}} XS_{ie} = \sum_{e \in \mathscr{E}} XC_{ie} = 1 \qquad i \in V \tag{10}$$

$$XS_{00} = XC_{00} = 1, \quad t_0 = 0 \tag{11}$$

$$t_{e+1} - t_e \geq 1 \qquad e \in \mathscr{E} \setminus \{E-1\} \tag{12}$$

$$t_e + p_i\, XS_{ie} - p_i \left(1 - XC_{if}\right) \leq t_f \qquad e, f \in \mathscr{E} : f > e \quad i \in V \tag{13}$$

$$t_e + p_i\, XS_{ie} + 2\overline{d} \left(1 - XC_{if}\right) - \overline{d} \left(XS_{ie} - XC_{if}\right) \geq t_f \qquad e, f \in \mathscr{E} : f > e \quad i \in V \tag{14}$$

$$\sum_{f=0}^{e} XC_{if} + XS_{ie} \leq 1 \qquad e \in \mathscr{E} \quad i = 1, \ldots, n : p_i > 0 \tag{15}$$

$$ES_i\, XS_{ie} \leq t_e \qquad e \in \mathscr{E} \quad i \in V \tag{16}$$

$$LS_i\, XS_{ie} + \overline{d}\,(1 - XS_{ie}) \geq t_e \qquad e \in \mathscr{E} \quad i \in V \tag{17}$$

$$XC_{ie}\,(ES_i + p_i) \leq t_e \qquad e \in \mathscr{E} \quad i \in V \tag{18}$$

$$XC_{ie}\,(LS_i + p_i) + \overline{d}\,(1 - XC_{ie}) \geq t_e \qquad e \in \mathscr{E} \quad i \in V \tag{19}$$

$$t_f - t_e - \delta_{ij} \geq \left(XS_{ie} + XS_{jf} - 2\right) 2\overline{d} \qquad e, f \in \mathscr{E} \quad \langle i, j \rangle \in A \tag{20}$$

$$\Delta_{ke}^{+} \geq \sum_{i=1}^{n} r_{ik}\,(XS_{ie} - XC_{ie}) \qquad e \in \mathscr{E} \quad k \in \mathscr{R} \tag{21}$$

Constraints (10) ensure that the start (respectively the completion) of any activity i is assigned to exactly one event e. Equations (11) set the start time of the project to time zero. Constraints (12) ensure that two successive events are one time period apart. If activity i starts at event e and ends at event f, then $t_f = t_e + p_i$ must be satisfied (cf. (13)–(14)). Constraints (15) guarantee that the completion-event of real activity i has a greater number than the start-event. With inequalities (16)–(20) the given temporal constraints are fulfilled, and the positive adjustment of resource k at event e is estimated with constraints (21).

5 Computational Study and Conclusion

In order to analyze the quality of the MILP-formulations, problem instances introduced by [6] are taken into account that contain up to 50 activities and 1, 3, or 5 renewable resources. In addition, we considered different project deadlines $\overline{d} := \alpha\, ES_{n+1}$ with $\alpha \in \{1.0, 1.5\}$. For both formulations, the standard solver CPLEX 12.4 was used on an Intel Core(TM) i7 CPU X990 with 3.47 GHz and 24GB RAM under Windows 7.

Table 1 Computation times and numbers of instances solved within 3 hours

Instances	MILP-1		MILP-2		Instances	MILP-1	
	t_{cpu}	$\text{inst}_{<3h}$	t_{cpu}	$\text{inst}_{<3h}$		t_{cpu}	$\text{inst}_{<3h}$
rlp-10-1-10	0.020	40	135.819	40	rlp-15-3-10	0.127	40
rlp-10-1-15	0.481	40	5824.282	26	rlp-15-3-15	1076.231	40
rlp-10-3-10	0.023	40	372.839	40	rlp-20-3-10	4.107	40
rlp-10-3-15	5.429	40	8855.333	12	rlp-20-3-15	5277.980	24
rlp-10-5-10	0.034	40	547.793	40	rlp-30-3-10	61.037	40
rlp-10-5-15	8.725	40	9778.847	5	rlp-50-3-10	5118.426	23

Table 1 summarizes the results obtained, where each line contains solutions for 40 instances, denoted by the number of activities, resources, and the parameter α. Column $\text{inst}_{<3h}$ designates the number of instances solved to optimality within a time limit of 3 h and column t_{cpu} indicates the average computation times in seconds (each unsolved instance is considered with a run time of 3 h).

For all instances with 10 real activities, model MILP-2 performs very bad in contrast to model MILP-1. In particular, only 43 out of 120 instances with $\alpha = 1.5$ are solved to optimality. That is why we skip the results of MILP-2 for instances with more than 10 activities. Model MILP-1 performs well for all instances with tight project completion deadlines, even 23 out of 40 instances with 50 real activities and three renewable resources are solved to optimality within 3 hours.

Model MILP-1 involves $\beta = \sum_{i \in V} |W_i|$ binary decision variables and $\gamma = |V|^2 + |\mathscr{R}|\,|T|$ constraints. Both numbers, β and γ, depend on the length of the time horizon, therefore CPLEX need significantly more time to receive an optimum for instances with $\alpha = 1.5$. Model MILP-2 contains $\beta = 2|V||\mathscr{E}|$ binary variables and $\gamma = 2\,|V| + |\mathscr{E}|\,(5\,|V| + |\mathscr{R}| - 2) + |\mathscr{E}|^2\,(|V|^2 - |V| + 1) + 2$ constraints. Although the numbers β and γ are independent of the scaling of the time axis, the model works worse. Since each event can be positioned at any point in time, the possibilities for placing an event are very high. During our computational study, we found out that MILP-2 can be improved by setting the number of events to the project completion deadline $\overline{d}$. Then, the solver has the chance to assign one integer-valued start time to each event. However, in that case, MILP-2 is approaching to MILP-1.

Future work will include the development of other model formulations and heuristic solution approaches for the problem.

References

1. Ballestin, F., Schwindt, C., Zimmermann, J.: Resource leveling in make-to-order production: modeling and heuristic solution method. Int. J. Oper. Res. **4**, 50–62 (2007)
2. Burgess, A., Killebrew, J.: Variation in activity level on a cyclical arrow diagram. J. Ind. Eng. **13**(2), 76–83 (1962)

3. Koné, O., Artigues, C., Lopez, P., Mongeau, M.: Event-based MILP models for resource-constrained project scheduling problems. Comput. Oper. Res. **38**(1), 3–13 (2011)
4. Neumann, K., Schwindt, C., Zimmermann, J.: Project scheduling with time windows and scarce resources. Springer, Berlin (2003)
5. Pritsker, A., Watters, L., Wolfe, P.: Multi-project scheduling with limited resources: a zero-one programming approach. Manage. Sci. **16**, 93–108 (1969)
6. Rieck, J., Zimmermann, J., Gather, T.: Mixed-Integer Linear Programming for Resource Leveling Problems. Eur. J. Oper. Res. **221**, 27–37 (2012)
7. Zimmermann, J.: Heuristics for resource levelling problems in project scheduling with minimum and maximum time lags. Technical Report WIOR-491, University of Karlsruhe (1997).

Minimizing Weighted Earliness and Tardiness on Parallel Machines Using a Multi-Agent System

S. Polyakovskiy and R. M' Hallah

1 Introduction

This paper studies the parallel machine scheduling problem $Rm \left| d_j \right| \sum \alpha_j E_j + \beta_j T_j$, where a set $N = \{1, \ldots, n\}$ of n jobs, available at time zero, is to be scheduled on m non-identical parallel machines. A job $j \in N$ is characterized by its positive integer due date d_j, processing time p_{ij} on machine i, $i = 1, \ldots, m$, a per unit cost of earliness α_j, and a per unit cost of tardiness β_j. All n jobs are ready at time zero and job preemption is not allowed. Furthermore, machines can be idle. The objective is to identify a schedule that minimizes $\sum_{j=1}^{n} (\alpha_j E_j + \beta_j T_j)$, the sum of weighted earliness and tardiness (WET) of the n jobs, where $E_j = \max\{0, d_j - C_j\}$ and $T_j = \max\{0, C_j - d_j\}$ are the earliness and tardiness of job j, $j \in N$, and C_j is its completion time. Job j is early if $C_j < d_j$, on-time if $C_j = d_j$, and tardy otherwise. This problem is NP hard [1, 2]. It is approximately solved here using a multi-agent system (MAS). Section 2 details MAS. Section 3 displays the computational results. Finally, Sect. 4 is a summary.

2 A Multi-Agent System (MAS)

Agent technology is well suited for discrete optimization due to its modular structure [4]. In contrast to traditional optimization procedures which are serial, iterative or treelike, MAS is decentralized. Decentralization partitions a complex problem into

S. Polyakovskiy (✉)
FZI Forschungszentrum Informatik, Karlsruhe, Germany
e-mail: maxles@yandex.ru

R. M'Hallah
Kuwait University, P.O. Box 5969, 13060 Safat, Kuwait
e-mail: rymmha@yahoo.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_62,

smaller and simpler components (agents). An agent is a hardware or software-based self-contained problem-solving entity with the key properties of autonomy, social ability, responsiveness and pro-activeness [3]. An agent perceives its environment and interacts with other agents, exhibits a goal-directed but initiating behavior, and carries out actions independently. Each agent is defined by a state, decision rules and goals. It employs the most appropriate paradigm for solving its particular problem. Agents actively interact and negotiate rather than execute some fixed procedures. Their activity defines the process' dynamics, evolution, and final structure. Each agent applies a "greedy" strategy: at every time epoch, it assesses the potential of available actions and chooses one that optimizes its reward. By striving to achieve its goal, each agent contributes to the improvement of the overall objective function.

We adopt a MAS to approximately solve $Rm\,|d_j|\sum \alpha_j E_j + \beta_j T_j$. MAS is comprised of independent intelligent agents related either to groups of already assigned jobs (G-agents) or to free jobs (I-agents). A ***G-agent*** is tagged to a machine. It creates and gradually expands a group of jobs by placing arriving jobs to the free time slots of the machine. Its goal is to use the machine rationally minimizing its WET. During the solution construction process, it competes with other G-agents to attract the best I-agents to its group. Specifically, at each time epoch, it identifies the I-agents whose ideal processing period overlaps the group's processing interval and tries to attach the job with maximal overlap.

An ***I-agent*** is free if it does not belong to any group, and busy otherwise. Free I-agents compete with each other to join a G-agent's group. They react to incoming attachment offers made by G-agents. Their goal is to enter a G-agent group such that they cause the minimal increase of the WET of the machine schedule.

MAS acts as a global agent: it supports agents' communication, runs special algorithms and guarantees convergence to a feasible solution. It has a reactive type architecture [6] with agents' behavior based on "if-then" rules. Only agents of different types can interact. They use direct informational exchange.

MAS constructs a feasible solution in two stages. The first stage defines the sets of I-and G-agents; thus constructs a partial schedule. The second stage iteratively schedules the free jobs; thus constructs a complete feasible solution with no free I-agent. In the following, we detail these two stages using the Notation of Table 1.

2.1 MAS Initialization

MAS starts by creating n free I-agents: one for each of the n jobs. It then builds its G-agents taking into account bottleneck time intervals where a large number of jobs compete for the same time slots. Jobs whose ideal processing period belongs to these bottleneck intervals are to be scheduled on a priority basis if their WET is to be minimized. Thus, MAS organizes its I-agents' competition by centering a G-agent for each cluster of I-agents with close due-dates. This choice of G-agents offers a rich variation of potential assignments of I-agents to G-agents. Specifically, MAS recruits G-agents by employing the peak clustering algorithm [7] which fixes

Table 1 Adopted notation

$\boldsymbol{G}$	The set of created G-agents.
$\boldsymbol{G_i}$	The set of G-agents assigned to machine i, $G_i \in G$.
$\boldsymbol{G[i][j]}$	A G-agent j assigned to machine i, $G[i][j] \in G_i$.
$\boldsymbol{J_{G[i][j]}}$	The set of jobs, attached to $G[i][j]$, sorted in non-decreasing order of completion times.
$\boldsymbol{S_{G[i][j]}}$	The starting time of $G[i][j]$; i.e., of the first job in $J_{G[i][j]}$.
$\boldsymbol{C_{G[i][j]}}$	The completion time of $G[i][j]$; i.e., of the last job in $J_{G[i][j]}$.
$\boldsymbol{N}$	The set of free I-agents.
$\boldsymbol{N_i}$	The set of jobs assigned to machine i, $N_i = \cup_{j \in G_i} J_{G[i][j]}$.
WET($\boldsymbol{N_i}$)	The WET of all jobs of N_i; i.e., the total WET for machine i.
WET($\boldsymbol{f, G[i][j]}$)	The minimal WET of jobs of $N_i \cup \{f\}$ when f joins $G[i][j]$.
INC($\boldsymbol{f, G[i][j]}$)	The increment of WET of $N_i \cup \{f\}$ when f is attached to $J_{G[i][j]}$.

the bottleneck intervals. It identifies the center of a cluster of jobs having close due dates by evaluating the peak function $\phi_f^1 = \sum_{f' \in N} \exp\left(-\frac{\|d_f - d_{f'}\|^2}{(.5r)^2}\right)$ for each job $f \in N$. The neighborhood size r is set empirically for different distributions of due dates or using the graphical information deduced from the peak functions whose x—axes correspond to due date values and their y—axes to their frequency.

A large ϕ_f^1 indicates a dense cluster around d_f; that is, several jobs are competing for the same time slots around d_f. MAS finds f_1^*, the center of the first ($k = 1$) cluster. f_1^* is the job having the largest peak function value: $\phi_{f_1^*}^1 = \max_{f \in N}\{\phi_f^1\}$. Then, MAS applies the procedure `Create`($G[i'][f_1^*]$) which proceeds as follows. It defines machine i' that processes f_1^* fastest: $p_{i'f_1^*} = \min_{i=1,\ldots,m}\{p_{if_1^*}\}$ and creates a new G-agent $G[i'][f_1^*]$. $G[i'][f_1^*]$ schedules itself on i' during the time interval $[d_{f_1^*} - p_{i'f_1^*}, d_{f_1^*}]$; so that its WET is nil. $G[i'][f_1^*]$ enters $G_{i'} = G_{i'} \cup \{G[i'][f_1^*]\}$, and sets its group $J_{G[i'][f_1^*]} = \{f_1^*\}$. Finally, f_1^*, which is an I-agent, marks itself busy and leaves N.

Subsequently, MAS starts an iterative search for other G-agents. On each round, it finds $f_{k+1}^* \in N$, the center of the next densest cluster. For $f \in N$, it evaluates a modified peak function $\phi_f^{k+1} = \phi_f^k - \phi_{f_k^*}^k \exp\left(-\frac{\|d_f - d_{f_k^*}\|^2}{(.5r)^2}\right)$. $\phi_f^{k+1} = 0$ for $f = f_k^*$, and takes small values for jobs with due dates near $d_{f_k^*}$. f_{k+1}^* is the job whose $\phi_{f_{k+1}^*}^{k+1} = \max_{f \in N}\{\phi_f^{k+1}\}$. If $\phi_{f_{k+1}^*}^{k+1} \geq 1$ and a machine i' can schedule f_{k+1}^* during $[d_{f_{k+1}^*} - p_{i'f_{k+1}^*}, d_{f_{k+1}^*}]$, MAS runs `Create`($G[i'][f_{k+1}^*]$) to obtain an additional G-agent. MAS stops this iterative process when $\phi_{f_{k+1}^*}^{k+1} < 1$; i.e., when f_{k+1}^* is already a part of a cluster; thus, can not be a candidate for a new cluster's center.

2.2 Computing the Incremental WET

An I-agent f computes INC(f, $G[i][j]$) via a mixed integer program (MIP)-based algorithm `ComputeMinimalWET`($G[i][j]$, $UB(z_f)$). Every time it receives an offer to join a $G[i][j]$'s group, f solves an MIP —via Cplex—to determine z_f = WET(f, $G[i][j]$) and the position $y \in Y = \{0, \ldots, \left|J_{G[i][j]}\right|\}$ where it will be inserted within $J_{G[i][j]}$. MIP has integer variables T_l, E_l and C_l, corresponding to the tardiness, earliness and completion time of job $l \in N_i \cup \{f\}$, and binary variables x_y, $y \in Y = \{0, \ldots, \left|J_{G[i][j]}\right|\}$, where $x_y = 1$ if f is inserted in position y, and 0 otherwise. In addition to standard precedence constraints, MIP puts an upper bound $UB(z_f)$ on z_f. If no upper bound is available, $UB(z_f)$ is set to ∞ in `ComputeMinimalWET`($G[i][j]$, $UB(z_f)$).

2.3 Final Solution Construction

To finalize the initial solution, MAS asks G-agents to undertake **group formation**. It iterates its requests until no I-agent is free. Initially, it sets G-agent $G[1][1]$ "active" and asks $G[1][1]$ to undertake group formation.

Consequently, $G[i][j]$ extends an invitation to a subset of free I-agents to join its group $J_{G[i][j]}$. First, $G[i][j]$ updates the set $V_{G[i][j]}$ of free I-agents whose zero-WET processing periods intersect the group's processing interval $[S_{G[i][j]}, C_{G[i][j]}]$. Second, it sorts the elements of $V_{G[i][j]}$ in non-decreasing order of the length of overlapping periods, and selects the first element of $V_{G[i][j]}$ as the I-agent f. Third and last, it invites f to join its group by sending it an attachment offer.

When invited to join $G[i][j]$, free I-agent f decides whether or not to accept the offer. To reach its decision, f undertakes a **group-joining action**. If this is its first attachment offer, f checks whether it can be scheduled on any machine $i' = 1, \ldots, m$, $i' \neq i$ during $[d_f - p_{i'}, d_f]$ (i.e., such that it has zero WET). If such a placement is possible on at least one machine, f applies procedure `Create`($G[i'][f]$) to create an additional G-agent. Otherwise (i.e., if f has already received an offer or no machine i' can schedule f with zero WET), f analyzes the $G[i][j]$'s offer. It computes INC(f, $G[i][j]$) = WET(f, $G[i][j]$) − WET(N_i), where WET(f, $G[i][j]$) = `ComputeMinimalWET`($G[i][j]$, $+\infty$) is the solution value of the MIP-based algorithm. In addition, f considers all possible potential offers that may emanate from other G-agents. It reviews every G-agent in $G \backslash \{G[i][j]\}$ whose $\left[S_{G[i'][j']}, C_{G[i'][j']}\right]$ overlaps $[d_f - p_{i'}, d_f]$. To assess the potential of its attachment to $G[i'][j']$, f runs the MIP-based function WET(f, $G[i'][j']$)= `ComputeMinimalWET`($G[i'][j']$, $UB(z_f)$) with $UB(z_f)$ = WET($N_{i'}$) + INC(f, $G[i][j]$).

A feasible WET(f, $G[i'][j']$) implies that $G[i'][j']$ can schedule f with less WET increment than $G[i][j]$. In this case, f declines the attachment offer of $G[i][j]$ and informs $G[i][j]$ of this decision. If WET(f, $G[i'][j']$) is nil then f proceeds to examine the next appropriate G-agent. In fact, if f does not obtain any value except

Table 2 MAS Results

$(CV_d, \bar{W}) \rightarrow$		(0.1,6)			(0.1,100)			(0.2,6)			(0.2,100)		
m	n	$\frac{\bar{z}}{\bar{z}_{MAS}}$	$\frac{\bar{RT}}{\bar{RT}_{MAS}}$	RT_{MAS}	$\frac{\bar{z}}{\bar{z}_{MAS}}$	$\frac{\bar{RT}}{\bar{RT}_{MAS}}$	RT_{MAS}	$\frac{\bar{z}}{\bar{z}_{MAS}}$	$\frac{\bar{RT}}{\bar{RT}_{MAS}}$	RT_{MAS}	$\frac{\bar{z}}{\bar{z}_{MAS}}$	$\frac{\bar{RT}}{\bar{RT}_{MAS}}$	RT_{MAS}
2	50	1.03	150.92	1.07	1.14	133.79	1.17	1.01	177.21	0.86	1.11	168.47	0.88
	100	1.08	103.04	4.60	1.21	91.53	5.30	1.09	113.48	3.80	1.29	103.32	4.10
	250	1.16	52.16	45.90	1.35	46.77	50.90	1.27	51.12	42.40	1.57	45.01	48.40
	350	1.20	34.39	132.40	1.43	32.25	140.30	1.35	30.53	135.00	1.85	26.48	153.90
	Average	1.12	85.13		1.28	76.08		1.18	93.09		1.45	85.82	
3	50	1.00	118.27	1.50	1.08	108.39	1.57	0.98	146.88	1.10	1.01	132.81	1.11
	100	1.06	79.68	6.30	1.16	72.95	6.90	1.05	106.12	4.30	1.18	93.83	4.70
	250	1.16	48.05	49.60	1.39	41.67	57.00	1.27	52.73	39.90	1.72	49.54	42.00
	350	1.21	37.88	117.70	1.44	34.67	128.20	1.39	39.88	97.00	1.89	34.56	111.60
	Average	1.11	70.97		1.27	64.42		1.17	86.40		1.45	77.68	
4	50	0.99	100.21	1.94	1.02	91.79	2.04	0.93	117.47	1.42	0.98	119.64	1.32
	100	1.07	67.40	8.00	1.11	68.96	7.90	1.05	87.03	5.30	1.08	85.99	5.20
	250	1.17	40.12	60.70	1.39	34.62	69.70	1.29	50.38	42.00	1.66	45.70	45.60
	350	1.23	30.03	147.70	1.49	28.26	155.90	1.41	36.99	105.10	1.88	34.37	110.70
	Average	1.11	59.44		1.25	55.91		1.17	72.97		1.40	71.42	
5	50	0.97	91.97	2.35	1.00	87.91	2.27	0.90	109.21	1.61	0.89	110.86	1.51
	100	1.04	62.21	9.20	1.11	59.96	9.20	0.99	82.36	5.80	1.08	82.82	5.70
	250	1.16	34.58	72.90	1.36	32.76	76.30	1.27	49.11	43.60	1.61	45.88	46.80
	350	1.23	26.53	169.60	1.48	25.19	179.00	1.41	38.44	100.60	1.82	35.40	108.20
	Average	1.10	53.82		1.24	51.46		1.14	69.78		1.35	68.74	
10	50	0.93	79.39	3.11	0.95	87.88	2.61	0.83	113.95	1.84	0.77	96.98	1.84
	100	1.00	54.71	13.40	1.03	58.80	12.10	0.89	75.52	7.60	0.92	74.71	7.60
	250	1.15	31.86	94.00	1.30	32.30	91.90	1.18	52.67	45.90	1.35	52.82	45.90
	350	1.22	25.29	203.00	1.41	25.26	202.30	1.35	42.88	98.30	1.64	42.52	98.30
	Average	1.07	47.81		1.17	51.06		1.06	71.26		1.17	66.76	
Overall average		1.10	63.43		1.24	59.79		1.15	78.70		1.36	74.08	

nil after processing the whole set G of available G-agents, then the offer of $G[i][j]$ is the best available for f; therefore, f accepts it and passes to $G[i][j]$ the completion times of all jobs in $N_i \cup \{f\}$ along with its position $y \in Y$ in $J_{G[i][j]}$ as obtained by WET(f, $G[i][j]$).

The action of $G[i][j]$ depends on the response of f. If f declines the attachment offer, $G[i][j]$ initiates a new attachment offer and sends it to the next I-agent of $V_{G[i][j]}$. If no I-agent in $V_{G[i][j]}$ is attached, $G[i][j]$ ends its group formation action. On the other hand, if f accepts the invitation, $G[i][j]$ inserts it into $J_{G[i][j]}$ in position y and marks it as busy. Subsequently, MAS updates the schedule of i according to WET(f, $G[i][j]$) : it modifies the completion times of the scheduled jobs and recomputes WET(N_i).

When $G[i][j]$'s group-formation action is completed successfully, MAS merges every pair of consecutive G-agents of machine i into a single G-agent if there is no idle time between the completion time of the first and the starting time of the second. It then applies a local search which repositions jobs of $J_{G[i][j]}$ to further reduce the machine's WET. For all pairs l and k of jobs in $J_{G[i][j]}$ such that $C_l < C_k$, the local search considers the three alternatives: (k, J_{lk}, l), (J_{lk}, k, l) and (k, l, J_{lk}), where $J_{lk} \subseteq J_{G[i][j]}$ is the sequence of jobs positioned between l and k. It retains the sequence with the least WET. It stops its iterative search over all possible pairs (l, k) when no further reduction of WET($G[i][j]$) is possible. Subsequently, MAS asks the next agent in G to perform a group formation.

It is possible that none of the G-agents attaches a job during a complete cycle of group-formation despite the existence of free jobs. This occurs when none of the processing intervals of the G-agents overlaps the ideal processing interval of any

free job. In this case, MAS makes the first free job $l \in N$ an additional G-agent via `Create`($G[i'][l]$), and asks the first G-agent of G to undertake group formation.

3 Results

Ten problems are generated for each of the 80 sets of instances: $n = 50, 100, 250, 350$; $m = 2, 3, 4, 5, 10$; p_{ij} are normally distributed with mean 100 and standard deviation 10; d_j are normally distributed with mean $\mu_d = \frac{50n}{m}$ and standard deviation σ_d; the coefficient of variation of d_j is $CV_d = \frac{\sigma_d}{\mu_d} = 0.1$, or 0.2; α_j and β_j, $j = 1, \ldots, n$, follow a uniform$(1, \bar{W})$, $\bar{W} = 6$ and 100. Table 2 gives the average of $\frac{\bar{z}}{z_{\text{MAS}}}$, $\frac{\bar{RT}}{RT_{\text{MAS}}}$, and RT_{MAS} where z_{MAS} and RT_{MAS} (seconds) are the MAS solution value for the best fitting neighborhood size r and average runtime of MAS whereas $\bar{z}$ and $\bar{RT}$ are their counterparts obtained by the hybrid heuristic H4 of [5]. For most tested instances, the average ratios are larger than 1 indicating that the solution quality of MAS is superior to that obtained by H4 and that it was obtained much faster. Moreover, MAS outperforms H4 significantly when the penalties get larger. It is more successful in handling large instances.

4 Conclusion

This paper addresses the minimum weighted earliness tardiness parallel machine scheduling problem. The problem is approximately solved using a decentralized multi-agent system. The proposed approach is superior to existing ones in terms of solution quality and run time. It can be extended to other optimization problems.

References

1. Lee, C.Y., Kim, S.J.: Parallel genetic algorithms for the earliness-tardiness job scheduling problem with general penalty weights. Comput. Ind. Eng. **28**, 231–243 (1995)
2. Lawler, E.L.: A pseudopolynomial algorithm for sequencing jobs to minimize total tardiness. Ann. Discrete Math. **1**, 331–342 (1977)
3. Jennings, N.R., Wooldridge, M.J.: Applying agent technology. Int. J. Appl. Artif. Intell. **9**(4), 351–369 (1995)
4. Parunak, H.V.D.: Agents in overalls: experiences and issues in the development and deployment of industrial agent-based systems. Int. J. Coop. Inf. Syst. **9**(3), 209–228 (2000)
5. M'Hallah, R., Al-Khamis, T.: Minimising total weighted earliness and tardiness on parallel machines using a hybrid heuristic. Int. J. Prod. Res. **50**(10), 2639–2664 (2012)
6. Wooldridge, M.J., Jennings N.R.: Agent theories, architectures and languages: A survey. In: Proceedings of ECAI94 Workshop on Agent Theories Architectures and Languages, pp. 1–32. Amsterdam, Netherlands, 1994
7. Yager, R., Filev, D.: Essentials of Fuzzy Modeling and Control. Wiley, New York (1984)

Multiprocessor Scheduling with Availability Constraints

Liliana Grigoriu

1 Introduction

The multiprocessor scheduling problem, whether it is possible to nonpreemptively schedule a set of independent tasks on m processors to meet a given deadline (with m considered to be an input parameter), is strongly NP-hard [5], and so are most related problems. As a consequence, the study of this area has been mainly concentrating on approximation algorithms: the *largest processing time first* (LPT) algorithm was first proposed [6] and shown to generate schedules the makespans of which are within $\frac{4}{3} - \frac{1}{3m}$ the optimal makespan, and later the Multifit algorithm was introduced in [2], and shown to generate schedules which end within 13/11 the optimal schedule's makespan in [14].

Due to maintenance or failures, machines may exhibit periods of unavailability. A review of scheduling with availability constraints is given by Sanlaville and Schmidt in [12]. We focus on the static offline variant of the problem, when downtimes and tasks are known in advance. A special case of this problem is the case when all downtimes are at the beginning of the schedule, that is, when the machines start processing at different times. Lee [11], Chang and Hwang [1] give worst-case analyses for this problem when using LPT and Multifit respectively.

Given that all downtimes could be infinite, the NP-hardness of multiprocessor scheduling results in the NP-hardness of the problem of finding an approximation algorithm that ends within a multiple of the time needed by the optimal schedule, unless assumptions about the downtimes are made.

In [9], the authors make the assumption that no more than half the machines are unavailable at any time, and show that for this situation the LPT algorithm ends within twice the time needed by the optimal schedule. In [10], the result is generalized to the case when an arbitrary number of machines, $\lambda \in \{1, .., m-1\}$, can be unavailable at

L. Grigoriu (✉)
University Siegen, Hölderlinstr.3, 57076 Siegen, Germany
e-mail: liliana.grigoriu@gmail.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_63,

the same time. In that case the makespan generated by the LPT schedule is not worse than the tight worst-case bound of $1 + \frac{1}{2}[[m/(m - \lambda)]]$ times the optimal makespan.

In [13] Scharbrodt et al. give a polynomial-time approximation scheme for the problem of scheduling with "fixed" jobs, that is, jobs that have to execute at certain predefined times. These are equivalent to downtimes, except that the optimal schedule also needs to execute them. The approximation scheme is for minimizing the makespan of the schedule for all the jobs, it does not consider the number of processors as a part of the input, and there can be more than one fixed job on one machine. They also show that, when considering the processors a part of the input, the best possible worst-case bound of a polynomial approximation algorithm for this problem is greater than or equal to 1.5 unless $P = NP$. The proof presented there also applies to the situation when there is at most one fixed job (or downtime) on each machine [8].

In this paper we consider the problem of nonpreemptive scheduling of a set of independent tasks on multiple machines with periods when they are unavailable (downtimes). A polynomial algorithm which uses a polynomial time approximation scheme for the subset sum problem as a subroutine, the schedules of which end within $1.5 + \varepsilon$ for any $\varepsilon \in (1, 1/2]$ times the end of the optimal schedule when scheduling with fixed jobs, was given by Didrich and Jansen in [3].

We give a simple Multifit-based polynomial-time approximation algorithm the schedules of which end within 1.5 the end of an optimal schedule or 1.5 the latest end of a downtime when there are at most two downtimes on each machine. The complete upper bound proof is contained in [7], but we do provide details amounting to a brief outline of it in this work. We also present an LPT-based algorithm which has the same property when there is at most one downtime on each machine.

If the downtimes represent jobs that the optimal schedule also needs to execute, our algorithms guarantee a maximum completion time within 3/2 the maximum completion time of the optimal schedule. When the unavailability periods can not be considered to be jobs, but it can be determined that the latest end of a downtime occurs before the end of the optimal schedule, for example because the sum of all tasks is greater than the sum of the available processing times on all processors before the latest end of a downtime, the maximum completion time of the schedules of our algorithms is also less than 3/2 the maximum completion time of the optimal schedule.

2 Simple Polynomial Algorithms

The Multifit algorithm The Multifit algorithm was first introduced by Coffman, Garey and Johnson in [2], as a heuristic for multiprocessor scheduling that uses the bin packing algorithm first fit decreasing (FFD). It works as follows:

1. Order tasks in non-increasing order of their duration

2. Assign upper bound (ub) and lower bound (lb) for end of schedule; (example: lb = sum of task durations/number of processors, ub = sum of task durations)
3. Assign $b = (ub + lb)/2$ as deadline
4. FFD: Assign tasks in the given order in the first time slot in which they fit
5. If all tasks are successfully scheduled decrease the upper bound: $ub = b$;
6. Else increase the lower bound: $lb = b$;
7. If $ub - lb \geq \varepsilon$ loop back to step (3).

Multifit-based algorithms for scheduling with availability constraints To be used for scheduling in the presence of downtimes, Multifit needs to be enhanced by a heuristic that orders the available time slots. After the ordering, FFD can be used to assign the tasks, while the binary search can be used to determine the schedule. Two likely effective orderings emerge when considering the problem:

- Order all time slots in increasing order of their duration (we called this algorithm FFDL Multifit), and
- Order all time slots ending in a downtime (pretimes) in increasing order of their duration and then append all other time slots (posttimes) also in increasing order of their duration (we called this algorithm MMultifit).

In Fig. 1 the basic terminology is recapitulated, while in Fig. 2 the main difference between FFDL Multifit and MMultifit schedules is exemplified.

The schedules of all pretimes stay unchanged while MMultifit assigns deadlines which are greater than the latest end of a downtime. FFDL Multifit does not have this property, but it may be that ordering all time slots in increasing order is beneficial for

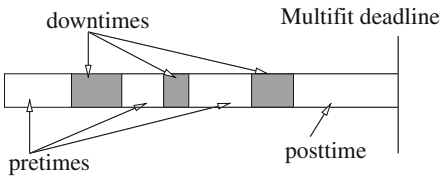


Fig. 1 Terminology

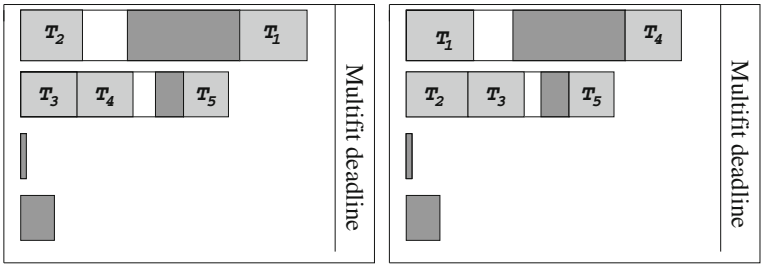


Fig. 2 Example of FFDL Multifit (*left*) and MMultifit schedules. The tasks fulfill $T_1 > T_2 > T_3 > T_4 > T_5$. FFDL Multifit puts T_1 in the smallest time slot while MMultifit assigns T_1 to the smallest pretime

the purpose of obtaining a feasable schedule. In analogy, Multifit-based algorithms for scheduling on uniform processors order all time slots in increasing order of their duration multiplied by the speed factor of the processor they are on [4, 7]. In fact, the generalization of FFDL Multifit to the problem of scheduling on uniform processors with at most one downtime on each machine reaches the 1.5 approximation bound when comparing to the end of the optimal schedule or the latest end of a downtime.

We were able to show that FFDL Multifit reaches the 1.5 approximation bound when scheduling on same-speed processors with at most two downtimes on each machine. The complete upper bound proof is given in [7], while a brief outline and some details of it are given in Sect. 3.

The time complexity of our Multifit-based algorithms is $O(n \log \mathrm{n} + (n\mathrm{km} + \mathrm{km}\log(\mathrm{km}))\log(\text{sum of task durations}))$, where n is the number of tasks and k the maximum number of downtimes on one machine. As is usual, the task durations are assumed to be given as integer multiples of some unit u. Then we can choose $\varepsilon = u$, and $lb =$ the earliest end of a downtime, and $ub =$ the earliest end of a downtime + the sum of task durations. If k is 2 or if we consider MMultifit, which only needs to generate a schedule for the pretimes once, and if $\log(km) < n$, the time complexity becomes $O(n\log n + n\mathrm{km} + nm\log(\text{sum of task durations}))$.

The LPT algorithm First introduced in [6], the LPT algorithm generates schedules which end within $\frac{4}{3} - \frac{1}{3m}$ the makespan of the optimal schedule for the problem of multiprocessor scheduling. LPT takes an ordering of the processors, orders the tasks in non-increasing order of their duration and assigns them in this order to processors as soon as they become available. The mentioned ordering of processors is used to break ties when 2 processors become available at the same time, and it is given or arbitrarily chosen at processing start.

LPT-based algorithm for scheduling with at at one downtime on each machine As described above, LPT can also be used for scheduling in the presence of machine downtimes. In fact, Hwang and Chang use LPT with no changes in their work [9, 10].

In [8], we have shown that a changed version of LPT, which establishes the processor ordering in a certain way, generates schedules which end within 1.5 times the end of the optimal schedule or within 1.5 the latest end of a downtime when there is at most one downtime on each machine. We also argued that the proof by Sharbrodt et al. [13] that no better bound can be achieved in polynomial time for a more general version of this problem if $P \neq NP$ applies to this case too.

The ordering of processors we used is described as follows: first, processors with nonempty time slots before downtimes (pretimes) are ordered in non-decreasing order of the duration of the pretimes and placed in the processor list. Then all other processors are ordered in non-decreasing order of the time at which their downtimes end and appended in this order to the processor list. Processors with no downtimes are considered to have downtimes ending at time 0. In general LPT and its variants generate a more balanced distribution of tasks to processors than Multifit and its variants, but Multifit worst-case bounds tend to be better.

3 Concepts and Methods Used in the Upper Bound Proofs

While proving the upper bound results we used several argument types, but the structure of the proof was similar in all situations.

We used the well-known method of assuming that there exists a *minimal counterexample*, that is, a problem instance for which our algorithm's schedule does not obey the upper bound, with a minimal number of processors, tasks, downtimes, and in one case minimal task lengths. A minimal counterexample exists whenever there is a counterexample, thus showing that it does not exist proved our theorems.

We also used the concept of a *compensating processor*, that is, a processor the optimal schedule of which has tasks with a greater total duration (sum of task durations) than those in the schedule produced by our algorithm, when our algorithm did not schedule the smallest task. In a minimal counterexample, the last task to be scheduled by both LPT and FFD (when failing to successfully schedule all tasks for a deadline $d \geq 3/2opt$), is the smallest task ($\overline{X}$). Here, opt stands for the maximum among the end of the optimal schedule and the latest end of a downtime. For LPT, $\overline{X}$ breaks the 3/2 bound and FFD is not able to schedule it when the deadline is d.

Thus, when considering the schedules of our algorithms for a minimal counterexample, it is meaningful to exclude $\overline{X}$. When our algorithm's schedule without $\overline{X}$ is considered, a compensating processor exists, since the optimal schedule must include all tasks.

We were then able to prove properties of the minimal counterexample and of the compensating processor which contradicted each other, resulting in our upper bound theorems being true. In most cases, weighing arguments were also used in the process.

4 Conclusion

In this paper we presented simple polynomial-time approximation algorithms for the problem of scheduling on multiple processors in the pesence of downtimes. A Multifit-based algorithm produces schedules which end within 1.5 times the end of the optimal schedule or 1.5 times the latest end of a downtime when there are at most two downtimes on each machine. We also presented an LPT-based algorithm which achieves the same bound when there is at most one downtime on each machine, and described the main concepts and methods used in the upper bound proofs. The 1.5 bound is tight in the class of polynomial algorithms for the considered problems unless $P = NP$.

Acknowledgments This research was partly funded by the Sectoral Operational Programme Human Resources Development 2007–2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/88/1.5/S/60203.

References

1. Chang, S.Y., Hwang, H.: The worst-case analysis of the MULTIFIT algorithm for scheduling nonsimultaneous parallel machines. Discrete Appl. Math. **92**, 135–147 (1999)
2. Coffman Jr, E.G., Garey, M.R., Johnson, D.S.: An application of bin-packing to multiprocessor scheduling. SIAM J. Comput. **7**(1), 1–17 (1978)
3. Diedrich, F., Jansen, K.: Improved approximation algorithms for scheduling with fixed jobs. In: Proceedings of 20th ACM-SIAM symposium on discrete algorithms (SODA), 675–684, 2009
4. Friesen, D.K., Langston, M.A.: Bounds for multifit scheduling on uniform processors. SIAM J. Comput. **12**(1), 60–69 (1983)
5. Garey, M.R., Johnson, D.S.: 'Strong' NP-completeness results: Motivation, examples, and implications. J. ACM **25**(3), 499–508 (1978)
6. Graham, R.L.: Bounds on multiprocessing timing anomalies. SIAM J. Appl. Math. **17**(2), 416–429 (1969)
7. Grigoriu, L.: Multiprocessor scheduling with availability constraints. Ph.D. thesis, Texas A&M University, 2010
8. Grigoriu, L., Friesen, D.K.: Scheduling on same-speed processors with at most one downtime on each machine. Discrete Optim. **7**(4), 212–221 (2010)
9. Hwang, H., Chang, S.Y.: Parallel machines scheduling with machine shutdowns. Comput. Math. Appl. **36**, 21–31 (1998)
10. Hwang, H., Lee, K., Chang, S.Y.: The effect of machine availability on the worst-case performance of LPT. Discrete Appl. Math. **148**(1), 49–61 (2005)
11. Lee, C.Y.: Parallel machine Scheduling with nonsimultaneous machine available time. Discrete Appl. Math. **30**(1), 53–61 (1991)
12. Sanlaville, E., Schmidt, G.: Machine scheduling with availability constraint. Acta Informatica **35**(9), 795–811 (1998)
13. Scharbrodt, M., Steger, A., Weisser, H.: Approximability of scheduling with fixed jobs. J. Sched. **2**(6), 267–284 (1999)
14. Yue, M.: On the exact upper bound of the multifit processor scheduling algorithm. Ann. Oper. Res. **24**(1), 233–259 (1990)

First Results on Resource-Constrained Project Scheduling with Model-Endogenous Decision on the Project Structure

Carolin Kellenbrink

1 Introduction

In the classical resource-constrained project scheduling problem (RCPSP, cf., e.g., [2]) the project structure is given exogenously, i.e. all jobs have to be implemented. In this paper, the RCPSP is extended by a model-endogenous decision on the project structure in order to include alternative modalities which activate specific jobs and precedence relations.

In this context, as well as in the RCPSP (cf., e.g., [6]), a project consists of $j = 1, ..., J$ jobs with a given duration of d_j periods. Among these jobs, precedence constraints have to be met. In addition, resource constraints must be considered: A job utilizes k_{jr} units of a resource r. A distinction is made between renewable and non-renewable resources. Renewable resources, e.g., machines or employees, are available with K_r units in each period. In contrast, non-renewable resources can be used with a total of K_r units only once within the whole planning horizon. An example of this kind of resource is a budget that is disposed for the whole project. The typical aim of the RCPSP is to create a schedule which minimizes the total makespan of the project.

Contrary to the multi-mode extension of the RCPSP (MRCPSP, cf., e.g., [3]), in the presented research approach not only the modes in which the jobs are implemented are determined, but it is also decided model-endogenously which jobs are implemented at all. Hence, this new class of problems is a generalization of the MRCPSP.

C. Kellenbrink (✉)
Leibniz Universität Hannover, Institut für Produktionswirtschaft,
Königsworther Platz 1,30167 Hannover, Germany
e-mail: carolin.kellenbrink@prod.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_64,

2 Requirements on the Project Structure

In projects with alternative modalities decisions have to be made regarding the implementation of the jobs. However, some requirements on the project structure need to be fulfilled so that a feasible combination of jobs is implemented.

As known from the RCPSP, some jobs are mandatory, i.e. they have to be implemented in any case. In addition,

- choices between alternative jobs have to be made, possibly
- causing the (non-)implementation of further jobs and/or
- activating further choices.

The precedence constraints are defined in the same manner as in the RCPSP. However, they only have to be met if the predecessor as well as the successor is implemented.

The scheduling of the project and the decision on the project structure depend on each other. Therefore, it is not reasonable to separate these two planning steps. Instead, the decision regarding the project structure should be made model-endogenously.

The basic idea of a mathematical model for formally describing the stated problem is to assign the jobs to suitable sets in order to consider the requirements on the project structure. These sets are also used in the genetic algorithm. Due to space limitations these sets and the model formulation are not presented in this paper.

The described problem occurs, e.g., with the regeneration of high-value capital goods. Different modes of regeneration are the cause of alternative jobs to be implemented. Another example is the turnaround at an airport, which is explained in the following.

3 Practical Example

The steps an aircraft passes through between its arrival at an airport and its next departure are called the turnaround. There are different modalities of how this turnaround can be organized, cf. [9]. The decision of which alternative is the best in a particular case depends on the provided resources and/or the available time horizon at this time. Therefore, it is expedient to determine the modalities and thereby the project structure model-endogenously.

Requirements on the project structure for a simplified turnaround are presented in Table 1. Some jobs are mandatory, e.g., cleaning the aircraft, catering and boarding. There is a choice between alternative jobs concerning the arrival. The aircraft can arrive either at the apron of the airfield or at the terminal. If the arrival at the terminal is implemented, two jobs are caused by this choice: The deboarding has to be carried out by bridge and a push-back has to take place in order to transport the aircraft back to the airfield. If in contrast the arrival at the apron is implemented, another choice is activated: The deboarding and particularly the transport of the passengers to the air-

Table 1 Requirements on the project structure for a turnaround

Requirement	**Example**
Mandatory implementation of some jobs	Cleaning of the aircraft
	Catering
	Boarding
Choices between alternative jobs	Arrival at the apron of the airfield or at the terminal
	Deboarding by foot or by bus
	Fueling with or without fire service
Jobs caused by choices made	Arrival at the terminal causes deboarding by bridge
	Arrival at the terminal causes push-back
Choices activated by choices made	Arrival at the apron of the airfield activates choice on the way the deboarding is implemented
Varying precedence constraints due to different modalities	Simultaneous implementation of boarding and fueling only if the fire service supervises the procedure

port can be realized by foot or by bus. Finally, there is a choice concerning the fueling. If the fire service supervises the fueling, parallel boarding and fueling is possible.

4 Related Literature

In [11] an extension of the MRCPSP by rework activities is presented. The rework is necessary if the original job is implemented in a predefined mode. In this approach, the rework activity consists of only one job, which in addition has to be a direct successor of the original job.

In [1] the RCPSP is extended by logical dependencies among some jobs, e.g., AND- or OR-dependencies. It is, however, assumed that the logical dependencies always go along with precedence constraints.

In research approaches regarding stochastic project networks, the project structure is not known in advance, cf. [10]. In contrast to this paper, the decision on the implementation of jobs depends on random exogenous influences.

Kuster et al. [9] addresses disruption management problems at airports with alternative process execution paths. The approach only supports decisions concerning the rescheduling but does not address the problem of how to create a new schedule in the first place.

5 Genetic Algorithm

Genetic algorithms mimic inheritance processes in biology. In the course of this solution approach each individual represents a solution for the underlying problem. The genetic algorithm starts with the initialization of a start population. Afterwards, new individuals are generated by crossover and mutation.

The genetic algorithm for the stated problem is a modification and generalization of the genetic algorithm in [4]. In order to extend that approach by the model-endogenous decision on the project structure, the representation of the individuals, the initialization of the population, the crossover and the mutation were modified. Due to space limitations, only the most important adjustment concerning the representation is presented in this paper.

For solving resource-constrained project scheduling problems an activity list is often chosen to represent an individual, cf., e.g., [4]. The activity list contains all jobs in the sequence of scheduling. The activity list is feasible if all precedence constraints are met, i.e. if each job is positioned in the activity list behind all its predecessors. A feasible activity list can be decoded to a unique schedule by the means of the serial schedule generation scheme, cf. [5]. Each job in the sequence of the activity list is scheduled as early as possible under consideration of the available capacity and the precedence constraints. With this method, an active schedule is created, i.e. no job can be started earlier without delaying another job, cf. [7].

For solving the resource-constraint project scheduling problem with model-endogenous decision on the project structure an activity list representation is not sufficient. In addition to the determination of the sequential arrangement of the jobs, decisions on the actual implementation of jobs have to be made. Therefore, the activity list is supplemented by adding an implementation list. The elements of the implementation list are coded binary. If the associated job on the activity list is implemented, the element of the implementation list equals 1, otherwise 0.

For the explanation of this representation a simple project consisting of six jobs is introduced. In this project a choice has to be made between the implementation of job 3 and job 4. A feasible project structure as well as the precedence constraints for this scheduling problem are shown in Fig. 1. A possible solution is represented by the following individual:

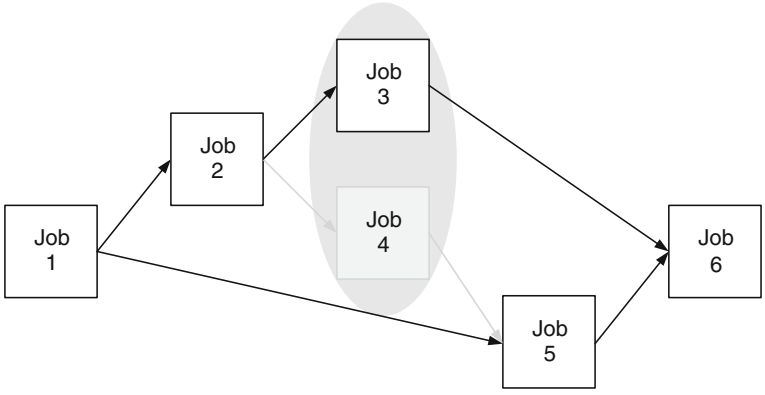


Fig. 1 Example for a project with alternative modalities

$$I = \begin{pmatrix} 1\ 5\ 2\ 4\ 3\ 6 \\ 1\ 1\ 1\ 0\ 1\ 1 \end{pmatrix} \begin{matrix} \text{activity list} \\ \text{implementation list} \end{matrix}$$

The implementation list contains the information that job 4 is not implemented in contrast to all other jobs. The activity list shows that the implemented jobs are scheduled in the sequence 1, 5, 2, 3, 6.

It should be noted that the feasibility of the activity list depends on the chosen implementation list. In case job 4 should be implemented instead of job 3, it would have been necessary to consider an (indirect) precedence relation between job 2 and job 5, cf. Fig. 1. The given activity list violates this constraint.

The fitness computation of the individuals is realized analogously to [4]. The fitness of individuals representing a feasible schedule equals the makespan of the project. This means that the lower the fitness value, the better the individual. If the capacity demanded by the implemented jobs exceeds the available capacity of the non-renewable resources, the individual does not represent a feasible schedule. In this case the additionally used capacity is added to the project horizon T to compute the fitness. Thereby it is ensured that individuals representing infeasible schedules always have a higher and, therefore, worse fitness value than feasible individuals.

6 Numerical Results and Outlook

In order to evaluate the developed genetic algorithm, it was first of all necessary to generate test instances. The instance generator ProGen (cf. [8]) was enhanced by the option to consider alternative modalities.

For the analysis in this paper 100 test instances each with 50 jobs have been created. The projects include 5–8 choices and in each choice 2–6 jobs can be chosen between. Additionally, in each test instance 2–4 jobs cause the implementation of each 1–3 further jobs.

Due to the novelty of the presented problem, a comparison with other solution approaches is not possible. Therefore, reference values have been computed using CPLEX 23.8. The instances were solved to optimality using a 2.00 GHz Intel Xeon machine with 16 GB of RAM and 4 threads. The average computational time was 84 s.

The genetic algorithm is implemented in Delphi XE and runs on a 2.66 GHz Intel Core2 Quad machine with 4 GB of RAM using one thread. Each generation of the genetic algorithm consists of 100 individuals and schedules, respectively. The performance of the genetic algorithm is visualized in Fig. 2. It shows a good solution quality in a short time. After 0.1 s ($\hat{=}$40 generations) the results deviate only 1 % from the optimum. After 0.2 s ($\hat{=}$80 generations) the average deviation amounts to 0.77 %. Afterwards only slight improvements of the solution quality can be observed.

Alternative modalities often go along with varying quality characteristics of the project. Therefore, the influence of the chosen project structure on quality should be

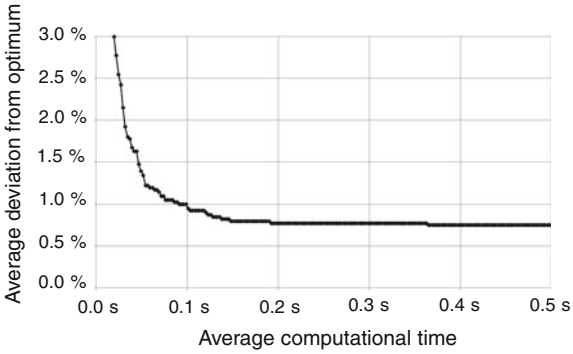


Fig. 2 Performance of the genetic algorithm

considered while deciding on the implementation of the jobs. This extension of the presented problem is an interesting topic for further research.

Acknowledgments The author thanks the German Research Foundation (DFG) for the financial support of this research project in the CRC 871 'Regeneration of complex durable goods'.

References

1. Belhe, U., Kusiak, A.: Resource constrained scheduling of hierarchically structured design activity networks. IEEE Trans. Eng. Manage. **42**, 150–158 (1995)
2. Brucker, P., Drexl, A., Möhring, R., Neumann, K., Pesch, E.: Resource-constrained project scheduling: notation, classification, models, and methods. Eur. J. Oper. Res. **112**, 3–41 (1999)
3. Elmaghraby, S.E.: Activity Networks: Project Planning and Control by Network Models. Wiley, New York (1977)
4. Hartmann, S.: Project scheduling with multiple modes: a genetic algorithm. Ann. Oper. Res. **102**, 111–135 (2001)
5. Kelley Jr, J.E.: The critical-path method: resources planning and scheduling. In: Muth, J.F., Thompson, G.L. (eds.) Industrial Scheduling, pp. 347–365. Prentice-Hall, Englewood Cliffs, NJ (1963)
6. Klein, R.: Scheduling of Resource-Constrained Projects. Kluwer Academic Pubslishers, Boston (2000)
7. Kolisch, R.: Serial and parallel resource-constrained project scheduling methods revisited: theory and computation. Eur. J. Oper. Res. **90**, 320–333 (1996)
8. Kolisch, R., Sprecher, A., Drexl, A.: Characterization and generation of a general class of resource-constrained project scheduling problems. Manage. Sci. **41**, 1693–1703 (1995)
9. Kuster, J., Jannach, D., Friedrich, G.: Extending the RCPSP for modeling and solving disruption management problems. Appl. Intell. **31**, 234–253 (2009)
10. Neumann, K.: Stochastic Project Networks. Springer, Berlin (1990)
11. Tiwari, V., Patterson, J.H., Mabert, V.A.: Scheduling projects with heterogeneous resources to meet time and quality objectives. Eur. J. Oper. Res. **193**, 780–790 (2009)

Part XVI
Simulation and System Dynamics

Market Penetration of Alternative Fuel Vehicles in Iceland: A Hybrid Modeling Approach

Ehsan Shafiei, Hlynur Stefansson, Eyjólfur Ingi Ásgeirsson and Brynhildur Davidsdottir

1 Introduction

System dynamics (SD) and agent-based (AB) modeling are two well-known dynamic simulation approaches, which have been widely used for studying the diffusion of alternative fuel vehicles (AFVs). Comparison of SD and AB methodologies and their strengths and weaknesses have been discussed in the literature. In general, AB and SD approaches differ in two ways: the fundamental relationships among components and the level of aggregation [1]. SD is a top-down approach that looks at the process of market development as a whole and facilitates understanding the interactions of stakeholders in complex systems. AB modeling is a bottom-up approach towards a systemic analysis of complex systems. AB models can capture heterogeneity of individuals in an interconnected network, but with increased computational costs. Due to conceptual and computational limitations of each approach, several researchers have recommended using the combined form of AB and SD approaches to study the diffusion of AFVs [2–4].

In this paper, we propose a framework for an integrated analysis of energy and transportation system in Iceland, which suggests a dynamic simulation model based on the approach combining SD structure with AB technique. Modeling framework

E. Shafiei (✉) · H. Stefansson · E. I. Ásgeirsson
Reykjavik University, Menntavegur 1, 101Reykjavik, Iceland
e-mail: ehsan@ru.is

H. Stefansson
e-mail: hlynur@ru.is

E. I. Ásgeirsson
e-mail: eyjo@ru.is

B. Davidsdottir
The Graduate Program in Environment and Natural Resources, University of Iceland, Reykjavik, Iceland
e-mail: bdavids@hi.is

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_65,

and the process of integration are discussed in Sect. 2. Description of case study and main simulation results are presented in Sects. 3 and 4. Finally, the paper is concluded in Sect. 5.

2 Model Description

In this section, the framework for the integrated top-down and bottom-up modeling is presented. The integrated model includes two components: an AB module for consumer choice behavior and an SD framework to represent the behavior of energy supply system.

We use AB methodology for studying the consumer behavior and market share evolution of light duty passenger vehicles (LDVs). We use a simplified form of the vehicle choice algorithm introduced in our previous work [5]. Consumers purchase the vehicles based on both their own preferences and imitating the behavior of other consumers. Consumer preferences are described on the basis of Multinomial Logit (MNL) model. In this framework, different vehicles are distinguished by various attributes and the consumers make choices among them so as to maximize their utility. We identify each type of vehicle by seven attributes: (1) vehicle price in $, (2) fuel cost in $ per km, (3) maintenance cost in $ per year, (4) vehicle range in km, (5) battery replacement cost in $, (6) fuel availability as the number of each fuel station relative to the number of gasoline stations, and (7) GHG emission in gCO2eq per km. Whether or not a consumer purchases a vehicle is determined by comparing the life of current vehicle with a stochastic vehicle lifetime. If the current age of the vehicle is greater than its lifetime, then consumer decides to purchase a new vehicle. The probability of purchasing vehicle j by consumer i at time t ($P_{i,j,t}$) is estimated as the following equation:

$$P_{i,j,t} = \frac{\exp\left(\sum_{a=1}^{A} \beta_{i,a,t} X_{a,j,t}\right)}{\sum_{j=1}^{V} \exp\left(\sum_{a=1}^{A} \beta_{i,a,t} X_{a,j,t}\right)} \tag{1}$$

where $\beta_{i,a,t}$ is the preferences of consumer i for attribute a, $X_{a,j,t}$ the value of attribute a for vehicle j, V the number of vehicle types and A the number of vehicle attributes.

The probabilities in Eq. (1) are aggregated into a cumulative distribution function of vehicle purchases. The function is the sum of probabilities for buying all of the vehicles up to and including a particular vehicle j:

$$Q_{i,j,t} = \sum_{h=1}^{j} P_{i,h,t} \tag{2}$$

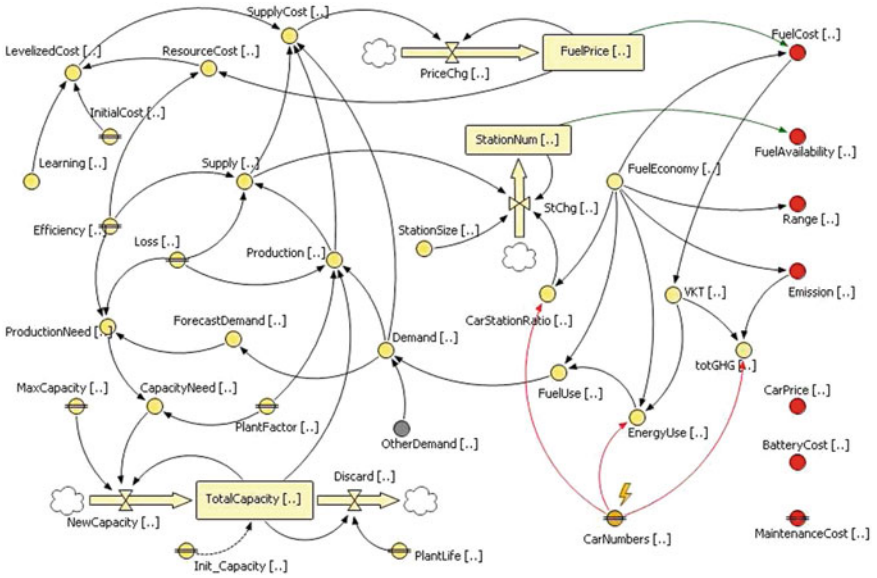


Fig. 1 SD structure of energy supply for transport sector

Should an agent decide to purchase a specific type of vehicle, a Monte–Carlo simulation is processed. A random uniform variable is generated and it is compared to the cumulative probability function. The input of Eq. (2) that yields the closest but higher outcome than the value of the uniform variable equals the type of vehicle that is purchased.

For simplicity and due to lack of suitable data, it is assumed that when a consumer decides to purchase a new vehicle, his old vehicle will vanish from the system. Consumers can periodically contact each other so that the users of each vehicle type may influence the potential users by sending a message to their neighbors.

For the simple SD framework for Iceland energy supply system, it is assumed that the security of energy supply and meeting transportation energy demand in the long-run is ensured by a central planner. This concept is shown in Fig. 1. The total existing capacity for any technology is increased by new installations and depreciated at a constant rate over technology lifetime. Total capacity, plant factor and total demand determine the amount of energy production. Fuel supply for transport sector is estimated based on production and energy losses. Final energy demand equals the sum of transportation demand (FuelUse) and the exogenous values of non-transport energy demand. It is assumed that the gap between energy production and forecasted energy demand can be reduced by additional energy production and, thus, new capacity installation. The levelized cost of energy production for each technology is calculated every time based on initial cost, resource cost and learning effects. Resource cost depends on input energy prices and technology efficiency. Energy price is controlled by price change, which is equal to the difference between the

current price and the average supply cost. Supply cost, which is calculated based on levelized costs, is influenced by supply and demand. Fuel price along with vehicles' fuel economy determine fuel cost attribute for vehicles. Number of fuel stations is controlled by station change flow, which is a function of vehicle-station-ratio, average station size and fuel supply. Fuel availability is configured as the number of each fuel station relative to the number of gasoline stations. Range and emission attributes are changed in proportion to fuel economy improvement over time. Vehicle price, maintenance cost and battery replacement cost are assumed to be improved over time exogenously. Vehicle-km traveled (VKT) is affected by fuel cost. Number of each vehicle type is extracted from the consumer AB module and is used to calculate total GHG emission by the vehicles, total energy requirement (EnergyUse) and car-station-ratio. Total energy use is calculated by VKT, fuel economy and car numbers and, then, is converted to fuel demand.

Total fuel consumption by various vehicles, which reflects the consumers' energy demand, is calculated by consumer AB module and is then transferred to the SD module of energy system. SD framework, on the other hand, determines both the fuel availability and price of various fuels to be sold to the consumers. Fuel cost attributes for vehicles is calculated based on consumers' driving patterns, vehicles' fuel economy and the fuel prices given by the SD component.

3 Description of Case Study

The integrated model, developed using AnyLogic software [6], is applied for simulation of Iceland LDV fleet. The time horizon of the study begins in 2013 and continues until 2050. The vehicle types considered in AB module are: gasoline internal combustion engine (ICE_P), diesel internal combustion engine (ICE_D), gasoline hybrid electric vehicle (HEV_P), diesel hybrid electric vehicle (HEV_D), plug-in hybrid electric vehicle with gasoline (PHEV_P), plug-in hybrid electric vehicle with diesel (PHEV_D), battery electric vehicle (BEV), bio-ethanol ICE (E85), bio-diesel ICE (B20), bio-gas ICE (BGV1), Dual fuel ICE with bio-gas and gasoline (BGV2), and fuel cell electric vehicle (FCV).

Consumers are heterogeneous with respect to their income attribute and word-of-mouth influences. Iceland income distribution is used to randomly assign an income attribute to each agent. Vehicle per capita index in Iceland in the base year is 0.65 and it is assumed to gradually approach the saturation value of 0.8. We assume that each vehicle represents an agent with independent decision behavior. Total vehicle population is scaled down in which each agent in the model represents 100 actual agents. Assuming such scaling factor, the population size of agents would be 2,100 in the base year. This value is increased corresponding to the growth rates of both vehicles per capita and population in Iceland.

The technologies represented on the energy supply system side are: electricity from hydropower, electricity from geothermal, bio-gas from municipal wastes, bio-diesel form waste oils, bio-ethanol from lignocelluloses biomass, hydrogen from

electrolysis, gasoline import and diesel import. Hydrogen and bio-ethanol production technologies have exogenous learning for the costs, while constant costs along the horizon are assumed for the other technologies. An annual growth rate of 3 % is applied for the prices of gasoline and diesel, while it is assumed a constant price along the horizon for typical biomass resources.

4 Simulation Results

The first figure in Fig. 2 shows the penetration rate and market share of various vehicles in total fleet. The shares of gasoline and diesel ICE vehicles are gradually reduced, replaced by HEVs, PHEVs and bio-fuel vehicles. The results show that biogas vehicle is an attractive alternative. However, its market growth is slowed down due to limited resource potential in Iceland. After period 2030, the share of BEV and FCV rise quickly on account of competitive fuel cost, fuel availability and decreasing purchase price. A minimum fuel availability of 10 % in 2020 has been assumed for hydrogen, bio-ethanol and bio-diesel. Sensitivity analysis showed that the initial fuel availability for these fuels should be at least 2 % to guarantee the market introduction of FCV, E85 and B20 vehicles. The second figure in Fig. 2 presents the corresponding trends for the projected fuel demand for LDVs in Iceland. Improvement of fuel economy and market penetration of more efficient AFVs are the main reasons for the decreasing pattern of LDVs' fuel demand after 2030.

5 Conclusions

In this paper, the general behavior of Iceland energy supply system was described by a simple SD structure. An AB paradigm was introduced to represent the consumer vehicle choice behavior. AB and SD modules for energy and transport sectors

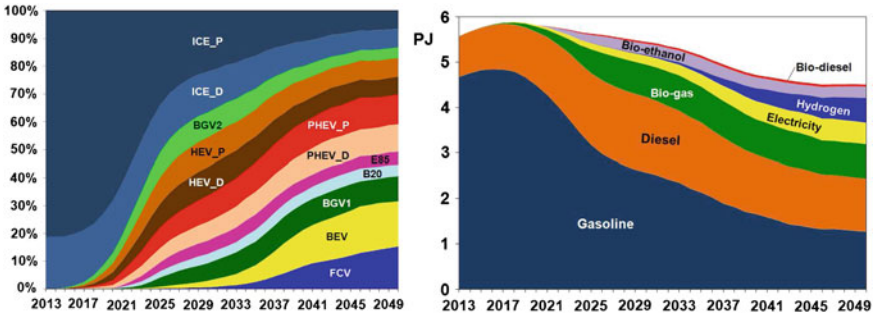


Fig. 2 The market share evolution of different vehicles in Iceland, and the trend of fuel demand for LDVs

work in a complementary way inside one integrated model. SD module includes a set of algebraic and differential equations, which are calculated continuously. The AB module is founded on state-charts, simple rules and events that occur when they are scheduled. Integration of components needs an endogenous specification of AB–SD interactions. Therefore, both AB and SD components are solved simultaneously. Integrated modeling structure gives the potential of building more accurate and computationally efficient model for simulating the transition to sustainable mobility. Hybrid AB–SD modeling can be more accurate than a pure SD structure as it can capture the interactions of heterogeneous agents. On the other hand, it can be solved faster than a pure AB framework as some components can be expressed continuously by simple algebraic and ordinary differential equations.

References

1. Parunak, H.V., Savit, R., Riolo, R.L.: Agent-based modeling vs. equation-based modeling: a case study and user's guide. In: Proceedings of Multi-agent Systems and Agent-based, Simulation, pp. 10–25 (1998).
2. Keenan, P., Paich, M.: Modeling general motors and the north American automobile market. In: The 22nd International Conference of the System Dynamics Society, Oxford, England (2004)
3. Kieckhäfer, K., Walther, G., Axmann, J., Spengler, T.: Integrating agent-based simulation and system dynamics to support product strategy decisions in the automotive industry. In: Proceedings of the 2009 Winter Simulation Conference, pp. 1433–1443 (2009)
4. Köhler, J., Whitmarsh, L., Nykvist, B., Schilperoord, M., Bergman, N., Haxeltine, A.: A transitions model for sustainable mobility. J. Ecol. Econ. **68**, 2985–2995 (2009)
5. Shafiei, E., Thorkelsson, H., Ásgeirsson, E.I., Davidsdottir, B., Raberto, M., Stefánsson, H.: An agent-based modeling approach to predict the evolution of market share of electric vehicles: a case study from Iceland. J. Technol. Forecast. Soc. **79**, 1638–1653 (2012)
6. AnyLogic simulation software, XJ Technologies Company, www.xjtek.com

Balancing of Energy Supply and Residential Demand

Martin Bock and Grit Walther

1 Introduction

Power demand fluctuates and has to be met by power supply. Increasing installation of power intense technologies, like BEVs and heat pumps will lead to an increase in total power demand, but more important, to an increase in demand fluctuations and peaks. It is beneficial to both utilities and customers to avoid fluctuations, as they increase costs. Balancing demand by influencing the energy consumption of the customers is known as demand side management (DSM), which takes place as an interaction between a utility and its customers.

Thereby, the utility can apply two kinds of instruments. Incentive-based instruments influence consumers' behavior based on price signals, e.g. offering lower prices during off-peak hours. Control-based schemes allow the utility to remotely control their customers' appliances, based on consumption peaks.

Regarding the customers, future fluctuations and peaks of the load curve depend on future equipment and usage of appliances. Thereby, appliances can be differentiated with regard to characteristics of energy consumption, e.g. constant demand (stand-by), pre-set programs (dish washer), cyclic demand (freezers). Reduction or shift of usage of appliances may be seen as discomfort by customers. Thus, reactions to DSM measures will differ depending on the choice between reducing costs and keeping the level of comfort. Additionally, it is expected that transformation and substitution between different forms of energy (e.g. generation of electricity by CHP [*combined heat and power*]) as well as storage of energy (BEVs, water tanks)

M. Bock (✉) · G. Walther
Chair of Production and Logistics, Schumpeter School of Business and Economics,
University of Wuppertal, Wuppertal, Germany
e-mail: martin.bock@wiwi.uni-wuppertal.de

G. Walther
e-mail: walther@wiwi.uni-wuppertal.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_66,

will gain importance. However, these options depend on the development of the distribution of such appliance types.

Against this background, the aim of the paper is to develop a concept for modeling demand side management as interaction between one utility and its customers. Thereby, incentive-based as well as control-based instruments can be applied, and reduction and shift of energy demand as well as transformation of energy is regarded at the demand side. Additionally, customer behavior with regard to price/comfort decisions is accounted for.

The paper is structured as follows. First, a brief literature review on demand side management in the residential sector is provided. Second, our integrated DSM model is presented. It is divided into two levels: a structural level, representing technical relations of customers' appliances, and a behavioral level, representing the interactions between the utility and its customers and the reactions of the customers to DSM. Finally, the paper closes with a discussion of further research.

2 Literature

Energy demand can be either modeled by top-down or bottom-up approaches [11]. Bottom-up approaches are used to analyze the effects of DSM, as they match appliances to load and determine the effect of changes in appliance usage on load curves. In early DSM attempts, only appliances with thermal functionality, such as space heating (e.g. see [6]) or HVAC *(heating, ventilation and air conditioning)* (e.g. see [1]), have been considered as being influenced by DSM. More recently, appliances have been grouped by the functionality they provide (e.g. see [4] and [12]) and different DSM measures are applied based on the group an appliance is assigned to. But these attempts do not allow substitution and transformation of different forms of energy, e.g. generation of electricity in a CHP. These aspects are considered by models of integrated energy flow networks, which explicitly regard conversion and storage of energy (e.g. see [5]). However, DSM effects are not considered, since energy demand is exogenous.

So far, DSM approaches often focus on only one of the instruments for consumption adaption. Utilities either control certain appliances directly (e.g. see [9]), or customers receive price incentives and alter appliance usage by weighing costs and comfort (e.g. see [3]), or appliances react automatically to changes in prices and/or net frequency (e.g. see [4, 12]). Based on the different groups of appliances, a combination of the before mentioned instruments seems more reasonable. With regard to modeling the utility's decisions, price incentives are determined based on costs (e.g. see [6]) or on supporting the integration of renewable energies (e.g. see [7]). Direct utility control is applied based on the power net status (e.g. see [8]) or based on costs (e.g. see [2]]). Customers differ in distribution and usage of appliances and therefore energy consumption. But aside of some agent-based approaches (e.g. in [10]), many models do not consider populations of heterogeneous customers or groups of customers.

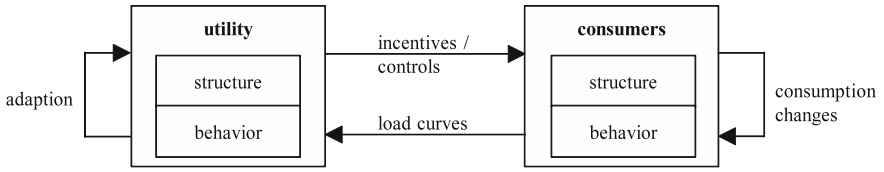


Fig. 1 Consumer and utility interaction

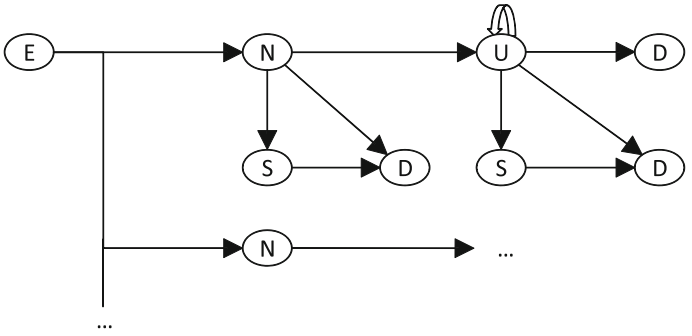


Fig. 2 Generic network flow model of the structural level

To the knowledge of the authors, the requirements as stated in part one of this paper have not been fulfilled completely by any one model. Therefore, a concept for integrated DSM will be presented in the following.

3 Model Description

Two kinds of actors are considered, the utility and a set of customers, and their interaction. Each actor is defined on two levels: technical characteristics and energy flows are modeled as the structural level, aspects of interaction and appliance usage as the behavioral level (see Fig. 1).

3.1 The Structural Level

The structural level is designed as a network flow model. It consists of edges, representing energy flows, and nodes, representing points of energy application. The network's purpose is to model the energy flows from the public nets to the points of energy consumption. In Fig. 2 the generic network flow model is shown.The model considers a set T of discrete periods, a set A of energy forms, a set J of nodes of different types, and edges (i, j) from node $i \in J$ to node $j \in J$.

The node *E (utility node)* represents the utility, which provides the energy demanded by the customers, denoted by e_{at}, and is connected to each customer. Thereon, the node *N (household node)* is the source in each household's network (see Eq. 2). Nodes denoted with a *D (demand nodes)* demand energy d_{jat} of type $a \in A$, for example power or heat. This energy needs to be provided in time with the correct type by the edges f_{ijat} leading into the node $j \in J$ (see Eq. 3). Nodes denoted with an *S (storage nodes)* store energy, using s^{in}_{jat}, with an efficiency of W^{in}_{ja} and unload the energy from the storage in later periods, using s^{out}_{jat} (see Eqs. 4 and 5). The state of the storage s^{s}_{jat} is determined by the loading and unloading, the storage efficiency W^{s}_{ja} and last period's state of storage s^{s}_{jat-1} (see Eq. 6). Nodes denoted with a *U (transformation nodes)* transform energy from one input type $a \in A^{i}_{j}$ into one or more output types $b \in A^{o}_{j}$, e.g. like a CHP-unit is fired by gas to provide power and heat. The transformed energy u_{jbt} is determined under consideration of the transformation efficiency W^{U}_{jab} (see Eqs. 7 and 8).

$$e_{at} = \sum_{j \in J^N} f_{jat} \qquad \forall a \in A; t \in T \quad (1)$$

$$f_{jat} = \sum_{k \in J: j \neq k} f_{jkat} \qquad \forall j \in J^N; a \in A; t \in T \quad (2)$$

$$\sum_{i \in J: i \neq j} f_{ijat} = d_{jat} \qquad \forall j \in J^D; a \in A; t \in T \quad (3)$$

$$s^{in}_{jat} = \sum_{i \in J: i \neq j} f_{ijat} * W^{in}_{ja} \qquad \forall j \in J^S; a \in A; t \in T \quad (4)$$

$$s^{out}_{jat} = \sum_{k \in J: j \neq k} f_{jkat} \qquad \forall j \in J^S; a \in A; t \in T \quad (5)$$

$$s^{S}_{jat} = s^{S}_{jat-1} * W^{S}_{ja} - s^{out}_{jat} + s^{in}_{jat} \qquad \forall j \in J^S; a \in A; t \in T \setminus \{1\} \quad (6)$$

$$\sum_{i \in J: i \neq j} f_{ijat} = \frac{u_{jbt}}{W^{U}_{jab}} \qquad \forall j \in J^U; a \in A^{i}_{j}; b \in A^{o}_{j}; t \in T \quad (7)$$

$$u_{jbt} = \sum_{k \in J: j \neq k} f_{jkbt} \qquad \forall j \in J^U; b \in A^{O}_{j}; t \in T \quad (8)$$

3.2 The Behavioral Level

The utility's target is to maximize its profits. Its revenue is determined by the retail price c^{N}_{jat}, which it defines in advance, and the resulting customer energy demands f_{jat}. It has to purchase the energy of form a as demanded by the households at period t at the market price C^{E}_{at}:

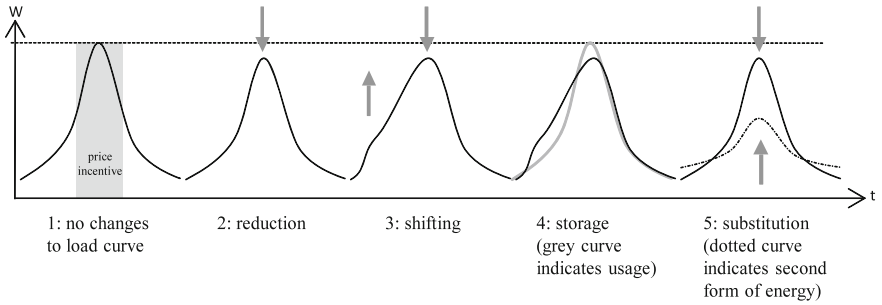


Fig. 3 Load curve adaptions due to DSM

$$max \sum_{j \in J^N} \sum_{a \in A} \sum_{t \in T} c^N_{jat} * f_{jat} - \sum_{a \in A} \sum_{t \in T} C^E_{at} * e_{at} \tag{9}$$

With regard to consumers' behavior, there are five effects DSM can have on appliance usage, and by considering all appliances of a customer, on load curves (compare Fig. 3). These effects are: (1) the load curve is unchanged, (2) reduction of energy demand, (3) load shift, i.e. a reduction takes place at one time but the load is increased at another time, (4) storage, i.e. time of energy consumption and energy usage is decoupled using the stored energy and (5) substitution, i.e. different forms of energy can be used or can be transformed in order to fulfill demand.

We assume that a customer's target function is to minimize costs and loss of comfort (see Eq. 10). The customer pays the time-varying price c^N_{jat} as a price per energy unit f_{jat} provided by the utility.

The discomfort or loss of comfort of a customer, denoted by μ is modeled based on the difference of usage of appliances without DSM influence and with DSM influence. Adjustability of usage and loss of comfort depend on the groups of appliances. We also assume that each household has a preference value λ, by which it weighs price and comfort. The higher the preference, the more a customer will accept loss in comfort to reduce costs.

$$min \sum_{j \in J^N} \sum_{a \in A} \sum_{t \in T} (c^N_{jat} * f_{jat}) * \lambda + \mu * (1 - \lambda) \tag{10}$$

4 Outlook and Further Research

The presented demand side model considers utility decision making and household energy consumption. Further research will be directed to heterogeneity of households in appliance equipment and usage, both concerning the situation at present and future development, the weighing of price and comfort and the resulting changes in

consumption behavior. On the utility side, decision and adaption behavior need to be studied, possibly utilizing artificial intelligence or heuristic approaches.

This model will be implemented as an agent-based simulation, with the utility and the customers represented as agents. The customer heterogeneity is well supported by agent modeling. The behavioral level will be realized as agent interaction and decision making, while the structural level will be designed as a network flow model.

References

1. Chan, M.L., Marsh, E.N., Yoon, J.Y., Ackerman, G.B., Stoughton, N.: Simulation-based load synthesis methodology for evaluating load-management programs. IEEE Trans. Power Apparatus Syst. **100**(4), 1771–1778 (1981)
2. Du, P., Lu, N.: Appliance commitment for household load scheduling. IEEE Trans. Smart Grid **2**(2), 411–419 (2011)
3. Eßer, A.; Kamper, A.; Franke, M.; Möst, D.; Rentz, O. (2006): Scheduling of electrical household appliances with price signals. In: Proceedings of the Operations Research 2006, Part VIII, pp. 253–258
4. Gottwalt, S., Ketter, W., Block, C., Collins, J., Weinhardt, C.: Demand side management: A simulation of household behavior under variable prices. Energy Policy **39**(12), 8163–8174 (2011)
5. Geidl, M., Andersson, G.: A modeling and optimization approach for multiple energy carrier power flow. In: Proceedings of IEEE PES PowerTech, St. Petersburg, Russ. Federation, In (2005)
6. Hämäläinen, R. P.; Mäntysaari, J.; Ruusunen, J.; Pineau, P. O. (1999): Consumption strategies and tariff coordination for cooperative consumers in a deregulated electricity market. In: Proceedings of the 32nd Hawaii International Conference on System Sciences
7. Hillemacher, L., Eßer-Frey, A., Fichtner, W.: Preis- und Effizienzsignale im MeRegio Smart Grid Feldtest - Simulationen und erste Ergebnisse. IEWT, Wien (2011)
8. Kaschub, T.; Mültin, M.; Schmeck, H.; Fichtner, W.; Kessler, A. (2010): Intelligentes Laden von batterieelektrischen Fahrzeugen im Kontext eines Stadtviertels. In: VDE-Kongress 2010: E-Mobility, VDE-Verlag
9. Paatero, J.V., Lund, P.D.: A model for generating household electricity load profiles. Int. J. Energy Res. **30**(5), 273–290 (2005)
10. Roop, J.M., Fathelrahman, E.M., Widergren, S.E.: Price response can make the grid robust: an agent-based discussion. IEEE 2005 Power Eng. Soc. General Meet. **3**(12–16): 2813–2817 (2005)
11. Swan, L.G., Ugursal, V.I.: Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew. Sustain. Energy Rev. **13**, 1819–1835 (2009)
12. Zeilinger, F. (2011): Simulation of the effect of demand side management to the power consumption of households. In: Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE)

IT-Based Decision Support for Turning on Germany's Energy Transition

The Impact of the Nord.Link: Complex Decision Support with System Dynamics

Bo Hu, Armin Leopold and Stefan Pickl

1 Introduction

Already for two decades Norway and Germany is thinking about a power cable across the North Sea and finally in June 2012 the agreement to realize an electricity interconnector between the two countries was started. This interconnection with a maximum capacity of 1,400 Megawatt will improve not only the security of energy supply, but also more stable energy prices for both nations [7]. Furthermore it has to be mentioned that this statement was given by the CEO of Statnett, Auke Lont, who says that it will be also of great importance exploiting the diversities in both nations' electricity system. Statnett, as the Norwegian Transmission System Operator (TSO), TenneT as the TSO on the German side and KfW as the Germany Bank for development and reconnstruction have agreed with the goal to start this subsea electricity interconnector in 2018 [12]. In details this means that Statnett owns the transmission grid and holds 50 %, while KfW holds at least 25 % and the remaining interest is held by TenneT. However, it is planned that this project needs about 3 years to be realized. In details it can be said that the 640 km long connection will start at the transformer station Tonstad and will then be as an 540 km long subsea electricity cable until it reaches the shores in Büsum and afterwards the transformer station in Wilster, Germany [8]. In close connection to this Nord.Link some information from the German energy efficiency congress 2011 will be stated in the following paragraph [6] as the Vice-president of Statkraft reported as an further option to build further pumped hydro power storage in connection with already existing reservoirs with

B. Hu (✉) · A. Leopold · S. Pickl
Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
e-mail: bo.hu@unibw.de

A. Leopold
e-mail: armin.leopold@unibw.de

S. Pickl
e-mail: stefan.pickl@unibw.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_67,

an approximated capacity of 15–20 GW. With other words, she shows the potential of pumped storage hydropower as an additional generation capacity. Nevertheless she also discussed the main existing barriers and the connected incentives related to the upcoming construction of new production capacity in Norway. The public has to understand the benefits of the emission reductions by investing in hydro power stations. Therefore the local environmental consequences have to be accepted to enable an trans-national power grid. To do so relevant economic compensations to ground owners and local communities are necessary [10]. Tennet [16] as the project-related transmission system operator on the German side strengthens the Nord.Link project with the following fact. The trans-national cable connection enables both countries to transfer power. On the one hand if there are strong winds in Northern Germany the electricity can be exported to Norway, where it can be stored until it is needed again in Germany. On the other hand in some month of the year Norway needs the stored power produced by Germany's Offshore wind mill parks for its own consumers. Generally speaking it has to be mentioned that hydropower is beneficial to compensate some of the existing short-term fluctuations caused by wind power [18]. Nevertheless this project can be seen as one further part of a sustainable North-West European trans-national power grid which is mainly based on renewable energy sources [4]. Nevertheless the German Advisory Council on Environment (SRU) stated in its report to reach a totally renewable energy supply the fact that there has to be a power transmission capacity which is nearly 30 times larger than the Nord.Link capacity [11].

2 System Dynamics Modeling and Simulation

In this paper we use our computer model [2] to assess the Nord.Link project regarding its contribution to a reliable, efficient and sustainable electricity supply in Germany. This System Dynamics model focuses on the electricity grid from which electricity is delivered for consumption at variable load. Electricity is generated dispatchably or non-dispatchably and fed into the grid. Dispatchable power generation uses mainly chemical fuels while wind and photovoltaic powered sources (WDPV) depend on the weather and time of day. One possible pathway to facilitate a renewable-centered electricity supply is thus to establish a storage subsystem, as shown in Fig. 1. In addition new technologies like synthesized natural gas (SNG) [13] and load management are also included in the model. See [2] for more details.

The so-called Stock and Flow model depicted above represents an integral equation system which can be solved using computational methods provided by a System Dynamics modeling and simulation environment like Vensim PLE [3]. In this way an electricity supply concept can be presented by this model using suitable parameters and assessed regarding their reliability and resource consumption. To do this we first enter the characterizing key parameters of the concept and then try to find the minimal dispatchable capacity which still provides reliable electricity supply under a given load profile and WDPV profile for an entire year. A concept is

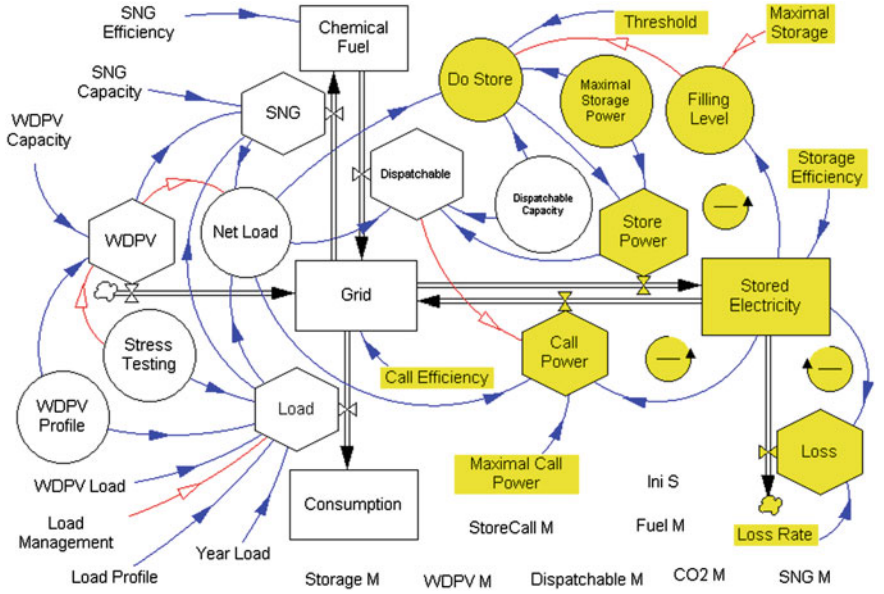


Fig. 1 A system dynamics model for Germany's electricity supply including a storage subsystem which consists of highlighted elements

considered as reliable if the cumulative energy shortage is less than 2.6 TWh during the entire year or 0.3 GW in average. Notice that possible excess electricity occurring at another point of time does not offset the cumulative shortage in the calculation. The total production cost which includes investment, operating, fuel costs and emission permits is then calculated for the concept.

As shown in Fig. 2 which depicts the essential results of our numeric simulations the project Nord.Link contributes to mitigate CO_2 emissions only if the installed renewable capacity is above 100 GW. At a capacity of 150 GW Nord.Link may help German electricity industry reduce the CO_2 emission by 0.1 % compared to 1990. The transport capacity of 16 such projects alone with 8 TWh or approximately one-fifth of the estimated total storage capacity of Norway [1] in combination of 200 GW installed renewable capacity in Germany would reduce the total CO_2 emission of electricity production by 2 % or by 5.36 million tons ("Emission 16 × NL" in Fig. 2). However, this would also increase the total production cost by 12–16 MWh. Figure 3 shows the parameters used for the simulations.

3 Concluding Discussions

With the help of System Dynamics we started to analyse how the upcoming Nord.Link project will contribute to a reliable, efficient and sustainable electricity supply in Germany. The System Dynamics model focuses mainly on the storage of energy.

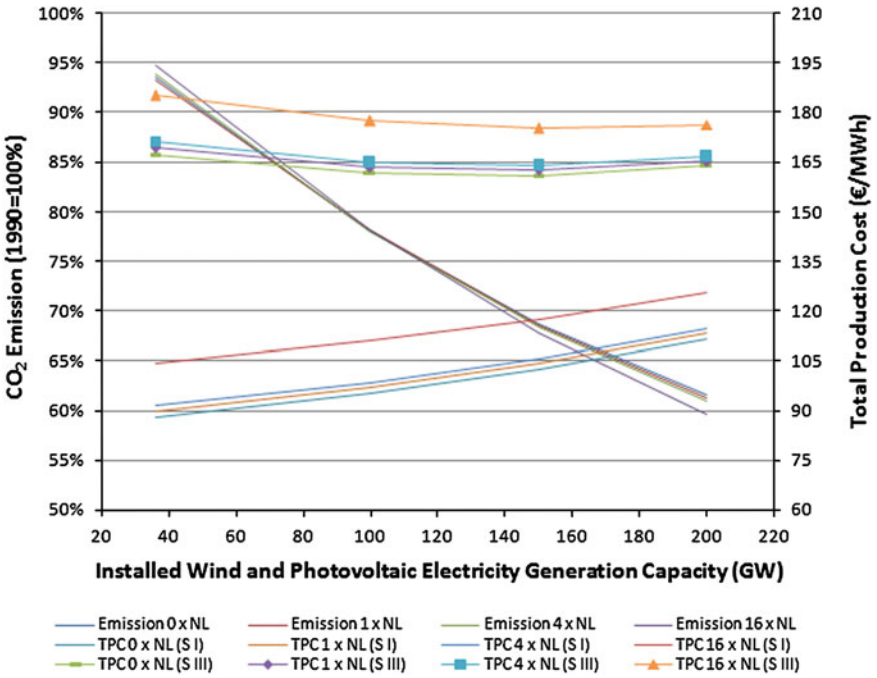


Fig. 2 CO_2 emission and total production cost depending on installed renewable capacity and storage. The total production cost (TPC) is calculated under cost scenarios I and III. Notice that the expansion of WDPV is largely cost-neutral under the cost scenario III (S III).

Investment costs				
- Dispatchable (B€/GW)	1.00	- Renewable (B€/GW)		1.80
- Storage power (B€/GW)	0.80	- Call power (B€/GW)		0.80
- Storage capacity (B€/GWh)	0.01			
Operation costs				
- Dispatchable (€/MWh)	8.50	- Other (B€/GW·Year)		0.03
Cost scenarios	I	II	III	IV
Fuel (€/MWh_e)	50	75	100	100
CO2 permits (€/tCO2)	25	37.5	100	100
Annual capital cost/interest	7.26%/6.00%	7.26%/6.00%	7.26%/6.00%	10.61%/10.00%

Fig. 3 Parameters used for the simulations. Data sources for the estimated specific costs: [5, 9, 11, 14, 15, 17]. The annual capital cost is calculated as the sum of interest and repayment within a period of 30 years.

It explains how to transfer mainly the wind power overflow from Germany with the help of the new trans-national cable to the Norwegian water reservoirs. Moreover, System Dynamics modeling helps to understand the fact that Nord.Link contributes more to reliability and to sustainability but not strongly to economic efficiency.

Generally speaking Nord.Link can be regarded as a starting point for further investments in transmission technologies of all kinds of renewable energy sources. On the one hand it strengthens the security of energy supply and on the other hand it reduces the import dependency of fossil fuels. Although the water reservoirs in Norway seem to be gigantic, the low interconnector's transmission capacity seems to be the bottleneck in Nord.Link. Furthermore the severe costs for the construction of this electricity interconnector increase the average price for renewable electricity dramatically. As a consequence the price for electricity seems to become more expensive for all consumers in the near future. Therefore it can be said, that System Dynamics modeling enables the policy maker to communicate this part of the German energy transition process in a more understandable and objective way to the public.

References

1. Alne, J.: Offshore wind energy storage in Norwegian Power Plants. http://tm-info.no/getfile.php/tm-info.no/Presentasjoner/TM04052010/07_Statkraft_Alne.pdf 2010
2. Arto, K., Dolgopolova, I., Hu, B., Leopold, A., Pickl, S.: Germany's electricity industry in 2025: evaluation of portfolio concepts. In: The 2012 International Conference of the System Dynamics Society, Proceedings, St. Gallen, 22–26 July 2012
3. Free Download—Vensim PLE. Ventana Systems. http://www.vensim.com/freedownload.html (2009) (19.07.2009)
4. Fuchs, M.: Eine zuverlssige Stromversorgung—Anforderungen aus Sicht der bertragungsnetze. Tennet (2010)
5. Groscurth, H.-M., Bode, S.: Anreize für Investitionen in konventionelle Kraftwerke Reformbedarf im liberalisierten Strommarkt. arrhenius Institut für Energie- und Klimapolitik, Februar 2009. http://www.arrhenius.de/uploads/media/arrhenius_DP2_Investionen_in_konventionelle_Kraftwerke.pdf (12.11.2011)
6. Holmen, M.: Reservoir hydropower to balance intermittent electricity generation. dena-Energieeffizienkongress 2011 (2011)
7. Mihm, A.: Berlin drngt auf Stromanschluss in Norwegen. FAZ.NET (2012)
8. Prognos: Das Nord. Link Projekt. http://www.prognos.com/Projekt.813.0.html (2012)
9. Reina, P.: Subsea cables tie European electric grids more closely. ENR.com, 23 Jan 2008. http://enr.construction.com/news/powerIndus/archives/080123.asp (2008) (06.05.2012)
10. Seidler, C.: Kabelprojekt Nord. Link die Nordsee-Stromautobahn kommt. Spiegel Online, 21 Juni 2012, 16:41 Uhr. http://www.spiegel.de/wissenschaft/technik/nord-link-ab-2018-stromkabel-zwischen-norwegen-und-deutschland-a-840229.html (22.06.2012)
11. SRU:100 % erneuerbare Stromversorgung bis 2050: klimavertrglich, sicher, bezahlbar. Sachverstndigenrat für Umweltfragen, SRU (2010)
12. Statnett: Agreement to realize electricity interconnector between Germany and Norway. http://www.statnett.no/en/News/News-archive-Temp/News-archive-2012/Agreement-to-realize-electricity-interconnector-between-Germany-and-Norway/ (2012)
13. Sterner, M., Jentsch, M., Holzhammer, U.: Energiewirtschaftliche und Ökologische Bewertung eines Windgas-Angebotes. Fraunhofer-Institut für Windenergie und Energiesystemtechnik (IWES), Feb 2011. http://www.greenpeace-energy.de/fileadmin/docs/sonstiges/Greenpeace_Energy_Gutachten_Windgas_Fraunhofer_Sterner.pdf (05.05.2012)
14. Sterner, M.: Bioenergy and renewable power methane in integrated 100 % renewable energy systems. Dissertation, Universitt Kassel, 23 Sept 2009. http://www.uni-kassel.de/upress/online/frei/978-3-89958-798-2.volltext.frei.pdf (06.05.2012)

15. Stromerzeugungskapazitten, Bruttostromerzeugung und Bruttostromverbrauch Deutschland. Bundesministerium für Wirtschaft und Technologie (2011)
16. Tennet: Tennet plant neue Stromkabelverbindung nach Norwegen. http://www.tennettso.de/site/binaries/content/assets/company/news/2012/juni/21062012-pm-tennet-plant-seekabelverbindungen-nach-norwegen.pdf (2012)
17. Umweltbundesamt Deutschland: Energieziel 2050: 100 % Strom aus erneuerbaren Quellen. Umweltbundesamt (2010)
18. Vogstad, K.-O.: A system dynamics analysis of the Nordic E-Market. Doctoral thesis, ISBN 82-471 6890-1, Norwegian University of Science and Technology (2004)

Part XVII
Software Applications and Modelling Systems

The Multiple Watermarking on Digital Medical Image for Mobility and Authenticity

Adiwijaya, T. A. B. Wirayuda, S. D. Winanjuar and U. Muslimah

1 Introduction

The development of digital technology and the internet has been rapid enough to give easy access and distribute a variety of multimedia information in digital form, such as text, images, audio and video. At this time, patient's data in the hospital can be stored in electronic media. Thus allows the data in digital medical images form such as X-ray image, mammogram form and others can be very easily manipulated by the rapid development of information technology today [5]. Medical images in digital form must be stored or transmitted via internet in a secure way to preserve stringent image quality standards and prevent unauthorized disclosure of patient data [6]. There are two things that must be concerned in digital medical images such as the authority of ownership and the authenticity of the image (originality). As consequences to ward these cases it is necessary to apply watermarking techniques. The principle of watermarking itself is inserting digital data (either text or image) into the original digital medical image to maintain the authority of ownership and to detect the authenticity of the medical image.

The primary applications of watermarking are to protect copyrights and integrity verification (authentication) [3]. The main reason for protecting copyrights is to prevent image piracy when the transmitter sends it on the internet. For integrity verification, it is important to ensure that the medical image originated from a specific source and that it has not been changed, manipulated or falsified. In the single watermark technique, there is one purpose that can be achieved for instance, maintaining the authority of possession, detecting the authenticity of the image. Since

Adiwijaya (✉)
Faculty of Science, Telkom Institute of Technology, Jl. Telekomunikasi no.1, Bandung 40257, Indonesia
e-mail: adiwijaya@ittelkom.ac.id

T. A. B. Wirayuda · S. D. Winanjuar · U. Muslimah
Faculty of Informatics, Telkom Institute of Technology, Jl. Telekomunikasi no. 1, Bandung 40257, Indonesia

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_68,

the basic purpose of this technique is to maintaining the authority of ownership, using robust watermarking techniques which the embedded data is not easily damaged if it is done manipulating the image, while to detect the authenticity of the image use fragile watermarking technique where the embedded data must be easily damaged to detect any manipulation done of the image. These two things are very contradictory, so they cannot be implemented with a single watermark. But it is possible with multiple watermarks. Robust watermark and fragile watermark can be implemented simultaneously on an image so that the authority of ownership and authenticity of the image can be maintained at the same time. This is the fundamental difference between single watermarks with multiple watermarks [3].

Multiple watermark contain two parts, namely signature watermark and reference watermark. Both of watermark benefitted robust watermark and a fragile watermark properties, respectively [6]. Signature watermark is used to maintain the authority of ownership because it has robust characteristic that is not prone to damage if the embedded image is manipulated so the data remains safe. While the reference watermark is used to investigate the authenticity of the image. This watermark reference is highly vulnerable to the manipulation of imagery, however due to its fragile that it will easily detect the manipulation of the image so that the authenticity of the image can be maintained [2, 7]. Methods that used to insert a watermark in digital medical images are very diverse. The first is a watermark on the image method which does not allow embedding in the image areas that are considered important (Region of Interest). Although this method produces good image quality in this area, but the main problem is easy to do copy attack in the area (areas that are not embedded watermark). Another method is to create a virtual line by inserting pixels that make up an additional line around the image border. This method will make the file size becomes larger and took a lot of space on the storage media. Besides the two above, there is also a watermarking scheme based on wavelet for inserting multiple watermarks in medical images.

In [4], the method of multiple watermarking of medical images for content authentication and recovery are wavelet and LSB. Agustina et al. [1] provide a watermarking approach using integer wavelet and block truncation coding for image temper detection. In this paper, we divide medical image into two parts, namely Region of Interest (ROI) and Region of Non-Interest (RONI). The reference watermark which is used to detect the authenticity of medical images (integrity control) is embedded in the ROI image. Meanwhile, the signature watermark which is used for authority of ownership (proprietary rights) is embedded in the RONI image on the wavelet domain.

To increase durability, we propose a method of signature watermark embedding using error correcting code, namely Reed–Muller Code. While the method used for reference watermark embedding is a Hash Block Chaining that worked on the spatial domain. Therefore, it takes multiple watermarking techniques such that two aims do watermarking on medical image can be fulfilled.

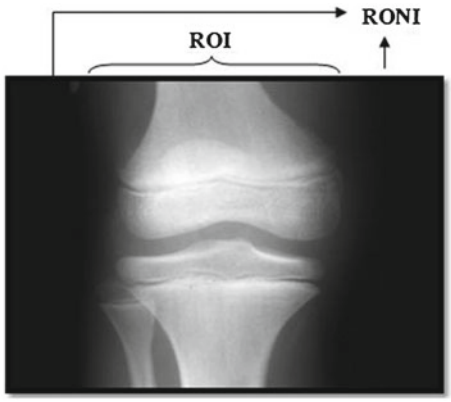


Fig. 1 ROI-RONI splitting of digital image

2 Proposed Multiple Watermarking

Embedding process of the reference watermark and signature watermark is done separately. Therefore, first carried out the extraction process of medical images into ROI (central part of image) and RONI (the outside of ROI).

As in Fig. 1, each margin of top, bottom, left and right parts that be used is 64 pixels (12.5 % of size of the original image, 512 × 512 pixels). These margins were considered as a border or side of the image. Top and bottom borders size 64 × 512 pixels are combined into RONI size 128 × 512 pixels. Left and right borders size 384 × 64 pixels are combined into RONI size 384 × 128 pixels. While the center of the image size 384 × 384 pixels serve as the ROI.

Having obtained RONI and ROI image, then the signature watermark will be encoded first by using the Reed–Muller Code. The codeword (message that result of encode signature watermark) will be embedded in the coefficients of decomposition Wavelet Discrete Transform (DWT) of RONI image by using Daubechies Mother Wavelet and reference watermark will be embedded into ROI image using the Hash Block Chaining. Hash function to be used is SHA-256 with the MAC technique. After the embedding process at the RONI and ROI image, the next step is to merge the parts into a whole image that has been watermarked. Figure 2 show a flowchart of embedding process in this multiple watermarking system.

In the extraction process, the signature watermark and the reference watermark extracted separately. So to extract the signature watermark and the reference watermark, the watermark image (the result of multiple watermarking, either has or has not been subjected to attacks) will be separated in advance to get RONI and ROI image. In RONI image, the signature watermark will be extracted by Wavelet transform.

The results of the extraction process will be decoded again using the Reed–Muller Code in order to repair if there are bits of error in the extraction. The message of decode result is the signature watermark previously embedded on medical images. While the ROI image, reference watermark will be extracted by using the method of Block Chaining Hash. Extraction process of two watermarks is illustrated by Fig. 3.

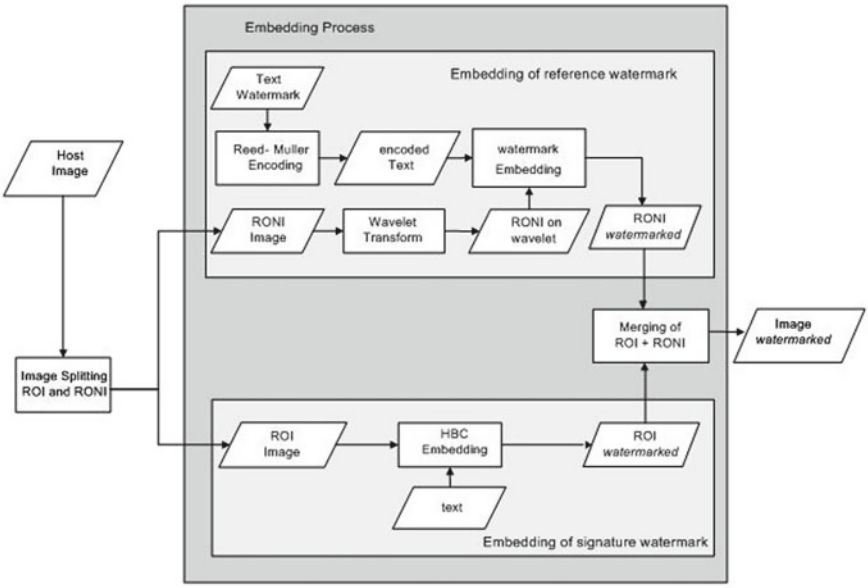


Fig. 2 Flow chart of embedding process of proposed multiple watermarking

3 Performance of System

The use of Error Correction Code (especially Reed–Muller Code) can reduce CER of 15.06 % (gaussian noise attack with SNR 20 dB) than not using the Error Correction Code at all. Robustness of signature watermark can be improved by using the Reed–Muller code, although not significantly. This is because the ability of the Reed–Muller Code that severely limited, the improvement of 1 bit of 8 bit errors that exist.

When signature watermark is embedded in the LL subband is more robust toward attacks of sharpening, blur, gaussian noise and JPEG compression rather than embedded in the subband LH, HL and HH. Reference watermark has a high level of fragility of being able to detect the low-level attacks, such as sharpening with alpha 0.02, blur with a radius of 0:55, gaussian noise with SNR 58 dB, JPEG compression with quality factor 99 %.

Fragility level of the reference watermark with MD5 higher than the MAC. The PSNR average of reference watermark MD5 is 0.6 dB lower (down 11.8 %) than MAC, but MAC is more secure because it uses a secret key during insertion and extraction. Multiple watermarking technique based on Wavelet and HBC has a good performance, it is seen from the PSNR average reaches of 46.6 dB.

If the level of Daubechies is increased, then the quality of the watermarked images were good, although not significantly so. The resulting PSNR ranged 47–50 dB with an average of 48.9 dB. MAC block size used in the HBC did not affect the quality of watermarked image. Watermarked image quality with MAC was similar to MD5. The

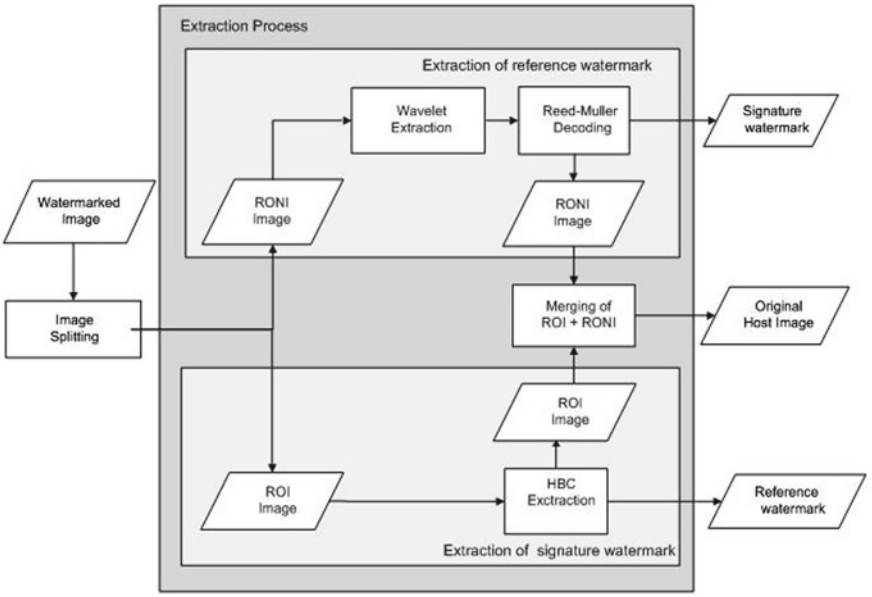


Fig. 3 Flow chart of extraction process of proposed multiple watermarking

resulting PSNR ranged from 47–49 dB. But the reference watermark embedding time by using the MAC faster than MD5. The difference time between 1 and 2 s, meaning MAC can reduce the time of MD5 to an average of 17.8 %. With a host of type JPEG image, the quality of watermarked images was generated on each color component has the same PSNR values. PSNR values for each color component of this value is not much different from the value of PSNR as a bitmap image of the host type. The resulting PSNR ranged from 46–49 dB.

Acknowledgments The authors would like to thank Telkom Institute of Technology for financial supporting in this research.

References

1. Agustina, R., Adiwijaya, Barmawi, A.M.: Pendeteksian dan Perbaikan Citra Termanipulasi yang Disisipi Watermark Menggunakan Block Truncation Coding (BTC) Berbasis Wavelet. Jurnal Telekomunikasi **15**, 116–122 (2010)
2. Boato, G., Natale, F.G., Fontanari, C.: Digital image tracing by sequential multiple watermarking. IEEE Trans. Multimedia **9**(4), 677–686 (2007)
3. Kallel, M., Lapayre, J.C., Bouhlel, M.S.: A multiple watermarking scheme for medical image in the spatial domain. GVIP J. **7**(1), 35–42 (2007)
4. Memon, N.A., Gilani, S.A.M., Qayoom, S.: Multiple watermarking of medical images for content authentication and recovery. In: Proceedings of IEEE 13th International Multitopic Conference, pp. 1–6 (2009)

5. Mostafa, S.A.K., El-Sheimy, N., Tolba, A.S., Abdelkader, F.M., Elhindy, H.M.: Wavelet packets-based blind watermarking for medical image management. Open Biomed. Eng. J. **4**, 93–98 (2010)
6. Pham, B., Woo, C.S., Du, J.: MultipleWatermark method for privacy control and tamper detection in medical images. In: Proceedings APRS Workshop on Digital Image, Computing (WDIC2005), pp. 59–64 (2006)
7. Wakatani, A.: Digital watermarking for ROI medical images by using compressed signature image. In: Proceedings of the 35th Hawaii International Conference on System Science, pp. 2043–2048 (2002)

A Novel Approach to Strategic Planning of Rail Freight Transport

Reyk Weiß, Jens Opitz and Karl Nachtigall

1 Introduction

In especially complex networks, timetabling is a protracted process, which, despite computer aided methods, comes along with a high manual effort. The reason is based on the huge amount of technical, operative and economical requirements and its dependencies among each other. Hence, the underlying infrastructure, the train characteristics and its resulting driving dynamic properties as well as the conflict-free positions of the individual train paths exert decisive influence on the time table. Additionally, the operational requirements, for example connections, symmetry and locomotive crew changes, play a major role. Not negligible are economical factors, such as the demand of specific train paths and their qualities (running times, waiting times, transfer times). Consequently, timetabling must be regarded as the core decision for an economically reasonable business.

Due to the large amount of constraints and the operator's obligatory required knowledge about geographic circumstances as well as the infrastructure, the manual editing is only possible for small subnetworks.

Particularly planning rail freight train paths based on an existing operation program for personal transport trains poses an outstanding challenge. The resulting rail freight train paths are highly dependent on the already existing train paths. Additional restrictions, like minimum number of train paths and a given set of quality factors, increase effort of the time tabling process as well. Furthermore, the manual

R. Weiß (✉) · J. Opitz · K. Nachtigall
Chair for Traffic Flow Science, TU Dresden, Dresden, Germany
e-mail: Reyk.Weiss@tu-dresden.de

J. Opitz
e-mail: Jens.Opitz@tu-dresden.de

K. Nachtigall
e-mail: Karl.Nachtigall@tu-dresden.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_69,

methods for determining rail freight train paths for different time table variants, in case of a predetermined time quota and budget, rapidly reaching their limits.

This presents the starting point for the software system TAKT [3–5]. In recent years, the group of the chair for traffic flow science at TU Dresden in close collaboration with the DB Netz AG have successfully developed a software system, which supports the decision maker in his tasks for the strategic planning phase. The software system TAKT automatically calculates and optimizes periodic train paths for complex railway networks by innovative approaches in strategic passenger and freight train timetabling.

In this work, the focus is set on the novel approach of strategic planning of rail freight transport by using the already implemented FLOWMASTER module of the software system TAKT. This results in different variants of strictly synchronized and conflict-free time tables with a maximum set of high quality rail freight train paths.

2 FlowMaster

The creation of the passenger's time table follows fixed standards like clearly defined paths and their runtimes and defined stops for each train. It exists a set of restrictions between the different trains, such as headway, symmetry and connections. Subsequently, the passenger operating program is fixed the degrees of freedom are only the departure times. The generation of rail freight trains has less restrictions. The path depends on an economic route between a defined start and end on a track which is open for rail freight trains. In addition, those trains only need to stop on specific stations for personnel changes or to allow passings of faster trains. For economic reasons, stops should have a minimum distance between each other.

The difficulty of the rail freight train generation is the determination of the taken tracks and all necessary stops between the start and the end based on the calculated passenger time table.

The novel approach of solving those problems lays in the creation of a set of sub paths, so called InfraAtoms, which can be combined to a conflict-free rail freight path. The creation and the reasonable combination of the InfraAtoms will be described in the following section.

2.1 InfraAtom Creation

In order to reach the goal of creating the InfraAtoms, we need to generate a set of sub paths which can be combined to almost all possible track variants.

Firstly, the generation of single rail freight trains with correct train characteristics as well as length, weight and traction units is done. The path will be bounded by the given start and end station for the train. By the built-in routing algorithm [6] of the software system TAKT, the best economic track will be found automatically.

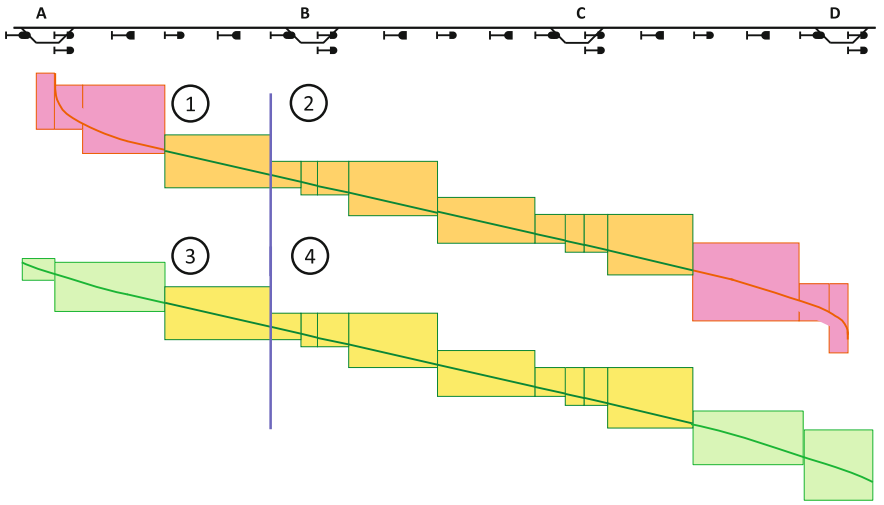


Fig. 1 The cutting of the train with and without stops at the intersection line results in four sub paths.

Additionally, it is possible to find a set of possible paths within a given detour factor which will generally cause a higher number of InfraAtoms. The driving dynamics of this train will be calculated without any stop of the train. This intends hat the train will start with a given speed and leaves the scope with the current speed.

Based on a calculated passing train an associated train will be generated, which stops on each possible station on the track that matches the requirements like minimum length. At this point, it is very important to choose the correct stop for the train in a station. Usually, it exists different possible stop nodes in a station. The passing train uses the through track. For this reason, the passing stop node should be excluded to prevent stops on the through track. The stop nodes on the opposing direction should be excluded as well, in order to avoid capacity degradation. Finally, choosing the stop node on the siding track is the best solution. If there is no suitable siding track, no stop will be automatically generated.

In the following step the trains with and without stops are split into sub paths. The intersection is a point on the track between two stops, where the blocking time of both trains is similar. As shown in Fig. 1, the intersection line is before station B at the end of a similar block. This cut results in a set of four sub paths. The halt train is cut into part 1 and 2 and the passing train into part 3 and 4. By the combination of sub path 1 and 4 and sub path 3 and 2 it is possible to create two additional InfraAtoms.

The routing algorithm and the driving dynamic calculation generate trains, which are traveling by using the maximum possible train speed which is limited by the maximal velocity of the traction unit and the maximal track velocity. For capacity reasons it could be a better solution to choose more than one slower train instead of one fast train. For these cases, it is useful to create additional InfraAtoms which

are adjusted to the velocity of the trains associated to the current track. In order to calculate the new maximum velocity, a graph-theoretical approach will be used.

Based on the determined path of the passing train, a two dimensional network will be created with a set of nodes and edges. The x-axis defines the distance from 0 Km to track length and the y-axis defines the time interval from 0 min to the train period. For each dynamic distance step and static time step a node will be created. The connection between two nodes has the necessary speed as a property which is determined by the distance and time delta of the start and end node. In this way, it is also possible to model additional stops by creating connections between nodes at the same distance.

The resulting network can be reduced by eliminating all nodes whose time values are intersecting the blocking time of an existing train. Edges will not be generated, if the corresponding speed values do not match the current train characteristics.

Finally, a shortest path algorithm can determine the best path with the highest possible train velocity and necessary additional stops. The resulting train with the adjusted velocity will be used as a base train for the InfraAtom creation.

2.2 InfraAtom Combining

The InfraAtom creation results in a directed acyclic graph (DAG) like illustrated in Fig. 2 with a set of InfraAtoms $K = \{1, \ldots, 15\}$ which can be used to construct a path from a source InfraAtom in $Q = \{1, 2, 3, 4\}$ to a sink InfraAtom in $S = \{13, 15\}$. For modeling the InfraAtom connections, the function

$$\begin{aligned} I : K &\to 2^K \\ s &\mapsto I(s) \end{aligned}$$

maps each InfraAtom $s \in K$ to the set of its incoming InfraAtoms and the function

$$\begin{aligned} O : K &\to 2^K \\ s &\mapsto O(s) \end{aligned}$$

maps each InfraAtom $s \in K$ to the set of its outgoing InfraAtoms, respectively. Based on the graph shown in Fig. 2 $I(13) = \{8, 9\}$, $I(3) = \emptyset$ and for outgoing related InfraAtoms $O(1) = \{5, 6\}$, $O(3) = \{8\}$.

The remaining task is the determination of a train path defined as a set of InfraAtoms $T \subseteq K$ such that

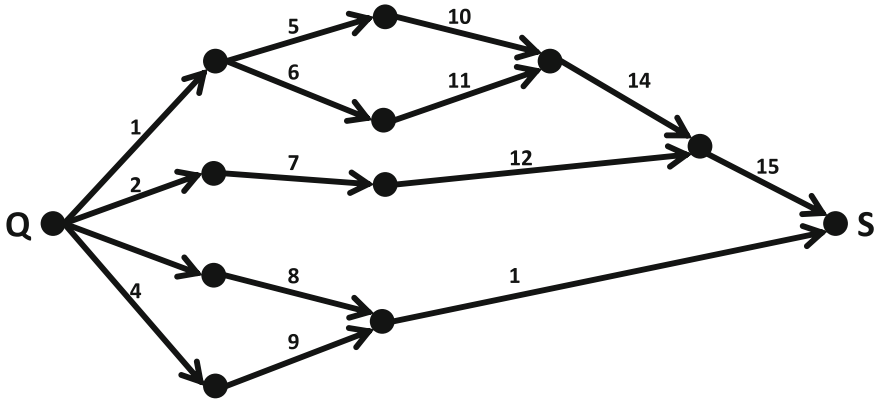


Fig. 2 Illustration of an example graph with a set of 15 InfraAtoms $K = \{1, \ldots, 15\}$ including 4 source InfraAtoms $Q = \{1, 2, 3, 4\}$ and 2 sink InfraAtoms $S = \{13, 15\}$.

$$\exists! s \in Q : s \in T$$
$$\exists! s \in S : s \in T$$
$$\forall s \in T \backslash Q : |I(s) \cap T| = 1$$
$$\forall s \in T \backslash S : |O(s) \cap T| = 1$$

Finding T can be done by a polynomial time algorithm. However, each InfraAtom is bound to constraints of a Periodic Event Scheduling Problem (PESP), which is NP-complete [7]. Hence, finding such T is NP-complete as well. Subsequently, the resulting PESP regarding the InfraAtom graph will be encoded to the Boolean Satisfiability Problem (SAT) [1, 2] and be solved by a state-of-the-art SAT solver. This encoding is not covered in this work.

3 Results

In a real-world example the FLOWMASTER was tested to determine and maximize the freight train paths on a highly frequented corridor. On this track, with a total distance of about 230 km, an operation program for passenger transportation was given. In a base period of 120 min, 88 trains with different periods from 30 min up to 120 min are traveling on the track. In addition, 10 connection restrictions between trains were given as well. In the first step, the FLOWMASTER module generated a set of 96 InfraAtoms. The quality of a rail freight train can be measured by the parameter $BFQ = t_{\text{run}}/t_{\text{fastestrun}}$, which is the quotient between the runtime and the fastest runtime. A train whose BFQ is higher than 1.4 is not acceptable. The FLOWMASTER was set up to determine 10–16 rail freight train paths. By the combination of the InfraAtoms, the algorithm could automatically find 10–15 rail

freight train paths with a ranging BFQ from ≈ 1.23 to ≈ 1.8. A higher number of paths results in more stops on the track with longer halt times, which actually cause a higher BFQ. Finally, the best time table was calculated for 12 rail freight train paths with a total BFQ of about 1.3. The experiment has successfully shown that the algorithm could find five different time tables, each for every train path count, in half an hour. Creating such a set of time table variants manually would take at least half a year, but only by a worker with a deep knowledge of the infrastructure. By comparison, the automated generated rail freight train path with the manual created ones, the correctness and usability as well as the high performance of the novel approach could be proved.

4 Conclusion

In this work it is shown how to create rail freight train paths based on an existing passenger operation program. Firstly, a useful set of sub paths (InfraAtoms) based on the current time table is created. In terms of capacity reasons, it is useful as well to create InfraAtoms with different velocity, fitting the track requirements and the trains traveling on the track. By combining the InfraAtoms, it is possible to create new train paths which will be accepted if the Periodic Event Scheduling Problem is feasible for the restriction system with the additional restrictions for the new train path. Although the number of InfraAtoms is still quite limited, experimental results have shown that this technique can be successfully applied to real-world scenarios.

References

1. Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.): Handbook of Satisfiability. IOS Press (2009)
2. Großmann, P., Hölldobler, S., Manthey, N., Nachtigall, K., Opitz, J., Steinke, P.: Solving periodic event scheduling problems with SAT. In: IEA/AIE. LNAI, vol. 7345, pp. 166–175. Springer (2012)
3. Großmann, P., Weiß, R., Opitz, J., Nachtigall, K.: Automated generation and optimization of public railway and rail freight transport time tables. MTM **5**, 23–26 (2012)
4. Nachtigall, K.: Periodic Network Optimization and Fixed Interval Timetable. Habilitation thesis, University Hildesheim (1998)
5. Opitz, J.: Automatische Erzeugung und Optimierung von Taktfahrplänen in Schienenverkehrsnetzen. Logistik,Mobilität und Verkehr, Gabler Verlag(2009)
6. Pöhle, D., Weiß, R., Opitz, J., Großmann, P., Nachtigall, K.: Innovative automatische Laufwegsuche mittels neuartigem Routing-Algorithmus zum Nutzen der Fahrplanoptimierung. VWT, Dresden (2012)
7. Serafini, P., Ukovich, W.: A mathematical model for periodic scheduling problems. SIAM J. Discrete Math. **2**(4), 550–581 (1989)

Proximal Bundle Methods in Unit Commitment Optimization

Tim Drees, Roland Schuster and Albert Moser

1 Introduction

The residual load can be described as difference between demand for electrical energy and feed-in of renewable energy sources. Their prioritized feed-in is highly volatile and will increase the volatility of the residual load in the future. It has to be covered by thermal and hydro power plants. At this, the unit commitment has to be determined under optimal costs and a macroeconomic solution for the optimal power plant schedule is estimated by using a unit commitment optimization.

Due to the increased load gradients it can be observed that the rate of convergence of existing methods based on Lagrangian Relaxation is slowing down. Therefore new methods with a higher rate of convergence need to be applied. This paper intends to exemplify the use of Proximal Bundle Methods in unit commitment optimization in hourly resolution forcasting one year. Hence, in a first step the mathematical model based on a Lagrangian Relaxation is illustrated. Afterwards, the iterative approach is demonstrated. Finally it is compared to the existing approach based on a gradient descent method in terms of robustness, rate of convergence, computation time and number of iterations. Besides, the article gives an outlook on necessary and possible extensions of the method.

2 Mathematical Model

This paper focusses on unit commitment optimization of the European power system. Due to the geographic field of observation the system is modeled in several single market areas. Between market areas cross-border exchange of electrical energy is possible under limited capacities. Every market area is defined by hourly demand of

T. Drees (✉) · R. Schuster · A. Moser
Institute of Power Systems and Power Economics (IAEW), RWTH Aachen, Schinkelstrasse 6, 52062 Aachen, Germany
e-mail: tim.drees@iaew.rwth-aachen.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_70,

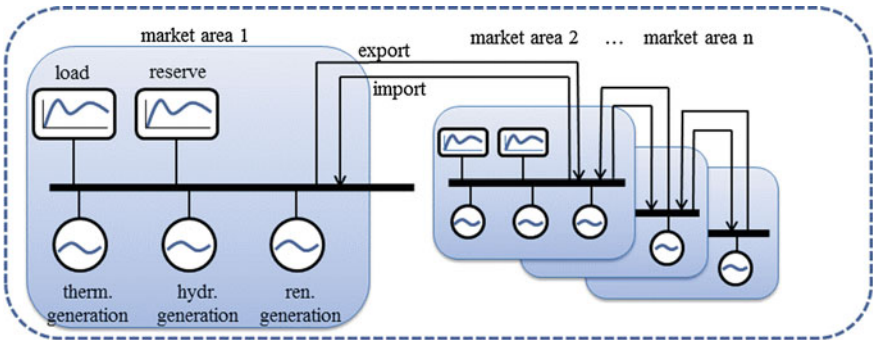


Fig. 1 System model

electrical energy and feed-in of renewable energies. The difference between demand, feed-in of renewable energies and cross-border exchange gives the residual load. Furthermore a market area consists of a hydrothermal generation system to cover the residual load (Fig. 1).

2.1 Objective Function

The unit commitment optimization is formulated as a minimal-cost problem. At this, the operational costs of the generation system need to be minimized (see 2). These costs consist of fuel costs FC and start-up costs SC. In this context, costs of hydro power plants are zero and do not need to be modeled. While the fuel costs depend on the current power plant output p, the start-up costs increase with the downtime of the plant. Hence, the objective function (OF) can be constituted as the sum of FC and SC over all thermal power plants $i \in I$ and time intervals $t \in T$:

$$OF(p,e) = \min_{e,p,s,w} \sum_{t=1}^{T} \sum_{i=1}^{I} (FC(p_i^t, e_i^t) + SC(e_i^t))$$

In the modeling of thermal power plants it has to be taken into account that thermal power plants have technical restrictions like start-up procedures or minimum up- and down-times. Therefor the binary variable e is implemented to model the operation mode and approves the feed-in only between a minimal and maximal power level $e_i^t \cdot p_{min} \leq p_i^t \leq e_i^t \cdot p_{max}$. During operation e is set to one, otherwise it is zero.

2.2 Constraints

In order to ensure operation of the grid and to guarantee security of supply, the residual load d^t and a reserve r^t have to be covered. These two requirements are added to the minimization problem as linear constraints. The load is covered, if the feed-in of all power plants equals the residual load d^t:

$$\sum_{i=1}^{I} e_i^t \cdot p_i^t + \sum_{j=1}^{J} (s_j^t - w_j^t) = d^t$$

In this definition J indicates the number of hydro power plants, s_j represents the feed-in and w_j (if existing) the demand of the pumped-storage hydro power plant j. The reserve coverage constraint can be formulated as:

$$\sum_{i=1}^{I} e_i^t \cdot (p_i^{max} - p_i^t) + \sum_{j=1}^{J} (s_j^{max} - s_j^t) + \sum_{j=1}^{J} w_j \geq r^t$$

Besides these two major constraints, several other system and time coupling constraints are modeled, like minimum up- and down-times of power plants and the continuity equation of hydro power plants.

2.3 Lagrangian Relaxation

The time horizon of the optimization is up to one year in hourly resolution t. Due to the complexity of the resulting problem a solution by mixed integer algorithms is not applicable. Hence, in this paper a solution based on Lagrangian Relaxation similar to *3* is presented. At this, the most complicated constraints of the problem, the load and reserve coverage constraints, are multiplied with lagrangian factors λ and μ and are written into the objective function:

$$L(e, p, s, w, \lambda, \mu) = \sum_{t=1}^{T} \{ \sum_{i=1}^{I} (FC(p_i^t, e_i^t) + SC(e_i^t)) + \mu (r^t - \sum_{i=1}^{I} (e_i^t \cdot p_i^{max} - p_i^t) - \sum_{j=1}^{J} (s_j^{max} - s_j^t) - \sum_{j=1}^{J} w_j) + \lambda (d^t - \sum_{i=1}^{I} e_i^t \cdot p_i^t + \sum_{j=1}^{J} (s_j^t - w_j^t)) \}$$

Due to this conversion it is necessary not only to minimize over e, p, s and w but to maximize over the lagrangian factors λ, μ in order to meet the relaxed constraints. Therefor the new formulation of the problem can be written as

$$\max_{\lambda, \mu} \min_{e, p, s, w} L(e, p, s, w, \lambda, \mu),$$

where $\min_{e,p,s,w} L(e, p, s, w, \lambda, \mu)$ is called the dual lagrangian function D(λ, μ). Due to the non-continuous primal function, the problem does not exhibit strong duality. Nevertheless, a small gap can be supposed, if the number of variables e, p, s and w is large (see 4).

2.4 Solving the Relaxed Problem

The dual lagrangian function can be divided into two independent subproblems: a thermal subproblem and a hydraulic subproblem. The thermal subproblem can be solved by Dynamic Programming and the hydraulic subproblem as a Network Flow Problem. The aim is the maximization of D(λ, μ), which is solved iteratively. As gradient descent methods show decreasing speed of convergence by increasing volatility of the residual load d^t, in this paper a solution based on Proximal Bundle Methods is presented.

2.5 Proximal Bundle Method

The Proximal Bundle Method (PBM) of K. C. Kiwiel (see 1) generates a polyhedral model $\hat{D}(\omega)$ of the dual lagrangian function to approach the maximum of the concave function D(ω). It generates a sequence of points $\{\psi^k\}_{k=1}^{\infty}$, that converges to the optimum in finitely many steps. In every iteration k a trial point ω^k is selected and a linearization plane $\bar{D}(\omega^k)$ is calculated. $\bar{D}(\omega^k)$ will improve the polyhedral model $\hat{D}(\omega) = \max\ \bar{D}(\omega^a),\ a = 1, \ldots, k$. A new trial point can be calculated by solving the quadratic problem:

$$\omega^{k+1} = \arg\min\,(\hat{D}(\psi^k) + \frac{s}{2} \cdot ||\omega - \psi^k||^2)$$

Herein s is a choosable step size, which can be adapted by different methods. Afterwards ω^{k+1} evaluated to ψ^k with

$$D(\omega^{k+1}) \geq D(\psi^k) + m \cdot (\hat{D}(\omega) - D(\psi^k)),$$

where $m \in [0, 1]$ is a fixed parameter. If ω^{k+1} is "better" than ψ^k, ω^{k+1} becomes the new best point ψ^{k+1}. Otherwise ψ^{k+1} remains ψ^k. This method can be applied to the unit commitment problem, if ω is defined as a vector including all lagrangian factors λ^t and μ^t.

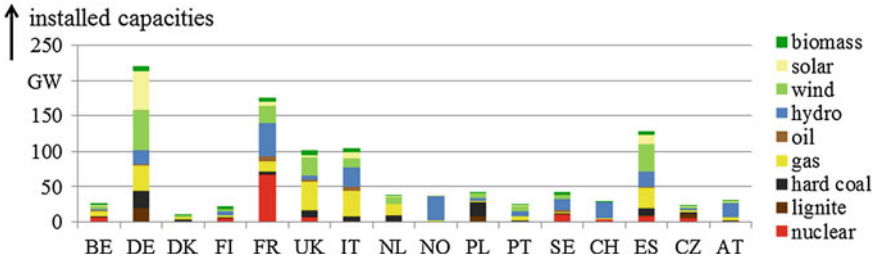


Fig. 2 Assumed capacities Europe 2022

3 Case Study

The introduced Proximal Bundle Method is now being applied in a market simulation for Europe in the year 2022. Due to the hourly resolution and two relaxed constraints the problem consists of 17,520 dual variables. The results are compared to the ones based on a gradient descent method (GDM). The used scenario is based on the Scenario Outlook and Adequacy Forecast (SOAF) of the ENTSO-E for the year 2020 (Scenario B) and is expanded by the grid development plan for Germany for the year 2022 (Scenario B). The number of considered units over all market areas amounts to 1,140 thermal and 380 hydro power plants. Figure 2 gives an overview of the assumed capacities of the year 2022 in Europe.

4 Exemplary Results

Figure 3 illustrates the dual objective function value over the iterations for PBM and GDM in the market area of Germany. In the first 20 iterations the rate of convergence of GDM is higher than the one of PBM. However, from iteration 21–43 the PBM

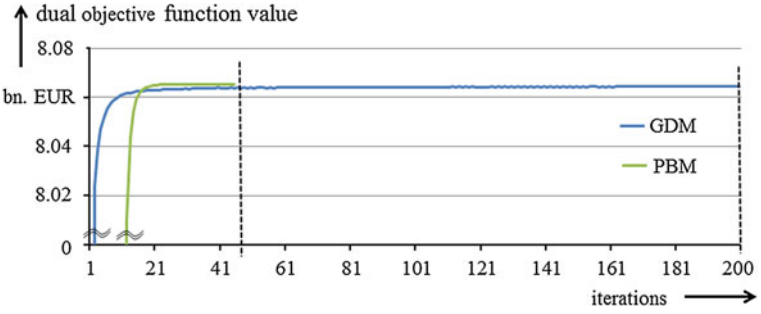


Fig. 3 Dual objective function value

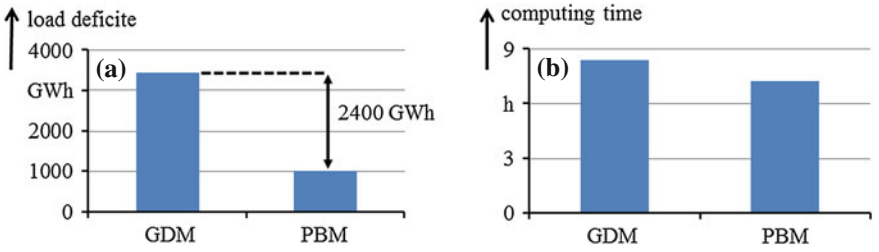


Fig. 4 **a** Load deficite. **b** Computing time

converges to a maximum of 8.0656 bn. EUR. At iteration 43 an internal stopping criterion (see 1) ends the iteration. The GDM can not reach this maximum value within 200 iterations. This example shows the advantage of generating a polyhedral model, which accelerates the rate of convergence with any additional linearization plane. The results of the European optimization are illustrated in Fig. 4. Figure 4a shows the calculated load deficite of PBM and GDM. The difference between both methods is about 2,400 GWh. Even though the generation costs of both methods are at level, it becomes clear that the PBM has found a better solution with a lower violation of the load coverage constraint. In Fig. 4b the computing time of both methods is compared. The analysis shows, that the computing time is decreased by PBM due to the reduced number of iterations in every market area.

5 Summary and Outlook

This paper aimed at illustrating the use of Proximal Bundle Methods in large scale unit commitment optimization based on a Lagrangian Relaxation. In comparison to a gradient descent method the Proximal Bundle Method can significantly reduce computation time due to an increased rate of convergence. In addition the generation of a polyedric model leads to a minor violation of the load coverage constraint. Future work will focus on the development of a Lagrangian Heuristic to generate feasible solutions by little changes in the schedule after the Lagrangian Relaxation. Furthermore additional constraints on the unit commitment for scenarios with a very high share of renewable energy sources will be investigated and modeled.

Hybrid Algorithm for Blind Equalization of QAM Signals

Abdenour Labed

1 Introduction

In digital communication systems, adaptive equalizers play a central role in combating signal distortions. Due to bandwidth limitations, the transmission channel may exhibit inter symbol interference (ISI) and an equalizer is used at the receiver to correct the signal distortion caused by the ISI.

In supervised equalization approaches a pilot sequence known by the receiver is periodically sent to help for the minimization of the distance between the desired and the actual output. Whereas, in blind (or unsupervised) equalization ones only some statistical properties of the transmitted signals are needed, to adapt the tap weights of equalizers.

One of the most known algorithms of adaptive blind equalization is the constant modulus algorithm (CMA) [1]. It is known to be robust and requires simple hardware implementation [2].

However, for QAM modulations (non constant modulus), used in high data-rate communications' systems, the CMA leads to not sufficiently low residual errors for a correct recovering of transmitted signals. Other algorithms, known to be more adapted to QAM symbols, have been proposed in the literature. For instance, the multi-modulus algorithm (MMA) [3] and the extended constant modulus (ECMA) [4] algorithms have been shown to achieve better equalization of QAM signals than does the CMA.

Unfortunately, for high-order QAM modulations, even these algorithms are unable to provide low residual ISI that allow reliable detection of the correct symbols. A way to improve performance CMA, MMA or ECMA, for dense QAM constellations, is to penalize their criteria by terms based on alphabet-matching functions

A. Labed (✉)
Computer science, Ecole Militaire Polytechnique, BP 17 Bordj El Bahri, 16111 Algiers, Algeria
e-mail: abd.labed@gmail.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_71,

(AMF) which leads to what is called hybrid adaptive blind equalization algorithms. They perform better than their classical counterparts.

One of the first AM criteria has been proposed by Barbarossa and Scaglione [5]. Beasley and Cole-Rhodes combined the MMA and this AM cost functions [6] and showed, using 16-QAM symbols, that the hybrid algorithm outperforms the original dual-mode CMA/AMA [5] and the MMA. In a similar way, He et al. [7] proposed an AM cost function based on even powers of sine functions and combined it with the CMA and demonstrated that hybrid algorithms have higher performance than conventional ones. Blind equalization algorithms are generally based on the minimization of cost functions that measure the closeness of the equalizer output to the constellation points and the stochastic gradient algorithm is commonly used to achieve numerically this minimization.

Our contribution consists of a combination of the ECMA cost function and a modified version of the AM cost function proposed in [5]. The modification is made to account for amplitude and phase information of the transmitted signals, and hence, improve the local convergence which leads to improved global performance. The performance of the derived algorithm has been studied for 16-QAMs (low order) in [8]. But in this work, its effectiveness is shown through simulation for 512-QAM (high order) signaling where performance is compared to those of the conventional MMA and ECMA together with a penalized adaptive version of the extended constant modulus algorithm, based on a squared cosine function [7].

2 Blind Equalization Model

Let us assume that a sequence $(s_n)_{n\in\mathbb{Z}}$ of independent-identically distributed (i.i.d.) symbols, drawn from a QAM constellation, is transmitted over the time invariant channel **h** of length K. The mathematical model for the receiver input at time n is

$$x_n = \sum_{l=0}^{K-1} h_l s_{n-l} + v_n \tag{1}$$

We introduce vectors $\mathbf{w} = [w_0, \ldots, w_{L-1}]^T$ and $\mathbf{x_n} = [x_n, \ldots, x_{n-L+1}]^T$ where $(\cdot)^T$ denotes transpose, then the equalizer output is

$$z_n = \mathbf{w}^{-H}\mathbf{x}_n = \sum_{l=0}^{L-1} w_l^* x_{n-l}. \tag{2}$$

with $(\cdot)^*$ and $(\cdot)^H$ the conjugate and transpose conjugate operators, respectively.

3 MMA and ECMA Algorithms

The idea of the multi-modulus, is to split the cost function into a real and an imaginary part to take account not only for amplitude information but also for phase information, which helps in better equalizing QAM signals than the CMA does.Whereas the extended constant modulus is based on the generalization of the modulus of a complex number.

3.1 MMA Algorithm

The cost function of the MMA to be minimized w.r.t $\mathbf{w}$ is given by

$$J_{MM}(\mathbf{w}) = \mathbb{E}\left\{\frac{1}{4}\left(\left(z_{nr}^2 - R_m\right)^2 + \left(z_{ni}^2 - R_m\right)^2\right)\right\}, \tag{3}$$

where $\mathbb{E}\{\cdot\}$ denotes expectation and z_{nr} and z_{ni} are the real and imaginary part of z_n, respectively. R_m is a dispersion constant given by

$$R_m = \frac{\mathbb{E}\left\{(s_{kr})^4\right\}}{\mathbb{E}\left\{(s_{kr})^2\right\}} = \frac{\mathbb{E}\left\{(s_{ki})^4\right\}}{\mathbb{E}\left\{(s_{ki})^2\right\}}. \tag{4}$$

with $s_k = s_{kr} + j s_{ki}$ a QAM symbol. Letting M denote the order of the modulation, we have $s_{nr},\ s_{ni} \in \{a_k;\ a_k = a_{kr} + j\, a_{ki},\ k = 1, .., M\}$, with $a_{kr},\ a_{ki} \in \left\{\pm 1, \pm 3, \ldots, \pm(\sqrt{M} - 1)\right\}$.
Equalizer coefficients are updated from formulas

$$\begin{cases} \mathbf{w}_{n+1} = \mathbf{w}_n - \mu\ \varphi_n \mathbf{x}_n \\ \varphi_n \quad\ = \left(z_{nr}^2 - R_m\right) z_{nr} - j\left(z_{ni}^2 - R_m\right) z_{ni}. \end{cases} \tag{5}$$

3.2 ECMA Algorithms

The ECMA algorithm is a particular case of the generalized constant modulus algorithms proposed in [4]. Introducing the generalized complex modulus as

$$|z_n|_p = \left(|z_{nr}|^p + |z_{ni}|^p\right)^{\frac{1}{p}}\ , \tag{6}$$

for $p \geq 1$, z_{nr} and z_{ni} denoting the real and imaginary part of z_n respectively, the ECMA criterion is given by

$$J_{ECMA}(\mathbf{w}) = \mathbb{E}\left\{\frac{1}{4}\left(|z_n|_4^2 - R_E\right)^2\right\}, \tag{7}$$

with the dispersion constant

$$R_{4,2} = \frac{\mathbb{E}\left\{|s_k|_4^4\right\}}{\mathbb{E}\left\{|s_k|_4^2\right\}}. \tag{8}$$

The learning rule for this algorithm is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - 4\mu\left(|z_n|_4^2 - R_{4,2}\right)\frac{z_{nr}^3 - jz_{ni}^3}{|z_n|_4^2}\mathbf{x}_n\ . \tag{9}$$

4 Hybrid Algorithms

They combine the CMA, the MMA or the ECMA cost function with an AM-based criterion. The AMF has to vanish at each constellation point, to be symmetric around these points and uniform with respect to them. The former property means that in the case of perfect equalization, the equalizer output coincides with one of the constellation points (zero error). The latter implies that the AMF functions do not favor any data symbol in the alphabet over another.

The AM cost function proposed in [5] is

$$J_{\text{Gauss}}(\mathbf{w}) = \mathbb{E}\left\{1 - \sum_{k=1}^{M} e^{-\frac{|z_n - a_k|^2}{2\sigma^2}}\right\} \tag{10}$$

where $a_k (k = 1, \ldots, M)$ are the constellation points and the parameter σ controls the width of the nulls.

The squared cosine AM cost function proposed in [7] has the form

$$J_{\text{Cos}}(\mathbf{w}) = \mathbb{E}\left\{\cos^2\left(\frac{z_{nr}\pi}{2d}\right) + \cos^2\left(\frac{z_{ni}\pi}{2d}\right)\right\}, \tag{11}$$

with $2d$ the minimum distance between symbols.

In our scheme the AM term given by (9), is modified such that the terms of the exponential are split into real and imaginary parts instead of the modulus, as in (10). It transforms to

$$J_{\text{MGauss}}(\mathbf{w}) = \mathbb{E}\left\{\left(1 - \sum_{k=1}^{M} e^{\frac{-(z_{nr} - a_{kr})^2}{2\sigma^2}}\right) + \left(1 - \sum_{k=1}^{M} e^{\frac{-(z_{ni} - a_{ki})^2}{2\sigma^2}}\right)\right\} \tag{12}$$

We refer to the resulting hybrid algorithms as Cos-ECMA and MGauss-ECMA, if the AM is that given by (10) and (11), respectively.

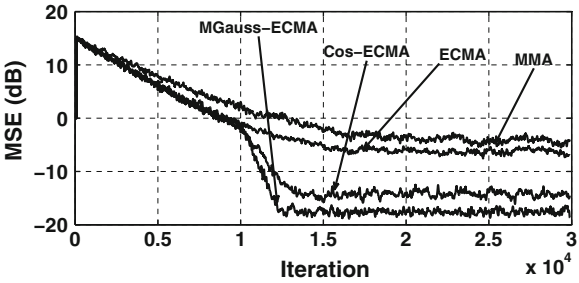


Fig. 1 Comparison of MSE for 512-QAM signals (SNR = 40 dB)

Hence, combining the ECMA cost function with the modified Gaussian AM term we get

$$J_{\text{MGauss}}(\mathbf{w}) = J_{\text{ECMA}}(\mathbf{w}) + \beta J_{\text{MGauss}}(\mathbf{w}) \tag{13}$$

with β a parameter that trades off between the ECMA and the MGauss terms.

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \nabla J_{\text{MGauss}} \tag{14}$$

5 Numerical Experiment

The mean square error (MSE) of the MGauss-ECMA is compared to that of the Cos-ECMA, together with that of the MMA and the ECMA. For comparison, high order 512-QAM symbols are randomly drawn, the equalizer length is set to L = 11 and channel

$h = [1 \quad 0.1294 + 0.483j]^T$ is used to transmit these symbols under a signal to noise ratio; SNR = 40 dB.

The values of the parameters (μ, β and σ) used in the numerical implementation are $\mu = 6\times10^{-9}$ and $\mu = 2\times10^{-9}$ for MMA and ECMA, respectively. Whereas, for Cos-ECMA $\mu = 2\times10^{-9}$ and $\beta = 1,000$ while for MGauss-ECMA $\mu = 2\times10^{-9}$, $\beta = 100$ and $\sigma = 0.4$.

From Fig. 1, we can notice that the hybrid algorithm Gauss-ECMA has the lowest residual error at steady state (−18 dB). It can also be seen that the Cos-ECMA has a lower residual error than that of ECMA and MMA.

6 Conclusion

The hybrid adaptive blind equalization algorithm combining the ECMA cost function and the modified Gaussian AM cost function (MGauss-ECMA) has been compared to the conventional MMA and ECMA algorithms as well as to hybrid algorithm

combining the ECMA and squared cosine function (Cos-ECMA). The simulations performed for the high-order QAM modulation, namely 512-QAM, shows that the two hybrid versions perform much better than conventional ones. However, the MGauss-ECMA lead to lower MSE than the Cos-ECMA.

References

1. Godard, D.N.: Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE Trans. Commun. **28**(11), 1867–1875 (1980)
2. Johnson, C.R., et al.: Blind equalization using the constant modulus criterion: a review. Proc. IEEE **86**(10), 1927–1950 (1998)
3. Yang, J., et al.: The multi-modulus blind equalization algorithm. In: Proceedings of IEEE 13th International Conference on, Digital Signal Processing (2006).
4. Li, X.L., Zhang, X.D.: A family of generalized constant modulus algorithms for blind equalization. IEEE Trans. Commun. **54**(11), 1913–1917 (2006)
5. Barbarossa, S., Scaglione, A.: Blind equalization using cost function matched to the signal constellation. In: Proceedings of Asilomar, pp. 550–554 (1997).
6. Beasley, A., Cole-Rhodes, A.: Blind adaptive equalization for QAM signals using an alphabet-matched algorithm. IEEE GLOBECOM (2006).
7. He, L., et al.: A hybrid adaptive blind equalization algorithm for QAM signals in wireless communications. IEEE Trans. Sig. Process. **52**(7), 2058–2069 (2004)
8. Labed, A., et al.: New hybrid adaptive blind equalization algorithms for QAM signals. Proc. ICASSP 2809–2812 (2009)

Automatic Scheduling of Periodic Event Networks by SAT Solving

Peter Großmann

1 Introduction

Computing a time table for a given railway network is based on periodic events and constraints imposed on these events. Events and their constraints can be modeled by so-called periodic event networks. The periodic event scheduling problem (PESP) consists of such a network and is the decision problem, whether all the events can be scheduled such that a set of constraints—specified by the network—is satisfied. The problem is *NP*-complete [9] and the currently best solutions are obtained by constraint-based solvers [7, 8] or by LP solvers, which solve linearized PESP instances by introducing modulo parameters [6]. However, these solvers are still quite limited in the size of the problem that they can tackle, which will be discussed in the results section.

In recent years, the performance of SAT solvers has been significantly increased and SAT solvers are now applied in real-world settings such as hardware verification or planning [1]. Hence, the question arises how state-of-the-art SAT solvers perform on encoded periodic event scheduling problems.

After introducing SAT and PESP in the preliminaries presented in Sect. 2, the encoding is discussed in Sect. 3, followed by a soundness and completeness proof. In Sect. 4, several real-world instances are presented to a state-of-the-art PESP solver [7] and a state-of-the-art SAT solver. The results indicate, that the SAT approach is by far superior to the PESP solver. The paper is concluded in Sect. 5 by a short discussion and an outline of future work.

P. Großmann (✉)
TU Dresden, Chair of Traffic Flow Science, Dresden, Germany
e-mail: peter.grossmann@tu-dresden.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_72,

2 Notations and Preliminaries

Both the satisfiability problem (SAT) and the periodic event scheduling problem (PESP) will be introduced in this section.

2.1 Satisfiability Problem

A *satisfiability problem* (SAT) consists of a propositional formula F and is the decision problem whether it exists an interpretation (or assignment) J from the set of propositional formulas to the set $\{\top, \bot\}$ of truth values such that J assigns $\top$ to F. In such a case, J is called *model* for F ($J \models F$) and F is said to be *satisfiable*. SAT is *NP*-complete [2].

It is well-known, that each propositional formula can be transformed into a semantically equivalent formula in conjunctive normal form (CNF), where a formula $F = \langle C_1, \ldots C_m \rangle$ is in CNF if it is a conjunction of m clauses, a clause $C = [L_1, \ldots, L_n]$ is a disjunction of n literals, and a literal L is either an atom p or the negation of an atom $\neg p$. Most modern SAT solvers accept SAT instances in CNF. For more details about SAT can be found elsewhere [1].

2.2 Periodic Event Scheduling Problem

Let $l, u \in \mathbb{Z}$ and $t \in \mathbb{N}$. $[l, u] := \{x \in \mathbb{Z} \mid l \leq x \leq u\}$ denotes the interval from l to u and $[l, u]_t := \bigcup_{z \in \mathbb{Z}} [l + z \cdot t, u + z \cdot t] \subseteq \mathbb{Z}$ the interval from l to u modulo t, which is called *periodic extension*.

Let (V, E) be a directed graph, $t \in \mathbb{N}$, and a a mapping, which assigns to each edge $(i, j) \in E$ an interval modulo t. $N = (V, E, a, t)$ is called *periodic event network* (PEN) with t being its period, V its set of (periodic) events, and $a(i, j)$ its constraint for each edge $(i, j) \in E$. The function $\Pi : N \to [0, t - 1]$, called *schedule*, assigns to each event in N an integer called the event's potential. In the sequel, let N be a PEN and Π a schedule for N.

Let $a(i, j) = [l, u]_t$ be a constraint for some edge $(i, j) \in V$. This constraint specifies a time consuming process such that the time between event i and j must be in $[l, u]_t$. Formally, $[l, u]_t$ *holds under* Π iff $\Pi(j) - \Pi(i) \in [l, u]_t$. A schedule Π is *valid* for N ($\Pi \models N$) iff all constraints occurring in N hold under Π.

The *periodic event scheduling problem* (PESP) consists of a PEN and is the decision problem whether it exists a valid schedule. More details about the definition and conclusions of the PESP can be found in the literature [3, 4, 7–9].

3 Encoding PESP as SAT

This section describes the encoding of a PESP instance to a SAT problem. Similarly as converting constraint satisfaction problems into SAT, the finite domains of the events can be translated in very different ways. Due to the lack of space, we present in this work only the *order encoding* [10] in details.

Because domains in PESP are subsets of $\mathbb{Z}$, especially they are sets of intervals, it is natural to apply the order relation $\leq$ on $\mathbb{Z}$. For the proofs of the lemmas presented in this section we refer to [3].

Let $q_{n,i}$ be a propositional variable with $i \in [-1, t-1]$ and $n \in V$. Then the variable $q_{n,i}$ is interpreted as $\Pi(n) \leq i$ and $\neg q_{n,i}$ as $\neg(\Pi(n) \leq i)$ which is equivalent to $\Pi(n) \geq i+1$. The function *enc* maps the set of events V to a propositional formula in CNF, such that this ordering holds:

$$enc(n) = (\neg q_{n,-1} \wedge q_{n,t-1}) \bigwedge_{i \in [0,t-1]} (\neg q_{n,i-1} \vee q_{n,i})$$

This encoding for variables of finite ordered domains is discussed in details elsewhere [10]. In order to encode all events' potentials we define $\Omega_N := \bigwedge_{n \in V} enc(n)$.

To encode a constraint $c = [l, u]_t \in a(i, j)$ we take a deeper look at all feasible pairs $(\Pi(i), \Pi(j))$, that hold under c. Uniting all these pairs are called the *feasible region* $S_c := [0, t-1] \times [0, t-1] \setminus P_c$ respectively the union of every other pair the *infeasible region* $P_c := \{(\Pi(i), \Pi(j)) \mid \Pi(j) - \Pi(i) \notin c\}$. Figure 1 shows the feasible and infeasible regions for the constraint $[3, 5]_{10}$ and all pairs of a particular rectangle, that shall be excluded. In general, let $c \in a(i, j)$ be a constraint. Then we can exclude all infeasible pairs of a rectangle $[i_1, i_2] \times [j_1, j_2]$ that is a subset of the infeasible region P_c with a single clause:

$$enc\_rec([i_1, i_2] \times [j_1, j_2]) = [\neg q_{i,i_2}, q_{i,i_1-1}, \neg q_{j,j_2}, q_{j,j_1-1}]$$

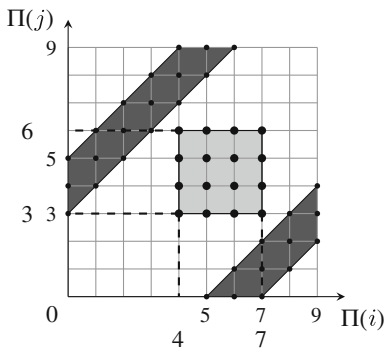


Fig. 1 Feasible (*dark grey*) and infeasible (*white*) regions for constraint $a(i, j) = [3, 5]_{10}$. The *light grey* square shows an infeasible rectangle.

With $\lfloor\cdot\rfloor$ being the round down function, $\lceil\cdot\rceil$ being the round up function and the integers $u, l \in \mathbb{Z}$ with $u < l$, we can define

$$\delta(l,u) := l - u - 1, \quad \delta y(l,u) := \left\lfloor \frac{\delta(l,u)}{2} \right\rfloor, \quad \delta x(l,u) := \left\lceil \frac{\delta(l,u)}{2} \right\rceil - 1.$$

These definitions allow us to determine a rectangle between u and l, such that it has maximum area having minimum perimeter. Basically each rectangle has a width of $\delta x(l,u)$ and a height of $\delta y(l,u)$.

In order to cover each infeasible pair in the area between $u - t$ and l, we need approximately t rectangles. The function $\zeta : \mathscr{P}(\mathbb{Z}) \to \mathscr{P}(\mathscr{P}(\mathbb{Z}) \times \mathscr{P}(\mathbb{Z}))$,

$$\zeta([l,u]_t) = \{H \times G \mid |H| = \delta x(l, u-t), |G| = \delta y(l, u-t), (H \times G) \cap S_c = \emptyset\},$$

with $H, G \in \mathscr{P}(\mathbb{Z})$ being intervals, maps to the set of all infeasible rectangles of constraint $c = [l,u]_t$. The sufficiency that all infeasible pairs of a constraint are excluded, is given by the following lemma.

Lemma 1 *Let $c = [l,u]_t$ be a constraint. Then the following holds*

$$\textit{(i)}\ \ P_c \subseteq \bigcup_{A \in \zeta(c)} A\,, \qquad\qquad \textit{(ii)}\ \ S_c \cap \bigcup_{A \in \zeta(c)} A = \emptyset.$$

Let the encoding of all edges $e \in E$ be $\Psi_N := \bigwedge_{e \in E} \bigwedge_{A \in \zeta(a(e))} enc\_rec(A)$.This results in the encoding of a PESP instance such that $enc\_pesp(N) := (\Omega_N \wedge \Psi_N)$.

After a model J has been found for this formula, extracting the value of the event n by J is done by the function $\xi_n(J)$, where $\xi_n(J) = k$ such that $J \not\models q_{n,k-1}$, $J \models q_{n,k}$ and $k \in [0, t-1]$. ξ_n is well-defined due to the following lemma.

Lemma 2 *Let $N = (V, E, a, t)$ be a PEN, $n \in V$ be an event and J an interpretation. Then $J \models enc(n) \Leftrightarrow \exists! k \in [0, t-1] : \forall i \in [-1, k-1] : J \not\models q_{n,i} and \forall j \in [k, t-1] : J \models q_{n,j}$.*

Extracting the schedule Π can be done on a per-element basis from a model J with $\forall n \in V : \Pi(n) = \xi_n(J)$.

Lemma 3 *Let $r \subseteq P_c$ be a rectangle in the infeasible region of constraint $c = a(i,j)$. Then $J \models enc\_rec(r) \Leftrightarrow (\xi_i(J), \xi_j(J)) \notin r$ with J being an interpretation.*

Theorem 1 (Soundness and Completeness)
Let $N = (V, E, a, t)$ be a PEN and $F := enc\_pesp(N)$ be the order encoded propositional formula of N. Then $\exists J : J \models F \Leftrightarrow \exists \Pi : \Pi \models N$ with J being an interpretation and Π being a schedule of N.

Proof

$$
\begin{aligned}
& \exists J : J \models F \overset{\text{Def. } enc\_pesp}{\Leftrightarrow} \exists J : J \models (\Omega_N \wedge \Psi_N) \Leftrightarrow \exists J : J \models \Omega_N, J \models \Psi_N \\
\overset{\text{Def. } \Omega_N}{\Leftrightarrow} & \exists J : J \models \textstyle\bigwedge_{n \in V} enc(n), J \models \Psi_N \overset{\text{Def. } \xi_\eta, \text{Lem } 2}{\Leftrightarrow} \forall n \in V : \Pi(n) := \xi_n(J), \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad J \models \Psi_N \\
\overset{\text{Def. } \Psi_N}{\Leftrightarrow} & \exists J : \forall n \in V : \Pi(n) := \xi_n(J), J \models \textstyle\bigwedge_{e \in E} \bigwedge_{A \in \zeta(a(e))} enc\_rec(A) \\
\Leftrightarrow & \exists J : \forall n \in V : \Pi(n) := \xi_n(J), \forall e \in E \; \forall A \in \zeta(a(e)) : J \models enc\_rec(A) \\
\overset{\text{Lem } 3,1}{\Leftrightarrow} & \exists J : \forall n \in V : \Pi(n) := \xi_n(J), \forall e \in E : e \text{ holds under } \Pi \\
\Leftrightarrow & \exists J : \forall n \in V : \Pi(n) := \xi_n(J), \Pi \models N \overset{\text{Def. } \xi_\eta, \text{Lem } 2}{\Leftrightarrow} \exists \Pi : \Pi \models N
\end{aligned}
$$

□

4 Experimental Results

The PESP instances, which are used for comparison, model public railway transport networks of up to 500 trains which are instances like the whole inter city network of Germany (*fernsym*), as well as subnetworks like south west ($swg_i, i \in \{1, \ldots, 4\}$, *mb13mg*) and south east (seg_1, seg_2, *we*) Germany, respectively. Each of which has a period of two hours. Table 1 shows the sizes of each PESP instance as well as the resulting formula size after reducing the problem to SAT.

The native domain PESP solver is described and analyzed by Opitz [8] and Nachtigall [7]. The authors claim that this solver is the best known algorithm to solve PESP instances. The solver is equipped with a decision tree method with respect to the events' potentials. Additionally, constraint propagation techniques are employed by propagating the current valid assignments of the events across the network. This solver will be referred to as PESPSOLVE.

Table 1 PESP instances, corresponding encodings and solving times

	PESP $N = (V, E, a, t)$		Order encoding F		Solving times	
Instance	V	E	$vars(F)$	F	PESPSOLVE	*Ordered* + RISS
k	18	730	2160	26536	47112	26
swg_2	60	1145	7140	83740	512	2
fernsym	128	3117	15232	353276	2035	7
swg_4	170	7107	20230	399191	912	8
swg_3	180	2998	21420	214011	66	2
swg_1	221	7443	26299	462217	Timeout	7
mb13mg	231	9805	26775	777894	Timeout	986
seg_2	611	9863	72709	1115210	Timeout	11
we	846	14690	78659	2049188	Timeout	Timeout
seg_1	1483	10351	176477	1348045	Timeout	10

The comparison has been performed on an Intel Xeon processor with 3 GHz and 16 GB of main memory with a runtime timeout of 24 h. The chosen SAT solver is RISS, developed by Manthey [5]. The runtime for the SAT approach is the sum of the reduction time and the solving time. Table 1 shows that the SAT approach outperforms the native domain solver PESPSOLVE by far, although in this work the parameters of the solver have not been adjusted to the specific application domain. Nonetheless, the speedup increases up to at least 12 342 (instance swg_1).

5 Conclusion

In this work it is shown that the periodic event scheduling problem can be reduced to satisfiability testing efficiently. Solving the SAT instance with a state-of-the-art SAT solver requires a very short time frame compared to a state-of-the-art PESP solver. The reduction to SAT allows to solve a whole set of larger, more complex instances, which could not be handled before. The speedup that has been measured is up to four orders of magnitude compared to the PESP solver.

Based on the presented method scaling the PESP instances becomes more interesting. From the used application railway networks the largest possible instance is to combine all subnetworks of Germany to a single network. Solving this instance in reasonable time is a huge open goal in transport engineering.

The final conclusion that can be drawn from the presented work is the following: the currently best PESP solver is now SAT-based.

References

1. Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.): Handbook of satisfiability. IOS Press (2009)
2. Cook, S.A.: The complexity of theorem-proving procedures. In: Harrison, M.A., Banerji, R.B., Ullman, J.D. (eds.) STOC, pp. 151–158. ACM (1971)
3. Großmann, P.: Polynomial reduction from PESP to SAT. Tech. Rep. 4, Technische Universität Dresden, Germany (Oct 2011)
4. Großmann, P., Hölldobler, S., Manthey, N., Nachtigall, K., Opitz, J., Steinke, P.: Solving periodic event scheduling problems with SAT. In: IEA/AIE. LNAI, vol. 7345, pp. 166–175. Springer (2012)
5. Hölldobler, S., Manthey, N., Saptawijaya, A.: Improving resource-unaware SAT solvers. In: Fermüller, C., Voronkov, A. (eds.) Logic for Programming, Artificial Intelligence, and Reasoning. LNCS, vol. 6397, pp. 357–371. Springer (2010)
6. Liebchen, C., Möhring, R.H.: The modeling power of the periodic event scheduling problem: Railway timetables-and beyond. In: Proceedings of the 4th international Dagstuhl, ATMOS conference. pp. 3–40. ATMOS'04, Springer (2007)
7. Nachtigall, K.: Periodic network optimization and fixed interval timetable. University Hildesheim, Habilitation thesis (1998)
8. Opitz, J.: Automatische Erzeugung und Optimierung von Taktfahrplänen in Schienenverkehrsnetzen. Logistik, Mobilität und Verkehr, Gabler Verlag (2009)
9. Serafini, P., Ukovich, W.: A mathematical model for periodic scheduling problems. SIAM J. Discrete Math. **2**(4), 550–581 (1989)
10. Tanjo, T., Tamura, N., Banbara, M.: A compact and efficient SAT-encoding of finite domain CSP. In: SAT. pp. 375–376 (2011)

The Importance of Automatic Timetabling for a Railway Infrastructure Company

Daniel Poehle and Werner Weigand

1 Introduction

The DB Netz AG is the most important railway infrastructure company in Germany and operates a railway network with a total length of about 34,000 km and more than 5,000 stations. This is the greatest and most complex network in Europe. 48,000 employees make about 30,000 train runs daily possible. The DB Netz AG has three mayor tasks to solve. At the one hand there is the signaling and control of the trains. This is managed locally in signal towers or centrally in operations centers. At the other hand the huge network has to be maintained and repaired permanently and the network has to be developed. The third task is to generate timetables to sell train-paths to the customers, both long-term and short-term.

Forecasts say that there will be a big growth in freight traffic by over 60 % during the next twenty years. To handle the predicted traffic volume in future it is necessary to develop and expand the network and to optimize the use of the existing infrastructure (see [3]). An optimized use of the infrastructure can be achieved if for example the train's velocity is harmonized or if the demand can be guided to less loaded routes and time.

The capacity management at DB Netz contains the forecast, capacity-analysis and capacity-design and is a continuous process for the development of the infrastructure. The main challenge results from the extensive need of time from planning till commissioning of the infrastructure on the one hand and the dynamically changing infrastructure requirements.

A central element of the capacity-management is a long-term railroad schedule. Based on the forecasts and concepts, for example the long-distance, local rail systems

D. Poehle (✉) · W. Weigand
DB Netz, Theodor-Heuss-Allee 7 , 60486 Frankfurt, Germany
e-mail: daniel.poehle@deutschebahn.com

W. Weigand
e-mail: werner.weigand@deutschebahn.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_73,

and freight-trains, a real timetable is generated and could be valid for example in six years. This special timetable will not be used in reality then but is an assumption for a possible schedule in future and is used as a base for the capacity-analysis. Although it is planned long-term, all important and relevant conditions are considered, e.g. cadences or connections. In summary the long-term schedule and its restrictions lead to the necessary infrastructure.

The long-term railroad schedule ensures more customer focused train-paths in passenger and freight transport and allows a growth in traffic because the need to develop infrastructure is early identified. So the network capacity and the efficiency will be increased as well as quality of the train-paths is considered and improved.

The creation of a timetable for the biggest European railway-network is naturally complex and time-consuming so there are efforts to aid the personnel by data processing or automate this process in essential parts. In the next section a program-system will be introduced which was designed to help in this topic.

2 Program-System TAKT

The software-system TAKT is a development of the professorship "Verkehrsstroemungslehre" at the TU Dresden [1]. It has been upgraded an enhanced during the last years in many research projects in cooperation with the DB Netz AG. TAKT automatically generates and improves timetables for complex railway networks on the base of periodic event scheduling problems (PESP[1]) by using specialized algorithms. After providing the input data a timetable is calculated automatically and can be evaluated in detail afterwards by the user.

TAKT consist of several modules, which are connected by interfaces as shown in Fig. 1 and will be described next.

The ROUTER prepares the input data for the generation of timetables. On the one hand an infrastructure database is necessary and on the other hand modeltrains for the timetable have to be defined. A modeltrain corresponds for example to a line of a long-distance train. The ROUTER consists of an innovative routing-algorithm and a calculator for the travel time and produces the modeltrains. The routing-algorithm finds a well-suited train-path in the network automatically.

All modeltrains with their characteristics and further conditions like connections for the passengers are called plan of operation. This is the base for the PESP-Solver, which schedules a feasible strictly periodic timetable. To get an enormous speed-up, the PESP is converted to a satisfiability-problem (SAT). After the solution with an up to date SAT-Solver, the result is retransformed and gives the train's feasible arrival and departure times. This transfomation provides practically useful solutions in appropriate runtimes which former solvers could not achieve.

The Optimizer reduces the waiting time for the passengers in the network. On the one hand there are passengers in the trains who don't want to wait too long at stations

[1] The PESP is described in Chap. 3.

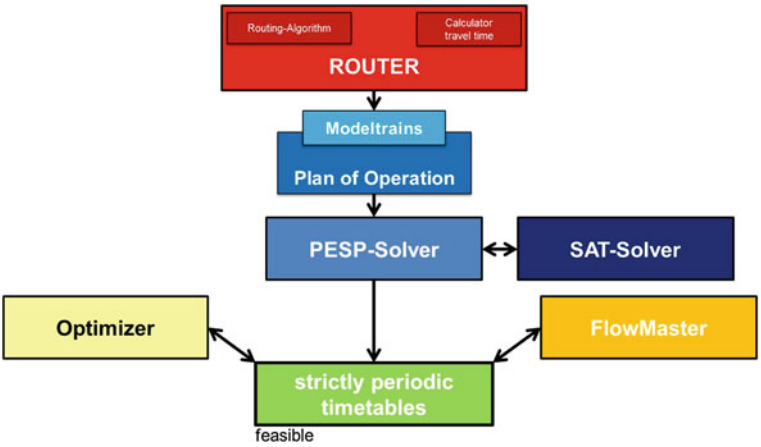


Fig. 1 Modules of the program-system TAKT

because of connections. On the other hand there are passengers at the stations who want to change trains and therefore a correspondence time is needed. All waiting minutes are weighted and only important connections will be arranged to reduce the total amount of waiting time in the railway network.

The FlowMaster generates as much as possible freight train-paths to an existing periodic timetable. This is done by connecting parts of a train-path until whole train paths between origin and destination are finished and no more capacity is left. Origins and destinations can be for instance great harbours or freight-stations.

In summary the program-system TAKT can strongly aid the employees at DB Netz to create a long-term timetable for the capacity management. The next section intoduces and explains the base model PESP.

3 Periodic Event Scheduling Problem

The automatic timetabling in the program-system TAKT is based on a periodic event network (see [2] for more details). Before introducing the PESP it is necessary to characterize the periodic event network and to differentiate it from a non-periodic event network. In the periodic event network important events are represented by a node, for example the train's arrival or departure at a station. The nodes in the network are connected with edges which represent a time span. There are for example edges for the travelling time or distance spacing. The periodic event network distances from the infrastructure and deals only with the modeltrain's relations in time. With the network's nodes and edges all processes and conditions in railway operation can be described. Figure 2 shows two events in a non-periodic network which are connected by an edge.

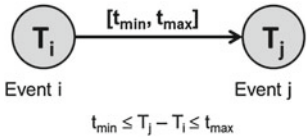


Fig. 2 Two events in a non-periodic network

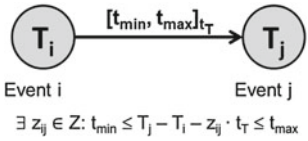


Fig. 3 Two events in a periodic network

A solution is feasible if the points in time T_i and T_j are between the time span $[t_{min}; t_{max}]$. Every point in time takes place exactly once. Thus the two events' order is fixed, T_j always follows T_i. For the periodic case we have to define a new variable to the event network, the modulo parameter z. It has to be an integer value. In Fig. 3 two periodic events with one restriction are displayed.

In the periodic case every event k is repeated indefinitely often at the points in time $T_k + z \cdot t_T$, so the condition must also be fulfilled in every period. So a feasible solution has to fulfill the advanced inequality in Fig. 3.

z_{ij} is the modulo operator of condition between event i and j. The events' order is no longer fixed and will be determined by the modulo operator. In the periodic event scheduling problem has to be decided if there is a feasible schedule **T** which assigns a point in time to every event so that all restrictions are met. Failing this the system is infeasible. Degrees of freedom in this model are the time spans on the conditions and the events' order.

On the base of the periodic network the PESP-Solver calculates a periodic timetable or finds infeasibility. However, this means not that there is no possible schedule for the plan of operation. The PESP-Solver diagnoses which events and restrictions cannot be fulfilled automatically. There is the opportunity to extend some conditions, for instance the stop time of particular modeltrains, to achieve a feasible timetable. Increasing a stop time in a station can be used to change the order of trains for example. The process of widening some conditions is repeated iteratively until a feasible solution is found or no expandable condition is in the set of restrictions which cannot be fulfilled. Then there would be no feasible schedule possible.

Due to the internal transformation from PESP to a satisfiability problem an enormous advantages in the solver's run time were possible. The program-system TAKT provides practically useful solutions in appropriate runtimes as can be seen in the next section.

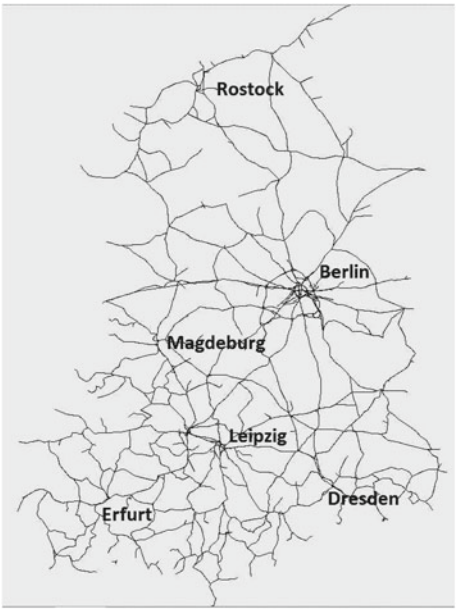


Fig. 4 The investigated area

4 Example

In this section some results of a typical long-term railroad schedule are presented. The investigated area covers two of the 7 regional districts of DB Netz and is shown in Fig. 4.

All together are in this area more than 4,000 stations and stops, 43,000 tracks and the total length of the infrastructure is about 18,000 km. Within this network 204 lines run (12 long distance, 176 regional and 16 local). The resulting periodic network contains 2,811 nodes and about 14,000 edges. To solve the PESP and get a feasible strictly periodic schedule the PESP-Solver needs 176 iterations to widen conditions and a total run time of only 3 hours. For comparison: If an average employee of DB Netz manually constructed a comparable timetable manually, this would last several weeks!

Figure 5 shows the train graph of a train running from Stralsund via Berlin and Halle/Saale to Erfurt.

5 Summary

The long-term railroad schedule is an effective instrument for capacity management at DB Netz and ensures more customer focused train-paths in passenger and freight transport. It identifies the need to develop infrastructure early and allows to adapt the infrastructure to a growing traffic.

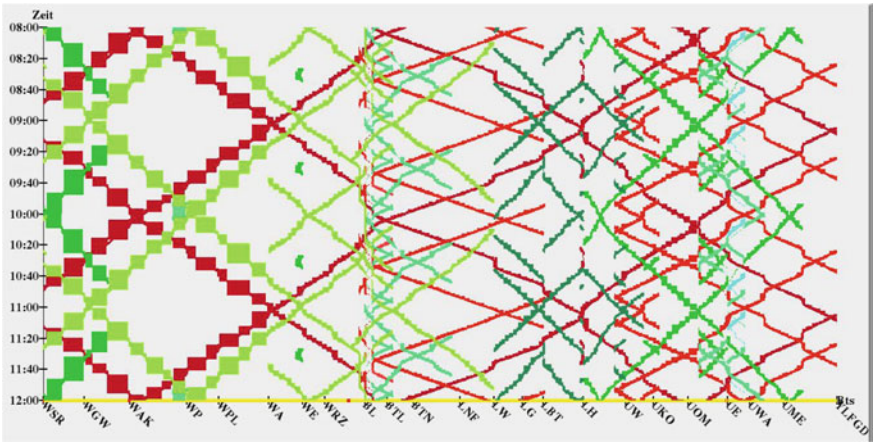


Fig. 5 Example timetable

The program-system TAKT is a great support for the creation of a long-term schedule because it saves a lot of time due to helpful automations with innovative components. It provides practically useful solutions in appropriate runtimes and improves the long-term schedule's quality substancially. Due to an easy variation of parameters and the automatic solvers it is initially possible for DB Netz to create and optimize more than one variant of a timetable for great areas.

References

1. Nachtigall, K.: Periodic Network Optimization And Fixed Interval Timetables. Habilitation at University Hildesheim (1998)
2. Opitz, J.: Automatische Erzeugung und Optimierung von Taktfahrplaenen in Schienenverkehrsnetzen. Gabler, Dresden (2009)
3. Weigand, W.: Langfristfahrplan und Kapazitaetsuntersuchungen. Deine Bahn, 28–31 May 2012 (2012)

Study of a Process of Task Accumulation and Distribution in a GRID-System

Zoia Runovska

1 Preliminaries

Along with operating systems and middleware to control GRID—Systems, additional means are used, whose main function is to provide a mechanism of tasks distribution scheduling. The task scheduling is based on an algorithm, which the effectiveness of task control largely depends on.

Many modern algorithms either assume the accumulation of task package before resources are allocated, or realize a rectangular packaging of tasks into a strip of a given width with the aim of queuing time minimization, what also requires that a certain amount of tasks is accumulated prior to their packaging.

In such situations, a typical problem arises, namely, what number of tasks has to be accumulated before the resources are allocated or before a rectangular task packaging is performed, in order to minimize the costs associated with the organization of task scheduling in a GRID-System.

The effectiveness of the algorithm depends essentially on task flow characteristics and on a size of indicated quantities as well as on estimated costs of the algorithm calculations. The algorithm does not have an essential practical value, if the calculations are too expensive.

2 The Problem and Its Formalisation

In this paper, a model of the process of task accumulation and distribution in a GRID-System is introduced. The problem is reduced to the construction and the analysis of a semi-Markov process in a transition and a stationary phases. An algorithm to choose an optimal package size is introduced.

Z. Runovska (✉)
HSHL, 59063 Hamm, Germany
e-mail: runovska@yahoo.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_74,

Let us describe the system considered here. The time intervals between the task entrances are mutually independent random values. The function of distribution of their durations is denoted by $F(x)$. The number of simultaneously entering tasks is some integer valued random value with a distribution $\{c_i\}$, $(i = 1, 2, \dots)$.

Task package is accumulated as long as its size does not exceed a certain limit value $N > 0$. As soon as the package size exceeds the value N, the resources are allocated for the task performance in accordance with the algorithm.

As a state of a considered system at the time moment t we take the number of tasks $v(t)$ waiting for a resource to be allocated at this moment of time. When studying the system, it is often needed to identify such characteristics as the distribution $p_i(t)$, $(i = 0, 1, \dots, N)$, the amount of tasks $v(t)$ at the time moment t, i.e. $p_i(t) = P\{v(t) = i\}$, the stationary distribution $\{p_i\}$ of task amount $v(t)$, etc.

The distribution $\{p_i\}$ can be used in one of the possible approaches of determining the limit value N that minimizes costs consisting of expenses for task package accumulation and expenses for calculations provided by scheduling algorithm.

In order to calculate the distribution $\{p_i(t)\}$, $(i = 0, 1, \dots, N)$ of task amount $v(t)$, we analyze the random process $v(t)$. The process $v(t)$ is a semi-Markov process [2], the duration of its stay in each stage i, $(i = 0, 1, \dots, N)$ is distributed according the distribution law $F(x)$, and the probabilities of transitions from stage to stage are:

$$p_{i0} = \sum_{k=N+1-i}^{\infty} c_k, \quad i = \overline{1, N};$$
$$p_{ij} = \begin{cases} c_{j-i}, & i = \overline{0, N-1};\ j = \overline{i+1, N}. \\ 0, & i = \overline{1, N};\ j = \overline{1, i}. \end{cases} \tag{1}$$

Let us introduce the following notations: $\gamma(t)$—the time interval from the moment of time t to the next time moment of task entrance; $dA(i, t)$—the probability of the fact that the task entrance will occur within the time interval $(t, t + dt)$, then the number of tasks $v(t)$ will become equal to i, i.e.

$$dA(i, t) = P\{\gamma(t) < dt;\ v(t + dt) = i\}.$$

We assume that at the moment $t = 0$, a task has entered, so that as a result the number of tasks became equal to v_0. The next task entrance is expected in a time interval, whose duration is distributed by a law $F(x)$. Having this assumption in mind, we can show that the function $dA(i, t)$ satisfies the equations:

$$dA(i, t) = p_{v_0 i} dF(t) + \sum_{j=0}^{N} p_{j,i} \int_0^t dA(j, \tau) dt\, F(t - \tau), \quad i = \overline{0, N}, \tag{2}$$

where p_{ij} are calculated by (1).

Functions $dA(i,t)$ allow us to calculate the distributions of our interest $\{p_i(t)\}$, $(i = 0, 1, \ldots, N)$, of the stages of the process $v(t)$ as follows:

$$p_i(t) = [1 - F(t)]\delta_{iv_0} + \int_0^t [1 - F(t-\tau)]dA(i,\tau), \quad i = \overline{0,N}, \tag{3}$$

where the Kronecker symbol is

$$\delta_{ij} = \begin{cases} 1, & i = j. \\ 0, & i \neq j. \end{cases} \tag{4}$$

We apply Laplace-Stieltjes transforms [1] to solve the systems (2) and (3):

$$a(i,s) = \int_0^\infty e^{-st} dA(i,t), \quad Re\, s > 0.$$

Transferring to the integral transformations in (2), we obtain

$$a(i,s) = f(s)[p_{v_0 i} + \sum_{j=0}^{N} p_{ij} a(j,s)], \quad i = \overline{0,N}. \tag{5}$$

The system (3) in the Laplace-Stieltjes transforms:

$$\varphi(i,s) = \frac{1}{s}[1 - f(s)][\delta_{iv_0} + a(i,s)], \quad i = \overline{0,N}, \tag{6}$$

where δ_{iv_0} is defined by (4).

The obtained integral characteristics $\{\varphi(i,s)\}, i = \overline{0,N}$ of the process $v(t)$ can be used to find an explicit representation of the distribution $\{p_i(t)\}, i = \overline{0,N}$. In case it is difficult, we can calculate the moments of a random process $v(t)$ and stationary distribution of stages of the process $v(t)$ with help of functions $\{\varphi(i,s)\}$.

3 The Approach to Minimize Task Scheduling Costs

Let us consider one of the approaches to choose the optimal number of tasks N, which ensure the minimization of costs associated with the organization of task scheduling in a GRID-system.

For this purpose, we introduce the following notation:

- r The costs per time unit, which associated with the accumulation of a task package;

$q(i)$ The costs of computing, provided by scheduling algorithm when the task amount i in the package has exceeded the limit value N, $(i = N + 1, N + 2, \dots)$;

n The average amount of tasks entering per time unit;

a The mathematical expectation of a time interval between two successive moments of task entrances;

m_i Duration of task waiting for the moment of resource allocation, if by entrance of this task, the amount of tasks in the package became i.

Using this notation, we can write the costs per time unit, which are made up of package accumulation costs and the costs of calculations provided by the scheduling algorithm, in a form of a function of the limit value N:

$$R(N) = \frac{nr \sum\limits_{i=0}^{N} p_i m_i + \sum\limits_{i=0}^{N} p_i \sum\limits_{j=N+1-i}^{\infty} c_j q(i+j)}{a + \sum\limits_{i=1}^{N} c_i m_i} \tag{7}$$

The denominator $R(N)$ represents the average value of a time interval between two successive moments of package processing to allocate the resources in accordance with the algorithm. The first term of the numerator represents the costs of task package forming; the second term represents the costs of a single processing of a task package.

The limit value of a package size, which minimizes costs, is defined as a point of minimum of the expression $R(N)$.

Let us show how to calculate the characteristics of the process $v(t)$ given in (7):

1. stationary distribution of the stages of the process $\{p_i\}$, $\quad i = \overline{0, N}$

$$p_i = \lim_{s \to 0} s\varphi(i, s), \ (i = 0, 1, \dots, N),$$

where $\varphi(i, s)$ are defined in (6);

2. time of the stay of the process $v(t)$ in a certain stage i, $(i = 0, 1, \dots, N)$:

$$a = \int_0^{\infty} x dF(x);$$

3. We find the time of the stay of the process $v(t)$ in a set of stages $E = \{1, 2, \dots, N\}$ with an initial stage i as a solution of a system of algebraic equations:

$$m_i = a + \sum_{k=1}^{N-i} c_k m_{i+k}, \quad i = 1, 2, \dots, N. \tag{8}$$

The system (8) has a triangular matrix of coefficients, hence there is no difficulty to calculate successively on $i = N, N-1, \ldots, 1$ the parameters m_i.

References

1. Romanovsky, P.I.: Fourier series. The field theory. Analytical and special functions. In: Nauka, M. (ed.) Laplace Transforms, p. 336 (1980)
2. Silvestrov D.S.: Semi-Markov processes with a discrete set of stages (basis of calculation of the functional and reliability characteristics of stochastic systems). p. 272. Sov. Radio, Moscow (1980)

Balancing Load Distribution on Baggage Belts at Airports

Frank Delonge

1 Introduction

Baggage belts at commercial airports are cumulative resources. Depending on aircraft sizes, passenger luggage of up to eight flights can be processed simultaneously by one single belt. While utilizing baggage belts up to their capacity limits is unavoidable during peak hours, it is usually undesired at other times. Nevertheless, we often face situations of imbalanced load distribution throughout belts in practice: Passengers are unnecessarily crowded tightly around some belts whereas other belts remain unoccupied at the same time, which is an unpleasant situation for passengers as well as for airport staff. We will refer to this first kind of imbalance as *local imbalance*. Secondly, imbalanced load distribution over a longer period of time leads to unsynchronized maintenance intervals. As a consequence, this kind of imbalance usually increases maintenance costs and should therefore also be avoided. This second kind of imbalance will be referred to as *global imbalances*.

Both types of imbalances already arise at the planning stage. As an example, Fig. 1a shows the planned schedule of one day of operation at a small German airport, in which the task-to-resource assignment is mainly driven by a user-defined preference rule system, e.g.

if	carrier = LH	and	origin = international	⇒	{ Belt1:100 %, Belt2: 60 %}
if	carrier = BA	and	type = charter	⇒	{ Belt2:100 %, Belt3: 80 % }

(see [2] for further details). If maximising assignment preferences is the dominating optimization criterion, imbalances in the definition of the rule system yield undesired imbalances in the resource load distribution. Thus, the resulting plan contains several situations in which resource load concentrates tightly on belt 2 but belts 1 and 3 remain (almost) empty at the same time. Furthermore, the overall load in terms of

F. Delonge (✉)
Airport Systems Division, INFORM Institute for Operations Research GmbH, Aachen, Germany
e-mail: frank.delonge@inform-software.com

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_75,

task minutes per resource on belt 2 is three times higher than the load on belt 3 for the displayed day.

We address these problem by extending our existing multi-objective constraint Programming (CP) framework by additional objectives to reduce both kinds of imbalances. Upon clients' request, both objectives can be invoked separately or in combination at user-defined weight in relation to other existing optimization goals.

2 Model

2.1 Measures for Imbalances

The reduction of local and global imbalances can both be mathematically formulated as variance minimization. In order to formalize this, we use the following notations:

$\mathscr{T}$	the time horizon,
$\mathscr{R}$	the set of all resources,
$l_r(t)$	the load density function over time $t \in \mathscr{T}$ for resource $r \in \mathscr{R}$
$\bar{l}(t)$	the average load density over time $t \in \mathscr{T}$, i.e. $\bar{l}(t) = \frac{1}{\lvert\mathscr{R}\rvert} \sum_{r \in \mathscr{R}} l_r(t)$,
L_r	the total load on resource $r \in \mathscr{R}$ in terms of task minutes i.e. $L_r = \sum_{t \in \mathscr{T}} l_r(t)$,
$\bar{L}$	the average total load per resource, i.e. $\bar{L} = \frac{1}{\lvert\mathscr{R}\rvert} \sum_{r \in \mathscr{R}} L_r$,

For *local load balancing*, we minimize the variance of the load density functions l_r throughout all resources $r \in \mathscr{R}$, i.e.

$$\min \ \big(Var(l)\big) \text{ with } Var(l) = \sum_{r \in \mathscr{R}} \|l_r - \bar{l}\|^2$$

We use the $\mathscr{L}^2$-norm as function measure, i.e. $\|f\|^2 = \int_{\mathscr{T}} f^2(t)\,dt$. For discrete time space $\mathscr{T}$ this yields

$$Var(l) = \sum_{r \in \mathscr{R}} \sum_{t \in \mathscr{T}} \big(l_r(t) - \bar{l}(t)\big)^2 \quad . \tag{1}$$

For *global load balancing*, we minimize the variance of the total load functions L_r:

$$\min \ (Var(L)) \text{ with } Var(L) = \sum_{r \in \mathscr{R}} \big(L_r - \bar{L}\big)^2 \quad . \tag{2}$$

2.2 Measure Transformation

Equations (1) and (2) can already be used as such to measure the balance of load distributions for existing plans. However, they do not allow for effective lower and upper bound computations needed by a Constraint Programming approach. Therefore, we apply the following transformations:

For *local load balancing*, starting with Eq. (1):

$$\begin{aligned} Var(l) &= \sum_{r\in\mathscr{R}} \sum_{t\in\mathscr{T}} \left(l_r(t) - \bar{l}(t) \right)^2 \\ &= \sum_{t\in\mathscr{T}} \left(\sum_{r\in\mathscr{R}} l_r^2(t) \;-\; 2\,\bar{l}(t) \sum_{r\in\mathscr{R}} l_r(t) \;+\; \sum_{r\in\mathscr{R}} \bar{l}^2(t) \right) \quad . \end{aligned}$$

As all our tasks are fixed in time, the sum of all tasks $\sum_{r\in\mathscr{R}} l_r(t) =: C(t)$ is a constant for all times t. Thus, also $\bar{l}(t) = \frac{1}{|\mathscr{R}|}\, C(t)$ is a constant function over time, which leads to the simplification

$$Var(l) = \sum_{t\in\mathscr{T}} \sum_{r\in\mathscr{R}} l_r^2(t) \;+\; \hat{C} \quad .$$

Therefore, minimizing the variance of l is equivalent to

$$\min \left(\sum_{r\in\mathscr{R}} \sum_{t\in\mathscr{T}} l_r^2(t) \right)$$

as arguments of the minimization functions only differ by a constant. Our objective z_{ltd} to balance the local task distribution can thus be defined as

$$z_{ltd} := \sum_{r\in\mathscr{R}} \hat{l}_r \text{ with } \hat{l}_r := \sum_{t\in\mathscr{T}} l_r^2(t) \; . \tag{3}$$

For *global load balancing*, starting with Eq. (2):

$$\begin{aligned} Var(L) &= \sum_{r\in\mathscr{R}} \left(L_r - \bar{L} \right)^2 \\ &= \sum_{r\in\mathscr{R}} L_r^2 \;-\; 2\,\bar{L} \sum_{r\in\mathscr{R}} L_r \;+\; |\mathscr{R}|\, \bar{L}^2 \; . \end{aligned}$$

Similar to local load balancing, the sum of all load functions $\sum_{r\in\mathscr{R}} L_r$ as well as the average total load per resource $\bar{L}$ are constants. For our objective z_{gtd} to balance

the global load distribution, this means:

$$z_{gtd} := \sum_{r \in \mathscr{R}} L_r^2 \,. \tag{4}$$

2.3 Implementation

The advantage of the transformations above is that $\hat{l}_r$ and L_r^2 are strictly monotonous increasing with growing number of task assignments to resource r and strictly monotonous decreasing with decreasing possibilities to assign tasks to r. Obviously, their sums $\sum_{r \in \mathscr{R}} \hat{l}_r$ and $\sum_{r \in \mathscr{R}} L_r^2$ have the same monotony properties, which are essential when integrating the load balancing objectives into our existing Constraint Programming (CP) framework. Within CP, numerical variables such as objective values are represented as so-called *range variables*, which store lower and upper bounds of the actual target value for the current search tree node. While descending a branch-and-bound tree by assigning tasks to resources, the domain of these variables gets reduced by tightening bounds in response to these assignments, until variable domains are reduced to single values at the leave nodes (refer for instance to [1] for a detailed introduction to basic CP concepts). Intelligent assignment decisions within our CP framework are performed by a one-step concept as explained in [1], in which possible developments of the objective's lower bound are compared. Thus, providing tight bounds increases the quality of decisions.

In this sense, Eqs. (3) and (4) allow for straight forward and effective computation of lower and upper bounds for z_{ltd} and z_{gtd}: For z_{ltd}, we allocate two histogram variables for each resource r. The first one stores the histogram of all tasks that are already assigned to r, whereas the second one stores the histogram of all tasks that can possibly still be assigned to r. These histograms represent lower and upper bound for the load density function $l_r(t)$ and can thus be used to compute bounds for $\hat{l}_r := \sum_{t \in \mathscr{T}} l_r^2(t)$ (3). Whenever the CP framework assigns a task activity a to some resource r_a, the lower histogram of r_a must be updated by incrementing values between the start and completion times of a. During this update procedure, $\hat{l}_{r_a}^{lower}$ can be partially recomputed by replacing the contribution of l_{r_a} for the changed portion of the histogram. Similarly, the upper histograms along with corresponding upper bounds for $\hat{l}_r$ must be updated for all other resources $r \neq r_a$.

In order to compute z_{gtd}, less effort is required. Per resource r, we only need to keep track of the total amount of task minutes that are already assigned to r and those that can still be assigned to r. In the same manner as for z_{ltd}, this range variable needs to receive updates with every task assignment. Summing up the squares of all these range variables yield the range for z_{gtd} (4).

Finally, we let z_{ltd} and z_{gtd} contribute to the total objective z, which is a weighted sum of all single objective criterions z_i, i.e.

Table 1 Statistics for the results of several variants of our balancing approach

variant	#tasks on 100 %-belt	Var(l)	Var(L)	Runtime
Original	534	262	880	0.8 s
Local	370	59	89	0.69 s
Balancing	−31 %	−77 %	−90 %	−14 %
Global	210	110	0	1.2 s
Balancing	−61 %	−58 %	−100 %	+50 %
Local + global	200	61	0.1	0.94 s
Balancing	−73 %	−76 %	−99.99 %	+17 %

$$z := \sum_i \lambda_i z_i \ . \tag{5}$$

The weights λ_i represent a user-defined relative importance in between criterions.

3 Results

As discussed in the introduction, Fig. 1a displays the original, hugely imbalanced schedule for one day of operation at a small German airport without considering load balancing.

Upon activation of the local load balancing criterion as the dominating objective (by choosing goal weights λ_i from Eq. (5) accordingly), all tight packs of tasks on the center belt are resolved by re-distributing these tasks evenly throughout resources (Fig. 1b). Table 1, which displays the results for a whole week, confirms this visual impression: The variance of l as defined in Eq. (1) decreased by 77 % at the price of 31 % decreased assignment preference compared to the original schedule [1] Moreover, activating the local load distribution objective implicitly reduces the global imbalance significantly.

Figure 1c shows the resulting schedule for global load balancing. The variance of L from Eq. (2) dropped to 0, which means that the load per resource in terms of task minutes is exactly the same for all belts. Although we still observe several local imbalance in the schedule, the local load distribution was also enhanced significantly, but the costs in terms of lowered assignment preference to achieve this perfectly balanced global load distribution were fairly higher compared to local load balancing.

Local and global balancing can be combined almost without loss of optimality for both $Var(l)$ and $Var(L)$, and the resulting schedule is displayed in Fig. 1d. Moreover, we were also able to increase the planning stability towards flight delays: Avoiding to unnecessarily exploit belt capacities to their limits obviously minimizes the risk of exceeding the limits.

[1] For simplicity, we use the number of tasks assigned to their most preferred resource as an indicator for assignment preference, as the measure used in our existing model is rather complicated.

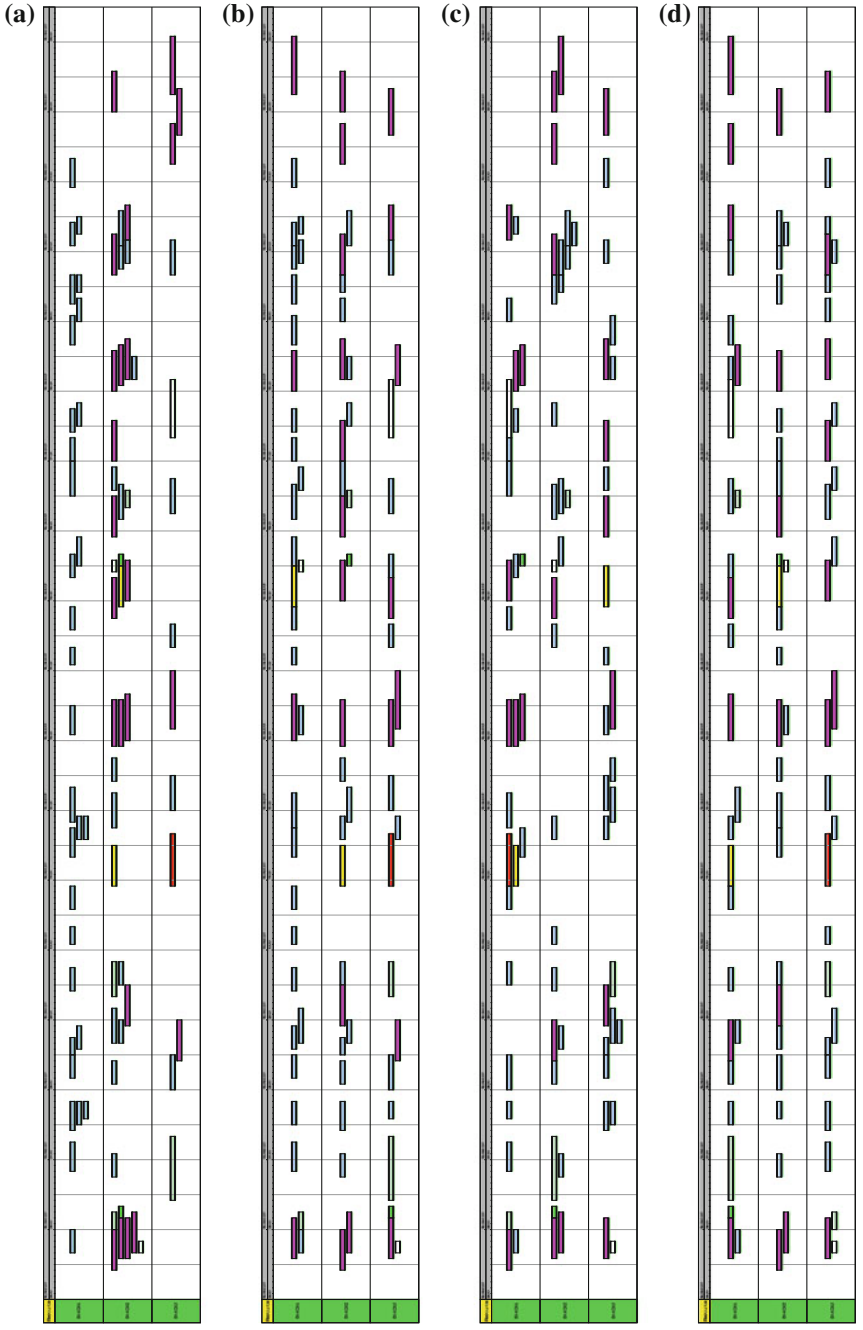


Fig. 1 Gantt charts of planned schedules for one day of operation at a small German airport: **a** original plan without consideration of load balancing, **b** with local load balancing, **c** with global load balancing, **d** with local and global load balancing

For all variants of invoking the load balancing objectives, the computational runtime remained at a magnitude of 1s per week or operation, which is fairly acceptable for commercial use.

References

1. Dechter, R: Constraint Processing, Morgan Kaufmann Publisher (2003) ISBN 1-55860-890-7
2. Dorndorf, U.: Staff and resource scheduling at airports. In: Operations Research, Proceedings 2006, Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), pp. 3–7. Springer, Heidelberg (2006)

Part XVIII
Supply Chain Management, Logistics and Inventory

On the Application of a Multi-Objective Genetic Algorithm to the LORA-Spares Problem

Derek Cranshaw, Raman Pall and Slawomir Wesolkowski

1 Introduction

An unfortunate fact of manufacturing is that products are often prone to failure. Whereas inexpensive products can be discarded and replaced upon failure, more expensive products are instead maintained by a repair by replacement policy: a failed component is removed from the product and replaced by a functioning spare part, if available. Otherwise, the replacement has to wait until a functioning component arrives.

Level of repair analysis (LORA) is an approach used during the design stage of complex equipment for the analysis of the cost effectiveness of competing maintenance strategies [1]. LORA is often defined as the problem of determining whether a component should be repaired or discarded upon its failure, and the location in the repair network to do such work. LORA is carried out as part of life cycle cost and cost of ownership analyses, and can play a significant role in minimizing these costs for capital equipment. This is important in defence, where maintenance requires complex support equipment and highly skilled personnel [1], and the unavailability of the equipment is especially undesirable.

The complexity of LORA is due to modular design and maintenance of equipment known as Prime Equipment (PE) which is carried out by exchanging components at locations placed at different hierarchy levels (i.e., there is a hierarchical

D. Cranshaw
Department of Applied Math, University of Waterloo, Waterloo ON N2L 3G1, Waterloo, Canada
e-mail: djcransh@uwaterloo.ca

R. Pall · S. Wesolkowski (✉)
Department of National Defence, Centre for Operational Research and Analysis, Defence R&D Canada, Ottawa ON K1A 0K2, Ottawa, Canada
e-mail: raman.pall@forces.gc.ca

S. Wesolkowski
e-mail: s.wesolkowski@ieee.org

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_76,

relationship between the operating locations and supporting depots). PEs are called *multi-indenture products* where each level that is removed from the main system is called an *indenture level*. That is, PEs are composed of components or *Line Replaceable Units* (LRUs) that can be broken down further into smaller subsystems called *Shop Replaceable Units* (SRUs) with the source of the failure being usually traceable to a single part.

All movement of broken components or restocking of working components is done through a defined system of supporting depots, called the repair network, since locating spares together with repair and test equipment close to each of the operating sites is usually expensive. We assume that the structure of the repair network forms a single tree with leaves called *Forward Operating Bases* (FOBs) which are the only locations where PEs can fail. Each FOB has a parent node called *Intermediate Depots* (IDs) from which it resupplies and to which it sends broken components, which in turn get resupplied from a single *Central Depot* (CD), the root of the repair tree. Once the component is received at the CD, it must be either repaired or replaced. Each of these classes of nodes (FOBs, IDs, and CD) forms a single level or *echelon* in the repair network hierarchy. In principle, any number of echelon levels is possible; however, in practice the limit is usually three. Each LRU is assigned an *Echelon Level of Repair* (ELOR), the echelon at which the part can be repaired. All SRUs must have a ELOR greater or equal to that of their parent LRU.

A related problem is the determination of the optimal number of spares for a given piece of equipment [2]. This is a major concern in defence and aircraft industries due to the enormous capital spent on spares every year. Recently, Ilgin and Tunali [5] developed an approach for joint optimization of spare part provisioning and maintenance policies for an automotive factory by integrating simulation with a genetic algorithm. The simulation model of the manufacturing line was used as an input to the genetic algorithm as part of the fitness function. However, this paper was limited to optimization based solely on cost.

The most common approaches in the literature on developing a possible spare provisioning decision model are simulation and mathematical programming. Sherbrooke [7] first applied mathematical programming to the spare parts inventory management problem in a multi-echelon setting. Furthermore, heuristics based on marginal analysis are employed to solve the sparing problem which requires very tight restrictions on the resource-component relations in the associated LORA models [4, 6].

For this problem of significant interest to the Canadian Department of National Defence, we adopt a multi-objective approach to extend the problem from the traditional approach of finding the best sparing policy to achieve a predetermined operational availability to that of identifying the relationship between the operational availability and resource cost. In this regard, an evolutionary algorithm is proposed that can identify non-dominated solutions to be used for a trade-off analysis. In Sect. 2, the multiobjective optimization method is explained. Section 3 describes the results. The paper concludes with Sect. 4.

2 A Multi-objective Approach

The approach is to use a Monte Carlo simulation with scenarios generated from a dataset detailing the expected failures of the equipment and their associated probabilities. Specifically the *Nondominated Sorting Genetic Algorithm II* (NSGA-II) [8, 3] was chosen due to its retention of the fittest solutions from one iteration to the next, and preservation of diversity amongst the solutions.

2.1 Objectives

The objectives that we consider in our model are the following:

- To minimize the total monetary cost of the maintenance strategy; and
- To minimize downtime of the equipment.

The costs included in the model that were captured by the first objective are defined as follows:

- *Shipping Cost*: The shipping cost of a failed part from one location to another (lower or higher in the hierarchy).
- *Holding Cost*: The daily cost of holding parts overnight at FOBs and IDs, which is directly proportional to the number of components at each base.
- *Repair Cost*: The cost of repairing a particular part at a particular echelon.

The algorithm sums all of these costs for each scenario and averages over all scenarios. Naturally, those solutions with the lowest average overall cost are the most fit with respect to this objective.

The second objective of the GA is to maximize the total availability of the system or, equivalently, to minimize the downtime of all PEs is the system. Each PE type has a "downtime cost" associated with it, which determines how much the system will suffer should the PE be out of commission. The total downtime cost is given by the sum over all PEs of the number of days each PE is out of commission, multiplied by the downtime cost of that PE. The algorithm averages the downtime costs over all scenarios, and the solutions with lowest average overall cost are the most fit with respect to this objective.

2.2 Chromosome Representation

In this paper, no PE has more than four indenture levels and the repair network has exactly three echelon levels. The solution chromosomes consist of two segments: the first detailing information about the components, and the seconds about the locations. For the first segment, each component has a target stock level and reorder threshold

for each echelon level. In addition, each component has an ELOR. Thus, this segment of the chromosome has $7n$ genes, where n is the number of components in the system. The second segment details the IDs from which the FOBs obtain their resupplies. The length of this segment is m, where m is the total number of FOBs. The total length of the chromosome is $7n + m$.

2.3 Scenario Generation

The fitness functions evaluate the efficiency of a repair management strategy based on a number of randomly generated failure scenarios. The algorithm used to generate a scenario operates as follows: for each component, a random number of failures is sampled from a Poisson distribution with mean equal to the *Average Yearly Failure* (AYF) of that component.

Each individual failure is randomly assigned a specific date in the year for the event (an integer from 1 to 365); as well as a location (i.e., FOB) where the failure occurred. The probability of failing on a certain date is uniform across all dates, while the probability of failing at the specific FOBs is taken as an input to the model. The resulting list of failures, sorted by ascending date, is a single scenario. A Monte Carlo approach is used to determine which maintenance strategy is truly optimal (i.e., through the use of many randomly generated scenarios).

2.4 Crossover and Mutation

The binary tournament selection procedure is used to choose parents for the crossover operator [3]. The crossover operator is divided into three stages. The standard crossover (random exchange of elements in the two parent chromosomes) is used for mating parts of the chromosome related to the stock levels and reorder points as well as the location of the FOBs. For the ELOR, given its tree structure the crossover operation must preserve the property that all components have a higher ELOR than their parent components. To do this a pivotal node is randomly selected. The tree with the higher ELOR at the pivotal node is labeled parent A, while the other is labeled parent B. The child tree inherits the entire branch starting from the pivotal node, including the pivotal node, from parent A, and all other nodes from parent B.

Like the crossover function, mutation has to be done in sections, since different sections of the chromosome mutate in different ways. The standard mutation operator [3] is used to randomly mutate the stock level. If the stock level for a part successfully mutates, the reorder point mutates to ensure that the reorder point remains below the stock level for all components at all echelons. The IDs associated to each of the FOBs are randomly mutated. Given that each component must always have a higher ELOR than its parent component, components may only mutate their ELORs to values between those of their parents and their children. The order in which parts are

considered for mutation is randomized to ensure that the structure of the tree after mutation is not biased. All PEs maintain an ELOR of 1.

3 An Application of the Technique

In order to illustrate the use of the methodology, we provide the results generated for a prototypical example consisting of a small repair network comprised of one central depot, two intermediate depots, and four FOBs. The algorithm was implemented in Matlab and run for 500 generations with a population size of 100, under 10 failure scenarios. A small example was used so that the true optimal values could be determined via exhaustive search, in order to test the validity of the algorithm. The population was randomly generated. There were twelve types of parts in the example of which two are PEs, four are LRUs, and six are SRUs. The hierarchy of the parts is depicted in Fig. 1 where their level of indenture is shown (several parts have sub-components common to other parts).

At each echelon and for each component, the target stock level varies from 1 to some high upper bound while the reorder threshold from 0 to the target stock level. The costs of shipping parts from one location to another ranged from $0 to $10,000 depending on the part number and the locations involved. Further, the shipping costs and times between location pairs were chosen such that the optimal maintenance strategy was obvious with little analysis. Each SRU was assumed to fail with equal probability at each FOB. Only the SRUs were given non-zero probabilities of failure (specified by their AYF)—meaning that the LRUs and PEs did not fail directly in the example. All failures could be traced to the failure of an SRU. Finally, the failure of each part was assumed to contribute equally to the downtime of the PE regardless of the cause of failure of the PE.

It was found that the algorithm converged to a non-dominated front of four solutions, as can be seen in Fig. 2. Moreover, it was found that this set of solutions was precisely the Pareto front for the example in question, illustrating that the algorithm converged to the required front as desired.

These non-dominated solutions can be used for a trade-off analysis, in which a decision-maker can specify their desired solution based on budgetary, political, or other motivations. This result indicates that a multi-objective genetic algorithm may be a viable method of addressing the LORA-Spares problem.

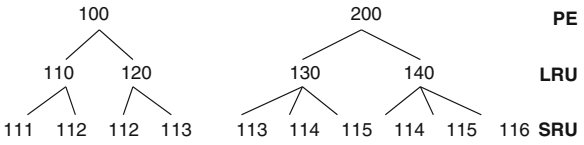


Fig. 1 Hierarchy of the parts in the example

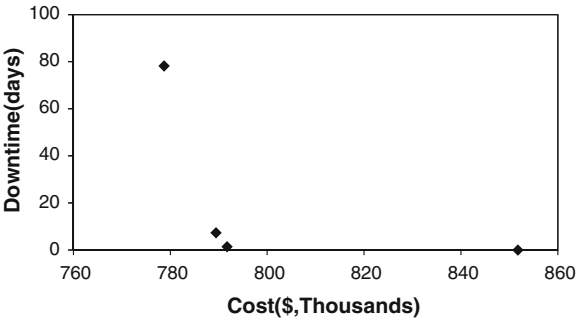


Fig. 2 The non-dominated front found in the example, consisting of four solutions

4 Conclusions

The need to address LORA and the sparing problem simultaneously is a problem of significant interest to the Canadian Department of National Defence. In this paper, we proposed an approach for the solution of this problem through an application of a multi-objective genetic algorithm, where both cost and availability are objectives to be optimized. An example was provided in which non-dominated solutions are identified that can be used for a trade-off analysis. Future work in this field could include testing the algorithm needs on larger, more complex problems to determine if it truly produces a good approximation of the Pareto-optimal front.

Acknowledgments This work was conducted as part of the Defence Research and Development Canada Technology Investment Funds project on LORA and sparing analysis.

References

1. Barros, L.: The optimization of repair decisions using life-cycle cost parameters. IMA J. Math. App. in Bus. Ind. **9**, 403–413 (1998)
2. Basten, R.: Designing logistics support systems: Level of repair analysis and spare parts inventories. University of Twente, Enschede, The Netherlands (2009)
3. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comp. **6**, 182–197 (2002)
4. Graves, S.C.: A multi-echelon inventory model for a repairable item with one-for-one replenishment. Manage. Sci. **31**, 1247–1256 (1985)
5. Ilgin, M.A., Tunali, S.: Joint optimization of spare parts inventory and maintenance policies using genetic algorithms. Int. J. Adv. Manuf. Technol. **34**, 594–604 (2007)
6. Muckstadt, J.A.: A model for a multi-item, multi-echelon, multi-indenture inventory system. Manage. Sci. **20**, 472–481 (1973)
7. Sherbrooke, C.C.: METRIC: a multi-echelon technique for recoverable item control. Oper. Res. **16**, 122–141 (1968)
8. Srinivas, N., Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms. J. Evol. Comp. **2**, 221–248 (1994)

Evaluation of a New Supply Strategy for a Fashion Discounter

Miriam Kießling, Tobias Kreisel, Sascha Kurz and Jörg Rambau

1 Introduction

The lot-type design problem LDP seeks for an optimal set of lot-types and a supply in terms of lots of such lot-types such that the resulting supply of sizes matches the branch- and size-dependent demand as closely as possible (see [1] for details). A *lot-type* is defined as a vector with a component for each size. This component specifies the number of pieces of that size in the lot. For example, for the sizes S, M, L, and XL, a lot of lot-type $(1, 2, 3, 1)$ contains one item of size S, two of size M, three of size L, and one of size XL.

In reality, excess supply is compensated by cutting prices. Therefore, we extended the model in [1] by a model for this recourse action. This resulted in the following stochastic mixed integer linear programm, in the following denoted by SLDP:

$$\max -\sum_{b\in B}\sum_{l\in L}\sum_{m\in M} c_{b,l,m}x_{b,l,m} - \sum_{i=1}^{n}\delta_i \cdot z_i + \mathbb{E}_\xi(Q(x,\xi)) \tag{1}$$

$$\text{s. t.} \qquad \sum_{l\in L}\sum_{m\in M} x_{b,l,m} = 1 \qquad \forall b \in B, \tag{2}$$

M. Kießling (✉) · T. Kreisel · S. Kurz · J. Rambau
Business Mathematics, University of Bayreuth, 95440 Bayreuth, Germany
e-mail: miriam.kiessling@uni-bayreuth.de

T. Kreisel
e-mail: tobias.kreisel@uni-bayreuth.de

S. Kurz
e-mail: sascha.kurz@uni-bayreuth.de

J. Rambau
e-mail: jorg.rambau@uni-bayreuth.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_77,

$$\underline{I} \leq \sum_{b \in B} \sum_{l \in L} \sum_{m \in M} m \cdot |l| \cdot x_{b,l,m} \leq \overline{I}, \tag{3}$$

$$\sum_{l \in L} y_l \leq \sum_{i=1}^{n} z_i, \tag{4}$$

$$\sum_{m \in M} x_{b,l,m} \leq y_l \qquad \forall b \in B, l \in L, \tag{5}$$

$$z_i \leq z_{i-1} \qquad \forall i \in \{2, \ldots, n\}, \tag{6}$$

$$x_{b,l,m} \in \{0, 1\} \qquad \forall b \in B, l \in L, m \in M, \tag{7}$$

$$y_l \in \{0, 1\} \qquad \forall l \in L, \tag{8}$$

$$z_i \in \{0, 1\} \qquad \forall i \in \{1, \ldots, n\}, \tag{9}$$

The meanings of the symbols are as follows: By L we denote the set of possible lot-types which can be delivered to a branch from the set B with multiplicity from the set M. If Branch b gets Lot-type l with Multiplicity m then the corresponding binary variable $x_{b,l,m}$ takes value 1, and 0 otherwise. At most n lot-types may be used. If we deliver Lot-type l to Branch b with Multiplicity m costs of $c_{b,l,m}$ arise. Every branch is supplied by exactly one lot-type with one multiplicity, see constraint (2). The overall supply has to be between the lower bound $\underline{I}$ and the upper bound $\overline{I}$ (3). By $|l|$ we denote the overall number of items in lot-type l. Using i lot-types implies costs of δ_i. Constraint (6) requires that also the costs for using $i-1, i-2, \ldots, 1$ are added if i different lot-types are used. Due to Constraint (5), y_l indicates whether some branch $b \in B$ is supplied by Lot-type l. Finally, (4) links y_l and z_i.

To compensate for excess supply – depending on a vector of random variables ξ describing the demand – price cuts are possible. The corresponding optimization problem seeking for optimal price-cut strategies is denoted by $Q(x, \xi)$ where x is the vector of all variables $x_{b,l,m}$. [1] The objective function is the revenue depending on supply x and the random vector ξ. Our goal is to maximize the total profit in expectation.

We performed a field study at our project partner using the SLDP. How can we find out, whether the new method outperforms the traditional manual planning? From an economic point of view, we should look at the actual revenues. However, the data showed that actual revenues during the field study are distorted too much by factors beyond our control. For this purpose we developed two different comparison methods which exclude such extern distortions and just reveil how well the size-dependent demand is met. We examine significance by using the Wilcoxon rank-sum test from statistics.

In the following we introduce our approaches and show results from the field study.

[1] An approximation of $Q(x, \xi)$ can be computed by solving a mixed integer linear programm or a dynamic program.

2 Comparison Methods

We developed two different indicators for demand consistency of the supply with sizes.

The first approach – called Normalized Sales Rate Deviation (NSRD) – compares the sales rates per size against each other in such a way that the popularity of the product itself has no dominating influence on the result anymore.

The simple and course idea is to observe the sales at the 50 %-*day* and look for each size at the fraction of supply that has been sold so far. The 50 %-day is the first day in the sales period where 50 % of the total supply of the product has been sold. We get an estimation of a normalized sales rate of each size relative to its supply, independent of the popularity of the product. We call this number the *normalized sales rate estimate* of a size, denoted by $\mathrm{NSR}(s)$.

For example, if there is a product with supply $(10, 20, 20, 10)$, and the size-dependent sales numbers up to the 50 %-day are $(2, 10, 15, 3)$, then the estimated sales rates are $(0.2, 0.5, 0.75, 0.3)$. This indicates that the supply of M was spot-on (i.e., relative sales in this size were the same as relative sales in total), wheras the supplies of S and XL were too small and the supply of L was too large in the considered branches.

Based on these observations, we are interested in how much the sales in a size deviate from the overall 50 %. To this end, we estimate the standard deviation of the observed normalized selling rates (relative to the supply). Because of multiple sales per day we can have an average NSR different from 0.5. Therefore, the standard deviation must be taken with respect to the sample average, which is[2]

$$\mathrm{SD}^{+} := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \hat{r})^2}, \tag{10}$$

where N is the sample size and $\hat{r}$ the sample average. In our case, we compute for N sizes

$$\mathrm{NSRD} := \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\mathrm{NSR}(s_i) - \widehat{\mathrm{NSR}})^2}, \tag{11}$$

the *normalized sales rate deviation* (for a single product).

The smaller the normalized sales rate deviation is, the more consistent is the supply with the demand for sizes.

The second approach uses ideas from [5]. Let P be the set of products in a given commodity group, S_p be the set of sizes for a product $p \in P$, b be a given branch, s be a given size, and $\theta_{b,s'}(p)$ be the first day at which Product p is sold out in Branch b and Size s' (here $\theta_{b,s'}(p) = \infty$ is possible). With this we can define the Top-Dog-Count $W(b, s)$ as

[2] We use the notation from [3].

$$\left|\left\{p \in P \;\middle|\; 0 = |\{s' \in S_p \mid \theta_{b,s'}(p) < \theta_{b,s}(p)\}|\right\}\right| \tag{12}$$

and the Flop-Dog-Count $L(b, s)$ as

$$\left|\left\{p \in P \;\middle|\; 0 = |\{s' \in S_p \mid \theta_{b,s'}(p) > \theta_{b,s}(p)\}|\right\}\right|. \tag{13}$$

We now consider the value $|W(b, s) - L(b, s)|$, in the following denoted as *top-dog-deviation of Size* s, denoted by TDD(b, s). We assume that a value closer to zero corresponds to a better supply strategy for s, because TDD(b, s) estimates the difference of the probability that s is sold out first and the probability that s is sold out last. In order to exclude numerical artefacts with too small integers, we restrict ourselves to samples that consist of a (random) subset of articles such that $W(b, s) + L(b, s)$ equals a given number. For each such sample we compute TDD(b, s).

Since for each branch and each size we get an individual measurement, the TDD is a finer measurement than the NSRD.

Now, any improvement measured by the above indicators in the field study could be a mere conincidence. To obtain statements about the significance of the observations NSRD and TDD(b, s), we apply the non-parametric *Wilcoxon rank-sum test* (see [2]). This method tests whether two sets of realizations stem from the same distribution.

Let us first sketch the Wilcoxon rank-sum test. Given two sets of realizations A and B, the null hypothesis is that A and B stem from the same distribution. The alternative says that distribution of A is shifted to the left.

In our context the alternative means: there is less deviation among normalized sales rates in the distribution of A than there is in the distribution of B. Or, respectively, A has smaller TDD(b, s)s than B.

The test is then conducted by sorting the values of both samples in ascending order, thereby assigning a rank to each value. In the next step the sum of the ranks for Family A is computed, which yields an observed rank sum.

In order to estimate the probability that the distribution of A is shifted to the left compared to the distribution of B, we have to consider the probability for getting rank sums smaller than or equal to the observed rank sum for A. If this probability lies below a predetermined significance level (5 or 10 %) we say that the observed difference can not be explained by chance, and we can assume that the A's distribution is shifted to the left compared to the distribution of B.

3 Comparison Results

We employed the SLDP on results of a field study with ladieswear shirts with sales periods from February to June 2011. For comparison we took historical data from the same commodity group in a time period from March until December 2006—at

Table 1 Standard deviation of selling rates (NSRD) for tested articles of ladieswear with ranks in braces

SLDP	8.58 (4)	16.92 (18)	13.49 (13)	17.32 (21)	17.29 (20)	14.37 (14)	12.83 (12)
	9.85 (5)	6.61 (3)	16.40 (17)				
Manual planning	18.15 (24)	21.89 (28)	22.71 (31)	12.44 (9)	19.93 (26)	22.98 (33)	27.46 (35)
	21.51 (27)	22.79 (32)	15.78 (15)	12.49 (10)	16.09 (16)	22.62 (30)	19.09 (25)
	17.06 (19)	10.52 (6)	17.53 (22)	11.66 (7)	12.35 (8)	5.09 (2)	17.65 (23)
	3.37 (1)	22.08 (29)	12.75 (11)	23.92 (34)	27.52 (36)	42.32 (38)	36.80 (37)

this time still all items were supplied by manual planning. We consider 26 branches and four different sizes, namely S, M, L and XL.

For the historical sample set only one lot-type was delivered, namely (1, 2, 2, 1). In the field study our system provided the lot-types (1, 1, 1, 1), (1, 1, 2, 2) and (2, 2, 3, 4). Now we would like to find out whether SLDP led to supplies that were significantly more consistent with the demand than the supplies suggested by manual planning.

At first we have a look at the normalized sales rate deviation. Aggregating over the tested branches and rounding to two decimal figures leads us to the observations for NSRD with the corresponding ranks in braces in Table 1.

We get a mean of 13.36 for the SLDP and 19.16 for the manual planning. So in average we get smaller deviations among the normalized sales rates per size for SLDP. Thus, we conclude that the systematic error in size distribution is smaller. To test significance, we apply the Wilcoxon rank-sum test with a predefined significance level of 5 %. The null hypothesis is that the distributions of NSRD for SLDP and the manual planning are identical, the alternative that for SLDP the distribution is shifted to the left. The test yields a rank sum of 127 for the SLDP. The probability of a rank sum lower than or equal to 127 is approximately 1.17 %, which is below the significance level. In other words: The probability that the observed improvements are the result of coincidence is well below the significance level 5 %, and the observations are significant.

As a next step we check the measurements TDD(b, s). Analogous to the selling rates, we compute mean values over all products, branches, and sizes, denoted by $\widehat{\text{TDD}}$. For the SLDP, $\widehat{\text{TDD}}$ is 2, whereas it is 3 for the manual planning. Again, we perform the Wilcoxon rank-sum test with a significance level of 5 %. We get the following result: With 208 observations (104 for each sample set) and a rank sum of 8,766 for the SLDP method, the probability for a shift to the left of the distribution of $\widehat{\text{TDD}}$ for SLDP method is less than 10^{-6} %. This also is well below the significance level: The observed improvements are significant.

4 Conclusion

We presented methods to compare the impact of different supply strategies for a fashion discounter in a real-world field study. The normalized sales rate deviation among sizes measures how evenly a product sells in the various sizes. The top dog deviation measures to what extent a size is sold out first more often than last, or vice versa. Both measures showed that our new supply strategy based on the stochastic lot design problem can significantly improve the demand consistency of the supply with respect to sizes, where significance is checked by the Wilcoxon rank sum test. Since the Wilcoxon text is robust and uses no assumptions on the distributions, we are confident that this result is practically relevant.

References

1. Gaul, C., Kurz, S., Rambau, J.: On the lot-type design problem. Optim. Methods Softw. **25**(2), 217–227 (2010)
2. H. Büning and G. Trenkler: Nichtparametrische statistische Methoden, Gruyter, (1994)
3. R. Purves, D. Freedman and R. Pisani: Statistics, WW Norton & Co, (1998)
4. S. Kurz and J. Rambau: Demand forecasting for companies with many branches, low sales numbers per product, and non-recurring orderings, Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, 22–24.10.2007, Rio de Janeiro, Brazil, 2007, pp. 196–201
5. S. Kurz and J. Rambau, J. Schlüchtermann, and R. Wolf.: The Top-Dog Index: A New Measurement for the Demand Consistency of the Size Distribution in Pre-Pack Orders for a Fashion Discounter with Many Small Branches, (submitted)

Heuristic Strategies for a Multi-Allocation Problem in LTL Logistics

Uwe Clausen and J. Fabian Meier

1 Introduction

We consider a "Less than truckload" (LTL) network with n depots and given shipping volumes between each two of them. For every connection between two depots, we also know the transport cost $d(i, j)$ per truck. Every shipping volume $w(i, j)$ can be transported directly from depot i to depot j or turned over at most twice at other depots. For this, depots have to be equipped with transshipment capacities (for which we have to pay); those depots are then called hubs. A priori every depot can be equipped with such capacities, but later we consider a restricted problem where only some of the depots have this ability.

Our aim is make strategic planning decisions (transshipment capacities, number of trucks in each depots, etc.) with the help of average data, so that we allow long computation times.

Since transport costs on an edge mostly depend on the number of trucks (and not just on the volume), we introduce integer truck variables. As the number of possible paths for transport is of order $O(n^4)$, we cannot get reasonable results for $n > 50$ by using Cplex. Therefore, we developed a heuristic approach using shipping trees (detailed in Sect. 3). It can be used in two ways: To get a primal bound and to get information about the "best hubs". If, in step 2, we restrict the problem to those best hubs (i.e. forbid turnover for all other hubs), we can improve the heuristic results.

Furthermore, this restricted problem allows us to find good feasible solutions with Cplex. In Sect. 4, we show how to find a lower bound from a strengthened LP relaxation. Evidence for the potential of our method will be given by the three real world test instances shown in Sect. 5.

U. Clausen
Fraunhofer-Institut für Materialfluss und Logistik, Joseph-von-Fraunhofer-Str. 2-4,
44227 Dortmund, Germany
e-mail: Uwe.Clausen@iml.fraunhofer.de

J. F. Meier (✉)
TU Dortmund, Institut für Transportlogistik, Leonhard-Euler-Str. 2, 44227 Dortmund, Germany
e-mail: meier@itl.tu-dortmund.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_78,

2 Formulation of the MIP

In the description below, we give a formulation for the problem we will call MAPIT (Multi-Allocation problem with integer trucks).

$$\text{Minimize} \quad \sum_{i,j} \mathit{Truck}(i,j) \cdot d(i,j) + \sum_{i,j} \big(\mathit{Transport}(i,j) - w(i,j)\big) \cdot t \tag{1}$$

$$\mathit{Transport}(i,j) = \sum_{k,l} \mathit{Route}(i,j,k,l) \cdot w(i,l) + \sum_{k,l} \mathit{Route}(k,l,i,j) \cdot w(k,j) + \sum_{k,l} \mathit{Route}(k,i,j,l) \cdot w(k,l) \quad \text{with} \quad \mathit{Transport}(i,j) \in \mathbb{R} \tag{2}$$

$$\mathit{Truck}(i,j) \geq \mathit{Transport}(i,j) \quad \text{with} \quad \mathit{Truck}(i,j) \in \mathbb{Z} \tag{3}$$

$$\sum_{k,l} \mathit{Route}(i,k,l,j) = 1 \quad \text{with} \quad \mathit{Route}(i,k,l,j) \in \{0,1\} \tag{4}$$

For the set I of depots, we define two parameters:

- $w(i,j), i,j \in I$, is the volume to be shipped from i to j (in fractions of a truckload)
- $d(i,j), i,j \in I$, is the cost for one truck to get from i to j.

Furthermore, we charge the transshipment capacities and the actual turnover with a factor t. The most important variable is $\mathit{Route}(i,k,l,j), i,k,l,j \in I$, which is binary and indicates the route $i \to k \to l \to j$ for the shipping volume $w(i,j)$ (k and l may be equal to each other or equal to j to indicate transport over less than two hubs). From this stem the variables $\mathit{Transport}(i,j), i,j \in I$, for the actual transport volume and $\mathit{Truck}(i,j)$ for the integral number of trucks to be used.

In (1) we first sum the transportation costs (trucks $\times$ costs) and then add the turnover costs (which can be calculated as the transport volume that leaves a point i without originating from it). In (2) we calculate the actual transport volume from the routing decisions stored in $\mathit{Route}(i,k,l,j)$. Equation (3) rounds up the transport volume to get the (integer) number of trucks. At last, (4) secures that we have exactly one routing for every pair (i,j).

Additionally, we should note that we can derive restricted problems RMAPIT by introducing binary parameters possiblehubs(i), $i \in I$, which restrict the possible turnover points. This reduces the number of possible values for $\mathit{Route}(i,k,l,j)$ and allows us to define a priori bounds for the *Truck*-variable. It is then at most $\lceil w(i,j) \rceil$ on nonhub-nonhub connections (i,j). To make a sensible choice for this binary parameter, we extract information about the "hub quality" from our heuristic.

3 Heuristic Approach

We divide a transportation plan (solution of MAPIT) into n trees T_j, $j \in I$, which represent the transport of the shipping volumes $w(i, j), i \in I$.

These trees have their root in j and have n nodes in total. The edges indicate transport with destination j. As at most 2 turnovers are allowed, a valid solution has at most 3 layers. Of course, not every possible solution can be divided into these trees, but we restrict the solution space to those for which this is possible. We easily construct a feasible initial solution by using direct transport for every connection (this is also the rounded solution of the LP relaxed problem). In our notation, it is represented by trees in which all $(n - 1)$ non-root nodes are directly connected with the root.

Why do we make this "tree assumption"? First of all, it is no strong restriction to assume that shipping volumes once gathered will not be divided anymore, so that we expect that tree-type solutions are not much worse than arbitrary solutions. Secondly, we simplify and speed up the computation: Every k in T_j represents a subtree in which the shipping volumes are already combined to one, so that between k and j we can consider all these commodities as "one big commodity".

3.1 The Algorithm

Our transportation graph is somehow the "sum" of the n sending trees T_j, sending shipping volumes to a specified depot j: The total transport volume on each edge $(k \to l)$ is the sum of all sending volumes on $(k \to l)$ in the different trees (notice that most of the trees will have zero sending volume on that edge since a tree uses only a small fraction of all edges). Our algorithm is a step by step improvement procedure:

1. We choose an edge $(k \to l)$ from the transportation graph with low capacity utilisation, i.e. containing a truck that transports a small volume over a long distance. A possible measure for that is $\big(Truck(k, l) - Transport(k, l)\big) \cdot d(k, l)$. For example, if we transport 3.1 truckloads on an edge costing 500 per truck, we transport 0.9 trucks of empty space for a total amount of $0.9 \cdot 500 = 450$ which might indicate a low capacity utilisation.
2. We look at all trees that contain this edge (usually only a few). Let T_j be such a tree: In this tree we try to replace the edge $(k \to l)$ by an edge $(k \to m)$ for which the overall transport and transshipment costs are lower. This has the computational advantage that we do not have to care whether k represents a huge or small subtree; we only have to compare the costs of transporting all shipping volumes in k (i.e. those that already reached k) on the given route to j with those on the alternative route. These costs consist of the linear turnover costs and the costs of additional trucks compared with the costs of saved trucks. Using an edge with low capacity utilisation we have a good chance of saving one truck on that

edge by rerouting a small shipping volume. For example, in Fig. 1 the shipping volume that has already reached depot 3 is $x(3) = w(1,4) + w(2,4) + w(3,4)$. The "tree assumption" now states that these three shipping volumes cannot be separated anymore on their way to depot 4. So we can consider $x(3)$ as unified shipping volume which can either be send directly to 4 or rerouted if this is cheaper.

3. After a (possible) replacement of the edge $(k \to l)$ in all trees which contain them, we go back to step 1.

To avoid looping we keep a list of recently chosen edges which are tabu for some time.

3.2 Example

Let us shed some light on the idea of the algorithm by discussing it for arbitrary starting values. We assume we have a graph with nine nodes, where T_4 is given by Fig. 1. At the moment, $Transport(7,6) = 3.16$, and, say, $d(7,6) = 40$. Then the capacity utilisation $(4 - 3.16) \cdot 40 = 33.6$ is very poor.

Assume that our checks gave the result that the capacity utilisation of $(7 \to 6)$ is the poorest of all edges. Then our job is now to go through all trees T_1 to T_9 and try to reroute the traffic on this edge (if there is any). We explain the mechanism for T_4: The edge $(7 \to 6)$ could be replaced by $(7 \to 5)$ or $(7 \to 3)$ to produce trees which are still valid. Of the three possibilities $(7 \to 6)$, $(7 \to 5)$ and $(7 \to 3)$ we choose the cheapest. For that notice that the rerouting decision only influences the tree T_4 (i.e. is made for every tree separately), but the cost function (which we need for finding the cheapest path) is a global function, i.e. incorporates the traffic of all trees.

After we have processed the edge $(7 \to 6)$ in all trees, we proceed with the next "worst edge". Experiments show that about $3n^2$ steps give a reasonable solution.

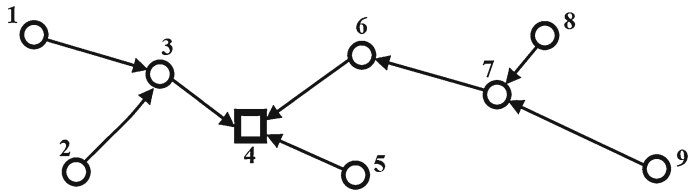


Fig. 1 Example tree T_4: $w(3,4)$, $w(5,4)$ and $w(6,4)$ are routed directly to 4, $w(1,4)$, $w(2,4)$ and $w(7,4)$ are routed through one hub, while the rest is routed through 2 hubs

3.3 Second Step Algorithm and RMAPIT

After we have run the algorithm described above we receive a medium good solution. From this solution we extract information about the usage of depots as hubs. Now, in the second step, we only keep those depots as possible hubs which have been assigned a high transshipment capacity. For all depots with low or non-existent transshipment capacity, we set the possiblehubs variable to zero, so forbidding transshipment completely. In this way we eliminate lot of variables from the problem which are probably of little use. We resulting problem will be called RMAPIT.

What do we do with this restricted problem? First, we run our tree algorithm again to produce a valid initial solution. After that we give RMAPIT to Cplex. Now the number of variables is low enough to allow the branch-and-cut algorithm to work properly. The numerical results from this steps can be found in Fig. 1.

4 Lower Bound

The naive lower bound for this problem is given by the solution of the LP-relaxed problem. This solution is obviously given by considering all transports to be direct (triangle inequality assumed) and so its value is $L_B = \sum_{i,j \in I} w(i, j) \cdot d(i, j)$.

To give a little additional strength to the LP relaxation we consider the following constraint: For every subset $K \subset I$ one can calculate the total shipping volume from K to $I \setminus K$. The number of trucks between K and $I \setminus K$ has to be at least the rounded up value of this total shipping volume. This is especially fruitful if the cardinality of K is 1 or 2: For example, if the total amount that leaves a fixed depot j is 6.4, then at least 7 trucks have to leave that depot. As you see from the computational results, the lower bound is still weak and has to be improved in the future.

5 Computational Results

We use real world data in three instances (index denotes the number of depots): I_{15}, I_{30} and I_{60}. Cplex was run for 12 h on a computer with 16GB RAM. The heuristic always finished in less than 10 min. The results are gathered in Table 1. They give evidence for the strength of the heuristic for high numbers of depots. In I_{15}, the heuristic is easily beaten by the algorithms of Cplex, but already for I_{30} we see that a clever restriction to the "best hubs" improves the quality of Cplex's best integer solution. For I_{60} even the worst heuristic result is a huge improvement, and the restricted problem gives reasonable solutions. Comparing the lower bounds, we see a slight advantage for the LP strengthening lower bound in bigger instances.

Table 1 Computational results

Computation	Instance 15	Instance 30	Instance 60
Cplex: best integer	45630.18	111587.68	951424.81
MAPIT Heuristic: best integer	88252.67	183288.10	372767.19
RMAPIT Heuristic: best integer	69866.43	136260.65	306832.32
RMAPIT Cplex: best integer	50866.71	100141.37	200920.49
Cplex: lower bound	38875.90	75754.94	118235.20
Lower bound from LP strengthening	38004.43	73681.57	121232.75
Cplex gap	15 %	32 %	88 %
Our gap	27 %	27 %	39 %

6 Directions for Further Research

The "tree approach" is up to now not very developed; theoretical observations and practical experiments will probably lead to a more sophisticated decision process and also adapted heuristic parameters (as the hub quality). By a closer investigation of the graph, especially by constructing mixed integer programs which are relaxations of MAPIT, we hope to improve the lower bound.

Acknowledgments This research project is part of the DFG project "Lenkung des Güterflusses in durch Gateways gekoppelten Logistik-Service-Netzwerken mittels quadratischer Optimierung".

Contracting Under Asymmetric Holding Cost Information in a Serial Supply Chain with a Nearly Profit Maximizing Buyer

Guido Voigt

1 Introduction

The purpose of the underlying contribution is to investigate how the design of screening contracts under asymmetric information can contribute to supply chain coordination when relatively small incentives are insufficient for aligning the actions of the supply chain parties. We use a stylized supply chain interaction model as presented by Inderfurth et al. [1] in order to analyze the impact of small pay-off differences on the supply chain performance. The model captures the basic supply chain conflict that suppliers typically prefer larger delivery lot-sizes in order to exploit economies of scale, while buyers tend to choose smaller delivery lots in order to have lower average inventories. We capture this situation with a lot-sizing decision in a serial supply chain facing deterministic end-customer demand. In this context, information asymmetry arises because the supplier (principal) cannot fully assess the buyer's (agent's) advantages of lowering the average inventories, i.e., the buyer's type (e.g., low cost type or high cost type) is unknown to the supplier.

While the screening theory is pretty much developed and established in the supply chain management literature (and other related areas) for fully rational and strictly profit maximizing supply chain parties, there are only a few contributions that analyze the supply chain behavior if these critical assumptions are not met. The strict profit maximization assumption postulates that all supply chain parties always take the (expected) profit maximizing action. The principal maximizes his expected profits by offering a menu self-selection contracts that provides incentives to reveal the private information. This revelation of private information, in turn, is also strongly linked to profit-maximization assumption, since revelation exploits the fact that there is a unique mapping between the private information and the profit maximizing contract. Since choosing the profit maximizing contract is assumed to be in the best interest

G. Voigt (✉)
Faculty of Management and Economics Operations Management,
Otto-von-Guericke University Magdeburg, Magdeburg, Germany
e-mail: guido.voigt@ovgu.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_79,

of the agent, the revealing contract choice is frequently denoted as self-selection. One direct implication of the profit maximization assumption is that the agent is always (weakly) indifferent between two contract alternatives (i.e., two contracts out of the menu-of-contracts). Due to the profit maximization assumption a distinct prediction on the contract choice is made, and the principal can theoretically infer the agent's private information from his contract choice (i.e., information revelation). We denote the strictly profit-maximizing contract as the self-selection contract and the next alternative to which the buyer is (weakly) indifferent as the indifference contract.

A recent laboratory study by Inderfurth et al. [1] shows that the strict profit-maximization assumption and, therefore, information revelation by self-selection is a critical assumption in the coordination literature under asymmetric information. In particular, their study shows that in 79 % of all observations the profit maximizing contract was chosen. However, there is also a non-negligible fraction of contract choices (i.e., 21 %) that cannot be explained by strict profit maximization. In 17 % of the observations, the agents chose the nearly profit maximizing indifference contract, i.e., only 4 % of the contract choices cannot be explained by strict or near profit maximization. Inderfurth et al. [1] are showing that these contract choices that are not strictly profit maximizing have a huge negative impact on the overall supply chain performance. The present work, therefore, presents a behavioral model that explicitly accounts for the fact that some agents might be insensitive to arbitrarily small pay-off differences between a self-selection and an indifference contract. Henceforth, we denote a contract that incorporates the buyer's insensitivity to arbitrarily small pay-off differences as a behavioral robust contract, since such a contract incorporates the buyers' behavior that is contrary to the standard assumption of strict profit maximization.

The approach presented in this contribution is closest to [2], Chap. 9.8.1, who are assuming that agents are making decision errors (so-called "trembling hand behavior") that can be described by a probability distribution. The principal accounts for these decision errors by adding a slack into the incentive constraint in order to increase the likelihood that the agent chooses the self-selection contract. The principal is then computing the optimal menu of contract by taking into consideration the probability with which a contract is chosen instead of the a-priori probability of the distribution of types, where types denotes agents having different realizations with respect to the private information (e.g., low cost type or high cost type). In contrast to [2] the present work extends the analysis to more than two types and allows for a different formulation of the agents decision errors. To this end, we are presenting a numerical study highlighting the impact of the insensitivity to small pay-off differences on the overall supply chain performance. Such a comparison has not been performed before but is especially important for the supply chain coordination point of view, since the second best outcome serves as a benchmark for the benefits of cooperation.

2 Outline of the Model

The strategic lotsizing model (see [1]) depicts a dyadic supply chain interaction in which the buyer (B) faces a deterministic and constant end-customer demand, d. Let f denote the setup costs for each delivery incurred by the supplier and R the buyer's unit costs when sourcing from an alternative supplier. It is assumed that the supplier makes zero profits if no trade takes place. We formalize the supplier's estimate regarding the buyer's cost with a probability distribution p_i $(i = 0, \ldots, n)$ over all possible holding cost realizations h_i $(h_i < h_j \; \forall \; i < j; i, j = 0, \ldots, n)$ that is assumed to be common knowledge.

Let the indices AI and FI refer to the situation under asymmetric information and full information, respectively. Under full information, the supply chain optimal outcome is achieved. Under asymmetric information, in contrast, the supplier maximizes his expected profits by offering a supply chain inefficient menu of contracts $A_i = \langle w_i^{AI}, q_i^{AI} \rangle$ $(i = 1, \ldots, n)$, where w_i^{AI} denotes the the wholesale price for each unit delivered to the buyer and where q_i denotes the respective order size. In the classical formulation of the problem, it is assumed that the buyer facing holding costs h_i chooses the order size q_i^{AI} as long as this minimizes his costs, i.e., as long as the incentive constraint $w_i \cdot d + \frac{h_i}{2} q_i \leq w_{i+1} \cdot d + \frac{h_i}{2} q_{i+1} \forall i = 1, \ldots, n$. However, in our approach we introduce the slack variable t_i that gives an additional incentive to the only nearly profit maxizing buyer to choose the self-selection contract. We refer to [2] for the concept of a nearly profit maximizing buyers. Note that t_i will not be included in the resulting contract offers, but will only be a fictional number controlling for the pay-off difference that will result between the contract offers.

Given a pay-off difference t_i, we are assuming that the buyer will choose the profit maximizing contract with probability $\alpha(t_i) = \alpha_i$, where α_i is monotonically increasing in t_i. In contrast, we assume that with probability $(1 - \alpha(t_i))$ the buyer is insensitive to arbitrarily small pay-off and therefore chooses the indifference contract. Furthermore, was assume upper and lower bounds (t_l and t_h) for the magnitude of the insensitivity to pay-off differences, i.e., the pay-off difference must be at least t_l for being tangible for the buyer, and is with certainty sufficient for a pay-off difference of t_h. Thus,the higher the pay-off difference, the higher the probability that the buyer chooses the profit-maximizing contract and vice versa.

In this case, the supplier maximizes the expected profits (EP) by offering a menu-of-contracts $A_i = \langle w_i, q_i \rangle$ that results from solving

$$\max EP\,[q_i, w_i, t_i] = \sum_{i=1}^{n} p_i \cdot d \left[\alpha_i \cdot \left(w_i - \frac{f}{q_i} \right) + (1 - \alpha_i) \cdot \left(w_{i+1} - \frac{f}{q_{i+1}} \right) \right]$$

$$s.t.$$

$$(w_i + t_i) \cdot d + \frac{h_i}{2} q_i \leq w_{i+1} \cdot d + \frac{h_i}{2} q_{i+1} \qquad \forall i = 1, \ldots, n-1$$

$$(w_n + t_n) \cdot d + \frac{h_n}{2} q_n \leq R \cdot d$$

$$q_{i+1} \leq q_i \quad \forall i = 1, \ldots, n-1$$
$$t_l \leq t_i \leq t_h \quad \forall i = 1, \ldots, n.$$

Note that the above formulated approach is already a reduced form of the classical optimization problem that already excludes some constraints that will not bind in the optimal menu of contracts. In the classical formulation of this problem, i.e., $\alpha_i = 1$ and $t_i = 0$, the objective function is convex and there is therefore only one critical point qualifying for an interior optimal solution. However, by introducing α_i, the objective function loses its convexity condition. However, for the numerical example presented in the following the optimality can be prooven by showing that there is only one critical point qualifying for the global optimum.

3 Numerical Example: Performance of Behavioral Robust Contract Parameters

The following numerical example depicts a situation with three buyer types. It is assumed that the buyer's insensitivity to small pay-off differences can be described by the following function, $\alpha_i = 1 + \frac{0.5 \cdot t_i - 0.3}{0.6}$, where $t_l = 0$ and $t_h = 0.6$. Moreover we set $f = 800$, $d = 100$, $(p_{low}, p_{med}, p_{high}) = (0.3, 0.4, 0.3)$, $h_{low}, h_{med}, h_{high} = (1, 3, 5)$, and $R = 15$.

The optimal parameters (see Table 1) still show the well known downward distortion of order sizes as known from the classical screening approach, i.e., $q_2 \leq q_2^{FI}$ and $q_3 \leq q_3^{FI}$. Moreover, the contract gives addtional incentives for the low and high cost type to self-select, i.e., $t_1, t_3 \geq 0$. In contrast, it is not optimal giving such an incentive to the medium cost type which may result in indifference contract choices, i.e., $t_2 = 0$.

In order to assess the impact of the buyer's insensitivity to small pay-off differences, we compare the behavioral robust contract to the a-priori screening contract with and without self-selection. The benchmark of the a-priori contract with self-selection depicts the classical situation in which all buyers are strictly profit maximizing, while the a-priori contract with insensitivity to small pay-off differences depicts the situation in which the buyer is in fact not strictly profit maximizing but the

Table 1 Optimal contract parameters/decision variables in the numerical example

Contract offer	Order size	Wholesale price	Additional incentive	Supply chain optimal order size
$A_1 = \langle q_1, w_1 \rangle$	$q_1 = 400.00$	$w_1 = 8.72$	$t_1 = 0.35$	$q_1^{FI} = 400.00$
$A_2 = \langle q_2, w_2 \rangle$	$q_2 = 182.58$	$w_2 = 10.15$	$t_2 = 0$	$q_2^{FI} = 230.94$
$A_3 = \langle q_3, w_3 \rangle$	$q_3 = 151.19$	$w_3 = 10.62$	$t_3 = 0.60$	$q_3^{FI} = 178.89$

Table 2 Performance benchmarks

	Supply chain performance	Coordination deficit	Coordination deficit in % of optimal supply chain cost
Behavioral robust contract	693.91	28.45	4.28
a-priori screening contract without indifference contract choices	685.88	20.42	3.07
a-priori screening contract with indifference contract choices	725.20	59.75	8.98

supplier nonetheless offers the classical a-priori screening contract. The supply chain performance is measured as the performance gap (coordination deficit) between the expected supply chain costs in the respective benchmarks and the expected optimal supply chain costs amounting to 665.46.

Tabel 2 summarizes the coordination deficits that arise in the respective benchmarks. If the classical screening approach is taken without considering the insensitivity to small pay-off differences, then the performance gap is substantially higher (3.07 % predicted by classical approach vs. 8.98 % when buyers are insensitive to small payoff differences). Note that a large impact on supply chain performance was also observed by [1]. However, the performance losses resulting from indifference contract choices can be substantially limited with a behavioral robust contract where the cost increase is only 4.28 % compared to the classical screening benchmark without indifference contract choices, i.e., 3.07 %. Thus, the behavioral robust contract shows a substantially higher performance compared to the setting in which the empirical finding of only nearly profit maximizing buyers is neglected.

4 Conclusion

Traditionally, it is assumed that all supply chain parties are strictly profit-maximizing. However, recent experimental work in the behavioral operations management area highlights that the strict profit maximization assumption is critical since buyers tend to be insensitive to small pay-off differences and, therefore, choose between strictly profit maximizing and only nearly profit maximizing contracts. Yet, such only nearly profit maximizing contracts have a substantially negative impact on the supplier's and overall supply chain's pay-offs, while the impact on the buyer's pay-offs is obviously negligible.

The present work relaxes the critical profit maximization assumption by assuming that the buyer in a serial supplier-buyer supply chain is only nearly profit maximizing. We show that the supplier can account for such behavior by increasing the pay-off

differences between the contract alternatives and/or by adjusting the order sizes in the respective menu of contracts. Therefore, we introduce a slack variable into the buyer's binding incentive constraints that allows the supplier to control for the pay-off differences between the contracts alternatives and, therefore, to influence the probabilities with which the respective contracts are chosen (see also [2]).

The analysis reveals that the second best benchmark that is predicted by the classical screening theory substantially overestimates the actual supply chain performance, since the nearly profit maximizing buyer will not choose the already downwards distorted profit maximizing order size, but sometimes the even more downwards distorted order size that is only nearly profit maximizing. However, if the supplier accounts for this behavior by designing a behavioral robust contract that accounts for insensitivity to small pay-off differences, then the performance losses can be substantially limited. Since the performance losses are mainly born by the supplier (note, the buyer is at least nearly profit maximizing), the supplier obviously has an incentive to anticipate such behavior and adjust the contract parameters accordingly. Nonetheless, we conclude that the inefficiencies arising from asymmetrically distributed information are underestimated in the coordination literature when strict profit maximization is assumed and along with this the benefits of cooperation (truthful information sharing and trusting information processing).

References

1. Inderfurth, K., Sadrieh, A., Voigt, G.: The impact of information sharing on supply chain performance in case of asymmetric information. Production and operations management, Forthcoming in (2012)
2. Laffont, J.-J., Martimort, D.: The Theory of Incentives: The Principal-Agent Model. Princeton University Press, Princeton/New Jersey (2002)

A Genetic Algorithm for the Integrated Scheduling of Production and Transport Systems

Jens Hartmann, Thomas Makuschewitz, Enzo M Frazzon and Bernd Scholz-Reiter

1 Introduction

The scheduling of production and transport processes in manufacturing supply chains is currently done separately. Thus, these schedules might lead to a local optimum of an objective pursued by the supply chain. Improvements of the operational supply chain performance might be achieved by an integrated consideration of operations [1]. The integrated production and transport scheduling problem (PTSP) can be formulated as a mixed-integer program (MIP) [3, 6]. The MIP comprises binary optimization variables that represent assignments, e.g. jobs to machines or transport devices, as well as continuous optimization variables like time and costs [2]. Since it belongs to the class of NP-hard problems, exact solutions are limited to small problem instances [7]. In particular, the considered production scheduling is based on a heterogeneous open flow-shop with several consecutive production levels [5]. Each production level n consists of several machines r, which feature a job-class specific processing time and cost. All jobs j have to be processed on one machine at each production level, which is denoted by $X_{j,n,r}$. The processing sequence of assigned jobs is given by $Y_{j,j',n,r}$. The jobs can be stored before the first production level, between production levels and before the assigned tour $A_{j,v}$ departs. Furthermore, jobs can be processed

J. Hartmann (✉) · T. Makuschewitz · B. Scholz-Reiter
BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Hochschulring 20,28359 Bremen, Germany
e-mail: hmn@biba.uni-bremen.de

T. Makuschewitz
e-mail: mak@biba.uni-bremen.de

B. Scholz-Reiter
e-mail: bsr@biba.uni-bremen.de

Enzo M. Frazzon
Industrial and Systems Engineering Department,Federal University of Santa Catarina (UFSC), Campus Universitário Trindade, Florianópolis-SC 88040-970, Brazil
e-mail: enzo@deps.ufsc.br

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_80,

externally (E_j) in a very short time, but causing a comparatively high cost. The transport of jobs from the production facility to their destination is performed by a homogeneous fleet of vehicles v, featuring a limited transport capacity. All considered tours start and terminate at the production facility. If at least one job is assigned to a tour, this tour is conducted (O_v). The tour can depart as soon as the processing of all assigned jobs is finalized. The routing of a tour v between the locations i is given by $Z_{v,i,i'}$. A performed tour involves fixed and variable costs. The variable costs depend on the duration of the tour. In addition, costs for an unpunctual delivery of orders occur. Furthermore, jobs can be shipped directly to their destination in time by a costly third party logistics provider (3PL), denoted by L_j. External processing and 3PL transport ensure the feasibility of the problem.

Genetic algorithms (GA) are efficient to explore a large solution search space. However, combinatorial problems can be difficult to solve with a GA since two feasible parents may not produce a feasible offspring. Furthermore the GA concept usually lacks a good local search capability [4]. This paper introduces a genetic algorithm for solving the PTPS of practically relevant size. The approach (depicted in Fig.1) decomposes the problem into a combinatorial and a continuous problem, which are tackled by different solution techniques. The GA features special operators for the mutation and crossover processes which ensure a feasible offspring, and local search capabilities for enhancing the fitness of the individuals.

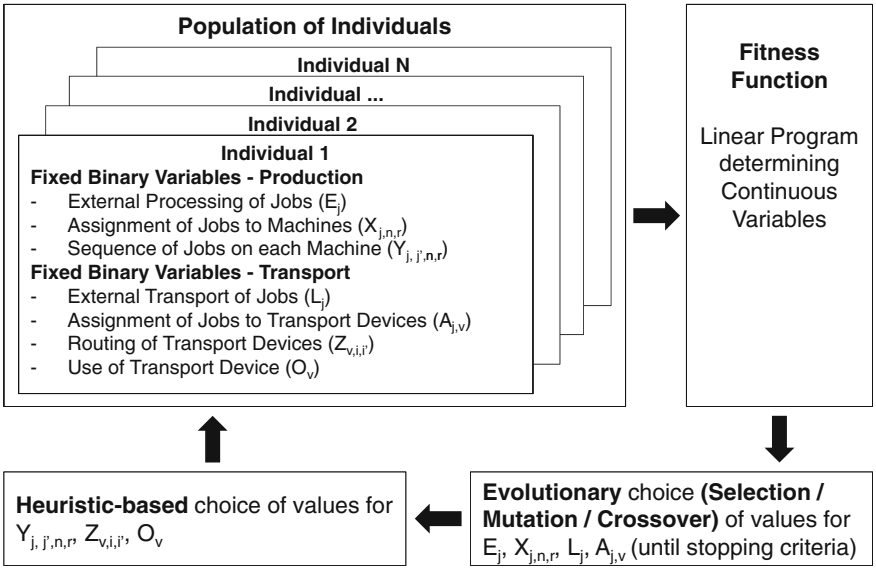


Fig. 1 Scheme of the genetic algorithm for the integrated scheduling of prodution and transport

The paper presents in Sect. 2 the chromosome representation and evolution, followed by the mathematical formulation of the fitness function in Sect. 3. Subsequently three heuristics for local search and results are sketched in Sect. 5.

2 Chromosome Representation and Evolution

A solution of the PTSP comprises binary and continuous variables. However, for fixed binary variables the optimization problem turns into a linear program (LP), which can be solved efficiently. Hence, the objective of the GA is to determine the best possible values of the binary variables in reasonable computation time. A *solution* built by the GA is encoded as a vector $(E_j, X_{j,n,r}, Y_{j,j',n,r}, L_j, A_{j,v}, O_v, Z_{v,i,i'})$ of the binary variables. Each binary variable is a gene of the chromosome that specifies an individual of the population.

This large number of binary variables and their dimensions would lead to a very large solution search space. However, not all combinations of values for the binary variables would lead to healthy individuals, representing a feasible solution of the planning problem. For the creation of healthy individuals the search space can be reduced by adding knowledge about the problem to the evolutionary operators. In this context it can be observed that a reasonable choice of values for some binary variables depends on other binary variables. In regard to the production system the choice for external processing for some jobs ($E_j = 1$) reduces the set of jobs that needs to be assigned to machines of the own production system. Furthermore, only a sequence of jobs assigned to the same machine has to be determined. In regard to the transport system only jobs not assigned for transport by a 3PL ($L_j = 0$) have to be assigned to tours $A_{j,v}$. Subsequent routing decisions only need to be taken for servicing the destinations of the assigned jobs.

Given these interdependencies the *mutation operator* is first applied to E_j and then to $X_{j,n,r}$. Hence, the number of 0 or 1 decisions in $X_{j,n,r}$ is reduced. In addition, knowledge about machine related parameters, e.g. processing time and costs, can be used to specify asymmetric probabilities for choosing or altering a certain machine on production level n for job j. Concerning the transport system first the assignment of jobs to a 3PL is done (L_j), afterwards the remaining jobs are assigned to tours $A_{j,v}$. The mutation of this assignment decision can be as well favored based on relevant parameters, e.g. destination of jobs or their due date. If at least one order is assigned to a tour the tour is conduced. This is indicated by O_v, which is set after the mutation is finalized. This mutation process ensures a feasible offspring. A *crossover operation* within a gene from two parent individuals would require the adaptation of the related genes in order to ensure a feasible solution, making such an operator computationally more demanding. Hence, in this approach crossover operations are limited to recombining the production related genes of one parent with the transport related genes from another parent. This procedure also ensures a feasible offspring, without excluding solutions from the general search space.

Based on interdependencies between variables and knowledge about the problem, two dependent binary variables are identified, which can be efficiently determined rather by a local search algorithm than by the evolutionary process. This applies to the sequence of jobs on each machine $Y_{j,j',n,r}$ and the routing decisions $Z_{v,i,i'}$ for each transport device. Appropriate heuristics are sketched in Sect. 4. The heuristic-based selection of $Y_{j,j',n,r}$ and $Z_{v,i,i'}$ reduces the solution search space of the GA and speeds up convergence. However, this might exclude the optimal solution.

3 Fitness Function

This section presents the mathematical formulation of the integrated production and transport scheduling problem for one production facility distributing jobs to various destinations. The fixed binary variables are denoted with a bar above.

3.1 Nomenclature

Sets

I	Locations of the transport network ($i, i' \in I$)		
I_i^C	Locations directly connected to i	V	Transport tours ($v \in V$)
I_v^{SL}	Start location of tour v	I^D	Location of the production facility
J	Jobs ($j, j' \in J$)	I_j^d	Destination of job j
K	Product classes ($k \in K$)	K_j	Assignment of class k to job j
N	Production levels ($n \in N$)	R	Resources/machines ($r \in R$)
N_j	Production levels n, on which job j has to be processed	R_n^e	Available resources/machines on production level n

Parameters

c_j^{ep}	Costs for external processing of j	$c_{j,k,n,r}^p$	Processing costs of j at r on n
c^d	Cost rate for unpunctual delivery	$d_{i,i'}$	Travel time between i and i'
c^{vv}	Cost rate of a transport tour	M	Large scalar
c^{fv}	Fixed costs of a transport tour	$pt_{j,k,n,r}$	Processing time of j on r at n
c^h	Storage cost rate for a job	$t_{j,n}^{aj}$	Supply date of j at n
c^{3PL}	Costs for the usage of a 3PL	t_j^{dd}	Desired delivery date of j

Positive continuous variables

$T_{j,n}^c$	Completion time of j at n	T_v^s	Start time of tour v
T_j^d	Tardiness of the delivery of j	$T_{v,i}^a$	Arrival time of tour v at location i
T_j^h	Total storage time of order j	T_v^{dv}	Duration of tour v

Binary variables

$\bar{A}_{j,v}$	Assignment of job j to tour v	$\bar{X}_{j,n,r}$ j	is processed at n on r
$\bar{E}_j$	External processing of job j	$\bar{Y}_{j,j',n,r}$ j	is processed before j' at n on r
$\bar{L}_j$	External transport of j by a 3PL	$\bar{Z}_{v,i,i'}$ i'	is visited after i by tour v
$\bar{O}_v$	Tour v is performed		

3.2 Mathematical Model

$$T^c_{j,n-1} + t^{aj}_{j,n} + \sum_{r \in R^e_n} pt_{j,k,n,r} \bar{X}_{j,n,r} \leq T^c_{j,n} \quad (j \in J; k \in K_j; n \in N_j) \tag{1}$$

$$T^c_{j,N} \geq t^{aj}_{j,n} - M(1 - \bar{E}_j) \quad (j \in J; n \in N) \tag{2}$$

The completion time of a job at a given production level has to be greater than or equal to the sum of the completion time at the previous production level and the required processing time of the assigned machine (1). The completion time of an externally processed job is at least its supply date (2).

$$T^c_{j,n} + pt_{j',k,n,r} \leq T^c_{j',n} + M(3 - \bar{X}_{j,n,r} - \bar{X}_{j',n,r} - \bar{Y}_{j,j',n,r}) \tag{3}$$
$$(j, j' \in J : j \neq j'; k \in K_{j'}; n \in N_j \wedge N_{j'}; r \in R^e_n)$$

$$T^h_j \geq T^s_v - M(1 - \bar{A}_{j,v}) - \sum_{n \in N_j} (t^{aj}_{j,n} - \sum_{r \in R^e_n} pt_{j,k,n,r} \bar{X}_{j,n,r}) \quad (j \in J; k \in K_j; v \in V) \tag{4}$$

Inequality (3) ensures that the completion time of a job on a machine is not less than the completion time of the previous job plus the processing time of the current job. Storage of a job may occur between its supply date and departure of its tour (4).

$$T^s_v \geq T^c_{j,N} - M(1 - \bar{A}_{j,v}) \quad (j \in J; v \in V) \tag{5}$$

$$T^s_v + d_{i,i'} - M(2 - \bar{Z}_{v,i,i'} - \bar{O}_v) \leq T^a_{v,i'} \quad (i, i' \in I : i \in I^{SL}_v \wedge i' \in I^C_i; v \in V) \tag{6}$$

$$T^a_{v,i} + d_{i,i'} - M(2 - \bar{Z}_{v,i,i'} - \bar{O}_v) \leq T^a_{v,i'} \quad (i, i' \in I \backslash (I^{SL}_v \wedge I^D) : i' \in I^C_i; v \in V) \tag{7}$$

$$T^a_{v,i} + d_{i,i'} - M(2 - \bar{Z}_{v,i,i'} - \bar{O}_v) \leq T^a_{v,i'} \quad (i, i' \in I : i' \in I^C_i \wedge i' \in I^D; v \in V) \tag{8}$$

$$\sum_{i' \in I : i \in I^C_{i'}} \bar{Z}_{v,i',i} M \geq T^a_{v,i} \quad (i \in I; v \in V) \tag{9}$$

A tour can start after all assigned jobs have been finalized (5). The arrival time at the destinations of the jobs is derived in (6), (7) and (8). The arrival time at a non-visited location equals zero (9).

$$T_v^{dv} \geq T_{v,i}^{a} - T_v^{s} - M(1 - \bar{O}_v) \quad (i \in I^D; v \in V) \tag{10}$$

$$T_j^{d} \geq t_j^{dd} - T_{v,i}^{a} - M(2 - \bar{A}_{j,v} - \bar{O}_v) \quad (j \in J; i \in I_j^d; v \in V) \tag{11}$$

$$T_j^{d} \geq T_{v,i}^{a} - t_j^{dd} - M(2 - \bar{A}_{j,v} - \bar{O}_v) \quad (j \in J; i \in I_j^d; v \in V) \tag{12}$$

In (10) the duration of a tour is determined. An unpunctual delivery of a job is captured by (11) and (12).

$$\begin{aligned} Min. \quad & \sum_{j \in J} T_j^d c^d + \sum_{j \in J} \sum_{k \in K_j} \sum_{n \in N} \sum_{r \in R_n^e} \bar{X}_{j,n,r} c_{j,k,n,r}^{p} \\ & + \sum_{j \in J} T_j^h c^h + \sum_{v \in V} (\bar{O}_v c^{fv} + T_v^{dv} c^{vv}) \\ & + \sum_{j \in J} (\bar{E}_j c_j^{ep} + \bar{L}_j c^{3PL}) \end{aligned} \tag{13}$$

The objective function (13) minimizes the costs for unpunctual deliveries, the processing and storage costs of jobs and as well the fixed and variable costs of each conducted tour. Furthermore, it takes the costs for external processing of jobs and transport to their destination by using a 3PL into account.

4 Heuristic-Based Choice of Binary Variables

Heuristics for job sequencing and vehicle routing can be applied to determine quickly a good solution for each individual. The sequence of all jobs assigned to a machine ($Y_{j,j',n,r}$) can be determined by priority rules, e.g. 'earliest due date first'. The initial assignment of jobs to tours and routing of tours is based on an adapted savings algorithm. During the iterations of the GA an adapted nearest neighbor algorithm is utilized for finding good tours. Both routing heuristics consider costs for an unpunctual delivery of jobs and are enhanced by a 3-opt improvement heuristic.

5 Results

The GA was implemented in MATLAB, using GAMS to determine the optimal LP solution. For small problem instances the GA finds optimal solutions. For larger instances only an upper bound for the MIP is known. A comparison with this bound showed that the GA computes solutions of 25–30 % less costs and is able to compute schedules for several hundreds of jobs. The convergence is speeded up compared to a GA without knowledge-based evolutionary operators and local search heuristics.

Acknowledgments This research was supported by CAPES, CNPq, FINEP and DFG as part of the Brazilian-German Collaborative Research Initiative on Manufacturing Technology (BRAGECRIM). The authors also thank Mr. Christoph Timmer for the implementation of the heuristics.

References

1. Chen, Z.-L., Vairaktarakis, G.L.: Integrated scheduling of production and distribution operations. Manage. Sci. **51**(4), 614–628 (2005)
2. Mula, J., Peidro, D., Diaz-Madroñero, M., Vicens, E.: Mathematical programming models for supply chain production and transport planning. Eur. J. Oper. Res. **204**(3), 377–390 (2010)
3. Pundoor, G., Chen, Z.L.: Scheduling a production-distribution system to optimize the tradeoff between delivery tardiness and distribution cost. Naval Res. Logistics **52**(6), 571–589 (2005)
4. Reeves, C.R.: Genetic algorithms. In: Potvin, J.-Y., Gendreau, M (eds.) Handbook of Metaheuristics, pp. 109–139. Springer, New York (2010)
5. Scholz-Reiter, B., Freitag, M., De Beer, C., Jagalski, T.: Modelling dynamics of autonomous logistic processes: Discrete-event versus continuous approaches. CIRP Ann. Manufact. Technol. **54**(1), 413–416 (2005)
6. Scholz-Reiter, B., Frazzon, E.M., Makuschewitz, T.: Integrating manufacturing and logistic systems along global supply chains. CIRP J. Manufact. Sci. Technol. **2**(3), 216–223 (2010)
7. Yuan, J., Soukhal, A., Chen, Y., Lu, L.: A note on the complexity of flow shop scheduling with transportation constraints. Eur. J. Oper. Res. **178**(3), 918–925 (2007)

A Dynamic Model for Facility Location in Closed-Loop Supply Chain Design

Orapadee Joochim

1 Introduction

At the strategic level, closed-loop supply chain (CLSC) management involves long-term decisions regarding the location and capacity allocation of forward/reverse logistics facilities, the assignment of products to facilities, and the distribution of products between facilities and their end users or suppliers. Both demand and return handling must be taken into account, and the overall problem becomes more complicated than a separate forward supply chain (FSC) or reverse supply chain (RSC). Thus, the system should be scalable enough for being able to facilitate different kinds of requirements without any potential disruption of supply chain activities.

In this paper, a mathematical model is developed to comprehensively determine strategic solutions for the capacitated facility location problem in closed-loop supply chains.

2 The Proposed Model

Figure 1 schematically illustrates the type of supply chain (SC) modeled in this paper. The mathematical symbols firstly introduced in Fig. 1 are provided along with a short description of each in Table 1. The presented SC system consists of four critical processes: (1) production, (2) distribution, (3) collection and (4) disassembly and remanufacturing.

The production and disassembly-remanufacturing centers could be located at the same site (as bidirectional facility) or different places (as unidirectional facility). It is possible to locate both the distribution center and collection center at the same

O. Joochim (✉)
Institut für Produktionswirtschaft, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover, Germany
e-mail: orapadee.joochim@prod.uni-hannover.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_81,

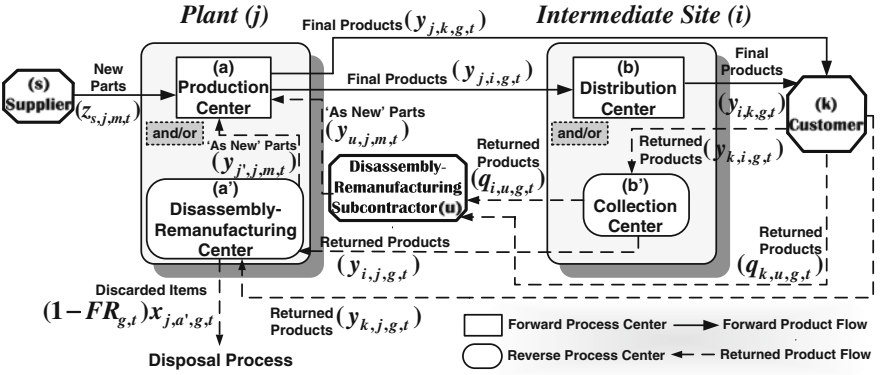


Fig. 1 Configuration of the proposed model

Table 1 Notation in the formulation of the proposed model

Indexsets			
$l \in \mathscr{L}$	Location sites	$c \in \mathscr{C}$	Center types for SC processes
$o \in \mathscr{O}$	Selectable location sites, $\mathscr{O} \subset \mathscr{L}$	$f \in \mathscr{F}$	Center types for FSC processes, $\mathscr{F} \subset \mathscr{C}$
$e \in \mathscr{E}$	Existing location sites, $\mathscr{E} \subset \mathscr{O}$	$r \in \mathscr{R}$	Center types for RSC processes, $\mathscr{R} \subset \mathscr{C}$
$n \in \mathscr{N}$	New location sites, $\mathscr{N} \subset \mathscr{O}$	$a \in \mathscr{A}$	Center types at plant sites, $\mathscr{A} \subset \mathscr{C}$
$j \in \mathscr{J}$	Plant sites, $\mathscr{J} \subset \mathscr{O}$	$b \in \mathscr{B}$	Center types at intermediate sites, $\mathscr{B} \subset \mathscr{C}$
$i \in \mathscr{I}$	Intermediate sites, $\mathscr{I} \subset \mathscr{O}$	$p \in \mathscr{P}$	Product types
$s \in \mathscr{S}$	External suppliers, $\mathscr{S} \subset \mathscr{L}$	$g \in \mathscr{G}$	Final products, $\mathscr{G} \subset \mathscr{P}$
$k \in \mathscr{K}$	Customers, $\mathscr{K} \subset \mathscr{L}$	$m \in \mathscr{M}$	Parts/components, $\mathscr{M} \subset \mathscr{P}$
$u \in \mathscr{U}$	External subcontractors, $\mathscr{U} \subset \mathscr{L}$	$t \in \mathscr{T}$	Periods in the planning horizon
Parameters			
$DP_{k,g,t}$	Demand of g by k in t	$KM_{o,c}$	Fixed expanding/relocating size for c at o
$RC_{k,g,t}$	Fraction of g returned from k in t	$AM_{m,g}$	Amount of m for assembling a unit of g
KO_o^{max}	Maximum capacity of o	$RM_{m,g}$	Amount of m from disassembling a unit of g
$KI_{o,c}$	Initial capacity of c at o	$FR_{g,t}$	Fraction of g satisfying specifications in t
$KC_{o,c}^{max}$	Maximum capacity of c at o	$FC_{o,c}$	Fraction of capacity of c allowed in o
$KC_{o,c}^{min}$	Minimum capacity of c at o	IR	Interest rate
Decisionvariables			
Non-negative integer			
$x_{j,a,g,t}$	Processing quantity	$exp_{o,c,t}$	Amount of expanded capacity
$y_{l,l',p,t}$	Transported products from l to l'	$mov_{o,o',c,t}$	Amount of relocated capacity
$z_{s,j,m,t}$	Purchased parts/components	$w_{o,c,t}$	Number of fixed sizes for expansion
$q_{l,u,g,t}$	Subcontracted returned products	$v_{e,n,c,t}$	Number of fixed sizes for relocation
Binary			
$\varphi_{o,t}$	1, if o is operated; 0 otherwise	$\rho_{e,c}$	1, if c is expanded at e; 0 otherwise
$\delta_{o,c,t}$	1, if c at o is operated; 0 otherwise		

site or to locate the distribution center and the collection center at different intermediate sites, i.e. bidirectional or unidirectional intermediate site.[1] In our framework, production centers have three alternatives for acquiring parts/components used to manufacture final products: (1) ordering the required parts/components from external suppliers, (2) re-processing the returned products and bringing those back 'as new' parts/components, and (3) outsourcing to subcontractors for disassembled and remanufactured parts/components. The manufactured products from production centers are initially transported to distribution centers and/or directly transported to customers. The distribution centers will then store the products until needed by customers. Whereas the collection points, which receive the used goods from customers, are used as storage for the reverse channel before the returned products are shipped to disassembly-remanufacturing centers and/or disassembly-remanufacturing subcontractors. Both disassembly-remanufacturing centers and disassembly-remanufacturing subcontractors can also receive the returned products straight from customers. The disassembly-remanufacturing centers are responsible for some essential activities of recovering, in which the returns are disassembled, tested, sorted and cleaned for reuse, repair and remanufacturing. Some discarded items from the disassembly-remanufacturing centers will be sent for the disposal process.

The objective function of the proposed model[2] as shown in (1) is based on cash flows. The total expenses in period t ($total\ expenses_t$) include purchasing expenses of parts/components, subcontracting expenses of disassembled and remanufactured parts/components, processing expenses at plant sites, transportation expenses of products, expenses of operating facilities, expenses of closing and opening facilities, expansion and relocation expenses, and disposal expenses.

$$\mathbf{MAX} \sum_{t \in \mathcal{T}} \frac{1}{(1+IR)^t} \left[total\ revenue_t - total\ expenses_t \right] \tag{1}$$

Forward and Reverse Flow Constraints Constraint (2) provides the amount of parts/components required for manufacturing. Constraint (3) assures the connection between the manufacturing process at any production center, and the outbound flows to distribution centers and directly to customers. Constraint (4) is the flow conservation at the distribution center. Constraint (5) ensures that all customer demands must be met. In constraint (6), the predefined return rate of final products is used as the return amount from customers. Constraint (7) is the flow conservation at the collection center. The volume of returned products sent to any disassembly-remanufacturing center is ensured by constraint (8). Constraints (9) and (10) establish the requirement for the outgoing flows of reusable parts/components from disassembly-remanufacturing process.

[1] These uni/bidirectional facilities are developed from the idea of Sahyouni et al. [3].

[2] The model is developed from the models of Demirel and Gökçen [1], and Melo et al. [2].

$$\sum_{s\in\mathscr{S}} z_{s,j,m,t} + \sum_{j'\in\mathscr{J}} y_{j',j,m,t} + \sum_{u\in\mathscr{U}} y_{u,j,m,t} = \sum_{g\in\mathscr{G}} x_{j,a,g,t} AM_{m,g} \ , \quad \forall j,\ a \in \mathscr{F},\ m,\ t \tag{2}$$

$$x_{j,a,g,t} = \sum_{i\in\mathscr{I}} y_{j,i,g,t} + \sum_{k\in\mathscr{K}} y_{j,k,g,t} \ , \quad \forall j,\ a \in \mathscr{F},\ g,\ t \tag{3}$$

$$\sum_{j\in\mathscr{J}} y_{j,i,g,t} = \sum_{k\in\mathscr{K}} y_{i,k,g,t} \ , \quad \forall i,\ g,\ t \tag{4}$$

$$\sum_{j\in\mathscr{J}} y_{j,k,g,t} + \sum_{i\in\mathscr{I}} y_{i,k,g,t} = DP_{k,g,t} \ , \quad \forall k,\ g,\ t \tag{5}$$

$$\left(\sum_{j\in\mathscr{J}} y_{j,k,g,t} + \sum_{i\in\mathscr{I}} y_{i,k,g,t}\right) RC_{k,g,t} = \sum_{i\in\mathscr{I}} y_{k,i,g,t} + \sum_{j\in\mathscr{J}} y_{k,j,g,t} + \sum_{u\in\mathscr{U}} q_{k,u,g,t} \ , \quad \forall k,\ g,\ t \tag{6}$$

$$\sum_{k\in\mathscr{K}} y_{k,i,g,t} = \sum_{j\in\mathscr{J}} y_{i,j,g,t} + \sum_{u\in\mathscr{U}} q_{i,u,g,t} \ , \quad \forall i,\ g,\ t \tag{7}$$

$$\sum_{k\in\mathscr{K}} y_{k,j,g,t} + \sum_{i\in\mathscr{I}} y_{i,j,g,t} = x_{j,a,g,t} \ , \quad \forall j,\ a \in \mathscr{R},\ g,\ t \tag{8}$$

$$\sum_{g\in\mathscr{G}} \left[FR_{g,t}\left(x_{j,a,g,t} RM_{m,g}\right)\right] = \sum_{j'\in\mathscr{J}} y_{j,j',m,t} \ , \quad \forall j,\ a \in \mathscr{R},\ m,\ t \tag{9}$$

$$\sum_{g\in\mathscr{G}} \left\{ FR_{g,t} \left[\left(\sum_{k\in\mathscr{K}} q_{k,u,g,t} + \sum_{i\in\mathscr{I}} q_{i,u,g,t} \right) RM_{m,g} \right] \right\} = \sum_{j\in\mathscr{J}} y_{u,j,m,t} \ , \quad \forall u,\ m,\ t \tag{10}$$

Capacity Expansion and Relocation Constraints It is possible to expand the capacity at some existing location sites e. Constraint (11) limits the capacity for further expansion at each center c of any existing location site e. Constraint (12) restricts the full expanded capacity at any existing location site e. Constraint (13) limits the capacity that can be relocated from each center c at any existing location site e to one or more new location sites n. Moreover, constraint (11) together with (13) make sure that an existing capacity can either relocate to new sites ($\rho_{e,c} = 0$) or expand its capacity ($\rho_{e,c} = 1$). Constraint (14) imposes that by period t center c has been established at the new location site n for expanding the additional capacity and/or relocating the capacity from one or more existing location sites e. The additional capacity allowed at any new location site n is restricted by constraint (15). For each time period t, the allowable amount of capacity added to every center c at any selectable location site o is bounded by constraints (16) and (17).

$$\sum_{t\in\mathscr{T}} exp_{e,c,t} \le \left(KC_{e,c}^{max} - KI_{e,c}\right)\rho_{e,c}\ , \quad \forall e,\ c \tag{11}$$

$$\sum_{c\in\mathscr{C}}\left[FC_{e,c}\left(\sum_{\tau=1}^{t} exp_{e,c,\tau} + KI_{e,c}\rho_{e,c}\right)\right] \le KO_{e}^{max}\varphi_{e,t}\ , \quad \forall e,\ t \tag{12}$$

$$\sum_{n\in\mathscr{N}}\sum_{t\in\mathscr{T}} mov_{e,n,c,t} \le KI_{e,c}\left(1-\rho_{e,c}\right)\ , \quad \forall e,\ c \tag{13}$$

$$\left(\sum_{\tau=1}^{t} exp_{n,c,\tau} + \sum_{e\in\mathscr{E}}\sum_{\tau=1}^{t} mov_{e,n,c,\tau}\right) \le KC_{n,c}^{max}\delta_{n,c,t}\ , \quad \forall n,\ c,\ t \tag{14}$$

$$\sum_{c\in\mathscr{C}}\left[FC_{n,c}\left(\sum_{\tau=1}^{t} exp_{n,c,\tau} + \sum_{e\in\mathscr{E}}\sum_{\tau=1}^{t} mov_{e,n,c,\tau}\right)\right] \le KO_{n}^{max}\varphi_{n,t}\ , \quad \forall n,\ t \tag{15}$$

$$exp_{o,c,t} = w_{o,c,t}KM_{o,c}\ , \quad \forall o,\ c,\ t \tag{16}$$

$$mov_{e,n,c,t} = v_{e,n,c,t}KM_{e,c}\ , \quad \forall e,\ n,\ c,\ t \tag{17}$$

Several additional constraints can not be provided in this paper due to space limitation. Among them are, e.g., maximum and minimum capacity constraints of facilities, facility configuration constraints allowing facilities to change their status (opened or closed) at most once, a logical constraint deciding whether facilities' capacity is to be added, and constraints enforcing non-negativity and binary conditions on the decision variables.

3 Numerical Example and Conclusion

The model is illustrated with a numerical example comprised of two existing plant sites (pl1 and pl2), one existing intermediate site (in1), one potential new plant site (pl3) and one potential new intermediate site (in2). Before the planning horizon starts, pl1 and pl2 have both production center (a1) and disassembly-remanufacturing center (a2), and in1 has both distribution center (b1) and collection center (b2). It is possible to open a1 and a2 at pl3, and b1 and b2 at in2. The model is solved using GAMS/CPLEX.

To evaluate the model, two scenarios are defined in terms of the percentage of products returned from customers over 10 year periods. In scenario L, low rates of returns are considered. Product demands of 10–30 % are assumed to return to the supply chain. Scenario H considers high rates of returns. Product returns of 70–90 %

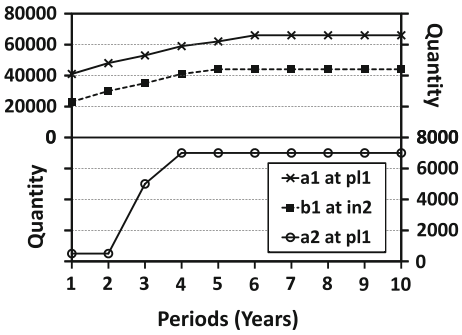


Fig. 2 Capacity expansion (scenario L)

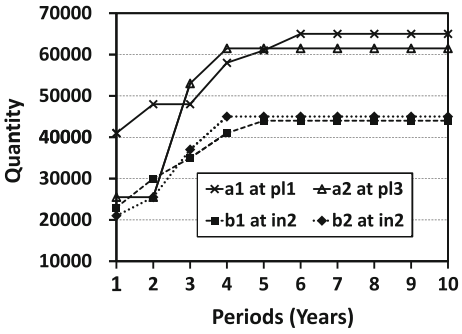


Fig. 3 Capacity expansion (scenario H)

from customers are assumed. Both scenario cases consider that customers' product demands of approximately 5–10 % gradually increase every year.

Figures 2, 3, 4, 5 show the capacity expansion and relocation at both existing and new location sites. There are investments in expanding the capacity of a1 at pl1 and b1 at in2 to meet increasing demands (both scenarios). For scenario L, the capacity of a2 at pl1 is expanded. In scenario H, it is more profitable to open a2 at pl3 and b2 at in2 for capacity expansion. There are also investments in capacity relocation from some existing facilities due to their high processing and shipping expenses. For

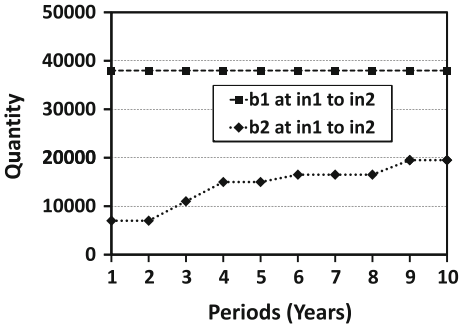


Fig. 4 Capacity relocation (scenario L)

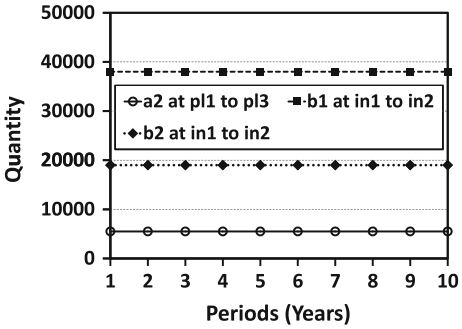


Fig. 5 Capacity relocation (scenario H)

both scenarios, the capacity of b1 and b2 are relocated from in1 to in2. Since returns increase in scenario H, the capacity of a2 is relocated from pl1 to pl3. The model selected to close a2 at pl2 due to high operating expenses at this site.

The configuration of both forward and reverse channels has a strong influence on the performance of each other. Bidirectional facilities eliminate substantial investment in infrastructure, equipment, and human resources. Only isolated, stand-alone forward and reverse facilities might be beneficial, if transportation and processing expenses are a large portion of the total expenses. The present model can be used to get better insight into the quantitative aspects of strategic planning within the CLSC context.

References

1. Demirel, N.Ö., Gökçen, H.: A mixed integer programming model for remanufacturing in reverse logistics environment. Int. J. Adv. Manuf. Technol. **39**, 1197–1206 (2008)
2. Melo, M.T., Nickel, S. Saldanha da Gama, F.: Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Comput. Oper. Res. **33**, 181–208 (2006)
3. Sahyouni, K., Savaskan, R.C., Daskin, M.S.: A facility location model for bidirectional flows. Transp. Sci. **41**, 484–499 (2007)

Pricing Indirect Behavioral Effects in a Supply Chain: From the Perspective of "Contracts as Reference Point"

Isik Bicer

1 Introduction

Effective supplier management is one of the key factors helping companies to promote operational excellence and to achieve substantial cost savings. A McKinsey study in 1993 highlighted the importance of the supplier integration and reported that some companies could reduce their manufacturing cost by 60–80 % and increase inventory turns from six to 50 a year by increasing supplier contribution [1]. Apart, integrating suppliers effectively into the internal operations would allow the companies to shorten the lead times and to improve customer satisfaction and product quality. However, having the full supplier support could not be possible without any cost, and hence, companies should be concerned about the satisfaction of their suppliers [10].

One of the important elements preventing the supplier satisfaction is the increasing order fluctuations in a supply chain as one moves to upstream. This phenomenon is known as Bullwhip effect, pioneered by Lee et al. [8], which is the source of tremendous inefficiencies for the upstream companies, e.g. excessive inventory, lost revenues, misguided capacity plan [8]. Although the downstream parties in a supply chain do not pay attention to the upstream inefficiencies, it is very important to measure the impact of such inefficiencies for effective supply chain management [3]. The empirical results in [9] provide evidence of the indirect effects of the retailers' behaviors on their own performances.

In this research, we analyze a supply-chain relationship in the presence of psychological costs as indirect behavioral costs by applying "contracts as reference points" model developed by Hart and Moore [5]. We consider a supplier and a retailer negotiating at time t_0 on a single-period and single-product contract which specifies the number of commitments and the number of options. The final order of the retailer is

I. Bicer (✉)
University of Lausanne,1015 Lausanne, Switzerland
e-mail: isik.bicer@unil.ch

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_82,

given at time t_1 ($t_1 > t_0$) which cannot be less than the number of commitments and more than the sum of the commitments and options. After the retailer's final order, the delivery of the products is done at time t_2 ($t_2 > t_1$) when actual demand is realized. Considering the two different contract enforcement models provided in [2], i.e. forced compliance and voluntary compliance, the indirect costs could be completely eliminated under forced compliance since the supplier should reserve a capacity to fulfill the retailer's maximum possible order quantity. However, under voluntary compliance, the elimination of the indirect behavioral costs would not be possible, and we show that there would be a trade-off for the retailer between decreasing the option cost and increasing the indirect behavioral cost that is imposed by the supplier when the uncertainty on her payoff function is increased by the retailer. Moreover, we demonstrate that the indirect cost of the retailer increases as the supplier becomes more critical to the retailer.

The rest of this paper is organized as follows. The next section describes the model, and briefly reviews the "contracts as reference points" theory [5]. Section 3 discusses the supplier-retailer relationship under voluntary compliance. Section 4 concludes.

2 Model Description

One of the assumptions of the "contracts as reference points" model is that each party is entitled to the maximum possible outcome when there is an uncertainty on their payoff function. If the parties receive less than what they are entitled to, there would be a psychological cost of feeling aggrieved which could only be eliminated by shading on consummate performance, imposing a cost burden on the other party [5]. Then, the payoff functions of the supplier and the retailer consisting of the psychological effects are obtained from [5] as follows.

$$\pi_s = u_s - \theta\alpha_r \tag{1}$$

$$\pi_r = u_r - \theta\alpha_s \tag{2}$$

where

- u_s and u_r are the gross payoff's of the supplier and retailer,
- $\alpha_s(\alpha_r)$ is the aggrievement level of supplier (retailer) which is the difference between the maximum possible outcome and the realized outcome [i.e. $u_s^{max} - u_s(u_r^{max} - u_r)$, the parties are entitled to the maximum attainable payoff since there is uncertainty on their payoff functions],
- θ is the psychological cost of 1$ of aggrievement to the supplier (or retailer). 1$ of aggrievement causes $\theta$$ to the supplier (or retailer).

On the other hand, our design of the flexible contract is same as the one described in [2], except that the final order including the firm commitments and options is given before the realization of demand. In our model, there are one supplier and one retailer negotiating on a contract at time t_0 for the delivery of a single product for one period where the contract specifies the number of firm commitments and options. Then, the retailer gives the final order at $t_1 > t_0$ depending on the number of commitments and options. After giving the final order, the retailer is delivered by the supplier at time t_2 ($t_0 < t_1 < t_2$) when the actual demand is realized. Using the same notation as in [2], we let $m \geq 0$ denote the number of commitments which is the minimum amount retailer has to buy; $o \geq 0$ the number of options; q the retailer's final order given at $t_1 > t_0$; and K the supplier's capacity reserved for the retailer's final order. Then, the following inequality should be satisfied as given in [2].

$$m \leq q \leq K \tag{3}$$

where $K = m + o$ under the forced compliance, and $K \leq m + o$ under voluntary compliance.

The range between the retailer's minimum and maximum possible order quantities in Eq. 3 determines the uncertainty of the supplier's payoff. Although the retailer could increase the number of options not to be out of stocks, the increase in the number of options might also increase the supplier's expected maximum profit causing shading on supplier's consummate performance.

3 Supplier-Retailer Relationship Under Voluntary Compliance

Supplier-retailer relationship is more complicated under voluntary compliance since the supplier is free to reserve a capacity between the retailer minimum and maximum possible order quantities, i.e. $m \leq K \leq m + o$. The supplier's underreservation decision would affect the retailer's payoff resulting in a loss of profit and causing psychic loss on the retailer. In addition, anticipating the supplier's underreservation tendency, the retailer has an incentive for information distortion that makes our analysis more complicated.

Since the supplier is free to choose the capacity level, she faces a newsvendor problem to determine the optimal amount. Having only the information of the number of commitments and options, we assume that the supplier solves the newsvendor problem by assuming uniform distribution between m and $m + o$ (i.e. maximum and minimum possible order quantities) for her capacity determination problem (see [7] for the detailed explanation of newsvendor model). Let w_m denote the price per commitment; w_o the price per option; w_e the exercise price per option exercised; c_k the supplier's capacity reservation cost per unit; and c_p the capacity conversion cost. Then, the optimal capacity reserved by the supplier for the final order of the retailer is obtained by the formula: $K = m + \left(1 - \frac{c_k}{w_e - c_p}\right) o$ [2, from Eq. 2 on p.

635]. Then, the supplier's maximum payoff would be:

$$\pi_s^{max} = w_m m + w_o o + w_e(K - m) - (c_k + c_p)K.$$

On the other hand, the actual payoff of the supplier for the final order quantity q, $m \leq q \leq K$, is obtained as

$$\pi_s^{act} = w_m m + w_o o + w_e(q - m) - c_k K - c_p q.$$

A careful reader might notice that the supplier's reserved capacity K is larger than m only if $w_e > c_p + c_k$. Otherwise, there will be no psychic loss since $K = m = q$. Thus, the supplier's psychic loss due to aggrivement under voluntary compliance could be derived as

$$\theta\alpha_s = \begin{cases} \theta(w_e - c_p)(K - q) & \text{if } w_e > c_p + c_k \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

As shown by Eq. 4, the retailer could only eliminate the supplier's psychic loss to prevent the supplier's shading on consummate performance when the reserved capacity of the supplier is equal to the number of commitments. But, this strategy would make the retailer be out of stocks. As a result, the elimination of the psychic loss of the supplier would not be possible in general. Anticipating the supplier's underreservation response for voluntary compliance, the retailer has also an incentive to amplify the number of options. If q_1^{max} and q_1^{min} denote the retailer's maximum and minimum possible order quantities at t_1, the number of commitments and options which ensure the maximum and minimum order quantities are obtained as follows.

$$m = q_1^{min} \tag{5}$$

$$o = (q_1^{max} - q_1^{min})\frac{w_e - c_p}{w_e - c_p - c_k} \tag{6}$$

where $w_e > c_p + c_k$. Apart, Eq. 6 shows that the number of options should be larger than $q_1^{max} - q_1^{min}$ to ensure the delivery of q_1^{max} quantities.

We analyze the retailer's option and shading costs for different values of w_e to understand the effect of the cost structure of the contract on the retailer's total cost. We allow different values of w_e while imposing two constraints: the first one is $w_o + w_e$ being equal to w_m, and the second is $w_e > c_p + c_k$. Hence, the retailer could decrease the option cost by increasing w_e and decreasing w_o by the same amount. However, this action increases the supplier's aggrievement level causing a shading on supplier's consummate performance. As in [5], the shading cost of the retailer is equal to the psychic loss of the supplier since the supplier shades on consummate performance to offset her psychic loss. In our analysis, we only focus on the option and shading costs since the cost of commitments is zero when $w_o + w_e = w_m$ [2, Lemma 1 on

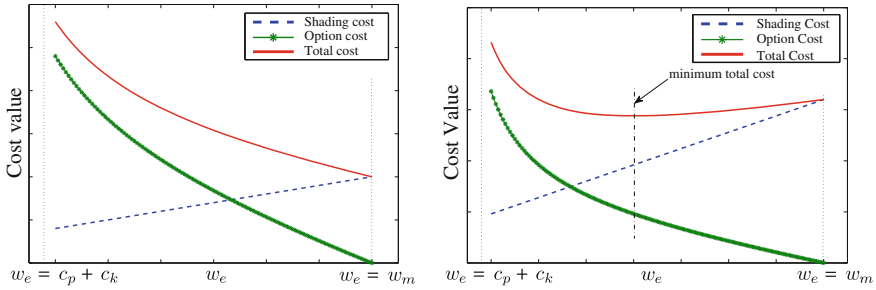


Fig. 1 Retailer's shading and option costs for two different θ values when $w_m = w_e + w_o$

p. 634]. Then, the sum of the retailer's option and shading costs based on Eqs. 5 and 6 could be derived as:

$$\psi = \theta(w_e - c_p)(q_1^{max} - q) + w_o(q_1^{max} - q_1^{min})\frac{w_e - c_p}{w_e - c_p - c_k}. \tag{7}$$

where q is the final order quantity Retailer's cost function, Eq. 7, depends on the retailer's minimum and maximum possible order quantities, i.e. q_1^{max} and q_1^{min}, which are determined at time t_0 while writing the contract. Apart, the cost function also depends on the optimal order quantity of the retailer which could be found by the model in [4].

Using Eq. 7 where $\theta = 1$ (i.e. psychological cost of aggrievement) and the maximum and minimum order quantities are obtained by assuming a flexibility component of 20 %, we get the retailer's cost curve for varying w_e with a constraint of $w_m = w_e + w_o$ as shown in Fig. 1. Given in the figure, the minimum cost is achieved for the minimum option value indicating that there is no trade off, and the retailer's action is not affected by the shading cost. Without any consideration of the indirect effects, this is also the optimal strategy for the retailer when $w_m = w_o + w_e$.

On the other hand, the shading cost would have a significant effect on the retailer's action for larger values of θ. As provided in Fig. 1, there is a trade off between decreasing option and increasing shading costs when $\theta = 4$. Thus, the retailer should pay attention to the supplier's profitability by not decreasing excessively the option cost when θ has a large value. Otherwise, the supplier shades on consummate performance, and the increase in shading cost of the retailer would overcome the decrease in the retailer's option cost. Since θ represents both the desire and the ability of the supplier to shade on consummate performance, it is expected to be larger for critical part suppliers. Therefore, the retailer should take a special care of the expectations of the critical part suppliers since any cost minimizing strategy could increase the total cost of the retailer.

4 Conclusion

We have examined the indirect behavioral effects in a supply chain by integrating "contracts as reference points" model into the quantity flexible contracts. Since the "contracts as reference points" model includes the quantitative measures for the psychological costs of the parties, our model allows to measure the impact of the indirect effects on the payoff function. In this Chapter, we demonstrate that the indirect behavioral costs would be significant for the retailer depending on how critical the supplier is to the retailer. Therefore, any cost minimization strategy of the retailer without any consideration of the supplier's payoff would increase the retailer's shading cost imposed by the supplier as a result of psychological cost of being aggrieved.

References

1. Asmus, D., Griffin, J.: Harnessing the power of your suppliers. McKinsey Quarterly, 63–78 (1993)
2. Cachon, G.P., Lariviere, M.A.: Contracting to assure supply: How to share demand forecasts in a supply chain. Manage. Sci. **47**(5), 629–646 (2001)
3. Davis, T.: Effective supply chain management. Sloan Manage. Rev. **3**, 35–46 (1993)
4. de Treville, S., Schurhoff, N., Trigeorgis, L., Avanzi, B.: Optimal sourcing and lead-time reduction under evolutionary demand risk. University of Lausanne, Working Paper (2012)
5. Hart, O., Moore, J.: Contracts as reference points. Ouarterly J Econ **123**(1), 1–48 (2008)
6. Hart, D.: Incomplete contracts and the theory of the firm. J. Econ. Organ. **4**(1), 119–139 (1988)
7. Hoop, W.J., Spearman, M.L.: Factory Physics, Ch. 2, pp. 65–68. McGraw-Hill, NewYork (2000)
8. Lee, H.L., Padmanabhan, V., Whang, S.: Information distortion in a supply chain: The bullwhip effect. Manage. Sci. **43**(4), 546–558 (1997)
9. Pagell, M., Sheu, C.: Buyer behaviours and the performance of the supply chain: An international exploration. Int. J. Prod. Res. **39**(13), 2783–2801 (2001)
10. Wong, A.: Integrating supplier satisfaction with customer satisfaction. Total Qual. Manag. **11**(4), 427–432 (2000)

Two-Stage Versus Single-Stage Inventory Models with or without Repair Ability

Ismail Serdar Bakal, Serkan Ozpamukcu and Pelin Bayindir

1 Introduction

We consider an inventory management of a single item under various settings. Our study is motivated by an application in military operations, and the item that we consider has critical importance. The current system has a two-echelon structure which consists of a central depot serving several bases. External demand stems from item failures and it occurs only at the bases.

Base i faces an external demand that follows a Poisson process with rate λ_i, independent of the other bases. The central depot and bases follow base-stock policies (the base stock levels are denoted by S_i, where 0 denotes the depot). The lead time between the depot and base i is T_i whereas it is L_0 between the depot and the outside supplier. All lead times are assumed to be deterministic and the objective is to minimize the long-run average cost of the system. The cost components are unit procument cost (c_p), backorder (b_i) and holding costs (h_i) per unit per unit time.

The military headquarters that operates this system see improvement opportunities that include acquiring repair ability and discontinuing the central depot. In order to investigate the implications of these opportunities, we consider four different inventory models, the first of which is the current system. The second system is the two-echelon model where the depot has repair ability. The third (fourth) model focuses on the single-echelon system without (with) repair ability (the bases replenish directly from the outside supplier with lead time L_i). For the models with repair ability, we assume that the repair facility has ample capacity.

I. S. Bakal (✉) · S. Ozpamukcu · P. Bayindir
Middle East Technical University, 06800 Ankara, Turkey
e-mail: isbakal@metu.edu.tr

S. Ozpamukcu
e-mail: sozpamukcu@gmail.com

P. Bayindir
e-mail: bpelin@metu.edu.tr

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_83,

The literature related to our work is very rich. Hadley and Whitin [1] is the first to give the exact costs for the single-echelon system. Phelps [2] considers the decision when to repair versus when to procure in a single facility system with multiple items. Allen and D'Eposo [3] consider a single facility model with (r, Q) policy. In this model the items are repairable and approximate steady state distributions are derived for stochastic demand and constant lead times. Simon [4] derives the exact results for this model. There are many studies that consider multi-echelon models with repair ability. Among them, METRIC can be regarded as a milestone study. Sherbrooke [5] presents METRIC for a two-echelon model with repair ability. The system is a conservative one, i.e. all items are repaired and none procured. Graves [6] considers a two-echelon system in which the depot acts like a central repair facility. There are few papers that consider the comparison of two and single-echelon models. Hausman and Erkip [7] consider a single-item system where all facilities follow $(S-1, S)$ policy. Axsäter [8] presents recursive procedures to determine exact costs and discusses the determination of optimal results for a two-echelon, single-item system. The item flow is very similar to that of our two-echelon models.

2 Analysis of the Models

In this section, we consider variations of the current system considering acquiring repair ability and removing the central depot. For the ease of analysis, we start with the single-echelon models (no central depot).

2.1 Single-Echelon Model without Repair Ability

We first consider the alternative setting where the central depot is removed from the system and the bases replenish directly from the outside supplier. Since the bases are independent, minimizing the total average cost reduces to minimizing the average cost of each base which is given as follows:

$$C_i(S_i) = h_i \sum_{j=0}^{S_i} (S_i - j) P(D_i = j) + b_i \sum_{j=S_i+1}^{\infty} (j - S_i) P(D_i = j) + c_p \lambda_i,$$

where D_i is the random variable denoting the lead-time demand for base i and it is Poisson distributed with rate $\lambda_i L_i$. It can easily be verified that the optimal base-stock level for base i can be characterized as $S_i^* = \min\{S_i \in 0, 1, 2, \ldots | \Delta C_i(S_i) \geq 0\}$, where $\Delta C_i(S_i)$ is the first order difference equation $(C_i(S_i + 1) - C_i(S_i))$.

2.2 Single-Echelon Model with Repair Ability

In this model, each base has a limited ability to repair the failed items. When an item fails (hence demand occurs), the base satisfies/backorders the demand and inspects the failed item. With probability ρ, the item is repaired in repair lead time R_i, incurring a repair cost of c_r. Holding costs apply for the items in repair as well. If the item cannot be repaired, the base places an order to the outside supplier.

Similar to Sect. 2.1, the long-run average cost of each base can be treated separately. Hence, we need to minimize

$$C_i(S_i) = h_i \left[\rho R_i \lambda_i + \sum_{j=0}^{S_i} (S_i - j) P(D_i^r = j) \right] + b_i \left[\sum_{j=S_i+1}^{\infty} (j - S_i) P(D_i^r = j) \right] + c_p \lambda_i (1 - \rho) + c_r \lambda_i \rho,$$

where D_i^r is the random variable denoting the inventory-on-order for base i and it is Poisson distributed with rate $\lambda_i[\rho R_i + (1 - \rho)L_i]$. The optimal base-stock level for base i can be characterized as $S_i^* = \min\{S_i \in 0, 1, 2, \ldots | \Delta C_i(S_i) \geq 0\}$.

2.3 Two-Echelon Model without Repair Ability

This model represents the current system where the central depot is responsible for replenishing the inventories of the bases. Since all locations follow base-stock policies, the demand faced by the central depot is Poisson distributed with rate $\lambda_0 = \sum_{i=1}^{N} \lambda_i$. It should also be noted that the items in-transit from the depot to the bases incur inventory holding cost at a rate of h_t, and there is no backorder cost at the depot. The cost incurred by the depot (including the holding cost for in-transit items) can be characterized as follows:

$$C_0(S_0) = h_0 \sum_{j=0}^{S_0} (S_0 - j) P(D_0 = j) + c_p \lambda_0 + h_t \sum_{i=1}^{N} \lambda_i T_i,$$

where D_0 is Poisson distributed with rate $\lambda_0 L_0$.

The long-run average cost of a base is more complicated since inventory-on-order is not equal to its demand during lead time (an order placed may not immediately be satisfied by the depot). However, we know through Graves [6] that the conditional distribution of the number of backorders at the depot belonging to base i, given the total backorders, is Binomial with success probability λ_i / λ_0. Hence, following the approach in Axsäter [8] and omitting the details, we have

$$C_i(S_0, S_i) = h_i \sum_{j=0}^{S_i} j \sum_{k=0}^{S_i-j} Y_{S_i-j-k} X_k - b_i \sum_{j=-\infty}^{-1} j \sum_{k=0}^{S_i-j} Y_{S_i-j-k} X_k$$

where $Y_{S_i-j-k} = \frac{(\lambda_i T_i)^{S_i-j-k} e^{-\lambda_i T_i}}{(S_i-j-k)!}$ and $X_k = \sum_{a=k}^{\infty} \frac{(\lambda_0 L_0)^{S_0+a} e^{-\lambda_0 L_0}}{(S_0+a)!} \binom{a}{k} \left(\frac{\lambda_i}{\lambda_0}\right)^k \left(1 - \frac{\lambda_i}{\lambda_0}\right)^{a-k}$.

When we consider the total cost of the two-echelon model without repair ability, we have $C(\overrightarrow{S}) = C_0(S_0) + \sum_{i=1}^{N} C_i(S_0, S_i)$. Unfortunately, $C(\overrightarrow{S})$ is not necessarily convex. However, given S_0, the optimal S_i values can easily be obtained using the first order conditions. Then, the optimal S_0 value can be obtained via complete enumeration for reasonable S_0 values.

2.4 Two-Echelon Model with Repair Ability

In this model, we consider an extension to the current setting by acquiring repair ability in the central depot. It is assumed that the inspection of the failed items is performed in the bases. Hence, failed items are transferred to the depot (in-transit time is T_i) only if they are repairable. The repair lead time is deterministic and denoted by R_0. Items in repair incur inventory holding cost at a rate of h_0.

We start the analysis by considering the cost of the depot which includes procurement, repair and inventory holding (items in-stock, being repaired, in-transit) costs:

$$C_0(S_0) = h_0 \sum_{j=0}^{S_0} (S_0 - j) P(D_0 = j) + c_p \lambda_0 (1-\rho) + c_r \lambda_0 \rho + h_I \sum_{i=1}^{N} \lambda_i T_i + h_0 \sum_{i=1}^{N} \rho \lambda_i R_0,$$

where D_0 is Poisson distributed with rate $(1-\rho)\lambda_0 L_0 + \rho \sum_{i=1}^{N} \lambda_i (R_0 + T_i)$.

The analysis for the long-run average cost of base i is almost identical to the analysis in Sect. 2.3. Omitting the details we have,

$$C_i(S_0, S_i) = h_i \sum_{j=0}^{S_i} j \sum_{k=0}^{S_i-j} Y_{S_i-j-k} X_k - b_i \sum_{j=-\infty}^{-1} j \sum_{k=0}^{S_i-j} Y_{S_i-j-k} X_k$$

Table 1 Parameter set for the base scenario (unit time: 1 week)

Demand	$\lambda_i = 3, i = 1, 2, 3$
Lead time	$L_0 = 3, R_0 = 2, T_i = 1, L_i = 4, R_i = 2, i = 1, 2, 3$
Costs (in thousand-dollars)	$c_p = 4, c_r = 1, h_0 = h_t = h_i = 0.02, b_i = 60, i = 1, 2, 3$
Repair ratio	0.4

where $Y_{S_i-j-k} = \frac{(\lambda_i T_i)^{S_i-j-k} e^{-\lambda_i T_i}}{(S_i-j-k)!}$ and $X_k = \sum_{a=k}^{\infty} \frac{(E[D_0])^{S_0+a} e^{-E[D_0]}}{(S_0+a)!} \binom{a}{k} \left(\frac{\lambda_i}{\lambda_0}\right)^k \left(1 - \frac{\lambda_i}{\lambda_0}\right)^{a-k}$. The total long-run average cost is then $C(\overrightarrow{S}) = C_0(S_0) + \sum_{i=1}^{N} C_i(S_0, S_i)$, and the discussion in Sect. 2.3 holds here as well.

3 Computational Findings and Concluding Remarks

In this section, our objective is to present our major findings through a detailed computational analysis. For this purpose, we establish a base scenario composed of 3 bases (The parameters of the base scenario is presented in Table 1). The parameters in Table 1 are in accordance with the actual data obtained from real life. The details of the computational analysis performed over the base case are omitted due to limited space. The reader may refer to Ozpamukcu [9] for further details.

We start with comparing the current system to a single-echelon model. As also noted in the literature (Muckstadt and Thomas [10], Hausman and Erkip [7]), the two-echelon model outperforms its counterpart single-echelon model due to the risk pooling effect. However, due to the nature of the military service that we consider, there is a considerable holding cost for in-transit items for the two-echelon system. Furthermore, the lead times may not be identical. Hence, the comparison is not straightforward.

Observation 1: For the base scenario, the two-echelon model is slightly better. This advantage gets more pronounces when h_t is smaller as expected.

Observation 2: Keeping the lead times of both models equal ($L_i = L_0 + T_i$), the benefits of the two-echelon model deteriorates as T_i and L_i increase, especially when h_t is significant.

Observation 3: Keeping the lead times of both models equal ($L_i = L_0 + T_i$), the benefits of the two-echelon model increase as L_0 and L_i increase.

Observation 4: As the demand rates increase, the costs of the two-echelon model increase faster in the presence of in-transit inventory holding cost.

Observation 5: As the number of bases increase (total demand rate remaining constant), the benefits of the two-echelon model increase.

We next consider the repair ability. It should be noted that the repair cost per unit is much less than the procurement cost in the base case, which complies with the real data. Hence, the system with repair ability always dominates the system without repair ability as expected. The effects of system parameters are discussed below.

Observation 1: The benefits of the repair ability increase (decrease) in L_0 (T_i). When the repair cost is much lower than the procurement cost, the increase/decrease is negligible.

Observation 2: The benefits of the repair ability decrease in R_0.

Observation 3: As c_r gets closer to c_p, the effects of system parameters become more pronounced. Furthermore, there is a threshold $c_r < c_p$ above which the repair ability is no longer beneficial.

3.1 Concluding Remarks

In this study, we considered single and two-echelon inventory models with or without repair ability. The data set that we used for our computational analysis represents a real life application. We have observed that the two-echelon model may no longer dominate the single echelon model as long as the inventory holding cost for in-transit items is significant. Furthermore, the repair ability is beneficial regardless of other system parameters when the repair cost is significantly lower than the procurement cost, which is the case for our work.

As future research directions, it is reasonable to consider the case where the bases have the opportunity of lateral transshipment, and/or the depot could be allowed to place emergency orders. Stochastic lead times can also be incorporated into the model.

References

1. Hadley, W., Whitin, T.M.: Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs (1963)
2. Phelps, E.: Optimal Decision Rules for the Procurement, Repair, or Disposal of Spare Parts. RM5678-PR, The Rand Corporation (1962).
3. Allen, S.G., D'Eposo, D.A.: An ordering policy for repairable stock items. Oper. Res. **16**, 669–674 (1968)
4. Simon, R.M.: Comments on a paper by S.G. Allen and D.A. D'Eposo, "An ordering policy for repairable stock items", P-3891-1, The Rand Corporation (1968).
5. Sherbrooke, C.C.: METRIC: a multi-echelon technique for recoverable item control. Oper. Res. **16**, 122–141 (1968)
6. Graves, S.C.: A multi-echelon inventory model for a repairable item with one-for-one replenishment. Manage. Sci. **31**, 1247–1256 (1985)
7. Hausman, W.H., Erkip, N.K.: Multi-echelon vs single-echelon inventory control policies for low-demand items. Manage. Sci. **40**, 597–602 (1994)
8. Axsäter, S.: Simple solution procedures for a class of two-echelon inventory problems. Oper. Res. **38**, 64–69 (1990)
9. Ozpamukcu, S.: An assessment of a two-echelon inventory system against alternative systems. M.S. Thesis, Middle East Technical University, Ankara (2011).
10. Muckstadt, J.A., Thomas, L.J.: Are multi-echelon inventory methods worth implementing in systems with low demand rates? Manage. Sci. **26**, 483–494 (1980)

Part XIX
Traffic and Transportation

Bi-Objective Network Equilibrium, Traffic Assignment and Road Pricing

Judith Y. T. Wang and Matthias Ehrgott

1 Traffic Assignment and User Equilibrium

Traffic assignment models the route choice of users of a road network. Given a set of origin-destination (OD) pairs and demand for travel between these OD pairs, it determines how many users choose each of the available routes, and thereby the amount of traffic on each section of the road network. Conventional traffic assignment is based on the assumption that all users want to minimise their travel time, or more generally, a generalised cost function

$$c(x_p) = m(x_p) + \alpha t(x_p), \tag{1}$$

where x_p represents traffic flow on route p, t is travel time, which is dependent on flow, and m is a monetary cost comprising of tolls, vehicle operating cost etc. that may also depend on flow and α is value of time. A user will choose the route between their origin and destination that has the least value of $c(x_p)$.

The traffic assignment problem is based on Wardrop's principle of user equilibrium [9], which can be stated as follows: *Under user equilibrium conditions traffic arranges itself in such a way that no individual trip maker can improve their generalised cost by unilaterally switching routes.* In other words, at equilibrium, the generalised cost of any used route between an OD pair must be equal and less than that of any unused route.

It is important to note that (1) is the linear combination of two components, time and monetary cost. In fact these are two different objective functions. Several authors

J. Y. T. Wang · M. Ehrgott (✉)
Department of Engineering Science, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
e-mail: j.wang@auckland.ac.nz

M. Ehrgott
e-mail: m.ehrgott@auckland.ac.nz

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_84,

have recognised this and suggested bi-objective traffic assignment models, see the references in [8]. However, these models are restrictive, by keeping the assumption of the existence of an additive generalised cost (or sometimes generalised time) function (1). Moreover, there is evidence, that users in reality do not behave according to this assumption, see references in [8]. In [8] we have suggested a more general bi-objective user equilibrium condition, that assumes that all users have the two objectives of minimising travel time and minimising toll cost.

Under bi-objective user equilibrium (*BUE*) *conditions traffic arranges itself in such a way that no individual trip maker can improve either his/her toll or travel time or both without worsening the other objective by unilaterally switching routes.*

We have shown that, even if considering all possible values of time, i.e. $\alpha \in [0, \infty)$, in (1), bi-objective models based on generalised cost provide only a subset of all possible solutions to traffic assignment that satisfy the BUE condition. Hence the definition of BUE provides an appropriate general framework for the study of traffic assignment in tolled road networks.

Furthermore, in [7] we have suggested the time surplus maximisation concept as a new route choice model that addresses the stochastic nature of route choice behaviour and the variability among users on their willingness to pay. It is based on the idea of time surplus. We assume that a user has in his mind a maximum time he is willing to spend in traffic, given any level of toll. If τ_p^k is the toll on route p for OD pair k and the travel time is $t(x_p^k)$ then the time surplus on route p for individual i is

$$t_{ip}^s = t_i^{max}\left(\tau_p^k\right) - t\left(x_p^k\right). \tag{2}$$

We assume that the higher the toll, the shorter the maximum time willing to spend, i.e. we assume that t_i^{max} is a strictly decreasing function of τ_p^k. This function t_i^{max} is an indifference curve between time and toll for user i. The time surplus maximisation concept stipulates that all users maximise their time surplus. This gives rise to a user equilibrium condition: *Under the time surplus maximisation user equilibrium* (*TSmaxBUE*) *condition traffic arranges itself in such a way that no individual trip maker can increase their time surplus by unilaterally switching routes.*

In order to find a solution of the TSmaxBUE problem, we employ a route-based formulation of the equilibrium condition and follow [4] to formulate this as a non-linear complementarity problem, which is solved by minimising an associated gap function. Notice that because time surplus is maximised, but the NCP formulation requires a cost function to be minimised, we need to write this cost function as $\eta_p^{ki} := M - t_i^{max}(\tau_p^k) + t(x_p^k)$ with a sufficiently large M in the NCP model.

2 Road Pricing

Road pricing is a common instrument to reduce congestion and has successfully been implemented in many cities around the world, e.g. in Singapore, Stockholm

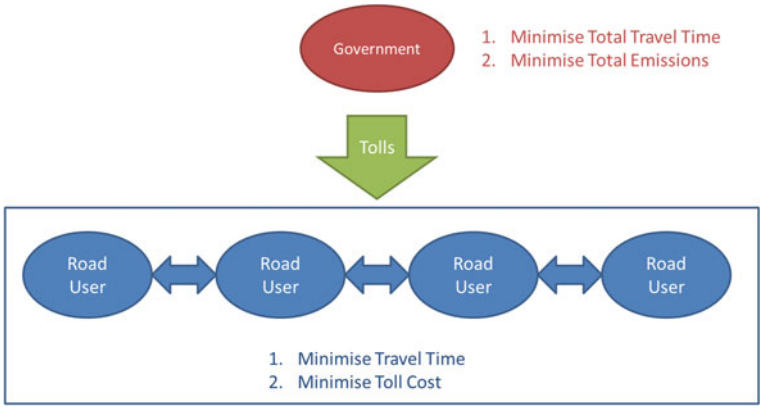


Fig. 1 The bilevel concept for road pricing

and London. The idea of congestion pricing is to charge tolls such that users are paying the marginal social cost rather than the average private cost for their trips. This induces changes in travel behaviour such that the total system travel time is minimised.

In todays world the idea of sustainable transport systems is gaining importance internationally. Sustainability encompasses the dimensions of economic, social and environmental sustainability [2]. The European Conference of Transport Ministers has defined a comprehensive catalogue of objectives of sustainable transport policy [3]. Among those, the objectives of creating wealth, reducing congestion, and reducing greenhouse gas emissions are relevant for this paper, the first in terms of economic sustainability, the second for both economic and environmental, and the last for environmental sustainability.

We suggest that, apart from considering tolls as a means to reduce congestion, road pricing can be an important instrument to reduce vehicle emissions. Hence the roading authority would pursue two objectives by charging road users: To minimise total travel time and to minimise total emissions. Road users on the other hand, will react to the imposed tolls and attempt to minimise their own travel time and toll cost. This framework is illustrated in Fig. 1.

At the government level there is, however, a dilemma. It is well known that tolls that minimise total travel time do not necessarily minimise emission levels [5, 10]. Hence the problem becomes that of the determination of efficient tolls such that neither the total travel time nor the total emissions can be reduced without worsening the other, which is a bi-objective optimisation problem.

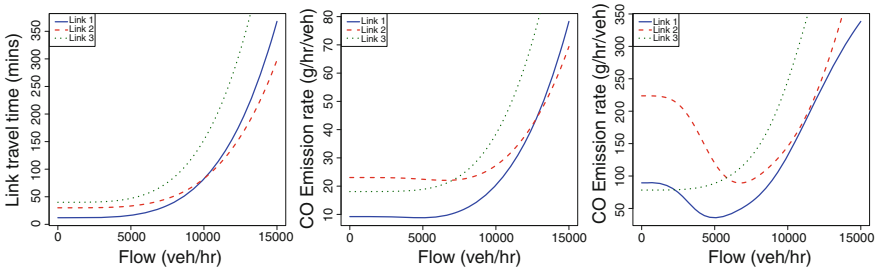


Fig. 2 Travel time (*left*) and CO emissions (*middle* and *right*) as functions of traffic flow

3 A Bi-level Bi-objective Optimisation-Equilibrium Model

Here, we first formalise the two objective functions at government level. The first objective function is to minimise total travel time:

$$\min z_t(x(\tau)) = \sum_{a \in A} x_a(\tau) t_a(x_a(\tau)), \tag{3}$$

where $t_a(x_a) = t_0 \left[1 + \alpha \left(\frac{x_a}{C_a}\right)^{\beta}\right]$, is a typical link travel time function [1]. Here t_a is the travel time on link a, which depends on link flow x_a. Also, t_0 is the free-flow travel time and C_a the practical capacity of link a. The values of $\alpha = 0.1$, $\beta = 4.0$ are typical, and we adopt them in the example of Sect. 4. The left plot of Fig. 2 shows three examples of travel time functions used in the example of Sect. 4.

The second objective function is the minimisation of total CO emissions.

$$\min z_e(x(\tau)) = \sum_{a \in A} x_a(\tau) e_a(v_a(x_a(\tau))). \tag{4}$$

Here, v_a is the traffic speed, which depends on link flow x_a and e_a is the CO emissions, which in turn depends on speed, on link a. Unfortunately, there is no consensus on the exact form of the emission function e_a. In Fig. 2 we show the functions proposed by Yin and Lawphongpanich [10] in the middle and by Niemeier and Sugawara [6] on the right.

4 A Three Link Example

We demonstrate our bilevel bi-objective-equilibrium model on a simple three link network. The three links (or routes) connect a single origin-destination pair and have the following characteristics. For route (link) 1, an expressway of 20 km length, we

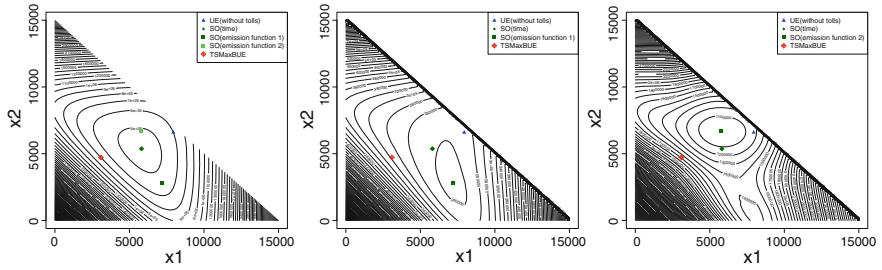


Fig. 3 Contour plots of travel time and emissions over feasible flows

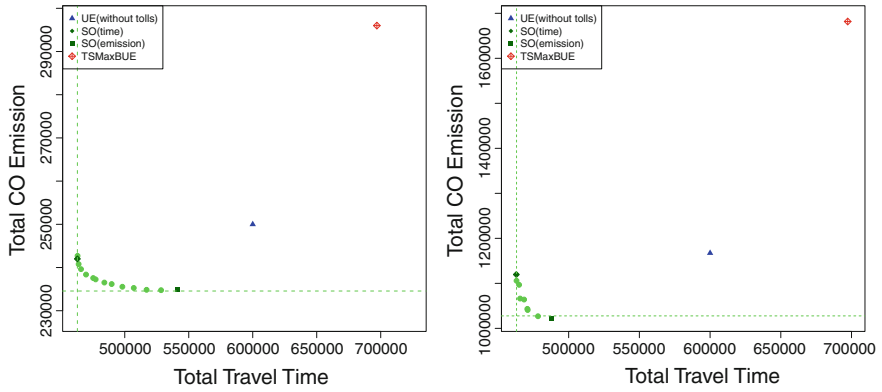


Fig. 4 The trade-off between total travel time and total CO emissions for efficient tolls

set $v_0 = 100$ km/h, $t_0 = 12$ min, and $C_a = 4,000$ vehicles per hour in the function (3). For route (link) 2, a highway of 50 km distance these values are $v_0 = 100$, $t_0 = 30$ and $C_a = 5,400$. Finally, route (link) 3, an arterial route of 40 km length, has $v_0 = 60$, $t_0 = 40$ and $C_a = 4,800$.

The travel time and emission functions for the three links of this network are illustrated in Fig. 2. In Fig. 3 we show contour plots of the travel time and emission functions over feasible flows together with the social optimum solutions for travel time and emissions, illustrating that these are indeed different. We also show the traffic flows at the untolled user equilibrium solution, and the TSmaxBUE solution at the toll values of $\tau_1 = 40$ and $\tau_2 = 20$.

To find the efficient tolls, we observe that we must have that the tolls are such that $\tau_1 > \tau_2 > \tau_3$ and assume that $\tau_3 = 0$. We assume that t^{max} is uniformly distributed between 10 and 25, 30 and 45, and 60 and 90 min on the three links, respectively. Moreover, we allow τ_1, τ_2 to be in the range between 1 and 40 in discrete steps of 1.

The resulting total CO emissions versus total travel time, for both emission functions, are plotted in Fig. 4, clearly illustrating the tradeoff between the two objectives as well as the difference to the untolled user equilibrium solution and the TSmaxBUE solution for $\tau_1 = 40$, $\tau_2 = 20$.

5 Conclusion

In this work we have proposed a bilevel framework for road pricing to support sustainable transport systems. On the upper level we consider a bi-objective optimisation problem of minimising total emissions and total travel time, whereas on the lower level we consider a bi-objective user equilibrium model with users who minimise their own travel time and toll cost. We have proposed the concept of time surplus maximisation as a way of dealing with the bi-objective user equilibrium. In future work, we will develop algorithms to solve the problem, based on the NCP formulation of the TSmaxBUE problem and using a multi-objective evolutionary algorithm to integrate this with the bi-objective optimisation problem on the upper level.

Acknowledgments This research was partially supported by the Marsden Fund project "Multiobjective network equilibria—From definitions to algorithms", grant number 9075 362506.

References

1. Bureau of Public Roads: Traffic assignment manual. U.S. Department of Commerce, Urban Planning Division, Washington D.C. (1964)
2. Commission, European: Planning and research of policies for land use and transport for increasing urban sustainablity PROPOLIS: Final report to European commission. European Commission, Brussels, Belgium (2004)
3. European Conference of Ministers of Transport: Sustainable transport policies. Available at www.cemt.org/online/council/2000/CM0001Fe.eps (2000)
4. Lo, H.K., Chen, A.: Traffic equilibrium problem with route-specific costs: Formulation and algorithms. Transp. Res. Part B Methodol. **34**(6), 493–513 (2000)
5. Nagurney, A.: Congested urban transportation networks and emission paradoxes. Transp. Res. Part D **5**, 145–151 (2000)
6. Niemeier, D.A., Sugawara, S.: How much can vehicle emissions be reduced? Exploratory analysis of an upper boundary using an emissions-optimized trip assignment. Transp. Res. Rec. **1815**, 29–37 (2002)
7. Wang, J.Y.T., Ehrgott, M.: Modelling stochastic route choice with bi-objective traffic assignment. In: International Choice Modelling Conference 2011 held in Leeds on 4–6 July 2011
8. Wang, J.Y.T., Raith, A., Ehrgott, M.: Tolling analysis with bi-objective traffic assignment. In: Ehrgott, M., Naujoks, B., Stewart, T., Wallenius, J. (eds.) Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems, pp. 117–129. Springer, Berlin (2010)
9. Wardrop, J.G.: Some theoretical aspects of road traffic research. In: Proceedings of the Institution of Civil Engineers, Part II, vol. 1, pp. 325–362 (1952)
10. Yin, Y., Lawphongpanich, S.: Internalizing emission externality on road networks. Transp. Res. Part D **11**, 292–301 (2006)

Contraction Hierarchies with Turn Restrictions

Curt Nowak, Felix Hahne and Klaus Ambrosi

1 Introduction

Modern digital road maps do not only model the bare geometry of the road network but also other real-life aspects like speed limits or turn restrictions. Algorithms solving the single source-single destination on such maps can therefore deliver a more precise estimation of driving distance and time of routes chosen by real drivers by taking into account such information. The demand for a concise estimation arises in many real world applications like vehicle routing, where tight time windows for delivery exist.

2 Related Work

Turn restrictions are normally modelled as a set forbidden combination of a pair of edges (TR) and can be approached in two different ways (see e.g., [4] in [5]):

- *Adaption of the graph*: Nodes with TR are split into complex nodes representing only allowed turns, thus strongly increasing both the memory needed for the storage of the graph and the number of iterations performed by the algorithm.
- *Adaption of the algorithm*: The originally node-oriented Dijkstra algorithm can be switched to edge-orientation by iterating over edges. This allows for an easy

C. Nowak (✉) · F. Hahne · K. Ambrosi
Institut für Betriebswirtschaft und Wirtschaftsinformatik,
Universität Hildesheim, Marienburger Platz 22, 31141 Hildesheim, Germany
e-mail: nowak@bwl.uni-hildesheim.de

F. Hahne
e-mail: ahne@bwl.uni-hildesheim.de

K. Ambrosi
e-mail: ambrosi@bwl.uni-hildesheim.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_85,

incorporation of TR into the algorithm. Yet it also leads to a much larger number of iterations (depending on the average node degree in the underlying graph) when compared to iterating over nodes.

We present an enhancement of Dijkstra's algorithm (*Dijkstra with adaptive TR*), which is TR-aware, iterates over nodes and adapts an edge-oriented point-of-view only when needed. As worst case, the number of iterations is equal to that needed by the edge-iterating version, but in many real-data cases it is smaller.

Our enhancement is also applicable when techniques like A* or Contraction Hierarchies (CH) are used. Geisberger and Vetter [2] presents an edge-based CH-algorithm accounting for turn-costs, modelling soft turn restrictions to which we will refer in the benchmarks in Sect. 4.

3 Dijkstra's Algorithm with Adaptive Turn Restrictions

3.1 Approach and Pseudocode

The existence of TR may cause nodes to appear more than once in the shortest path, even with all edge costs greater zero. See e.g., Fig. 1, where the shortest path from node a to node i is $\langle a, e, f, g, h, d, g, f, i\rangle$ (*note:* no U-turns are allowed). In other words: shortest paths may contain cycles and sub-paths do not have to be shortest paths by themselves.

In a TR-free graph, a node-oriented Dijkstra's algorithm maintains two sets:

- *Reached*: A set of reached nodes, queued to be expanded. Here the length of the best shortest path so far serves as (upper) bound for further paths reaching nodes in this set. Only one instance of each node may be in this set.
- *Settled*: After being expanded, a node is settled; no further improvement is possible. again, only one instance of each node may be in this set.

In a graph containing TR, more information has to be saved for a node j: the edge via which j has been reached, j.viaEdge, (only required if parallel edges are possible) and a flag j.TRfree. Node j is marked as j.TRfree = FALSE when it has been reached via a path containing TR or there is a TR starting from j.viaEdge. By means of this additional information, we can distinguish between several instances of a node in *Reached* and *Settled*.

As an example, let j become a member of *Reached* for the first time via a path w_1 with length $c(w_1)$ containing TR (j.TRfree $\leftarrow$ FALSE). Now let us assume, that j is reached again, now via a path w_2 not containing TR. If $c(w_2) > c(w_1)$, another instance j' of j must be added to *Reached*, as $c(w_1)$ must not serve as a bound for $c(w_2)$ (it may only serve as a bound for non-TR-free paths). If, however, $c(w_2) \leq c(w_1)$, a *better* (i.e., not longer and TR-free) path is found: j.distance and j.viaEdge are updated and j.TRfree $\leftarrow$ TRUE. Furthermore, all other instances (accessed via TR-free paths or not) j' with j'.distance $> j$.distance can be removed

from *Reached*. Should j be reached again afterwards via a third path w_3 containing TR and $c(w_3) < c(w_2)$, a new instance j'' must be added. Similar considerations apply for the set *Settled*.

As a summary, the modification of the algorithm allows the adding of multiple instances of a node via different edges both to *Reached* and *Settled* until the node is added to *Settled* via a TR-free path. In *Settled*, a maximum of one "TR-free" instance per node may be contained. The length of paths containing TR may serve only as bounds for TR-free paths when they end with the same edge. The pseudocode is given in Algorithms 1–3.

Algorithm 1 Dijkstra's algorithm with adaptive TR

Require: Graph $G = (V, E, TR)$, starting node s, target node t with $s, t \in V$

s.distance $\leftarrow 0$, s.predecessor $\leftarrow s$ ▷ Initialization
i.distance $\leftarrow \infty$, s.predecessor $\leftarrow \emptyset$, i.TRfree $\leftarrow$ TRUE , i.viaEdge $\leftarrow \emptyset$ $\forall i \in V \setminus \{s\}$
Settled $\leftarrow \emptyset$, *Reached* $\leftarrow \{s\}$

while *Reached* $\neq \emptyset$ **do**
  Chose h with h.distance $= \min\{i.\text{distance} \mid i \in \textit{Reached}\}$
  if $h = t$ **then** return shortest path to t **endif**
  Delete h from *Reached*
  Add h to *Settled*
  if h.$TRfree$ **then** Delete all (other) instances of h from *Reached* **endif**

  for all $j \in \mathcal{N}(h)$ **do** ▷ Expansion of node h
    if h.viaEdge, $(h, j) \in TR$ **then goto:** 10 **endif** ▷ Do not add j because of TR
    if $\exists j' \in \textit{Settled} \bigcup \textit{Reached}$ **and** (j'.TRfree **or** j'.viaEdge = (h, j)) **and** (h.distance + $c(h, j) \geq j'$.distance) **then**
      goto: 10 ▷ Shorter path to j already known (TR-free or reached via same edge)
    end if
    if h.TRfree **and** $((h, j), (j, k)) \notin TR$ **then** Encounter j via (h, j) TR-free (Algorithm 2)
    else Encounter j via (h, j) TR-aware (Algorithm 3) **endif**
  end for ▷ End expansion of node h
end while

Ensure: A shortest path from s to t respecting TR with length t.distance is recursively deducible from t.predecessor. If t.predecessor = $\emptyset$, no path exists between s and t.

Algorithm 2 TR-free encountering of an edge in Algorithm 1

Require: Edge (h, j) expanded most recently

Delete all intances of j from *Reached* having:
j.distance $\geq h$.distance + $c(h, j)$ ▷ Keeps non-TR-free instances of j reached via shorter paths

j.predecessor $\leftarrow h$; j.distance $\leftarrow h$.distance + $c(h, j)$
j.viaEdge $\leftarrow (h, j)$; j.TRfree $\leftarrow$ TRUE ▷ possibly overwriting FALSE
Add j to Reached

Algorithm 3 TR-aware encountering of an edge in Algorithm 1

Require: Edge (h, j) expanded most recently

1: **if** $\exists j' \in$ *Reached* **and** j'.viaEdge = (h, j) **then**
2: j'.distance $\leftarrow h$.distance + $c(h, j)$ ▷ shorter path to j' via edge (h, j) found
3: j'.TRfree $\leftarrow$ FALSE ▷ possibly overwriting TRUE
4: **else**
5: j.predecessor $\leftarrow h$; j.distance $\leftarrow h$.distance + $c(h, j)$
6: j.viaEdge $\leftarrow (h, j)$; j.TRfree $\leftarrow$ FALSE
7: Add j (possibly multiply) to *Reached*
8: **end if**

3.2 A Comprehensive Example

The example follows the algorithm calculating the shortest path from a to i in Fig. 1. Results are listed in Table 1. Distance are given in square brackets, an index denotes the viaEdge of nodes that have been added via non-TR-free paths. The following remarks describe selected iterations:

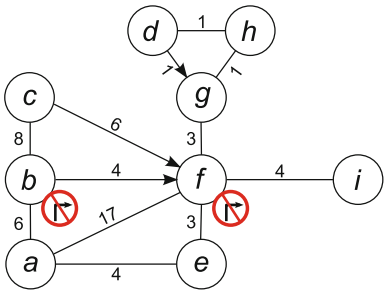


Fig. 1 Sample graph with turn restrictions

Table 1 Iterations performed while calculating the shortest path from node a to node i

Iteration	*Reached*	*Settled*
1	$\{a[0]\}$	$\emptyset$
2	$\{e[4], b_a[6], f[17]\}$	$\{a[0]\}$
3	$\{b_a[6], f_e[7], f[17]\}$	$\{a, e[4]\}$
4	$\{f_e[7], c_b[14], f[17]\}$	$\{a, e, b_a[6]\}$
5	$\{g_f[10], c_b[14], f[17]\}$	$\{a, e, b_a, f_e[7]\}$
6	$\{h_g[11], c_b[14], f[17]\}$	$\{a, e, b_a, f_e, g_f[10]\}$
7	$\{d_h[12], c_b[14], f[17]\}$	$\{a, e, b_a, f_e, g_f, h_g[11]\}$
8	$\{g_d[13], c_b[14], f[17]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h[12]\}$
9	$\{c_b[14], f_g[16], f[17]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h, g_d[13]\}$
10	$\{f_g[16], f[17]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h, g_d, c_b[14]\}$
11	$\{f[17], i_f[20]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h, g_d, c_b, f_g[16]\}$
12	$\{i_f[20], i[21]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h, g_d, c_b, f_g, f[17]\}$
13	$\{i[21]\}$	$\{a, e, b_a, f_e, g_f, h_g, d_h, g_d, c_b, f_g, f, i_f[20]\}$

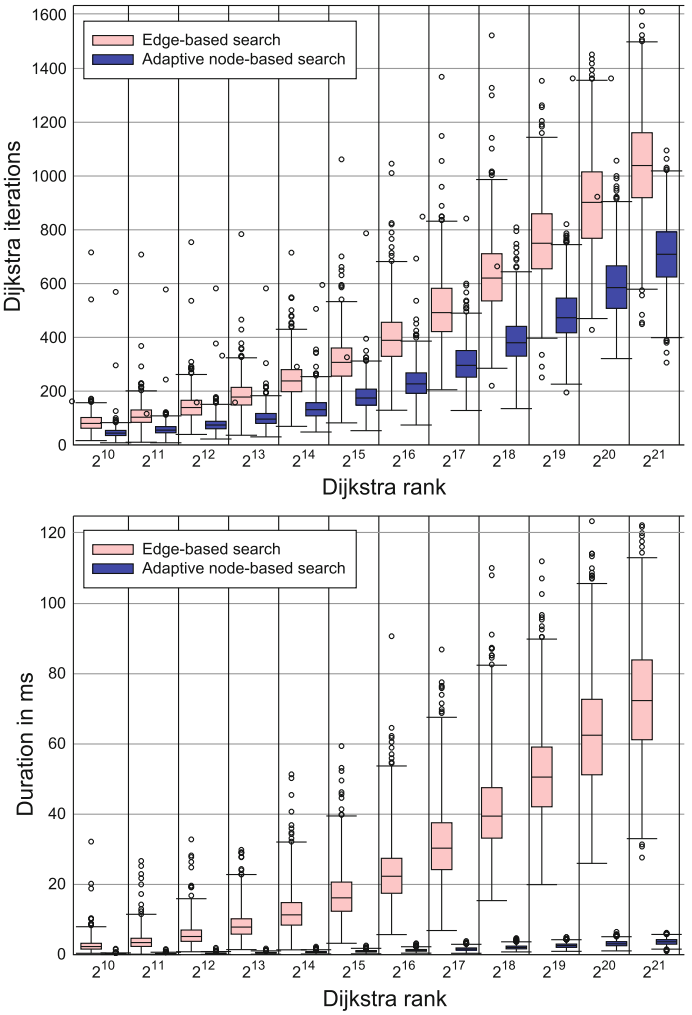


Fig. 2 Benchmark results. *The boxes* cover the second and third quartile of all queries; whiskers reach out for a total of 99 % of the queries; the remaining 10 outliers are plotted individually. The x-Axis denotes so called Dijkstra ranks: The number of iterations required by a standard edge-based Dijkstra search for finding the same paths. The edge-based search refers to our own implementation of [2]. Node/edge stalling was applied (see e.g., [1]). For the CH algorithm, the relative saving for running time is much higher than for number of iterations. The reason is, as mentioned in Sect. 3.3, that the finding of the bridging node between forward and backward search is time consuming. Very similar results were achieved when applying the algorithm to different maps. (Omitted in this paper due to space limitations.)

- *Iteration 3*: Multiple instances of f occur in *Reached*.
- *Iteration 5*: f is not expanded towards i due to TR.
- *Iteration 10*: f is not added to *Reached* via node c (length 20), as f is already a member of *Reached* and has been accessed *TR-free* with distance 17.
- *Iteration 12*: Multiple instances of f occur in *Settled* (and of i in *Reached*). As f has been settled TR-free, it will not be added to *Reached* again in later iterations even if there were other edges leading towards f.
- *Iteration 13*: Abort, as target i becomes a member of *Settled*.

3.3 Application to Algorithms Using CH

A second advantage of our algorithm over edge-based approaches unfolds in CH. Here, the search is performed in a bi-directional way and a bridging node between forward and backward search needs to be identified (see e.g., [1]). The node degree in CH may be very high. Thus edge-based searches need considerably greater effort when trying to find a common node between two settled edges whose ending nodes have high degrees. Our approach does not need to iterate over all adjacent edges when checking whether a node is settled in both search directions. This faster identification of bridge nodes allows much shorter query times (see Fig. 2).

4 Benchmarks and Results

Our benchmark uses a modified OSM data set of Germany, with 3,281,871 nodes, 4,087,526 edges and 30,958 TR. Upon this map, we built a CH using the parameters [2] use for their benchmarks. For every given Dijkstra rank, 1,000 random fastest-path queries were performed. Results are shown in the boxplots in Fig. 2.

5 Conclusion and Outlook

Our algorithm is an enhancement for Dijstra's algorithm (and its decendants), allowing to take into account turn restrictions while staying node-oriented as long as possible. Especially for CH, a massive reduction of computational effort can be achieved when implemented efficiently.

For brevity, the adaptation of the A*-algorithm to this approach as presented by Nowak et al. [3] is not included to this paper. Yet, it is easily applicable and equally beneficial.

References

1. Geisberger, R., Sanders, P., Schultes, D., Delling, D.: Contraction hierarchies: faster and simpler hierarchical routing in road networks. In: Proceedings of the 7th Workshop on Experimental Algorithms (WEA 08) 5038 of, Lecture Notes in Computer Science (2008), pp. 319–333
2. Geisberger, R., Vetter, C.: Efficient routing in road networks with turn costs. In: Pardalos, P.M., Rebennack, S. (eds.) Experimental Algorithms, pp. 100–111. Springer, Berlin (2011)
3. Nowak, C., Hahne, F., Ambrosi, K.: Contraction hierarchies with A* for digital road maps. In: Klatte, D., Lüthi, H.-J., Schmedders, K. (eds.) Selected papers of the International Conference on Operations Research, pp. 311–316. Springer, Berlin (2012)
4. Schaechterle, K.-H., Braun, J.: Vergleichende Untersuchung vorhandener Verfahren für Verkehrsumlegungen unter Verwendung elektronischer Rechenanlagen, vol. 222 of Forschung Straßenbau und Straßenverkehrstechnik. Typo Verlags-GmbH, Bonn (1977)
5. Schmid, W.: Berechnung kürzester Wege in Straßennetzen mit Wegeverboten. Universität Stuttgart, Stuttgart, Dissertation (2001)

Empirical and Mechanistic Models for Flight Delay Risk Distributions

Lucian Ionescu, Claus Gwiggner and Natalia Kliewer

1 Introduction

In transportation planning and scheduling one often has to deal with the issue of disruptions. Exemplary causes are bad weather conditions, technical failure, and congestion. These disruptions lead to delayed departures. In air transportation, these are called primary delays. In the sequel, insufficient buffer times cause propagation of primary delays on consecutive flights. In this paper we refer to this kind of delays as secondary or propagated delay. To prevent propagations, resource scheduling not only has to deal with low planning costs but also with the allocation of buffer times in a cost-efficient way. Here, one main question is the prediction of primary delays, so that buffer times can be scheduled where they are needed most.

Related work includes stochastic optimization for robust scheduling [1, 2] including the use of empirical delay distributions as inputs for their algorithms, e.g. [3]. Empirical results can be found in [4], where turnaround processes were analyzed and [5], where propagated delays were first simulated and then analyzed.

In this context this paper deals with an analysis of delay data of Lufthansa AG for a time span of four years with about 2.2 million flights. The main variables are fleet-type, day-time, delay reasons, season and departure and arrival airports. The purpose of the analysis is the understanding of primary delays, their describing parameters and propagation effects.

We give a first insight to the project by analyzing two important issues. The first one is the impact of airport congestion factors to delays as described in Chap. 2. Chapter 3

L. Ionescu (✉) · C. Gwiggner · N. Kliewer
Information Systems, Freie Universität Berlin, Garystr. 21,14195 Berlin, Germany
e-mail: lucian.ionescu@fu-berlin.de

C. Gwiggner
e-mail: claus.gwiggner@fu-berlin.de

N. Kliewer
e-mail: natalia.kliewer@fu-berlin.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_86,

deals with delay propagation and absorption regarding incoming and outgoing delays of consecutive flights in a rotation. We want to find out if there are dependencies between incoming and outgoing delays for delay reasons other than rotation delays. Finally we describe a mechanistic model for predictions of ground delays regarding critical paths during the turnaround process in Chap. 4.

2 Airport Utilization and Delays

We define airport utilization as the relative departure flow at an airport. Let λ_i be the number of departures in time interval i and μ the maximum λ over all time intervals. Then the utilization $\rho_i = \lambda_i/\mu$ is an indicator for the congestion at this airport. Moreover, let d_ρ be the average departure delay and p_ρ the probability of being delayed under utilization ρ.

The first graph of Fig. 1 shows the average departure delay and the number of delayed flights per hour, plotted against utilization. Blue lines describe the morning hours (5:00–9:00) and red lines the evening hours (18:00–23:00). In the morning the average delay increases with higher utilization. However, in the evening, the average delay only increases until an utilization level of about $\rho \sim 0.55$, then decreases. The 75 %—quantile is proportional to the average delay, although there are peaks at medium congestion levels (dotted lines). The ratio of the quantile value and the mean value is constant near 1.40 for all utilization levels. Thus, the data is not scattered with

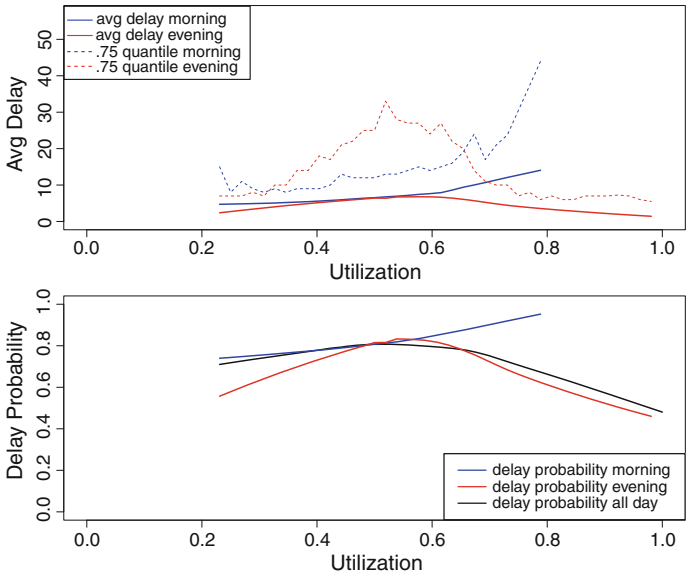


Fig. 1 Delays per utilization levels

constant variance around the conditional mean, as it would be required in standard regression models.

In the same way the second graph of Fig. 1 plots the delay probability against utilization. The same effect as above can be observed—a decreasing delay probability in the evening hours with high utilization. Additionally we show the delay probability over the whole day (black line). It becomes apparent that the highest utilization levels appear during the evening.

In general it is expected that delays increase with higher utilization due to queuing issues. This is clearly violated by our observations during the evening hours. One reason can be night flying restrictions, another might be longer buffer times during the evening. Yet another might be our definition of utilization, which includes less variation than the corresponding quantity from queueing theory.

3 Incoming Versus Outgoing Delays

In this section we check the dependencies between incoming and outgoing delays for several delay reasons at the hub airport in Frankfurt. An incoming delay is defined as the difference between the scheduled and actual arrival time of a flight. An outgoing delay describes the difference between the scheduled and actual departure time of the consecutive flight in a rotation. We expect that primary delay occurs independently of incoming delay. As an indicator for possible dependencies we use a Chi-Square test.

Figure 2 shows incoming and outgoing delays of consecutive flights in a rotation with turnaround in Frankfurt. The black lines are the 45-degree-line and the interpolated average. The upper left panel shows outgoing rotation delays (IATA Delay Code 93). In general, these marks are on or directly below the 45-degree-line due to the effect of delay absorption due to buffer times. A dependency is given by definition and a Chi-Square test indicates this dependency of incoming and outgoing delays ($\chi^2 = 3740.7$, df $= 36$, $p < 2.2 \times 10^{-16}$). The upper right panel shows reactionary delays by reasons other than rotation, e.g. by crew changing. Interestingly, the average outgoing delay seems to be independent of incoming delay, but the independence hypothesis in distribution has to be rejected again ($\chi^2 = 70.88$, df $= 36$, $p < 0.0005$). Turnaround delays (lower left panel) are independent on incoming delay ($\chi^2 = 14.06$, df $= 9$, $p < 0.12$). In the case of ATFM delays (lower right panel) the discretization has an impact on the null hypothesis ($\chi^2 = 56.46$, df $= 36$, $p < 0.016$ and $\chi^2 = 33.45$, df $= 9$, $p < 0.0001$).

In conclusion, rotation delays show a self-explanatory dependency. All other reasons are expected to be independent, but our tests are not yet conclusive. This is even more interesting as rotation delays are always recorded first, so that they dominate other resources that connect the considered flight pairs. Further investigation has to be made on this effect.

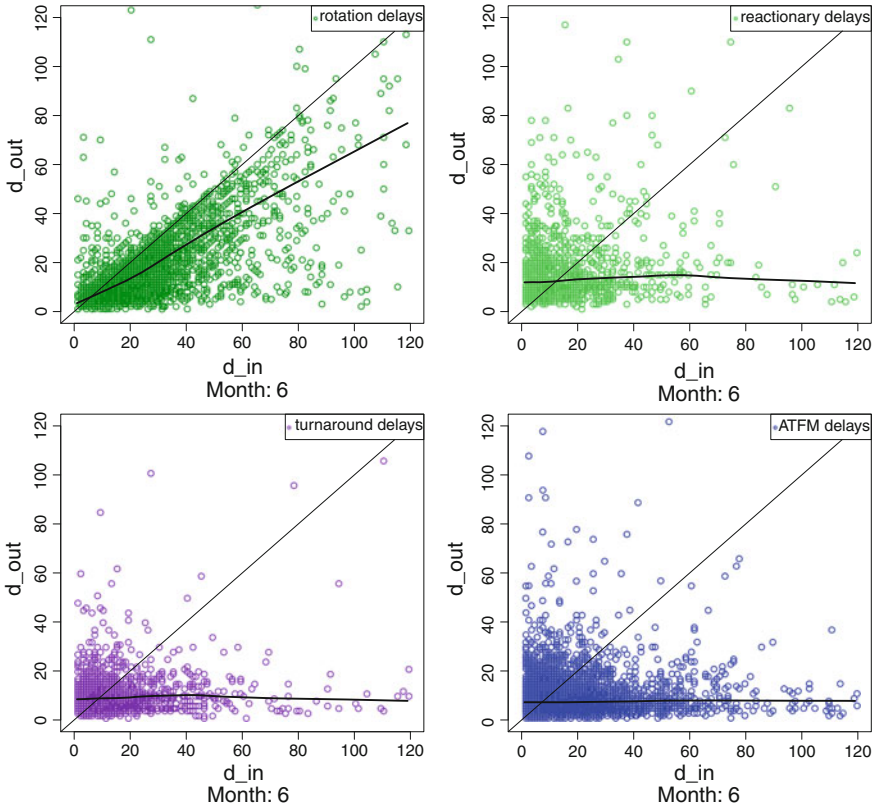


Fig. 2 Incoming versus outgoing delays in FRA

4 Mechanistic Model

The period in which an aircraft is at the gate is called turnaround. During a turnaround, about 15 activities, such as loading and unloading, are executed either sequentially or in parallel. The sequence of activities that determines the earliest possible completion for the turnaround is called the critical path. In general the activity durations depend on external factors and on each other: when one activity is delayed, it is likely that the following ones will be accelerated. As a consequence, the critical path for turnarounds cannot be determined a priori. But its essential components were identified to be de-boarding, one of cleaning, fuelling and catering, followed by boarding [4, 6].

More formally, for aircraft i we have

$$TA_i = Db_i + max \begin{pmatrix} Cl_i \\ Ct_i \\ Fl_i \end{pmatrix} + Bd_i \tag{1}$$

$$d_i = TA_i - TS_i, \tag{2}$$

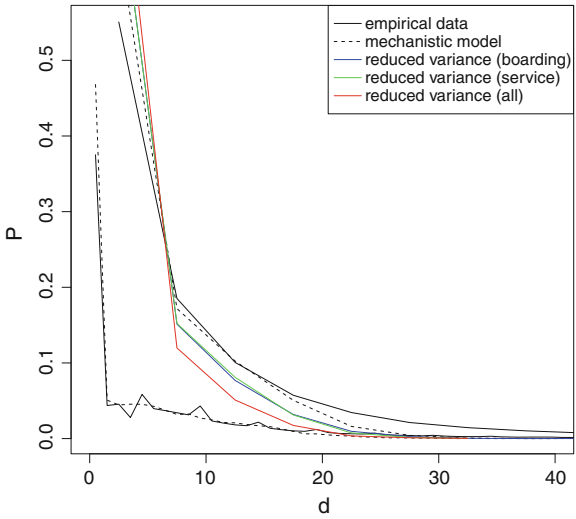


Fig. 3 Delay: Mechanistic model

where Db, Cl, Ct, Fl, Bd are random variables, TS is the scheduled duration of the turnaround and d is the departure delay.

In [4], Weibull and Gamma distributions were identified as reasonable candidates for the components. Both distributions occur frequently in reliability applications, for example when the probability of failure varies with time or the duration of an activity depends on the shortest duration of a large number of events [7]. As a first step we choose

$$X(\theta_x, \theta_e) \sim Weibull(\alpha(\theta_x, \theta_e), \beta(\theta_x, \theta_e)) \tag{3}$$

for all variables $X \in \{Db, Cl, Ct, Fl, Bd\}$. The parameter θ_x signifies that the distribution depends on external conditions, including the patterns discovered in the previous chapters. The parameter θ_e says that the distribution also depends on the other components. In the latter case, the resulting joint distribution of the critical path has to be modelled. The case for external dependencies is more simple.

Anyhow, let us now assume that once an aircraft arrived punctually, the durations of the activities on the critical path do not depend on each other. Moreover, we assume that the durations do not depend on the aircraft itself. This simple model already yields good results, as can be seen in Fig. 3. The bold lines show the empirical delay with binsize 1 and 5 min respectively. The black dotted lines are from Monte Carlo Simulations with ad-hoc parameter estimations. Except for the tail, the curves agree well. The fat tail is an indication for the exo- and endogenous dependencies. The blue dotted lines are simulations with 10 % reduced variances for de-boarding and boarding. The green ones are for cleaning, fuelling and catering respectively. The red line is the result of reductions of all component variances by 10 %. As expected, the resulting delay decreases.

5 Outlook

In this paper we presented new results on the analysis of aircraft delay data. This is useful for realistic robust scheduling systems. We first showed that during the evening hours, delays do not behave as expected by queueing theory. We then investigated dependency between incoming delay and the most important delay categories. Weak dependencies seem to exist, that need more detailed study for clarification. In the final part of the paper we developed a mechanistic model of turnaround delays. We used average conditions of the external factors. Once the effects in the previous chapters are fully understood, we can integrate them in our model, leading to a deeper understanding and to the possibility to predict situations that cannot be observed (as illustrated by the impact of variance reduction). Last but not least, similar effects are expected in other transportation systems, such as public transport.

References

1. Yen, Joyce W., Birge, John R.: A stochastic programming approach to the airline crew scheduling problem. Transp. Sci. **40**(1), 3–14 (2006)
2. Dück, Viktor, Ionescu, Lucian, Kliewer, Natalia: Increasing stability of crew and aircraft schedules. Transp. Res. Part C **20**, 47–61 (2012)
3. AhmadBeygi, S., Cohn, A., Lapp, M.: Decreasing airline delay propagation by re-allocating scheduled slack. IIE Trans. **42**(7), 478–489 (2010)
4. Fricke, H., Schultz, M.: Delay impacts onto turnaround performance. In: 8th USA/Europe ATM R&D Seminar. Napa. CA. US (2009)
5. AhmadBeygi, S., Cohn, A., Guan, Y., Belobaba, P.: Analysis of the potential for delay propagation in passenger airline networks. J. Air Trans. Manage. **14**, 221–236 (2008)
6. A380-Airplane characteristics for Airport planning AC. Airbus (2005)
7. Nikulin, M., Limnios, N. (eds.): Recent advances in Reliability Theory: Methodology, Practice and Inference, pp. 477–492. Birkhäuser, Boston (2000)

Energy-Optimized Routing of Electric Vehicles in Urban Delivery Systems

Henning Preis, Stefan Frank and Karl Nachtigall

1 Introduction

Forced by the german government, the keyword *Electromobility* encapsulates different efforts for sustainable transport systems. Main objectives are the diversification of energy resources and the reduction of the negative impacts, such as noise and air pollution in urban areas. One focus is set on the development and integration of battery electric vehicles (BEV). Along with technical innovations in terms of battery performance and public charging infrastructure appropriate operating are required [6].

Routing desicions for BEV are difficult issues due to a large number of influencing factors, especially the restricted energy supply [1]. In commercial delivery systems, where many customers must be delivered with a number of vehicles, the problem extends to a variant of the Vehicle Routing Problem (VRP), which has been extensively studied for a few decades [9]. Among a fast variety of model extensions there are a few considering aspects connected with electric vehicle operation and energy consumption. For example Kara et al. [4], Kuo [5] and Xiao et al. [11] propose adapted objective functions for minimizing fuel consumption depending on vehicle weight and payload. Further Erdogan et al. [2] and Schneider et al. [8] introduce distance minimization models with additional range restrictions and refueling on a given infrastructure.

H. Preis (✉) · S. Frank · K. Nachtigall
Institute for Logistics and Aviation, Dresden University of Technology,
01062 Dresden, Germany
e-mail: henning.preis@tu-dresden.de

S. Frank
e-mail: stefan.frank@tu-dresden.de

K. Nachtigall
e-mail: karl.nachtigall@tu-dresden.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_87,

2 Calculation of Energy Consumption

The amount of energy that a fleet of electric vehicles consumes while serving the customers is directly related with operational costs. In connection with storage capacity (i.e. battery capacity) it is also responsible for range restrictions. To provide reasonable schedules it is necessary to estimate the anticipated energy consumption, which depends on the driving resistances, i.e. rolling resistance, air resistance, gradient resistance and resistance to acceleration [10]. Excluding the last, these resistances can be calculated without knowledge of driving behavior. For an arc between two vertices i and j with the distance s_{ij} and an altitude difference h_{ij} we determine the energy consumption E_{ij}^F of the empty vehicle with a tare weight m_F, a frontal area A and a drag coefficient c_w driving at velocity v, air density ρ, with a coefficient of friction μ and gravity g as follows:

$$E_{ij}^F = m_F g(h_{ij} + s_{ij}\mu) + \frac{1}{2}\sqrt{h_{ij}^2 + s_{ij}^2}\, c_w A \rho v^2 \tag{1}$$

The additional energy consumption E_{ij}^L caused by the payload depends on the demands of the customers that will be allocated to the vehicles. Therefore we calculate a specific energy consumption per kilogram, which must be multiplied with the assigned loading weight:

$$E_{ij}^L = g(h_{ij} + s_{ij}\mu) \tag{2}$$

Following this, energy can be recuperated at descending sections, which depends on the technical implementation of the vehicles. We considered this by means of a recuperation factor κ. The energy consumption used in the optimization model is

$$c_{ij}^F = max(E_{ij}^F, \kappa E_{ij}^F) \tag{3}$$

respectively,

$$c_{ij}^L = max(E_{ij}^L, \kappa E_{ij}^L) \tag{4}$$

Driving behavior and special charcateristics that affect energy consumption at single arcs $(i, j) \in E$ may be included by adequate penalty factors.

3 Mixed Integer Programming Formulations

The optimization problem is defined on a directed, complete graph $G = (V, E, c)$, where the vertex set V consists of the depot $D = \{0\}$, the set of customers $K = \{1, 2, ..., n\}$ and the set of charging stops $L = \{n + 1, n + 2, ..., n + p\}$, which must be defined for multiple use of charging stations by different vehicles. All customers $i \in K$ are associated with demands b_i, service times t_i^S and service time

windows $\left[t_i^b, t_i^e\right]$. With each arc $(i, j) \in E$ we associate an energy consumption of the empty vehicle c_{ij}^F and a component c_{ij}^L depending on loading weight as described above.

We propose a two-index formulation involving the binary variables x_{ij} equal to 1 if arc $(i, j) \in E$ is used by a vehicle and 0 otherwise and the real-valued variables m_{ij} representing the allocated loading weight. Furthermore we define t_i to be the arrival time at vertex i and e_i as the remaining energy level on arrival at vertex i. The number of vehicles of the homogenous fleet is not restricted. The cargo capacity of a vehicle is given by CAP and the battery capacity by BAT. Vehicles leave the depot with a fully charged battery and can only be recharged at charging stops within a constant service time. The maximum duration until the arrival at the depot is denoted as T_{max}. The mathematical formulation of the Energy-optimizing-VRP is:

$$\text{Minimize} \quad \sum_{i \in V} \sum_{j \in V} \left(c_{ij}^F \, x_{ij} + c_{ij}^L \, m_{ij} \right) \tag{5}$$

subject to

$$\sum_{i \in V; i \neq j} x_{ij} = 1 \qquad \forall j \in K \tag{6}$$

$$\sum_{i \in V; i \neq j} x_{ij} \leq 1 \qquad \forall j \in L \tag{7}$$

$$\sum_{i \in V; i \neq j} x_{ij} - \sum_{i \in V; i \neq j} x_{ji} = 0 \qquad \forall j \in V \tag{8}$$

$$\sum_{i \in V; i \neq j} m_{ij} - \sum_{i \in V; i \neq j} m_{ji} = b_j \qquad \forall j \in V \backslash D \tag{9}$$

$$0 \leq m_{ij} \leq x_{ij} \, CAP \qquad \forall (i, j) \in E; i \neq j \tag{10}$$

$$0 \leq t_i \leq t_j - \left(t_i^S + t_{ij}^F \right) x_{ij} + T_{max} \left(1 - x_{ij} \right) \qquad \forall i \in V; j \in V \backslash D \tag{11}$$

$$t_i + \left(t_i^S + t_{i0}^F \right) x_{i0} \leq T_{max} \qquad \forall i \in V \backslash D \tag{12}$$

$$t_i^b \leq t_i \leq t_i^e \qquad \forall i \in K \tag{13}$$

$$e_j \leq e_i - c_{ij}^F \, x_{ij} - c_{ij}^L \, m_{ij} + BAT \left(1 - x_{ij} \right) \qquad \forall i \in K; j \in V \tag{14}$$

$$e_j \leq BAT - c_{ij}^F \, x_{ij} - c_{ij}^L m_{ij} \qquad \forall i \in V \backslash K; j \in V \tag{15}$$

$$0 \leq e_i \leq BAT \qquad \forall i \in V \tag{16}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i, j \in V \tag{17}$$

$$m_{ij} \in \mathbb{R} \qquad \forall i, j \in V \tag{18}$$

$$t_i \in \mathbb{R} \qquad \forall i \in V \tag{19}$$

$$e_i \in \mathbb{R} \qquad \forall i \in V \tag{20}$$

In this formulation, the objective function (5) minimizes the total energy consumption. Constraints (6) and (7) ensure, that each customer is served exactly once, while charging stops may be executed once. Constraints (8) impose flow conservation at all vertices. Constraints (9) indicate feasible sequences of cargo weight which satisfy the demands of the customers. Constraints (10) force, that only used arcs are allocated with payload, which is not allowed to exceed cargo capacity. Constraints (11)–(13) guarantee the compliance of service time windows, maximum route duration and furthermore the prevention of subtours. Constraints (14)–(16) define feasible energy levels which includes recharging at charging stops and restricts recuperation up to battery capacity.

4 Adapted Tabu Search Heuristics

A solution S for the vehicle routing problem is given by a set M of routes T_r, where each route contains a sequence of vertices $T_r = (0, ..., v_{r_i}, v_{r_j}, ..., 0)$ starting and ending at the depot and each vertex belongs to exactly one route. Each solution is connected with an objective value calcultated by

$$F(S) = \sum_{M \setminus Q} \sum_{(v_{r_i}, v_{r_j}) \in T_r} \left(c^F_{v_{r_i}, v_{r_j}} + c^L_{v_{r_i}, v_{r_j}} m_{v_{r_i}, v_{r_j}} \right) \tag{21}$$

where Q is a set of dummy routes only consisting of charging stations and depot, that is excepted from the objective function. Additionally, we calculate an extended objective function $F_+(S)$, that will be used in the search process. Here we add different penalty terms for violations of maximum route duration $P_{dur}(S)$, loading restraints $P_{load}(S)$, service time windows $P_{tw}(S)$ and energy restraints $P_{en}(S)$ weighted by coefficients, that will be updated during the search process

$$F_+(S) = F(S) + \alpha P_{dur}(S) + \beta P_{load}(S) + \gamma P_{tw}(S) + \delta P_{en}(S) \tag{22}$$

The searching procedure follows well known tabu search strategies reported by Gendreau et al. [3] and Osman [7], which provide a proven solution capability. F^* denotes the best value of $F(S)$ found in the search process so far. S^* is the related best solution. The best neighbor N^* of a solution S is determined by the best value F^*_+ of the extended objective function. The neighborhood of a current solution is defined by the *Relocate* Move, that removes one vertex from its original position and inserts it best possible into any other or the same route. The search procedure is guided by the paramters τ for the number of tabu declared moves , it_{max} for maximum number of non-improving iterations, the factor p for penalizing repetitions z of performed moves and a factor f for appropriate weigthing of violated contraints. The search routine involves the following successive processing steps:

Step (0): *Initial Solution*: Generate S by a savings heuristics, that starts with single routes for all vertices and merges them successivly without violating any constraints; set $F^* = F(S)$

Step (1): *Initialization Tabu Search*: Set paramters τ, it_{max}, p and f for controlling the search process; set number of iteration $it := 1$; set $\alpha := \beta := \gamma := \delta := 1$

Step (2): *Search*: Identify neighbor $N^*(S)$ with the lowest value of $(F_+^* + pz)$ reached by a *Relocate* move, that is not tabu or is tabu, but improves F^*

Step (3): *Update*: if $F_+^* < F^*$ and $N^*(S)$ is feasible then set $S^* := N^*(S)$, $F^* := F_+^*$ and set $it := 1$ else increase it; update the list of tabu declared moves; increase z for the performed move; recalculate α, β, γ and δ (e.g. $\alpha := \frac{\alpha}{f}$ if duration constraint is violated else $\alpha := \alpha f$)

Step (4): *Termination*: if $it > it_{max}$ than terminate else set $S := N^*(S)$ and go to step (2).

5 Illustrative Examples and Computational Results

To illustrate the problem we generated 160 test instances for a circular service area with a diameter of 10 km shaped as a basin. The one and only depot is situated in the center at an altitude of 0 m. Altitudes of randomly situated customers follow a sine function with maximum value at the edge of the service area. Different altitude scenarios are modelled by increasing the amplitude in steps of 200 m. The number of customers variates from 10 to 100, whereas the number of charging stations is fixed to 3. Service time windows of 2 h are randomly assigned to 50 % of the customers. We assume a homogenous fleet with a not restricted number of vehicles having an empty weight of 2 tons and carrying random customer demands up to 200 kg

Table 1 Energy consumption in (kWh) and changes in (%) compared to VRPTW

Max. altitude difference	0 m			200 m			400 m			600 m		
	Min.	Av.	Max.	Min.	Av.	Max.	Min.	Av.	Max.	Min.	Av.	Max.
10 customers	1.27	1.87	2.20	2.56	3.22	4.48	3.02	3.99	4.94	3.94	5.93	6.98
25 customers	2.93	3.43	3.89	4.48	6.06	6.91	7.63	9.60	11.76	11.03	12.58	14.54
50 customers	5.30	5.83	6.15	9.28	11.81	13.07	14.11	17.40	21.14	21.42	23.99	26.64
100 customers	8.89	9.99	10.95	17.88	20.32	22.86	29.65	32.98	37.13	39.89	44.32	46.94
Average energy saving in %	3.75			12.86			20.15				22.55	
Average increase distance in %	3.63			12.77			18.92				29.56	

with a maximum payload of 1.5 tons. Vehicle speed is assumed with 15 km/h. The recuperation factor is 0.1. All instances are solved by the introduced tabu search heuristics with parameter setting (n, 3000, 0.2, 1.05), which has been exposed to work well. Additionally, all instances with 10 customers and 25 customers are solved by a MIP-solver. The gaps between the exact solution and the approximation produced by the heuristics are within a range up to 8 %, on average 2.5 %. Table 1 shows the absolute objective values as well as the trade off between energy-based optimization and distance-based optimization, which has been calculated by a commonly used VRPTW [9] with the same tabu search framework. The energy savings grow linearly with the maximum altitude differences, where the generated detours increase to nearly the same extent and even more at altitude differences up to 600 m (Table 1).

Furthermore we explored effects of reducing battery capacity for all instances with 25 customers. Cutting the battery capacity to 50 % of the total energy consumption of the scenario did not have impact on recharging needs but generates more balanced routes with an increase by 2 % of total energy consumption. The reduction to 25 % forced the use of charging stations. In average 2.4 charging stops are performed, what generates detours with an increase by 36 % of energy consumption.

References

1. Artmeier, A., et al.: Efficient energy-optimal routing for electric vehicles. Proceedings of the twenty-fifth AAAI conference on artificial intelligence, San Francisco (2011)
2. Erdogan, S., Miller-Hooks, E.: A green vehicle routing problem. Transp. Res. Part E. **48**, 100–114 (2012)
3. Gendreau, M., et al.: A Tabu search heuristic for the vehicle routing problem. Manage. Sci. **40**, 1276–1290 (1994)
4. Kara, I., et al.: Energy minimizing vehicle routing problem. Lect. notes Compute. Sci. **4616**, 62–71 (2007)
5. Kuo, Y.: Using simulated annealing to minimize fuel consumption for the time-dependent vehicle routing problem. Comput. Ind. Eng. **59**, 157–165 (2010)
6. Menrath, M.: Elektrisierende Zukunft auf Deutschlands Straßen - Förderprogramme der Bundesregierung im Bereich Elektromobilität. In: Tagungsband Verkehrswissenschaftliche Tage 2012, Dresden University of Technology, Dresden (2012).
7. Osman, I.H.: Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem. An. Oper. Res. **41**, 421–451 (1993)
8. Schneider, M. et al.: The electric vehicle routing problem with time windows and recharging stations. Working paper (2012).
9. Toth, P., Vigo, D.: An overview of vehicle routing problems. In: Toth, P., Vigo, D.: The vehicle routing problem. Siam, Philadelphia (2002).
10. Woll, T.: Verbrauch und Fahrleistungen. In: Hucho, W.-H.: Aerodynamik des Automobils, Strömungsmechanik, Wärmetechnik, Fahrdynamik, Komfort. Vieweg, Wiesbaden (2005).
11. Xiao, Y., et al.: Development of a fuel consumption optimization model for the capacitated vehicle routing problem. Compute. Oper. Res. **39**, 1419–1431 (2012)

Integrated Network Design and Routing: An Application in the Timber Trade Industry

Julia Rieck and Carsten Ehrenberg

1 Introduction

In transport networks consisting of nodes with surpluses or deficits of products, central transshipment facilities (hubs), and interconnecting links, strategic and tactical network design problems, as well as operational routing problems have to be solved. Strategic network design problems determine the number and location of hubs within the network. Tactical network design problems assign nodes to one or more hubs in order to specify possible transportation paths between pairs of origins and destinations. Operational problems focus on route construction for vehicles moving in the network. Obviously, network design and vehicle routing problems are closely interlinked within the context of the design and management of transport networks. Interrelationships between network design and vehicle routing are particularly important and need to be studied if the different planning problems can be compared at one planning level. In cases when suppliers and customers are known in advance and preassigned routes are consistently traversed (e.g. at large trading companies), routing aspects may be integrated in the hub location planning process. Then, an *integrated network design and routing approach* leads to a more realistic cost estimation and excels general network design methods (cf. [7]).

J. Rieck (✉) · C. Ehrenberg
Operations Research Group, Clausthal University of Technology, Julius-Albert Street 2, Lower Saxony, Germany
e-mail: julia.rieck@tu-clausthal.de

C. Ehrenberg
e-mail: carsten.ehrenberg@tu-clausthal.de

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3_88,

2 Application Area and Problem Specification

The problem under consideration is derived from a real-life application in the timber-trade industry. A timber-trade company picks up wooden products (e.g. beams, pallets, door and window frames) from sawmills or wood-manufacturers, and supplies a number of distributors or wood-processing businesses. Usually, recipients periodically require the same mixture of products (e.g. ply wood, panels, and OSB-boards) for their production or retail activities. Since an increasing number of wood-processing companies works in a just-in-time environment, and tries to avoid extensive stock holding, aggregating products over time and then performing full truckloads from one origin to one destination is not required. Hence, most timber-trade companies provide their pick-up and delivery services using hub-and-spoke networks. Employing this network structure, economies of scale and scope can be realized by consolidating freight through one or more hubs. Since material handling at hub locations is especially difficult for large-sized, heavy, and non-standardized wooden products, the number of intermediate hubs on a transportation path is limited to no more than two. However, routing all freight through a hub is not always the best option. If vehicles are fully loaded with all products required by one or a few adjacent destinations, direct transports can be cost-effective. In practice, customers place recurrent orders for wooden products which allows constructing pick-up and delivery routes in advance, i.e. the routing problem may be considered under a strategic/tactical perspective. The overall network operating costs are composed of fixed costs for establishing and maintaining hub facilities, and variable costs for handling and transporting goods on routes.

Designing efficient transport networks for timber-trade companies involves solving the following three interdependent subproblems (cf. the survey of location-routing problems [5]): (a) How many hub facilities are required and where should they be positioned? (b) Should node-to-node links be directly connected or via one or two hubs? (c) Can nodes be combined to form vehicle routes, where every route starts and ends at an assigned hub?

In the course of this paper, we apply a simultaneous planning approach that incorporates hub location and vehicle routing in a static planning environment, i.e., we study one aggregate, representative planning period and average the data (cf. e.g. [1, 2, 8]). The planning problem is represented as a directed graph $G = (V, A)$. Vertex set $V = I \cup H \cup J$ consists of nodes I for supply points (i.e. places with a surplus of products), nodes H for potential hubs, and nodes J for delivery points (i.e. places with a deficit of products). We assume that $I \cap J = \emptyset$; consequently, nodes that act as both supply and delivery points are duplicated.
For a problem instance with four supply points, two potential hubs, and four delivery points, the resulting transport digraph contains ten vertices (cf. Fig. 1), namely, four supply nodes $i_1, \ldots, i_4$, two hub nodes h_1, h_2, and four delivery nodes $j_1, \ldots, j_4$. The subgraphs consisting of supply points and hubs, as well as hubs and delivery points, are complete. In order to simplify matters, forward and backward arcs between pairs of nodes within the subgraphs are described by edges. To connect the

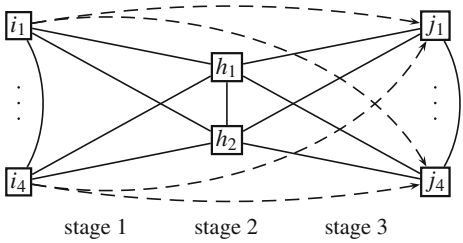


Fig. 1 Transport digraph G

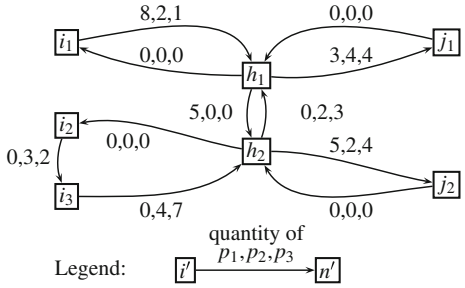
Table 1 Possible vehicle routes

route r_1 (stage 1)	route r_2 (stage 2)	route r_3 (stage 3)
$h - i \ldots i' - h$	$h - h' - h$	$h - j \ldots j' - h$
$h - i \ldots i' - j \ldots j' - h$	h	$h - i \ldots i' - j \ldots j' - h$
h		h

two subgraphs, G contains arcs $\langle i, j \rangle$, $i \in I$, $j \in J$. Each arc is associated with a non-negative weight d_{ij} that represents the distance between nodes i and j. In order to determine the transportation costs of an arc, distances are multiplied by a scaling factor γ [€/km]. If hub h is established, we consider fixed costs $f_h > 0$ for facility construction and management. Based on an existing hub h, a fleet of v_h homogeneous vehicles performs routes in order to serve the supply and delivery points. Each vehicle $k \in K_h$ starts its routes at an assigned hub h and returns to the same hub; $K := \bigcup_{h \in H} K_h$. The capacity C of a vehicle may not be exceeded. Supply points $i \in I$ have a surplus $a_{ip} \geq 0$ and delivery points $j \in J$ a deficit $b_{jp} \geq 0$ of products $p \in P$. We measure the different quantities in abstract transport units, which can be calculated from the dimensions and weights of the products considered. In order to consolidate products at hubs, pick-ups must precede deliveries. Only products that have arrived at a hub can be sorted and consolidated for further transport. We therefore take the existing three-stage network (cf. Fig. 1) into account, and assume that each vehicle traverses at most three routes r_1, r_2, r_3 starting and ending at hub $h \in H$; $R := \{r_1, r_2, r_3\}$.

Table 1 lists all possible vehicle routes ($i, i' \in I$, j, $j' \in J$, h, $h' \in H, h \neq h'$). Interhub routes are assumed to be direct, while all other routes are multi-stop. The first route might be an exclusively pick-up route, or a pick-up and delivery route; the second route will typically be an interhub route, and the third route might be an exclusively delivery route, or a pick-up and delivery route. Furthermore, in all cases, vehicles might be stationary at hub h. By considering one route for each network stage, we obtain $3 \cdot 2 \cdot 3 = 18$ possible routings for the vehicles.

In order to solve the network design and routing problem, we developed a mixed-integer linear programming model that makes use of decision variables $x_{ij}^{kr} \in \{0, 1\}$ to indicate if vehicle k traverses arc $\langle i, j \rangle$ on route r in an optimal solution, and decision



a_{ip}	1	2	3
1	8	2	1
2	0	3	2
3	0	1	5

b_{jp}	1	2	3
1	3	4	4
2	5	2	4

Fig. 2 Feasible solution and supply and delivery quantities

variables $y_h \in \{0, 1\}$ which specify if hub h is established. The following example illustrates the structure of a feasible solution to the specified problem. We assume that the underlying digraph consists of three supply points i_1, i_2, i_3, two delivery points j_1, j_2, and two potential hubs h_1, h_2. Pick-up and delivery quantities for products p_1, p_2, p_3 are listed in Fig. 2 (on the right). Just one vehicle is available at each hub, and its capacity is eleven transport units. Figure 2 depicts a feasible solution to the problem under consideration. The first vehicle traverses three different routes, $r_1 = (h_1, i_1, h_1)$, $r_2 = (h_1, h_2, h_1)$, and $r_3 = (h_1, j_1, h_1)$, and the second vehicle traverses two routes, $r_1 = (h_2, i_2, i_3, h_2)$, and $r_3 = (h_2, j_2, h_2)$. Both hub facilities are established and act as collection and consolidation centers.

3 Computational Results

Test computational runs were conducted on 960 randomly generated problem instances, duly allowing for control parameters in order to reproduce a wide range of practical applications. Since in the timber-trade industry the number of supply points is usually lower than the number of delivery points, we considered outbound network configurations with 12–17 nodes. Further, we constructed eight scenarios, where the variable transportation costs $c := \gamma\, d$, the fixed costs f for establishing hubs, and the tightness T of delivery matrices $(b_{jp})_{j\in J, p\in P}$ vary (cf. Table 2). In doing so, we are able to cope with uncertainties in future developments (i.e. increased energy costs, changes in wages, or modified pick-up and delivery quantities). The cost components are categorized as either small (s) or large (l), and the matrices may have a tight (t) or a non-tight (nt) setting. For each combination of network structure and scenario, six instances are generated.

Usually, the variable transportation costs, which are composed of costs for fuel, vehicle usage, driver wages, and toll, and the fixed costs, that consist of costs for depreciations, wages, and energy, refer to different planning horizons (short-term and long-term). In order to perform a comparative study, the cost components must be determined for one representative planning period. Using one week as reference, we

Table 2 Network configurations and scenarios

Network configurations $\|I\|$ - $\|H\|$ - $\|J\|$				$\|V\|$
2 - 2 - 8		4 - 2 - 6		12
2 - 3 - 8	3 - 2 - 8	4 - 3 - 6	5 - 2 - 6	13
2 - 2 - 10	3 - 3 - 8	4 - 2 - 8	5 - 3 - 6	14
2 - 3 - 10	3 - 2 - 10	4 - 3 - 8	5 - 2 - 8	15
2 - 2 - 12	3 - 3 - 10	4 - 2 - 10	5 - 3 - 8	16
	3 - 2 - 12		5 - 2 - 10	17
Scenarios c - f - T				
s - s - t	s - s - nt			
s - l - t	s - l - nt			
l - s - t	l - s - nt			
l - l - t	l - l - nt			

calculated a transportation cost rate of $c^s := 1.18$ € in the best case and of $c^l := 1.63$ € in the worst case. For the fixed costs, we obtained the values $f^s := 2,696.00$ € and $f^l := 3,135.00$ €. A demand matrix $(b_{jp})_{j\in J, p\in P}$ is denoted as tight if 75–100 % of matrix elements are larger than zero and is denoted as non-tight if 50–75 % of matrix elements are larger than zero (i.e. matrix (b_{jp}) in Fig. 1 is tight).

All problem instances are solved using GAMS/CPLEX 12.4 on an Intel i7-2760QM machine with 8 GB RAM. Since the considered problem is an *NP*–hard optimization problem (cf. [3, 4]), and therefore difficult to solve to optimality, we use a maximum computation time of one hour after which the best solution found is returned. In order to improve the bounding accuracy and accelerate solving, a heuristic upper bound on the objective function is supplied (cf. [6]).

Table 3 shows computational results obtained for different network configurations. In column #opt the numbers of instances solved to optimality within the time limit, in column t_{cpu} the average computation times [seconds], and in column *gap* the relative gap between the best integer solution found and the best lower bound is given. The results show that problem instances consisting of networks with up to 14 nodes and a small $|I|/|J|$–ratio are harder to solve than instances with a large ratio, i.e. an asymmetry in the numbers of supply and delivery points results in larger run times. Notice, that a contrary characteristic can be observed for instances with $|V| \geq 15$ due to an increase in potential routes that contain an arc $\langle i, j\rangle, i \in I, j \in J$. Additionally, instances with a non-tight demand matrix need significant less time to receive an optimum than instances with a tight demand matrix (cf. Table 3 on the bottom).

Based on the consideration of scenarios, we are able to evaluate cost-consequences of a certain network design decision with respect to uncertainty. Generally, a company implements the network structure associated with the scenario that is assumed to be most likely. Then, once hub facilities are established, this decision cannot be revised on a short-time-basis. The only variation possibility may be to re-optimize the vehicle routes for the realized (but sub-optimal) network design. For that reason, we construct

Table 3 Computational results for network configurations and scenarios

Configuration	#opt	t_{cpu}	*gap*	$\|V\|$	$\|I\|/\|J\|$
2 - 2 - 8	48	87.07	–	12	0.25
4 - 2 - 6	48	62.21	–	12	0.67
2 - 3 - 8	48	144.79	–	13	0.25
3 - 2 - 8	48	116.71	–	13	0.38
4 - 3 - 6	48	48.19	–	13	0.67
5 - 2 - 6	48	83.63	–	13	0.83
2 - 2 - 10	42	639.78	0.10	14	0.20
3 - 3 - 8	48	265.92	–	14	0.38
4 - 2 - 8	48	209.03	–	14	0.50
5 - 3 - 6	48	134.02	–	14	0.83
2 - 3 - 10	39	984.50	0.08	15	0.20
3 - 2 - 10	30	1,049.28	0.08	15	0.30
4 - 3 - 8	48	477.36	–	15	0.50
5 - 2 - 8	41	1,329.15	0.05	15	0.63
2 - 2 - 12	23	518.64	0.10	16	0.17
3 - 3 - 10	26	788.08	0.09	16	0.30
4 - 2 - 10	14	1,140.24	0.14	16	0.40
5 - 3 - 8	34	1,684.94	0.06	16	0.63
3 - 2 - 12	8	961.56	0.17	17	0.25
5 - 2 - 10	0	–	0.16	17	0.50

Scenario	#opt	t_{cpu}
s - s - t	82	510.15
s - l - t	84	599.83
l - s - t	80	593.31
l - l - t	85	639.55
s - s - nt	103	340.51
s - l - nt	100	312.85
l - s - nt	100	360.76
l - l - nt	103	449.63

Table 4 Decision matrix and application of minimax rule

Decision/Scenario	*s - s - t*	*s - l - t*	*l - s - t*	*l - l - t*	*s - s - nt*	*s - l - nt*	*l - s - nt*	*l - l - nt*
h_1	10,736	11,175	13,802	**14,241**	9,518	9,957	12,120	12,559
h_2	10,537	10,976	13,527	**13,966**	10,271	10,710	13,160	13,599
h_1, h_2	10,758	11,636	12,805	13,683	10,810	11,688	12,877	**13,755**

a decision matrix, where the reasonable location decisions are confronted with all scenarios (cf. Table 4). Then, an appropriate decision rule that reflects the decision makers risk attitude may be used. Following the minimax rule, hubs h_1 and h_2 should be established for the problem instance depicted in Table 4.

4 Conclusion

We have presented a network design problem that integrates strategic and operational aspects in order to determine the overall network costs. The problem is modeled as a mixed-integer linear program and solved with CPLEX. All instances with $|V| \leq 13$ are optimally solved and for the remaining instances the gap is comparatively small. Future work will include the explicit consideration of problem-specific preprocessing techniques and cutting planes.

References

1. Cunha, C.B., Silva, M.R.: A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil. Eur. J. Oper. Res. **179**, 747–758 (2007)
2. Karaoglan, I., Altiparmak, F., Kara, I., Dengiz, B.: A branch and cut algorithm for the location-routing problem with simultaneous pickup and delivery. Eur. J. Oper. Res. **211**, 318–332 (2011)
3. Lee, Y., Lim, B.H., Park, J.S.: A hub location problem in designing digital data service networks: Lagrangian relaxation approach. Location Sci. **4**, 185–194 (1996)
4. Lenstra, J.K.: Rinnooy Kan, A. H. G.: Complexity of vehicle routing and scheduling problems. Networks **11**, 221–227 (1981)
5. Nagy, G., Sahli, S.: Location-routing: Issues, models and methods. Eur. J. Oper. Res. **177**, 649–672 (2007)
6. Rieck, J., Zimmermann, J.: A hybrid algorithm for vehicle routing of less-than-truckload carriers. In: Geiger, M.J., Habenicht, W., Sevaux, M., Sörensen, K. (eds.) Metaheuristics in the Service Industry, Lecture Notes in Economics and Mathematical Systems 624, pp. 155–171. Springer, Berlin (2009)
7. Sahli, S., Nagy, G.: Consistency and robustness in location-routing. Stud. Locational Anal. **13**, 3–19 (1999)
8. Takano, K., Arai, M.: A genetic algorithm for the hub-and-spoke problem applied to containerized cargo transport. J. Marine Sci. Technol. **14**, 256–274 (2009)

Author Index

S. Helber et al. (eds.), *Operations Research Proceedings 2012*,
Operations Research Proceedings, DOI: 10.1007/978-3-319-00795-3,