Lecture Notes in Logistics
Series Editors: Uwe Clausen · Michael ten Hompel · Robert de Souza

Dirk Mattfeld
Thomas Spengler
Jan Brinkmann
Martin Grunewald *Editors*

Logistics Management

Contributions of the Section Logistics of the German Academic Association for Business Research, 2015, Braunschweig, Germany

Springer

Lecture Notes in Logistics

Series editors

More information about this series at http://www.springer.com/series/11220

Dirk Mattfeld · Thomas Spengler
Jan Brinkmann · Martin Grunewald
Editors

Logistics Management

Contributions of the Section Logistics of the German Academic Association for Business Research, 2015, Braunschweig, Germany

Springer

Editors
Dirk Mattfeld
Decision Support Group, Business Information Systems
University of Braunschweig
Braunschweig
Germany

Jan Brinkmann
Decision Support Group, Business Information Systems
University of Braunschweig
Braunschweig
Germany

Thomas Spengler
Production and Logistics Research Group, Institute of Automotive Management and Industrial Production
University of Braunschweig
Braunschweig
Germany

Martin Grunewald
Production and Logistics Research Group, Institute of Automotive Management and Industrial Production
University of Braunschweig
Braunschweig
Germany

ISSN 2194-8917 ISSN 2194-8925 (electronic)
Lecture Notes in Logistics
ISBN 978-3-319-20862-6 ISBN 978-3-319-20863-3 (eBook)
DOI 10.1007/978-3-319-20863-3

Library of Congress Control Number: 2015943369

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

In the past decades, logistics management turned out to be one of the most important success factors in managing the tremendous challenges of a more and more globalized economy. In industry and science, many new methods and tools for decision support in procurement-, production-, distribution-, resource-, and supply chain management have been developed and applied successfully. In September 2015 the conference on logistics management LM2015 took place at the Technische Universität Braunschweig, Germany, as the ninth event of a biannual logistics management series that started back in 1999. Founded and supervised by the special interest group on logistics in the German Academic Association for Business Research (VHB), this series of conferences focuses on academic achievements with respect to the management of the logistics function of firms.

LM2015 is held in conjunction with VHB's special interest group on production, thus, without loss of variety in the treatment of logistics topics, one can observe a bias to production logistics, i.e., facility layout, inventory management, line configuration, or production planning and scheduling. This emphasis is underlined by the invited keynote of Rainer Lasch and Roy Fritzsche, entitled "Condition-Based Maintenance Planning Within the Aviation Industry." The contribution aims at the increase of aircraft availability due to condition-based maintenance planning as well as the efficient supply of spare parts.

Apart from the aforementioned invited paper the other 19 contributions of this volume have undergone a thorough review by at least three referees each. The condition of acceptance has been set to an above average rating of at least two of the referees involved. Two of these papers address the management of supplies: "Considering Small and Medium-Sized Suppliers in Public Procurement—The Case of the German Defence Sector" discusses the practice of lot-wise calls in German defense procurements. The empirical study "Integration of Cultural Aspects in Supplier Evaluation" focuses on cultural differences in the cooperation between buyer and supplier.

The next three papers are devoted to the coordination of supply chains. "RoRA-SCM—Review of the Integration of Flexibility Planning in Supply Chains" addresses robust, resilient, and agile planning in the light of uncertainty introducing risk and opportunity. The paper surveys optimization models and their integration into the supply chain matrix. "Coordination in Multimodal Supply Chains: Drafting an Agent-Based Exchange" introduces a market place concept for agent-based freight exchange based on bidding auctions. "Flexibility of 3PL Contracts: Practical Evidence and Propositions on the Design of Contract Flexibility in 3PL Relationships" covers flexibility mechanisms in contract design. An analysis is performed in order to identify areas of prospective research.

A bunch of four papers deals with the integrated management of resources within a firm. "Integrated Facility Location, Capacity, and Production Planning in a Multi-Commodity Closed Supply Chain Network" introduces a mixed integer model for forward and reverse flows of multiple make-to-order products. "An Extended Model Formulation of the Facility Layout Problem with Aisle Structure" suggests a mixed integer model capable of handling layout arrangement and aisle structure simultaneously. "Integrated Make-or-Buy and Facility Layout Problem with Multiple Products" tackles a simultaneous make-or-buy decision while considering costs of outsourcing as well as production costs imposed by facility layout. The paper "Qualification and Competence Related Misfits in Logistics Companies: Identification and Measurement" is the only one in this volume addressing the important topic of human resources in a mixed method approach.

Flexible Production Management is considered by three papers. "A Lot Streaming Model for a Re-entrant Flow Shop Scheduling Problem with Missing Operations" examines the impact of increasing the number of sublots for this stochastic scheduling problem. The paper "Identifying Complexity-Inducing Variety: Adapting ClustalW for Semiconductor Industry" adapts the protein multiple sequence alignment program ClustalW for complexity measurement in workflow design. The "Consideration of Redundancies in the Configuration of Automated Flow Lines" performs a numerical analysis on redundant configurations aiming at ensuring throughput rates while avoiding costly buffer space.

Another four papers deal with distribution management. "Prepositioning of Relief Items Under Uncertainty: A Classification of Modeling and Solution Approaches for Disaster Management" combines facility location and inventory management under uncertainty. "Forecasting Misused E-Commerce Consumer Returns" predicts the ratio of returning items already used by customers. This is of particular importance since used items cannot be directly resold anymore. "A Rollout Algorithm for Vehicle Routing with Stochastic Customer Requests" maximizes the number of served requests within the time limit of a work shift. "A Real-World Cost Function Based on Tariffs for Vehicle Routing in the Food Retailing Industry" integrates the price structure based on tariffs as they appear for freight carriers. In particular, compartments of vehicles and time windows for delivery are considered.

The last three papers of this volume are devoted to transport management. "An Adaptive Large Neighborhood Search for the Reverse Open Vehicle Routing Problem with Time Windows" extends the OVRP by time windows and truck specific starting positions. A metaheuristic generates return trips to the depot for the case of dynamic vehicle routing problems. "A Two-Stage Iterative Solution

Approach for Solving a Container Transportation Problem" suggests a mixed integer model for inland container transportation under the assumption that one trucking company owns depots, a fleet, and a sufficient number of empty containers. The paper "Vehicle Routing and Break Scheduling by a Distributed Decision Making Approach" introduces a framework for distributed decision making such that drivers and planners coactively solve the integrated problem.

Acknowledgments

The editors of this volume express their gratitude to all of the authors and to everybody who has contributed to this volume. In particular, we thank Springer for the convenient and seamless collaboration in the editing and publishing process. We gratefully acknowledge the tremendous work carried out by the referees. In particular, we want to express our gratitude to:

Bierwirth, Christian, Martin-Luther-Universität Halle-Wittenberg,
Buscher, Udo, Technische Universität Dresden,
Corsten, Hans, Universität Kaiserslautern,
Darkow, Inga-Lena, Universität Bremen,
Dethloff, Jan, Hochschule Bremen,
Ehmke, Jan Fabian, Freie Universität Berlin,
Elbert, Ralf, Technische Universität Darmstadt,
Eßig, Michael, Universität der Bundeswehr München,
Fische, Kathrin, Technische Universität Hamburg-Harburg,
Fleischmann, Moritz, Universität Mannheim,
Fröhling, Magnus, Karlsruher Institut für Technologie,
Geiger, Martin, Universität der Bundeswehr Hamburg,
Geldermann, Jutta, Universität Göttingen,
Giesa, Frank, Hochschule Bremen,
Gronalt, Manfred, Universität für Bodenkultur Wien,
Grunewald, Martin, Technische Universität Braunschweig,
Grunow, Martin, Technische Universität München,
Günther, Hans-Otto, Technische Universität Berlin,
Gutenschwager, Kai, Ostfalia Hochschule für angewandte Wissenschaften,
Helber, Stefan, Leibnitz Universität Hannover,
Ivanov, Dmitry, Hochschule für Wirtschaft und Recht Berlin,
Jammernegg, Werner, Wirtschaftsuniversität Wien,
Kiesmüller, Gudrun P., Otto von Guericke Universität Magdeburg,
Kimms, Alf, Universität Duisburg-Essen,
Kliewer, Natalia, Freie Universität Berlin,

Kopfer, Herbert, Universität Bremen,
Kotzab, Herbert, Universität Bremen,
Large, Rudolf O., Universität Stuttgart,
Lasch, Rainer, Technische Universität Dresden,
Meisel, Frank, Christian-Albrechts-Universität Kiel,
Minner, Stefan, Technische Universität München,
Otto, Andreas, Universität Regensburg,
Pesch, Erwin, Universität Siegen,
Pflaum, Alexander, Universität Bamberg,
Schimmelpfeng, Katja, Universität Hohenheim,
Schönberger, Jörn, Technische Universität Dresden,
Schultmann, Frank, Karlsruher Institut für Technologie,
Schwindt, Christoph, Technische Universität Clausthal,
Seuring, Stefan, Universität Kassel,
Sucky, Eric, Otto-Friedrich-Universität Bamberg,
Volling, Thomas, Fern-Universität Hagen,
Voß, Stefan, Universität Hamburg,
Wallenburg, Carl Marcus, WHU—Otto Beisheim School of Management,
Walther, Grit, Rheinisch Westfälische Technische Hochschule Aachen,
Wäscher, Gerhard, Otto-von-Guericke-Universität Magdeburg.

Last but not least we also want to thank our scientific collaborators Tatiana Deriyenko, Patrick-Oliver Groß, Christoph Johannes, Felix Köster, Christoph Müller, Viola Ricker, Ninja Söffker, Marlin Ulmer, Katharina Wachter, and Christian Weckenborg for their support in editing the final version of this book.

Braunschweig
September 2015

Dirk Mattfeld
Thomas Spengler
Jan Brinkmann
Martin Grunewald

Contents

Part IV Resource Management

Part V Flexible Production Management

Part VI Distribution Management

Part I
Invited Contribution

Condition-Based Maintenance Planning Within the Aviation Industry

Rainer Lasch and Roy Fritzsche

Abstract Delays or cancellations of flights are mostly caused by avoidable, maintenance-related downtime of aircrafts. For a more efficient supply of spare parts for aircrafts of an airline, we examine options for improving the underlying logistics network integrated within an existing segment of the aviation industry. The objective is to guarantee a high supply of spare parts. Moreover, the paper presents effects of the choice of different maintenance strategies on total maintenance costs. By applying a condition-based maintenance strategy, unscheduled component failures can be reduced and an increase in availability and reliability is possible. A validation of the developed model is provided by means of a simulation study.

1 Introduction

Networks in the aviation industry consist of spare parts warehouses, repair bases, flight plans, and aircrafts. The major requirements for each airline operator are aircraft availability and operability, as well as reliability levels and product quality (Wang et al. 2007; Lee et al. 2008; Papakostas et al. 2009). Therefore, operators and decision makers are confronted with significant challenges to reduce their equipment maintenance costs as much as possible while providing excellent service without disruptions. To achieve high reliability of the entire fleet, condition-based maintenance strategies based on a proactive prediction of a possible failure in the future are necessary. The increasing interest in optimal maintenance strategies is influenced by rising costs, improved quality of spare parts, and an increasing pressure for reduced inventories of spare parts (Wheeler et al. 2010; Ayeni et al. 2011). Especially the penalty costs caused by delays, unnecessary downtime, and failure of the aircrafts continue to increase. To avoid these idle times, one of the

R. Lasch (✉) · R. Fritzsche
Lehrstuhl für Betriebswirtschaftslehre, insb. Logistik, Technische Universität Dresden, Münchner Platz 1/3, 01069 Dresden, Germany
e-mail: rainer.lasch@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_1

objectives of the airline network designers is to transfer unscheduled maintenance actions to scheduled variable actions through a simplified logistics planning.

The objectives of high reliability and availability of spare parts for the entire fleet can be realized by an improved prediction of failure times of components, i.e. an increase in the probability that the correct spare part is available at the right place and at the right time. The basis of good failure prognosis values is a comprehensive database which consists of a collection of historical failure data and measured sensor data. Prognostics is now recognized as a key feature in maintenance strategies as it should allow for the avoidance of inopportune maintenance spending. With the help of better prognosis, the component's useful lifetime is maximized by adjusting the failure rate downwards dynamically and extending the component's service lifetime significantly (Fritzsche 2012). The developed model uses the dynamic concept of maintenance-free operating period (MFOP) instead of the older and static mean time between failures (MTBF) concept to prevent future failures. MFOP is defined as a measurement for (aircraft) reliability (Kumar 1999). In this paper we want to present significant effects of chosen maintenance strategies on downtime, delay, and cancellation costs, and with it, on total maintenance costs in aviation industry maintenance planning. With the application of an improved (dynamically adjusted failure rate) condition-based maintenance strategy, total maintenance cost is minimized by decreasing downtime and delay time.

The remainder of this paper is organized as follows: In Sect. 2, literature related to preventive maintenance models is examined with respect to integrated logistics networks. To show the significant influences of the chosen maintenance policy on the availability of the entire network, we discuss all related cost functions and their constraints in Sect. 3. Furthermore, we identify cost influenced variables through the comparison of three different maintenance strategies. The aspired objective, continuous availability of spare parts at lowest cost, is examined using a simulation tool in Sect. 4. A conclusion, limitations of the model, and future research are presented in Sect. 5.

2 Literature Review

Maintenance is defined as all technical and managerial actions taken during the usage period to maintain or restore the required functionality of a product or an asset. Maintenance strategies can be divided into corrective (reactive/unscheduled) and preventive (scheduled) maintenance (Yardley et al. 2006). In corrective maintenance, the maintenance action is taken after some problems such as breakdowns in a product are found; in preventive maintenance, the performance of inspection and/or service activities are pre-planned in order to restore the functions of operating systems or equipment at a specific point in time. Preventive maintenance can be differentiated into time-based and condition-based maintenance

(CBM). While parts are replaced after a fixed time interval in the time-based methods, for condition-based methods an optimal replacement time is prognosticated, based on past data and measured sensor-state data or goal programing (Moghaddam 2013). With emerging technologies such as RFID, various types of sensors, micro-electro-mechanical systems, wireless telecommunication, and supervisory control and data acquisition product embedded information devices are expected to be used in the near future for gathering and monitoring the status data of products during their usage period (Prajapati et al. 2012). These data enable the diagnosis of the product's degradation status in a more exact way. CBM focuses on the prediction of the degradation process of the product, which is based on the assumption that most abnormalities do not occur instantaneously, and usually there are several kinds of degradation processes, from normal states to abnormalities. CBM provides the ability for the system to continue operating as long as it is performing within predefined performance limits. Based on condition-based information, maintenance will be performed only when the system needs maintenance and enables to identify and solve problems in advance before product damage occurs. In order to reduce downtime, delays, and cancellations of scheduled flights, a CBM strategy with predictions of installed components' failure times using the MFOP approach seems highly promising. In CBM a data-driven, model-based, or hybrid approach can be used. While the data-driven approach uses historical data to automatically develop a model of system behavior, the model-based approach has the ability to incorporate physical understanding of the target product and relies on the use of an analytical model to represent the behavior of the system, including degradation (Tobon-Mejia et al. 2012). The CBM can be done by gathering and monitoring product status data, making a diagnosis of a product status in a real-time way, estimating the deterioration level of the product including its repair cost or its replacement cost, predicting the time of the product's abnormality, and executing appropriate actions (e.g. repair, replace, left to use).

Integrated logistics models can be classified into (1) location and network models (Daskin 1995; Candas and Kutanoglu 2007; Snyder et al. 2007; Sourirajan et al. 2009; Jeet et al. 2009), (2) inventory models (Feeney and Sherbrooke 1966; Zipkin 2000; Sherbrooke 2004; Muckstadt 2005), and (3) integrated network models (Barahona and Jensen 1998; Nozick and Turnquist 2001). These existing models refer only to specific sub-problems, are too inflexible, and do not consider the overall view of the network. Furthermore, no inclusion of flight plan, location allocation, and dynamic failure rate adjustment is possible. The model presented in Sect. 3 fits best for aviation industry requirements to reduce total maintenance cost. The applied CBM strategy models spare part demand dynamically by adjusting the component's failure rate. Furthermore, a single socket is considered, all locations are regarded as identical, and it is possible to store spare parts at any location.

3 Model Description

Integrated logistics networks can be divided into (1) location selection (discrete, continuous), (2) objective function (minisum, minimax), (3) variables (binary, integer, mixed-integer), (4) cost consideration (no cost, variable cost, fixed cost), and (5) solution method (exact, heuristics, metaheuristics, simulation) (Kaviani 2009). Based on this classification scheme our developed model can be characterized as follows. The location selection is discrete because of the given flight plan. The objective function is a summation of total cost to be minimized. In the model, binary and integer variables are possible. Fixed costs are necessary to open a spare parts depot at a location. We use a heuristic approach for the initial solution and simulation-based optimizing to solve the problem. The integration of the parameters in a simulation-based optimization is implemented in the model by adjusting the MFOP calculated failure rates of the installed components. The adaptation of the failure rate is possible through the use of the learning effect of the MFOP method. MFOP guarantees a certain number of periods of operation without any interruption for unscheduled maintenance. Each MFOP period is followed by a maintenance recovery period, where the aircraft is repaired and prepared to complete the next MFOP period. Consequently, the optimization of durability of a working component is necessary and is implemented by the MFOP method (Relf 1999). Based on this concept, unscheduled maintenance is changed to scheduled maintenance.

3.1 Operational Requirements in the Aviation Industry

Necessary conditions for logistics networks are presented in the following (Candas and Kutanoglu 2007; Lee et al. 2008; Jeet et al. 2009).

1. General logistical requirements:

   - Two-echelon model: The model corresponds to a two-level model with depots and demand points
   - Potential depots: All airports in the network run potential spare parts depots
   - High-value parts: Ostensibly, high inventory costs in networks of the aviation industry are caused by high-quality and less demanded spare parts
   - Repair capacity: An infinite repair capacity of the hubs and the manufacturers of new parts is assumed
   - Repairable items: The considered components belong to serviceable units and they pass a repair cycle in case of a failure
   - Single item model: The proposed model is a single part model, but by repeated application of the algorithm multiple items could be examined
   - Multi-sourcing strategy: Depots can be served from other depots by lateral transshipments

- (S-1, S) Order policy: Because of low demand the replacement is done by a one-for-one policy
- Lost-sales case: For non-compliance, an order is considered as lost

2. Failure rate (Demand rate): The failure rate of installed components is initially assumed to be Poisson distributed with the possibility for adjustment based on excellent underlying prognosis data.
3. Fill rate: The fill rate is the percentage of all demanded spare parts that can be covered by the existing stock. It is therefore dependent on the existing stock of the depot and the associated demand.
4. Warehouse stock and inventory costs: The initial/safety stock will not be allocated deterministically to the depots. It is a decision variable in the model which can be adjusted during the optimization.
5. Transportation costs: Transportation costs for lateral transshipments of spare parts are assumed to be fixed.
6. Fixed costs for depot opening: The costs for opening a new depot are assumed to be fixed and identical for any location.

3.2 Input Parameters

For the comparison of the effects of different maintenance strategies on total maintenance costs in the airline industry, several input parameters are defined. The complete maintenance planning process is divided into the areas (1) airport/flight (departure, flight, and landing of the aircraft), (2) maintenance (repair activities, use of manpower, replacing defective components), and (3) logistics network (distribution and stock level of spare parts, sourcing non-available parts, cost-efficient lateral transshipments between warehouses). In the following the input parameters for the three areas are discussed:

1. Airport/Flight: To reduce extended downtime for components' replacement, the flight plan (including departure and arrival time, departure and arrival location) and the times and ranges of the A, B, C, and D checks should be coordinated. Based on the manufacturer's instructions about the lifetime of a component (later sensor data is used), a degradation of service life of installed components is performed.
2. Maintenance: To calculate an aircraft's turnaround time, it is necessary to know the waiting and transfer times for repair of the aircraft, downtime, and repair time of spare parts for A, B, C, D checks, and the installation times for parts. To generate inexpensive spare parts, the exchange repair cost, rental cost for repair facilities, personnel costs, and costs of opening a location are required.
3. Logistics Network: For optimization of the logistics activities, most input data is required here. This mainly includes inventory holding costs of the depots. For an improved spare parts pooling, the transfer times and costs of the components

(from a warehouse to the airport/repair facility, or between two main bases) are very important. In order to minimize the total maintenance costs of an airline, downtime, delay, and cancellation costs are required. Furthermore, costs to balance the fill rates, distribution costs of repaired components, penalty costs for unscheduled maintenance, and incorrect deliveries and maintenance costs are important. For a redistribution of spare parts in the network their price as well as the location's safety stock (for backfill) are required.

The MFOP concept is based on very high quality information about the component failures in order to make optimal decisions on the remaining lifetimes. The used Weibull distribution is suitable to describe all three phases of life (premature failures, random failures, and wear failures). In the known CBM strategies, an age-independent Poisson distributed and constant failure rate (based on MTBF) is assumed (Kutanoglu 2008). The CBM strategy presented here uses an adjusted failure rate based on measured sensor data.

3.3 Influences on Total Maintenance Cost

The total maintenance cost can be defined as a function dependent of MFOP. Using the notation

C_{Tot} total maintenance cost
C_I total inventory cost
C_D total downtime cost
C_T total transportation cost
C_R cost of scheduled and unscheduled repairs
C_P cost of prognosis

the total maintenance cost can be expressed as (Wong et al. 2005):

$$C_{Tot} := C_I + C_T + C_D \tag{1}$$

The objective is to minimize the total maintenance cost C_{Tot} using an improved CBM strategy with dynamic failure rate adjustment by utilizing historical failure rate data. Whereas the total inventory costs C_I include all the investments incurred due to stock keeping activities at a station, the total transportation costs C_T consist of capital, fuel, lubricants, and operational costs. Under an existing CBM strategy, inventory is already distributed to the locations. Therefore, costs are calculated for all previous transportation actions. For this reason, and in order to simplify the presentation, we assume that the total inventory cost and total transportation cost are given and constant. The total downtime costs C_D consist of C_R, including the costs of downtime, delays, or cancellation of services, and also the cost of prognosis, C_P, to realize CBM actions. Therefore, C_D can be expressed as an increasing function f in both values C_R and C_P:

$$C_D := f(C_R, C_P) \tag{2}$$

C_R is a function dependent on *MFOP*, an incidence probability *IP* that an event occurs, repair time t_R, and random/external influences *X*:

$$C_R := g(MFOP, IP, t_R, X) \tag{3}$$

MFOP is the expected period between two successional failures, so that the function *g* is decreasing for increasing *MFOP*. The function *g* is also decreasing with increasing *IP*, because of higher probability that failures occur at the predicted time of maintenance actions. Due to higher costs associated with longer maintenance intervals, the function *g* is increasing with increasing t_R. C_P is a function *h* dependent on *MFOP* and increases with increasing *MFOP*, because *MFOP* extension results in higher failure rate memorizing and analysis cost:

$$C_P := h(MFOP) \tag{4}$$

Using formulas (2), (3), and (4), the total maintenance costs are dependent on the *MFOP* value and can be expressed as:

$$C_{Tot} := f(g(MFOP, IP, t_R, X),\; h(MFOP)) + C_I + C_T \tag{5}$$

The objective is to calculate an optimal *MFOP* value through skillful analysis of historical and sensor data to minimize the total downtime costs. Depending on the specific practical situation being considered, if the cost functions in the above equation are continuous and differentiable, an optimal *MFOP* value can be found by differentiating the total cost function in (5) with respect to *MFOP* and setting it equal to zero:

$$\delta C_{Tot}/\delta MFOP = \delta f(g(MFOP, IP, t_R, X),\; h(MFOP))/\delta MFOP = 0 \tag{6}$$

For the determination of the *MFOP* value, several constraints based on conceptual and practical considerations are necessary. For example, *MFOP* should not exceed the time of permanency of the Original Equipment Manufacturer. Also, C_P should not be greater than the difference of the cost for unscheduled repairs and the cost of scheduled repairs and C_P should not exceed a given budget for the dedicated prognostic cost. If we assume *n* constraints in our model, expressed by $r_k(MFOP, IP, t_R, X)$, then the generic mathematical model to find the optimal *MFOP* value can be described as follows:

$$\text{Min} \quad f(g(MFOP, IP, t_R, X),\; h(MFOP)) + C_I + C_T \tag{7}$$

$$\text{s.t.} \qquad r_k(MFOP, IP, t_R, X) \le 0 \qquad \forall\ k \le n \tag{8}$$

$$MFOP, t_R \geq 0 \tag{9}$$

$$0 \leq IP \leq 1 \tag{10}$$

The general functions f, g, h, r_k can be linear or non-linear, continuous or discontinuous, and must be specified depending on the specific equipment and the company being considered. The optimal *MFOP* value can be found by solving the mathematical model represented by the objective functions (7) and the constraints (8), (9), and (10). However, practical considerations may prohibit such an easy solution to the problem because the discontinuous and irregular cost functions found in practice do not allow the use of differential calculus to find the derivatives needed in (6). Therefore, we suggest the following practical procedure to find an optimal or near optimal *MFOP* value:

1. Find or estimate all operational and cost parameters, especially those that vary with the length of *MFOP*.
2. Perform a simulation study and improve *MFOP* values to find downtime cost and total maintenance cost.
3. Identify and state the constraints on the length of the *MFOP* value.
4. As far as possible, use the data from the simulation study to fit various cost functions. If such cost functions can be developed, go to step 5; otherwise proceed to step 6.
5. Use the estimated cost functions to find an optimal *MFOP* value (either in closed form or using numerical approximation) and STOP.
6. Simulate with increasing values of *MFOP* until it is too high and repeated unscheduled failures occur. Select an appropriate *MFOP* value that minimizes total maintenance cost and STOP.

4 Experimental Results

To show significant influences of three maintenance strategies (reactive, time-based, condition-based) on total maintenance cost, the model described in Sect. 3 is used within an example from practice. In cooperation with an international aviation research company, a simulation study was created using the simulation framework Plant Simulation by Siemens PLM Software. A scenario network of four airlines (Quantas Airline QF, Virgin Airline VS, Korean Airline KE, and Thai Airline TG) with 45 aircrafts (20 for QF, 9 for VS, 7 for KE, and 9 for TG) and four main bases with ten outstations based on a real airline flight plan is simulated. On the basis of externally supplied software, an initial solution was calculated using a genetic algorithm. The overall spare parts amount of 22 pieces of one item was calculated and the parts were distributed in varying quantities to the airports (see Fig. 1). More spare parts are needed in the heavily frequented main bases than in the less frequented outstations. We considered a $50,000 component per aircraft,

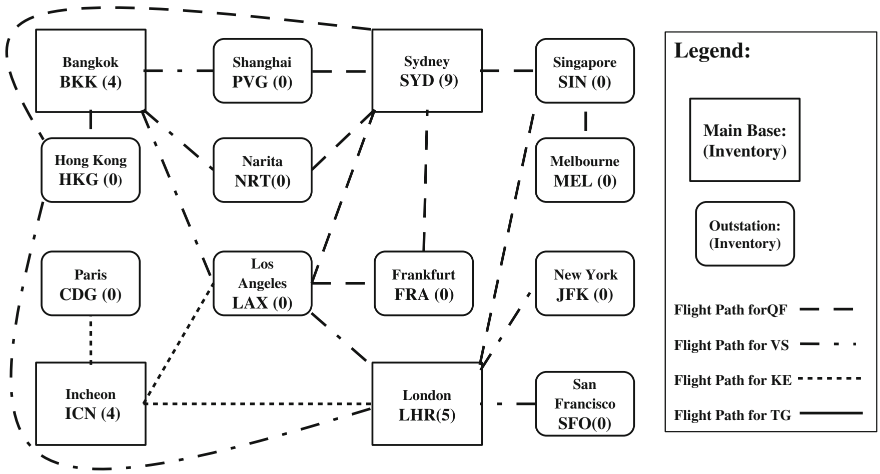


Fig. 1 Airline network information

with a mean time between unscheduled repairs (MTBUR) of 1500 h and a replacement time of 20 min. For the Weibull scale parameter we selected 63.2 % of the MTBUR (Stone and Van Heeswijk 1977), and the Weibull shape parameter of 1.7 (bigger than 1 for an increasing failure rate with time) is based on pre-tests. According to the manufacturer's data, the time-based exchange interval was set to 750 h. The repair time is 25 days and indicates how long it takes to restore the state "as good as new"; 60 days are required to replace a damaged component and spare parts can be replaced within 45 min turnaround time (Fromm 2009). Furthermore, annual inventory costs of over $10,000 and $2,000 for the logistics transportation of spare parts are assumed. The penalty cost per minute for unscheduled downtime of $90 dominates those of the downtime cost for scheduled maintenance of $50. The maintenance costs of the aircraft are valued at $300 per man-hour. Delay costs are calculated per seat per minute. Assuming 300 seats per aircraft (Airbus A330-200), the delay costs amount to $175 per min. For the cancellation of a scheduled flight due to aircraft damage, the costs amount to $60,000. For basic, middle-rate, and high-rate sensors the purchasing costs amount to $6,250, $8,500, and $11,000, respectively. The costs for data connection, linking, and transmission amount to $0.024 per data point. The investment for basic, middle-rate, and high-rate interface software and server system amount to $500,000, $650,000, and $1,000,000, respectively.

4.1 Simulation Process

In an initial solution, 45 aircrafts were distributed randomly to all airports. The initial use-time for every installed component was uniformly distributed $U(0, 948)$.

After starting the simulation, aircrafts depart as per their planned flight scheduled time and fly their given flight time until they arrive at the next airport. After arriving at the destination, airport reliability of the installed component is calculated using the following formula:

$$\text{rel(spare part)} = \exp\left(-\frac{\text{flight time for the trip}}{\alpha - \text{total flight time}}\right)^{\beta}$$

where α and β are the parameters of the Weibull distribution. If the reliability of a component is less than δ multiplied by the calculated reliability, the spare part for the aircraft is labeled for maintenance. Here δ is a B(1, 0.999999) binomial distributed parameter for random unscheduled maintenance. If $\delta = 0$, the spare part failed and unscheduled maintenance has to be done at the current airport. If $\delta > 0$, scheduled maintenance planning within a planning horizon is started and the aircraft starts the next scheduled flight. The planning for scheduled maintenance includes the identification of the best reachable airport within the planning horizon. The best reachable airport is defined as the cost minimizing airport for spare parts exchange.

During simulation the next scheduled destinations of an aircraft within the planning horizon are stored in an Excel spreadsheet. The farthest destination within the planning horizon storing the spare part is assigned as the exchange airport. If the spare part is not in stock at any of these following destinations, the last destination within the planning horizon is assigned as the exchange airport because of maximum planning time and the latest acceptable delivery time. The needed spare part from an airport with the shortest transportation time is moved to the exchange airport and the associated stock level is reduced by one. When the aircraft arrives at this exchange destination, the installed component is replaced by the delivered spare part. For a too short planning horizon or an unscheduled maintenance, the spare part may not be available at the exchange destination and waiting time occurs. If waiting time is longer than 12 h, the following flight is cancelled and another aircraft is arranged. To analyze network performance, all information about orders of spare parts, delivery time, and updated stock levels are recorded in an Excel spreadsheet. Failed components are committed into a repair cycle. The assumed repair time for a spare part is 25 days under 90 % scrap rate and 60 days under 10 % scrap rate, defined as the percentage of failed spare parts that cannot be repaired. 10 % scrap rate means that one out of ten spare parts has to be discarded. After exchanging the component, all parameters are set to “as good as new”, i.e., used time/overall flight time is zero and calculated reliability of spare part is one.

4.2 Simulation Results

The described scenario in Fig. 1 was simulated for two years (2011, 2012) and stopped quarterly to get appropriate values for the parameters. In Fig. 2 the

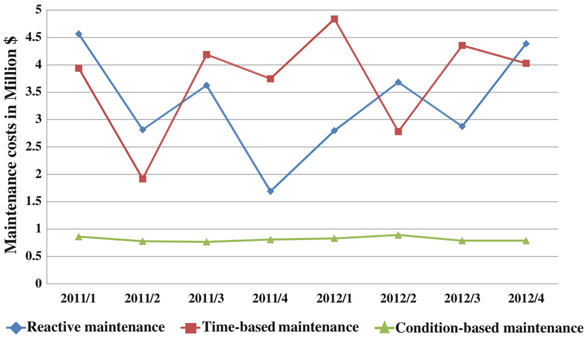


Fig. 2 Total maintenance cost per quarter in Million $

quarterly calculated total maintenance costs of the given network are presented. The total maintenance costs include the total downtime costs (cost for scheduled/unscheduled maintenance downtime, cost for penalty downtime, delay cost, installed spare part cost and prognosis cost), the total inventory costs, and the total transportation costs.

It can be seen that total maintenance costs hardly vary using the CBM strategy. Due to the continuous prognosis of possible failures and the reduction of unscheduled failures in the CBM strategy, the resulting total maintenance costs are predictable and projectable. The cost functions of the reactive and time-based maintenance strategies are more volatile and these maintenance strategies result in significantly higher total maintenance costs. Considering the two-year period, the cumulative condition-based total maintenance costs are about 22 % of the cumulative reactive total maintenance costs. The higher costs arise from randomly occurring failures of components, resulting in delays, or even cancellations, and therefore in very high penalty costs. The slightly lower costs for the reactive maintenance compared to the time-based maintenance result from the use of the full lifetime of the components, resulting in less maintenance actions. Another reason is that the spare parts are in stock at the highly frequented four main stations (BKK, ICN, SYD, LHR), and the associated deliveries of spare parts in case of unscheduled failures are relatively quick.

The downtime for scheduled and unscheduled maintenance actions consists of the repair time and the waiting time for spare part delivery. Based on optimal exchange times for spare parts as a result of good prognosis in the CBM strategy, unnecessary downtimes on the ground can be reduced. Due to continuously performed spare part exchanges resulting in more maintenance actions, the time-based maintenance strategy results in longer downtimes on the ground compared to reactive maintenance. Again, reactive and time-based maintenance strategies show more volatile downtimes (see Fig. 3).

The delay time is defined as the difference between real starting time and scheduled starting time of an aircraft. Delays result from unscheduled failures or missing resources not available in time at a station and have negative effects on the image of an airline. In Fig. 4, the CBM strategy results only in 2012/2 in a positive delay time. In all other quarters the planning horizon was long enough to determine

Fig. 3 Downtime in hours

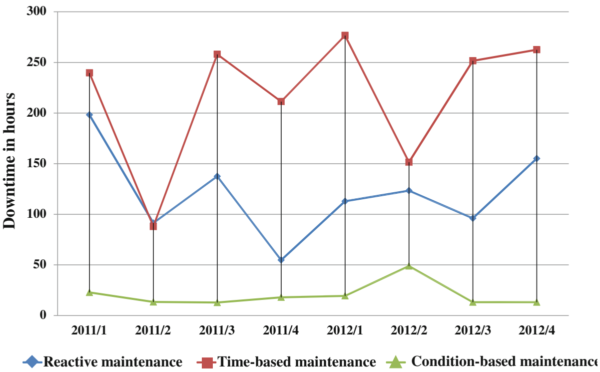
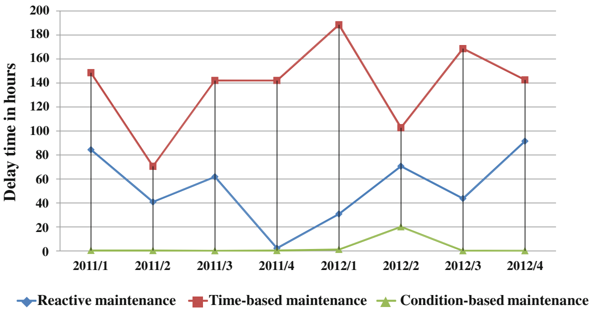


Fig. 4 Delay time in hours



an optimal station at which necessary resources for an exchange of the component are available. The reactive and time-based maintenance strategies lead to volatile and longer delays.

The results of the simulation study show clearly that the total maintenance costs can be significantly reduced using a CBM strategy based on the prediction of the degradation process of the component. The cost for prognosis should not exceed the cost difference between CBM strategy and time-based or reactive maintenance strategies. The simulation also emphasizes that downtime and delay time can be considerably reduced by applying CBM strategy. Therefore, a CBM strategy also increases the availability of the entire system. Since more flights under the existing conditions are possible, revenues are higher.

5 Conclusion

This paper presents effects of the choice of different maintenance strategies on total maintenance costs while ensuring appropriate quality and customer service. After developing expressions for various cost components, a mathematical model and a practical procedure are developed to find the optimal or near optimal *MFOP* value.

Due to a life cycle of more than 30 years, the profitability in the aviation industry is dominated by maintenance of aircrafts and not by sales of aircrafts. Analysts expect a growth of the operating budget for aviation maintenance, repair and overhaul of more than 54 billion dollars in 2015 (Lee et al. 2008). These costs emphasize the need for a reduction of total maintenance costs using a CBM strategy.

While the case study in the airline industry illustrates the benefits of a CBM strategy, it should be noted that the results are quite dependent on the quality of the available historical and sensor data. Therefore, future research should consider the use of appropriate information technology to provide better historical and sensor data in the era of big data. Furthermore, multi-objective optimization and the interdependencies between different variables should be an essential part of future research.

References

Ayeni P, Baines T, Lightfoot H, Ball P (2011) State-of-the-art of 'lean' in the aviation maintenance, repair, and overhaul industry. Proc Inst Mech Eng Part B J Eng Manufact 225:2108–2123

Barahona F, Jensen D (1998) Plant location with minimum inventory. Math Program 83:101–111

Candas MF, Kutanoglu E (2007) Benefits of considering inventory in service parts logistics network design problems with time-based service constraints. IIE Trans 39(2):159–176

Daskin M (1995) Network and discrete location: models, algorithms, and applications. Wiley-Interscience, New York

Feeney, G, Sherbrooke C (1966) The (s-1, s) inventory policy under compound Poisson demand. Manage Sci 12 (5):391–411

Fritzsche R (2012) Cost adjustment for single item pooling models using a dynamic failure rate: a calculation for the aircraft industry. Transp Res Part E Logistics Transp Rev 48(6):1065–1079

Fromm H (2009) Bewertung innovativer Instandhaltungsszenarien in den frühen Phasen des Innovationsprozesses in der Luftfahrt. Shaker

Jeet V, Kutanoglu E, Partani A (2009) Logistics network design with inventory stocking for low demand parts: Modeling and optimization. IIE Trans 41(5):389–407

Kaviani M (2009) Location-inventory problem. Physica

Kumar U (1999) New trends in aircraft reliability and maintenance measures. J Qual Maintenance Eng 5(4):287–295

Kutanoglu E (2008) Insights into inventory sharing in service parts logistics systems with time-based service levels. Comput Ind Eng 54(3):341–358

Lee S, Ma Y, Thimm G, Verstraeten J (2008) Product lifecycle management in aviation maintenance, repair and overhaul. Comput Ind 59(2–3):296–303

Moghaddam KS (2013) Multi-objective preventive maintenance and replacement scheduling in a manufacturing system using goal programming. Int J Prod Econ 146(2):704–716

Muckstadt J (2005) Analysis and algorithms for service parts supply chains. Springer, New York

Nozick L, Turnquist M (2001) Inventory, transportation, service quality and the location of distribution centers. Eur J Oper Res 129(2):362–371

Papakostas N, Papachatzakis P, Xanthakis V, Mourtzis D (2009) An approach to operational aircraft maintenance planning. Eur J Oper Res 199(3):781–792

Prajapati A, Bechtel J, Ganesan S (2012) Condition based maintenance: a survey. J Qual Maintenance Eng 18 (4):384–400

Relf MM (1999) Maintenance-free operating periods—the designer's challenge. Qual Reliab Eng Int 15(2):111–116

Sherbrooke CC (2004) Optimal inventory modeling of systems: multi-echelon techniques. Springer, New York

Snyder L, Daskin M, Teo C (2007) The stochastic location model with risk pooling. Eur J Oper Res 179(3):1221–1238

Sourirajan K, Ozsen L, Uzsoy R (2009) A genetic algorithm for a single product network design model with lead time and safety stock considerations. Eur J Oper Res 197(2):599–608

Stone G, Van Heeswijk R (1977) Parameter estimation for the Weibull distribution, Electrical Insulation. IEEE Trans Electr Insul 12(4):253–261

Tobon-Mejia D, Medjaher K, Zerhouni N, Tripot G (2012) A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models. IEEE Trans Reliab 61 (2):491–503

Wang L, Chu J, Wu J (2007) Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. Int J Prod Econ 107(1):151–163

Wheeler K, Kurtoglu T, Poll S (2010) A survey of health management user objectives in aerospace systems related to diagnostic and prognostic metrics. Int J Prognostics Health Manag 1–19

Wong H, Cattrysse D, Van Oudheusden D (2005) Stocking decisions for repairable spare parts pooling in a multi-hub system. Int J Prod Econ 93:309–317

Yardley R, Raman R, Riposo J, Chiesa J, Schank JF (2006) Impacts of the fleet response plan on surface combatant maintenance. vol 358, RAND Corporation, USA

Zipkin P (2000) Foundations of inventory management. McGraw-Hill, New York

Part II
Procurement

Considering Small and Medium-Sized Suppliers in Public Procurement—The Case of the German Defence Sector

Michael Eßig and Andreas H. Glas

Abstract The consideration of Small and Medium-Sized Enterprises (SMEs) in tender processes is an important goal of public procurement regulation. Public procurement law sets the rule to divide contracts in smaller lots as SMEs are expected to have better chances for a lot than for the whole task. This assumption is questioned with data from the German defence sector. This investigation determines the percentages of SMEs participating in and winning public tenders as well as the specific factors that influence award decisions. Key finding is that an increase in lot-wise calls will normally not lead to an increase in successful SME participation in public procurement processes. This is unexpected because lot-wise tenders are considered to be the main tool available to public procurement agents to increase the ability of SMEs to participate in and win public tenders.

1 Introduction

Small and Medium-Sized Enterprises (SMEs) are considered to be an important economic factor, for example, as employers and innovators (Zheng et al. 2006; Denes 1997). SMEs are of strategic importance for the economy as they together employ the majority of work force (up to 70 %) (Mudambi and Schründer 1996). Therefore, the promotion of SMEs is one of the social aspects attached to the procurement task of public institutions (Kidalov and Snider 2011; Carter 2004). This becomes even more important as an increasing volume of public tenders can be witnessed and the opening of the European market increased public procurement competition (Perlo-Freeman et al. 2009). However, further research on SME participation in public tendering is required, because academic discussions of this topic are rather rare and often rudimentary (Karjalainen and Kemppainen 2008;

M. Eßig · A.H. Glas (✉)
Research Center for Law and Management of Public Procurement,
Bundeswehr University Munich, 85577 Neubiberg, Germany
e-mail: andreas.glas@unibw.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_2

Zheng et al. 2006). This is surprising as SME promotion is a stated goal of both the European Committee and the German Federal Government (Van der Horst et al. 2000). It is the aim of this article to investigate to what extent public procurement regulation really promotes SMEs.

SMEs are defined in contrast to other, larger corporations. Most definitions refer to the number of employees, the annual turnover or the annual balance sheet total (Ellegaard 2006; Arend and Wisner 2005; Park and Krishnan 2001). Besides these quantitative definitions, independence from larger corporations is demanded, and only a small percentage of shares are allowed to be owned by one or more large enterprises (European Commission (EC) 2006). For the purpose of this study, both quantitative and qualitative considerations are taken into account when defining SME: An SME must employ fewer than 250 people and have an annual turnover of less than 50 million Euro and/or a balance sheet total of less than 43 million Euro (Van der Horst et al. 2000; EC 2006).

The guiding question of this article is, if public procurement legislation, in particular the award of contract through lots, really supports successful SME participation. To answer that question, public sector procurement is considered with an empirical data base of 387 awards of contract. The analysis follows a three-step approach. Firstly, a number of research hypothesis are derived from the literature considering the current state of SME participation in public procurement. Second, the peculiarities of the public procurement law are analyzed. Third, the empirical research is described and the results of the investigation are discussed. A summary of the findings with possible areas of future research conclude this article.

2 Literature Review

2.1 SME and Competing Objectives of Public Procurement

Public procurement has received increasing interest from the scientific community in recent years (Snider and Rendon 2008; Erridge 2005; Thai 2005; Harland et al. 2000). At present, public procurement accounts for roughly 13.5 % of the EU Gross Domestic Product (GDP) (Bovis 2007). For its management, Schapper et al. (2006) proposed a framework that shows the different considerations that must be regarded when making a sourcing decision (Fig. 1).

According to that framework, public procurement is supposed to achieve three main objectives: (1) the realization of a political agenda and business outcomes; (2) the accountability to the stakeholders (politicians and the public) in regard to conforming to public procurement rules; and (3) the efficient and effective management, trying to get the best value for the money spent. The framework illustrates that public procurement must satisfy a number of competing objectives. The primary goal is to satisfy specified public demands for goods or services while taking temporal, quantitative and financial constraints into consideration.

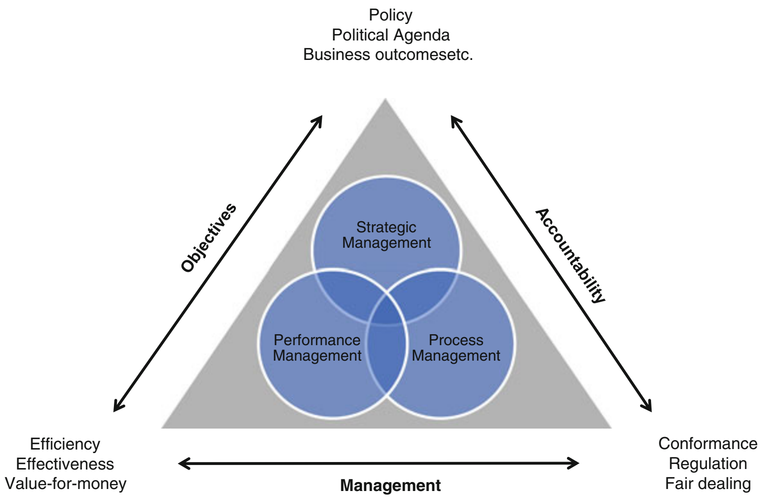


Fig. 1 Public procurement management framework, adapted from Schapper et al. (2006)

Besides, public procurement is characterized by a number of policy goals. Examples of these are environmental protection, the support of national enterprises or the promotion of SMEs (Reed et al. 2005; Erridge 2005). Distinct conflicts arise from trying to promote competition to ensure efficient procurement on one hand and to promote and support other aspects—here SMEs—on the other hand. Research regarding the consideration of SMEs in public procurement is rather limited (Karjalainen and Kemppainen 2008; Reed et al. 2005; Knight et al. 2003). The absence of qualitative and quantitative data is as much a challenge as the non-existence of factors influencing the procurement process from the perspectives of both the seller and the buyer (Zheng et al. 2006). This is in particular the case when looking at SME participation in individual sectors such as defence or health care. However, some studies already demonstrated that the consideration of SMEs can lead to a number of economic advantages, e.g. higher innovation rate, and that it is advisable to consider them (Carter et al. 1999; Erridge et al. 1998).

Studies of the EC (2004b, c) reveal that SMEs participate in 78 % of all award processes. In addition, results of Reichardt (2002) and NERA (2005) indicate that the percentage of successful SMEs is larger if the awarding institutions have a centralised procurement organisation structure. According to Piga and Zanza (2005), SME consideration in public procurement processes is primarily accomplished by dividing tenders into lots that can be bid for separately. SMEs are further considered by giving them awards as sub-contractors (Piga and Zanza 2005; EC 2004a).

The mentioned findings serve as the foundation of the two research hypotheses. H1 considers the specifics of the used data set, as it is expected that fewer SMEs successfully participate in public tenders in the defence sector. H2 generally refers to the observation that there is a positive relation between the number of

Table 1 Overview of main research hypotheses and research sources

Hypothesis		Source
H1	The percentage of participating and successful SMEs in defence is smaller than the percentage of 78 % for all public tenders in Europe	NERA (2005), EC (2004b, p. 31f), EC (2004c, p. 20), Reichardt (2002, p. 30f)
H2	Bidding success of SMEs is influenced by different company- and tender-specific factors. SME bidding success decreases with rising tender volumes, increases with the division of the tender into smaller lots (i.e., a lot number increase) and is influenced by the procurement procedure	Karjalainen and Kemppainen (2008, p. 231f). Bovis (2007, p. 228), EC (2006, p. 38f), Pfohl (2006, p. 20f), Piga and Zanza (2005, p. 187f), EC (2004b, p. 31 f), EC (2004c, p. 20), Reichardt (2002, p. 30f), Lee et al. (1999, p. 299f)

participating companies in an award process and the intensity of competition in the respective sectors of industry. Thus, H2 looks at the number of SMEs participating in and winning of the award processes (Table 1). It is expected that a number of aspects such as the selection of the award process, the use/non-use of lots, or the competitive situation have an impact (Bovis 2007). We further speculate that limited human and financial resources, meaning the number of employees as well as the total revenues, also influence the award decision (Karjalainen and Kemppainen 2008; Pfohl 2006; Lee et al. 1999). In the methodology section, H2 is further specified through sub-hypotheses.

2.2 *Specifics of German and European Law Regarding Public Procurement*

In German law, the consideration of the basic principles of competition, anti-discrimination, efficiency and transparency is based on European Public Procurement Directives 2004/17/EU, 2004/18/EU and lastly 2014/24/EU (2014). What cannot be found in public procurement guidelines is a rule giving preferential treatment to SMEs (Zheng et al. 2006). However, German regulation contains statements about SME participation, e.g. in the norms for award and contract General Services, Construction and Contractor Services (VOL/VOB/VOF), in the law against the restriction of free trade (GWB), and in defence security regulations (VSVgV).

According to German law, SMEs are considered as follows: "*The interests of small and medium-sized undertakings shall primarily be taken into account in an award procedure. Contracts shall be subdivided into partial lots and awarded separately according to the type or area of specialization (trade-specific lots) […].*" (§ 97 (3) GWB). It is expected that SMEs fit better for tenders divided into partial lots, while SMEs' degree of specialization is expected to fit better to tenders divided into trade-specific lots. However, the real effects of these regulations are doubted by the scientific literature (Schramm 2008; Blankart 2008). Recently, the use of micro lots in the United Kingdom was questioned (Loader 2014).

The defence sector is part of the public sector and is also bound to public procurement guidelines and law. Through the constitutional law of Art 296 EC as well as the secondary common law of the European procurement guidelines, deviations from these rules are permissible for the procurement of military goods and services for reasons of national security or confidentiality (EC 2009). Generally, the promotion of SMEs is equally an objective in the defence sector, as it is in other public sectors such as health care.

3 Method

3.1 Empirical Test of Hypotheses and Findings

The total number of public tenders in Germany cumulates to around 7.2 million contract awards per year (Kröber et al. 2008). Of these, the defence procurement agency since 2012 called BAAINBw is responsible for the central procurement of all significant requirements of the armed forces with a total volume of about Euro 8.74 billion today (Trybus 2005; BMVg 2015). According to Reichardt (2002), the percentage of award-winning SMEs in defence is smaller than in the overall public sector (EC 2004b, c).

To ensure that this is not a methodological problem, a sample was chosen that also includes the procurement of "off the shelf" civilian goods in addition to defence specific, military goods. The sample was based on the awards for the years 2006–2008. In the database tender and award data of a specific department, E4, were sampled. This department is responsible for the procurement of military specific goods and services as well as for commercial goods and services. Specifically, the procurement of spare parts for weapons systems of the armed forces is conducted there. E4 is responsible for roughly 8000 annual tenders with a yearly award volume of about Euro 360 million (Hoos 2009; BWB 2008; Petry 2008).

Statistical sampling was used for the investigation (Fig. 2). A complete and numbered list of all award procedures and contracts was obtained. Overall, 18,512 awards were retrieved in the three-year timeframe. The random sampling method allowed the determination of the minimum number of data sets required. The minimum size is 376 of randomly sampled awards. To achieve that size, SPSS statistics software was used and 420 completed awards were randomly selected. Of these 420 awards, 387 were suitable (completeness of data).

For the sample, the number of participating SMEs and if a SME succeeded in winning the award are assessed. Then, it is possible to estimate numbers for the total population by utilising point estimation procedures. For the determination of sample and population, random variables PA (participant) and SU (success) were

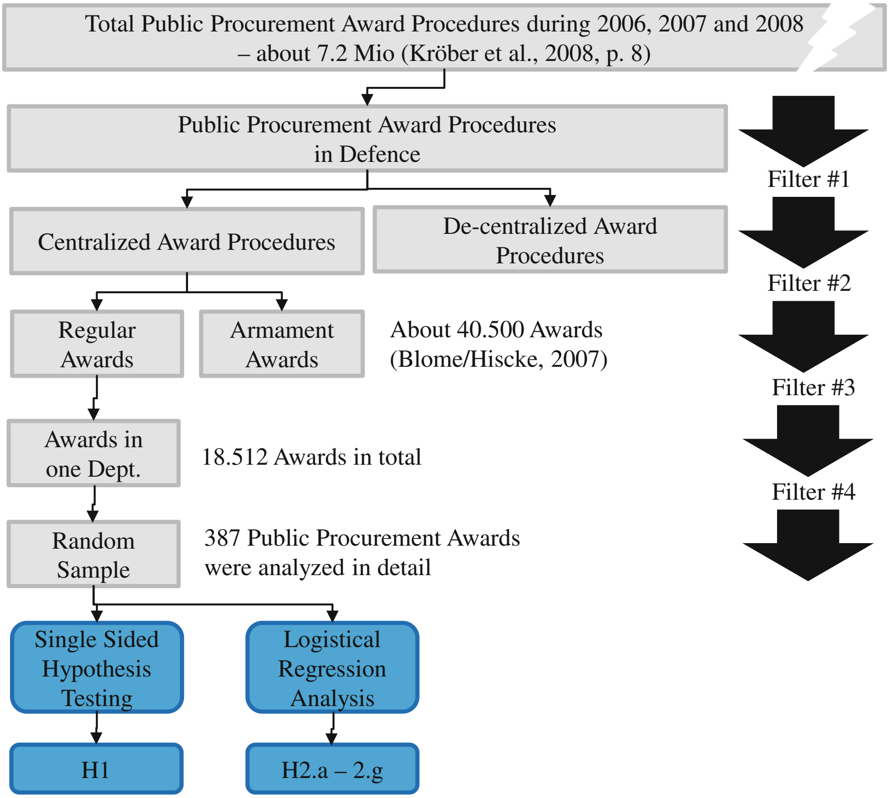


Fig. 2 Sampling method

selected (PA = 1 for SME participated in the tender and SU = 1 for SME received the award; PA and SU = 0 in other cases).

The necessary statistical values for the examined arithmetic mean μ based on hypothesis 1 are: $H1_{PA}: \mu < 0.78$ and $H1_{SU}: \mu < 0.78$. The null hypotheses are: $H0_{PA}: \mu \geq 0.78$ and $H0_{SU}: \mu \geq 0.78$. Through a determination of the *p*-value based on the sample, it was determined that the null hypothesis can be retained if it holds true that it is of a larger magnitude than the level of significance α of 0.05.

Hypothesis 2 is investigated by utilising the method of multivariate analysis. The selection of the most useful method will be carried out through the determination of the different variables (Backhaus et al. 2006; Hosmer and Lemeshow 2000; Menard 1995). The following factors were retrieved: The total number of participating SMEs (TSME), the number of employees (EM) and the annual turnover (TO) of the successful company, the award volume (VOL), the type of award (KIND), the type of the procured service or material (TYPE) and the number of lots (NLOTS). The following sub-hypotheses are derived:

H2.a. The more SMEs take part in the tender, the higher the likelihood of SMEs' bidding success
H2.b. The more employees an SME has, the higher the likelihood of SME's bidding success
H2.c. The higher the annual turnover of the SME, the higher the likelihood of SME's bidding success
H2.d. The higher the award volume of the contract, the lower the likelihood of SMEs' bidding success
H2.e. The more competitive the award procedure, the higher the likelihood of SMEs' bidding success
H2.f. The more commercial standard the good or service to be awarded, the higher the likelihood of SMEs' bidding success
H2.g. With increasing division of the award into smaller lots, the likelihood of SMEs' bidding success increases

The seven parameters are to be considered as independent variables that are scaled metrically, respectively nominally. To test for structure, it is necessary to determine an appropriate independent variable (here: SU, bidding success). On that basis, the method of logistic regression was executed. This method allows determining the likelihood that the event "an SME wins an award during a public tender" will occur. On the other hand, it gives the chance to indicate which factors, represented by the regression coefficient, influence the occurrence of this event. The resulting logistical regression equation for the observed event is k $(k = 1 \ldots n)$ when using the logistical function for the calculation of the probability P of the occurrence of SU = 1 and considering the defined parameters:

$$P_k(SU = 1) = \frac{1}{1 + e^{-z_k}} \text{ with} \tag{1}$$

$$\begin{aligned} z_k = \beta_0 &+ \beta_1 TSME_k + \beta_2 EM_k + \beta_3 TO_k + \beta_4 VOL_k + \beta_5 KIND1_k \\ &+ \beta_6 KIND2_k + \beta_7 KIND3_k + \beta_8 KIND4_k + \beta_9 KIND5_k \\ &+ \beta_{10} KIND6_k + \beta_{11} TYPE_k + \beta_{12} NLOTS_k \end{aligned}$$

After formulating the research model, it is necessary to estimate the logistical regression function to be able to interpret the regression coefficients and to test the whole model as well as the characteristic variables. In this investigation, the estimation will be accomplished using the statistics software SPSS. SPSS calculates the necessary parameters by maximising the likelihood of the event SU = 1. The variable KIND stands for the type of award and was coded as follows: (1) Limited tender/not public tender with public participation competition; (2) Limited tender without public participation competition; (3) Negotiation procedure; (4) Single tender action; (5) Single tender action in competition/with public participation competition; and (6) Competitive dialogue.

Table 2 Impact of positive and negative regression coefficients regarding the likelihood that the event SU = 1 will occur

Regression coefficient	Odds	Likelihood of SU = 1
$\beta_i > 0$	Increases by e^{β_i}	Increases
$\beta_i < 0$	Decreases by e^{β_i}	Decreases

A logistical regression allows the interpretation of the results in the sense that a negative regression coefficient implies a decreasing likelihood, and vice versa. The variables are used to compare the reference categories, meaning that a negative prefix implies that the likelihood of an event occurring is decreasing in relation to the reference category. On the other hand, an increasing value for the independent variable implies that the odds (relation of the likelihood of the events to each other, SU = 1 to SU = 0) change by the factor $e^{\beta_i}(i \in \{1\ldots12\})$. The same is true of the odds of the categories changing in relation to the variables in the reference categories (Table 2).

4 Empirical Results

The 387 sampled award documents indicated that a total of 1308 companies participated in the bidding processes, and in total 787 lot-wise or whole contracts were awarded. Based on the documents, it was possible to identify all relevant information: Company-specific key data were collected either based on the award documents or from the software tools available. The award-specific characteristics were extracted from the award documents.

For hypothesis 1, the means of the random sample $\overline{X}$ with regard to the random variables PA and SU were determined to be 56.8 %, respectively 56.5 % (Table 3), when conducting a t-test of a sample with a division value of 0.78.

The tests show a high degree of significance because the *p*-value α_1 is smaller than 0.001 and therefore smaller than the level of significance of $\alpha = 0.05$. The medium difference between the test value and the average sample value comes to PA of 21.2 % and SU of 21.5 %, taking an error tolerance of 5 % into account. The Kolmogorov-Smirnov adjustment test showed that variables had normal distributions (Table 4).

Table 3 Test results of the random variables PA and SU

Variable	N	Mean	Standard deviation	Standard error of mean
PA	1308	0.57	0.496	0.014
SU	787	0.57	0.496	0.018

Table 4 Test results of the random variables PA and SU

Variable	Test value = 0.78					
	T	df	Significance (two-sided)	Mean difference	95 % confidence interval for the difference	
					Lower	Upper
PA	−15.469	1307	0.000	−0.212	−0.24	−0.19
SU	−12.135	786	0.000	−0.215	−0.25	−0.18

The application of the logistical regression to the collected data to test hypothesis 2 shows the following results:

$$P_k(SU = 1) = \frac{1}{1 + e^{-z_k}} \text{ with} \tag{2}$$

$$\begin{aligned} z_k = {} & 2.338 + 0.075 * TSME_k - 0.004 * EM_k + 0.002 * TO_k - 4.24 \\ & * VOL_k - 1.127 * KIND1_k + 0.338 * KIND2_k - 1.348 \\ & * KIND3_k - 1.136 * KIND4_k + 4.273 * KIND5_k \\ & + 0.02 * TYPE_k - 0.002 * NLOTS_k \end{aligned}$$

The interpretation of the specific regression coefficients (Table 5) reveals the following: The probability that an awarded contract is given to an SME rises if more SMEs take part in a tender or the annual turnover increases. Concerning the type of the procured good or service, the probability and the odds ratio rise if the good or service is military specific. Also, the influencing factors KIND2 and KIND5 have higher odds, what implies that the selection of a restricted call for tenders without public contest or a negotiated award with public contest will raise the odds of an award for SMEs in comparison to a public call for tenders or an open procedure. The competitive dialogue was not applied and thus not considered. In contrast, the odds ratio decreases to a minimal level if the number of employees and the number of awarded lots and positions rise. A negative odds ratio or, rather, a worse influence will result from realising a restricted call for tenders, a restricted procedure with competition among participants, a negotiated procedure or a freely negotiated award in comparison to the reference category. The same is for the award volume; i.e., if the volume increases, the odd ratio decreases.

To assess the quality of the model, the above-defined criteria for the adaptation and pattern quality will be interpreted (Table 6). First, the value of the deviance, which is smaller than the χ^2 distribution, suggests good model fit. Similarly, the values of the Cox and Snell R^2 and Nagelkerke R^2 methods indicate good to very good model accuracy. The assessment of the classification results, i.e., the comparison of observed and calculated values, showed that 23 cases were misclassified. Thus, it appears that the values of the correctly stated cases amounted to 94.1 % and proportional chance probabilities of 54.52 and 50.41 %, respectively. The Press's

Table 5 Interpretation of regression coefficients

Influence factor	Regression coefficient	Regression coefficient	Odds	Probability ef = 1
TSME	0.075	>0	Increase by 1.078	Increases
TO	0.002	>0	Increase by 1.002	Increases
EM	−0.004	<0	Decrease by 0.996	Decreases
NLOTS	−0.002	<0	Decrease by 0.998	Decreases
TYPE	0.02	>0	Increase by 1.02	Increases
KIND1	−1.127	<0	Decrease by 0.324	Decreases
KIND2	0.338	>0	Increase by 1.402	Increases
KIND3	−1.348	<0	Decrease by 0.26	Decreases
KIND4	−1.136	<0	Decrease by 0.321	Decreases
KIND5	4.273	>0	Increase by 71.736	Increases
KIND6	Not occurred			
VOL	−4.24	<0	Decrease by 0.014	Decreases

Q-test showed a value of 301.057, which is higher than the tabulated χ^2 value. The Hosmer-Lemeshow test result leads to the conclusion that the null hypothesis is rejected, and it must be assumed that the difference between the empirical and calculated values is not equal to zero. It can be assumed that the results differ significantly from a random assignment. The method of standardised residuals substantiates the above-mentioned 23 outliers, whose values are between −3.343 and 137,318.788. These are special cases in which, for example, the number of employees corresponds to the average number of an SME, but the annual turnover was greater than 50 million Euro.

The attribute variables, TSME, TO, NLOTS, TYPE, KIND1, KIND2 and KIND3 indicate a lower value of the Wald statistic as the tabulated χ^2 value. Due to the addition of these high levels of significance between 25.3 and 96.4 %, it must be assumed that the factors have no significant effect. It can be assumed that the variable VOL has a significant effect if the level of significance is accepted at 6.6 %. Finally, it can be considered that the influence factors EM and KIND4 have a significant effect, as their significance is below 5 % (Table 7).

Table 6 Examination of the overall model

Quality criterion	Accepted range of values	Value from the analysis	Interpretation
Deviance (−2LL value)	−2LL value < χ^2 value	192.2 < 420.094	Good fit of the model
Cox and Snell R^2	Acceptable starting values >0.2; good starting values >0.4	0.586 > 0.4	Good fit of the model
Nagelkerke R^2	Acceptable starting values >0.2; good starting values >0.4; very good starting values >0.5	0.783 > 0.5	Very good fit of the model
Analysis of classification results	Higher than maximum and proportional random probability	0.941 > 0.5452 0.941 > 0.5041	Higher hit rate than random
Press's Q-Test	Highest possible χ^2 value; level of significance <5 %	301.057 > 3.84	significant difference from random assignment
Hosmer-Lemeshow test	Smallest possible χ^2 value; level of significance >70 %	1421.058 > 15.51 Level of significance: 0 %	Difference between the calculated and actual values not equal to 0

Table 7 Significance tests of variables

Influencing factors	Degrees of freedom	χ^2 value	Wald statistics	Significance	Interpretation
TSME	1	3.84	1.158	0.282	No significance
TO	1	3.84	1.206	0.272	No significance
EM	1	3.84	15.525	0.000	Significant
NLOTS	1	3.84	0.018	0.892	No significance
TYPE	1	3.84	0.002	0.964	No significance
KIND1	1	3.84	1.305	0.253	No significance
KIND2	1	3.84	0.161	0.688	No significance
KIND3	1	3.84	0.969	0.325	No significance
KIND4	1	3.84	6.984	0.008	Significance
KIND5 and KIND6					Without result
VOL	1	3.84	3.37	0.066	No significance

5 Discussion and Conclusion

H2.a is accepted what means that an increasing number of participating SME's positively effects the bidding success. This seems obvious, as at least one SME must award for a contract, and the more SME apply the higher is the likelihood.

The results also show that the shares of participating SMEs in the defence sector is less than 78 % and thus lower than in the overall public sector. This is not surprising and in line with H2.d, which describes the influence of the increasing award volume. In other words, the higher the overall award volume, the lower is the success likelihood of SMEs. As some defense systems (aircrafts, tanks etc.) are complex and expensive, companies able to produce them must have far higher turnover or balance sums than SMEs are per definition allowed to have. This can be connected with H2.c, which indicates that a higher level of the annual turnover have a positive effect on the bidding success of SMEs. Overall: Company-size matters in defence business.

H2.e shows that an increased competitiveness of the awarding procedure has a positive effect on the SME bidding success. While a restricted call for tenders without public competition reduces SMEs' success likelihood, a negotiated award or a competitive procedure increases the chances for SMEs. The more competition in the procedure, the higher is the success likelihood of SMEs. This result is also not unexpected as SMEs are praised for their high innovativeness and flexibility.

The most striking result of this study is the rejection of H2.g. The analysis revealed that a division of the award into lots does not lead to a higher success of participating SMEs. This is strictly in contrast to the legal regulations, which postulate lots as the driving force for SME promotion. The empirical data indicate that a further strengthening of lot-wise tenders is not likely to increase the chances of SMEs' success. Instead, other aspects should be improved if a higher rate of SME bidding success is desired. According to Loader (2013, 2014), the most common barriers reported are overly prescriptive qualification criteria, poorly written tender specifications and prohibitive resource requirements. On the other hand also SMEs struggle with the public tender participation and their competence in matters of procedure for participation in public tenders should be improved. SMEs face a lack of appropriate resources to engage in public biddings and their skills to complete a bid in the required standards are low. Overall, this is the cause that SMEs perceive public procurement procedures to be unfair and so they are reluctant to engage (Loader 2013, 2014; Pickernell et al. 2011).

In parts the results of this study reaffirms previous findings (e.g. Karlajainen and Kemppainen 2008). On the other hand they offer new insight into the effectiveness of public procurement regulation. SMEs (or their associations) call to unbundle contracts and to provide smaller tender opportunities. Responding, governments take actions to increase the use of lots (Loader 2014). The findings of this study stand in fundamental contrast to these public procurement practices. Even if this study is limited, as its scope only comprised defence, the key observation is striking: The division of larger tenders into smaller lot sizes and the lot wise award of contracts do not have an impact on SME public bidding success rates.

Acknowledgments The Authors specially thank the Federal Office of Bundeswehr Equipment, Information Technology and In-Service Support (BAAINBw) for their active support.

References

Arend RJ, Wisner JD (2005) Small business and supply chain management: is there a fit? J Bus Ventur 20:403–436

Backhaus K, Erichson B, Plinke W, Weiber R (2006) Multivariate analysemethoden: eine anwendungsorientierte Einführung, 11th edn. Springer, Berlin

Blankart CB (2008) Öffentliche Finanzen in der Demokratie: eine Einführung in die Finanzwissenschaft, 7th edn. München

BMVg (2015) Bericht des Bundesministeriums der Verteidigung zu Rüstungsangelegenheiten. März, Berlin

Bovis C (2007) EU Public Procurement Law. Cheltenham 2007

BWB (2008) Auftraggeber Bundeswehr: organisation, Auftragsvergabe und Vertragsgestaltung. http://www.bwb.org/fileserving/PortalFiles/02DB022000000001/W26DVFZP486INFODE/AGBw_de.pdf?yw_repository=youatweb. Accessed on 24 Apr 2008

Carter CR (2004) Purchasing and social responsibility: a replication and extension. J Supply Chain Manage 40(4):4–16

Carter CR, Auskalnis RJ, Ketchum CL (1999) Purchasing from minority business enterprises: key success factors. J Supply Chain Manage 35(1):28–32

Denes TA (1997) Do small-business set-asides increase the cost of government contracting? Public Adm Rev 57(5):441–444

Ellegaard C (2006) Small company purchasing: a research agenda. J Purchasing Supply Manage 12:272–283

Erridge A, Fee R, McIlroy J (1998) Involvement of SMEs in public procurement. Public Procurement Law Rev 7(2):37–51

Erridge A (2005) UK public procurement policy and the delivery of public value. In: Thai KV et al (eds) Challenges in public procurement. Boca Raton, USA, pp 335–352

European Commission (2004a) Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts. Off J Eur Union, L 134, pp 114–240

European Commission (2004b) The access of SMEs to public procurement contracts: final report. http://ec.europa.eu/enterprise/newsroom/cf/document.cfm?action=display&doc_id=3687&userservice_id=1&request.id=0. Accessed on 4 Nov 2009

European Commission (2004c) A report on the functioning of public procurement markets in the EU: benefits from the application of EU directives and challenges for the future. http://ec.europa.eu/internal_market/publicprocurement/docs/public-proc-market-final-report_en.pdf. Accessed on 4 Nov 2009

European Commission (2006) The new SME definition: user guide and model declaration. http://ec.europa.eu/enterprise/policies/sme/files/sme_definition/sme_user_guide_en.pdf. Accessed on 4 Nov 2009

European Commission (2009) Directive 2009/81/EC of the European Parliament and of the Council of 13 July 2009 on the coordination of procedures for the award of certain works contracts, supply contracts and service contracts by contracting authorities or entities in the fields of defence and security, and amending Directives 2004/17/EC and 2004/18/EC

European Commission (2014) Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement
Harland CM, Gibbs J, Sutton RY (2000) Supply strategy for the public sector: framing the issues. In: Proceedings of the 9th annual IPSERA conference and 3rd annual North American research symposium of purchasing and supply chain management, Toronto, Canada, pp 342–352
Hoos S (2009) In: Eßig M, Witt M (eds) Beitrag des Strategischen Einkaufs zur Optimierung der logistischen Leistungstiefe am Beispiel des Verteidigungsressorts, pp 97–124
Hosmer DW, Lemeshow S (2000) Applied logistics regression, 2nd edn. New York 2000
Karjalainen K, Kemppainen K (2008) The involvement of small- and medium-sized enterprises in public procurement: impact of resource perceptions, electronic systems and enterprise size. J Purchasing Supply Manage 14:230–240
Kidalov MV, Snider KF (2011) US and European public procurement policies for small and medium-sized enterprises (SME): a comparative perspective. Bus Politics 13(4)
Knight L, Caldwell N, Harland C, Telgen J (2003) Government reform and public procurement: academic report. University of Bath, UK
Kröber R, Fielseler J, Kirch T (2008) Kostenmessung öffentlicher Liefer-, Dienstleistungs- und Bauaufträge aus Sicht der Wirtschaft und der öffentlichen Auftraggeber. Berlin
Lee K, Lim G, Tan S (1999) Dealing with resource disadvantage: generic strategies for SMEs. Small Bus Econ 12(4):299–311
Loader K (2013) Is public procurement a successful small business support policy? A review of the evidence. Environ Plann C: Gov Policy 31(1):39–55
Loader K (2014) SME suppliers and the challenge of public procurement: evidence revealed by a UK government online feedback facility. J Purchasing Supply Manage 1–10. doi:10.1016/j.pursup.2014.12.003
Menard S (1995) Applied logistics regression analysis. Thousand Oaks, CA, Sage
Mudambi R, Schründer CP (1996) Progress towards buyer-supplier partnerships: evidence from small and medium-sized manufacturing firms. Eur J Purchasing Supply Manage 2(2/3): 119–127
NERA (2005) A study of the benefits of public sector procurement from small businesses. London
Park D, Krishnan HA (2001) Supplier selection practices among small firms in the United States: testing three models. J Small Bus Manage 39(3):259–271
Perlo-Freeman S, Perdomo C, Sköns E, Stalenheim P (2009) In: SIPRI (ed) Military expenditure, pp 179–211
Petry D (2008) Die Modernisierung der Beschaffung. Strat Tech 51:95–97
Pfohl HC (2006) Betriebswirtschaftslehre der Mittel- und Kleinbetriebe: Größenspezifische Probleme und Möglichkeiten zu ihrer Lösung, 4th edn. Berlin
Pickernell D, Kay A, Packham G, Miller C (2011) Competing agendas in public procurement: an empirical analysis of opportunities and limits in the UK for SMEs. Environ Plann C: Gov Policy 29(4):641–658
Piga G, Zanza M (2005) An exploratory analysis of public procurement practices in Europe. In: Thai KV et al (eds) Challenges in public procurement: an international perspective. Boca Raton, Florida, pp 179–206
Reed T, Luna P, Pike W (2005) Balancing socioeconomic and public procurement reform goals: effective metrics for measuring small business participation in public procurement. In: Thai KV et al (eds) Challenges in public procurement: an international perspective, Boca Raton, Florida, pp 81–100
Reichardt KD (2002) Wehrtechnischen Mittelstand konsequent fördern. Lage der Wehrtechnischen Industrie zu Beginn der 15. Legislaturperiode, Bonn, pp 115–120
Schapper P, Veiga Malta JN, Gilbert DL (2006) An analytical framework for the management and reform of public procurement. J Public Procurement 6(1):1–26
Schramm J (2008) Auftragsgegenstand und Leistungsbeschreibung. In: Müller-Wrede M (ed) Kompendium des Vergaberechts: Systematische Darstellung unter Berücksichtigung des EU-Vergaberechts, Köln, pp 307–328

Snider KF, Rendon RG (2008) Public procurement policy: implications for theory and practice. J Public Procurement 8(3):310–333

Thai KV (2005) Challenges in public procurement. In: Thai KV et al (eds) Challenges in public procurement: an international perspective. Boca Raton, USA, pp 1–19

Trybus M (2005) European Union law and defence integration. Bloomsbury Publishing, Oxford

Van der Horst R, Nijsen A, Gulhan S (2000) Regulatory policies and their impact on SMEs in Europe: the case of administrative burdens. In: Sexton DL, Landström H (eds) Handbook of entrepreneurship. Malden, pp 128–149

Zheng J, Walker H, Harland C (2006) The role of SMEs in public procurement: a review of literature and research agenda. In: Proceedings of the 15th annual IPSERA conference. Paper ID#57

Integration of Cultural Aspects in Supplier Evaluation

Stefan Winter and Rainer Lasch

Abstract In international sourcing and in strategic partnerships cultural differences can lead to problems in the cooperation of buyer and supplier. Therefore, cultural criteria should be included in supplier evaluation and selection processes in order to anticipate and to avoid possible problems. Although there are recommended cultural criteria in literature, we do not know much about the application of cultural criteria in supplier evaluation. In this paper, we present the results of an empirical study about the integration of cultural aspects in supplier evaluation. We explore company practices and problems with suppliers due to cultural differences. Furthermore, we give recommendations for the application of cultural criteria in supplier evaluation processes in practice.

1 Introduction

There are several studies which report about cultural problems in international sourcing and in partnerships between companies. For example, Handfield (1994) refers to cultural barriers as the most experienced problem by material managers in using international sources. Both national and corporate cultures have a negative effect on international joint venture performance (Pothukuchi et al. 2002). Furthermore, it is argued that a low level of cultural similarity increases coordination cost because of cultural adjustment (Kogut and Singh 1988; Barkema et al. 1996; Li et al. 2001; Heiman et al. 2008). On the positive side, shared values in corporate cultures shall decrease coordination costs between cooperation partners (Das and Teng 1998; Sarkar et al. 2001; Beimborn et al. 2009). Especially in cases where cooperation or flexibility is required or innovation is important, cultural factors have to be included in the supplier evaluation and selection process

S. Winter (✉) · R. Lasch
Lehrstuhl Für Betriebswirtschaftslehre, Insb. Logistik, Technische Universität Dresden, Münchner Platz 1/3, 01069 Dresden, Germany
e-mail: stefan.winter@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_3

(Cummins et al. 2012). Furthermore, the inclusion of national and organizational culture criteria in supplier evaluation and selection gets more important because of the increasing trends: global sourcing, outsourcing, supplier integration, and joint ventures.

In literature about international supplier and strategic partner selection cultural criteria are included in suggested evaluation methods (see Sect. 2.2). However, we do not know much about the application of cultural criteria in supplier evaluation. Therefore, the aim of this contribution is to examine the integration of cultural criteria in supplier evaluation. For this purpose, we conducted a literature review about contributions with cultural criteria in supplier evaluation and interviewed purchasing experts of various companies about cultural criteria and problems due to cultural differences.

The rest of this paper is organized as follows. Section 2 deals with the definition of culture and the literature on cultural criteria in supplier evaluation. In Sect. 3, the methodology for the empirical study is described. Furthermore, the results of the expert interviews are presented and discussed. A conclusion with a summary of results and future research directions is given in the end.

2 Literature Review

First, we define culture (Sect. 2.1). Second, we provide an overview about contributions with culture as a criterion in supplier evaluation (Sect. 2.2).

2.1 Culture

There are many definitions of culture in literature. Three famous ones are:

- "Culture consists in patterned ways of thinking, feeling and reacting (…)" (Kluckhohn 1951, p. 86).
- "[C]ulture is defined as shared motives, values, beliefs, identities, and interpretations or meanings of significant events that result from common experiences of members of collectives that are transmitted across generations" (House and Javidan 2004, p. 15).
- Culture is "the collective programming of the mind that distinguishes the members of one group or category from another" (Hofstede 2001, p. 9).

In general, culture seems to distinguish one group from another on the basis of a determined set of shared values, beliefs, behaviors, and attitudes (Bik 2010). Groups can be people in a society with a national culture or staff of an organization with an organizational culture (Leung et al. 2005). If cultural core values are identical, the level of cultural capability will be high (Rijamampianina and Carmichel 2005; Beimborn et al. 2009).

2.2 Culture in Supplier Evaluation

We could identify a lot of contributions with culture as a criterion in supplier evaluation. Whereas the national culture is mostly related to literature about international or global supplier selection (see Table 1), organizational culture is associated to literature about the selection of strategic partners (see Table 2).

If we examine the criteria about culture from international or global supplier selection contributions, we can identify the two main points communication and ethics as important issues. Communication problems can arise because of language barriers and different communication devices between buyers and suppliers of different countries. Furthermore, ethical standards and business customs can vary from country to country. Consequently, these criteria refer to the national culture.

In contrast, the suggested cultural criteria for the selection of strategic partners refer to the organizational culture. Feeling of trust, management attitude/vision/outlook for the future, strategic fit, top management compatibility, compatibility across levels and functions of buyer and supplier firms, and supplier's organizational structure and personnel are often included in the different contributions as originally suggested by Ellram (1990). Furthermore, the degree of (strategic) cooperation, business methodology and operating values (e.g., TQM), communication issues, and supplier's reputation were also recommended as evaluation criteria. As we can see from Table 2 there is an interaction between organizational culture and strategy.

Based on the different contributions we have recommendations for cultural criteria in supplier evaluation. However, we do not know details about the application of national and organizational culture criteria in supplier evaluation, especially what cultural differences and criteria have to be considered in practice as well

Table 1 Contributions including culture in international/global supplier selection

Author	Criteria and subcriteria/interpretation	Background
Min (1994)	Cultural and communication barriers (main criterion) [cultural similarity, ethical standards, EDI capability]	International supplier selection
Chan et al. (2007)	National cultural compatibility (subcriterion of perceived risks) [languages, business customs, ethics, and communication devices vary from country to country]	Supplier selection in the airline industry
Huang and Keskar (2007)	Cultural similarity (subcriterion of asset and infrastructure metrics) [cultural and language barriers between buyer and supplier]	Procedure for metrics development
Yücenur et al. (2011)	Cultural difference (subcriterion of risk factors) ['In a global supplier selection process, understanding each other is a key factor for both manufacturer and supplier. The same idea structure and the similar cultural structure can help them to communicate with each other easily and also have good relationships.']	Supplier selection in global supply chains

Table 2 Contributions including culture in strategic partner selection

Author	Criteria and subcriteria/interpretation	Background
Ellram (1990)	Organizational culture and strategic issues (main criterion) [feeling of trust, management attitude/outlook for the future, strategic fit, top management compatibility, compatibility across levels and functions of buyer and supplier firms, supplier's organizational structure and personnel]	Supplier selection decision in strategic partnerships
Briggs (1994)	Culture (key requirement) [underlying business methodology and operating values]	Supplier selection for partnerships
Sarkis and Talluri (2002)	Culture (subcriterion organizational factors) [see Ellram (1990)]	Strategic supplier selection
Humphreys et al. (2003)	Supplier's culture (main criterion) [TQM management culture, teamwork and cooperation, strategic fit culture of both firms]	Supplier criteria requirements for strategic purchasing
Bayazit et al. (2006)	Organizational culture (subcriterion of commercial structure) [long-term relationship (the supplier's willingness to develop longer-term relationships), reliability/trust (refers to acting responsibly and meeting performance expectations reliably), management capability (includes management's commitment, overall professional ability, and willingness to develop a closer working relationship with the buyer)]	Supplier selection for a construction company
Chan et al. (2007)	Organizational culture and strategic issue (subcriterion of company background/business structure) [management attitude/commitment and outlook for future, company organizational structure and personnel, company past record/reputation in the field]	Supplier selection of critical items in the airline industry
Huang and Keskar (2007)	Cultural similarity (subcriterion of asset and infrastructure metrics) [cultural and language barriers between buyer and supplier]	Procedure for metrics development

(continued)

Table 2 (continued)

Author	Criteria and subcriteria/interpretation	Background
Nukala and Gupta (2007)	Cultural and strategic issues (main criterion) [cooperation and information exchange, supplier's financial stability/economic performance, supplier's green image, flexibility]	Supplier selection in strategic planning of supply chains
Chou and Chang (2008)	Organizational culture and strategy (main criterion) [management capability (the DMs impression of the management capability of a supplier according to mutual interaction), strategic fit (strategic fit represents the fit between firm strategy and supplier strategy as judged by the DMs)]	Supplier/vendor selection from the perspective of strategic management of the supply chain
Thanaraksakul and Phruksaphanrat (2009)	Cultural congruence (subcriterion of learning and growth perspective) [cultural difference, feeling of trust]	Strategic supplier selection
Wen and Chi (2010)	Organizational culture and strategic issues (subcriterion of partnership issues) [compatibility across levels of buyer and supplier, communication system]	Green supplier selection
Chen (2011)	Organization culture (subcriterion of relationship strategy analysis) [-]	Supplier selection for relationships
Dogan and Aydin (2011)	Organizational culture and strategy (main criterion) [management vision (the view of the company for future regarding effective and collaborative process), strategic fit (compatibility of buyer and supplier in terms of organizational, cultural and strategic factors; whether supplier and buyer share a common vision of the future)]	Supplier selection analysis for strategic (: tier-1) suppliers
Hashemi et al. (2015)	Culture (subcriterion of economic supplier selection criteria) [feeling of trust, management attitude for the future, strategic fit, top management capability, compatibility among levels and functions, supplier's organizational structure and personnel, future strategy direction, degree of strategic cooperation]	Green supplier selection

as in what stage of the supplier selection and evaluation process culture criteria should be applied and how. Therefore, we conducted in-depth interviews with some companies.

3 Empirical Study

First, we outline the applied methodology (Sect. 3.1). Second, we describe our results in Sect. 3.2. Afterwards, we discuss our results (Sect. 3.3).

3.1 Methodology

To gain insight into the integration of cultural aspects in supplier evaluation in practice, interviews with purchasing experts were conducted and a multiple case study was developed. Objective of the empirical study was to examine how companies integrate cultural issues in supplier evaluation in practice. The research procedure, which is oriented on the procedure of Eisenhardt (1989), is described in the following.

After the research question was defined, cases were selected (see Table 3). Interviews with experts of the purchasing area from seven companies were conducted. Purchasing as investigation object is appropriate because it is responsible for the procurement of products and for supplier evaluation. All companies have strategic suppliers and most of them purchase abroad. The background of the interviewed experts differs in several terms in order to help identifying if the same phenomenon exists at certain sites but not at other sites (Stuart et al. 2002). This is also consistent with the argument of Eisenhardt (1989) that extreme situations or polar types should be considered because only a low number of possible cases can usually be studied. We included experts from different manufacturing industries, from retail, and from the service sector which are confronted with different

Table 3 Data about the different experts

Experts	Supplier countries	Purchasing items
1	Germany	Indirect materials, simple services
2	Europe	Complex services
3	Japan, USA, South Korea, Taiwan, China, Europe	Components, indirect materials
4	Germany, China, India	Components
5	Germany, worldwide	Components
6/7	Asia, America, Europe	Components, modules, systems, indirect materials, simple and complex services
8	Germany, France, England, Netherlands, USA, China	Raw materials, components, modules, systems, simple services

purchasing items in order to identify similarities and differences in practice. The number of companies conforms to the suggested number of Eisenhardt (1989) who recommends four to ten cases.

Data was collected by semi-structured interviews. Interviews are one of the most important sources of case study information (Yin 1989). As preparation for the interview an interview guide concerning the topic's general questions, supplier evaluation and selection, occurred cultural problems, importance of cultural criteria, and integration in supplier evaluation and selection was developed and provided to the interviewees before the conversation. Data collection refers to the theory of national and organizational culture as well as supplier evaluation (in relation to selection and controlling), which is suggested by Yin (1993) to set priorities. Interviews were recorded or notes were made and electronically saved.

After data collection data was prepared and analyzed regarding the integration of cultural aspects in supplier evaluation processes. Both within-case analyses and cross-case analysis were applied. Within-case analysis was used to identify unique patterns and to give the investigators rich familiarity with each of the cases. Cross-case analysis was applied to search for patterns to identify similarities and differences (Eisenhardt 1989). The selected procedure conforms to the requirements about validity and reliability of Yin (1989) to ensure a good research design. Results are presented in the following chapter.

3.2 Results

First, company approaches are introduced in Sect. 3.2.1 to give insights into different purchasing situations. Second, the integration of cultural criteria in supplier evaluation is presented in Sect. 3.2.2. Last, cultural problems with suppliers are described in Sect. 3.2.3.

3.2.1 Company Approaches

Expert 1 is only confronted with national suppliers for indirect material and simple services. The supplier selection starts with calls for tenders followed by negotiations. *Expert* 2 has to select regional suppliers for complex services from his customer countries. Communication problems are solved by means of translators. *Expert* 3 has to select suppliers for components already in the R&D phase. First, ideas are discussed with suppliers before selection. Countries differ regarding innovation. However, there is no ranking of countries. Principally, suppliers are determined regarding their competencies. *Expert* 4 selects suppliers for components after contacts. There are periodical evaluations and audits. Between suppliers there are perceptible cultural differences. *Expert* 5 has selected suppliers for projects in the country of the customer and cooperates with general suppliers over years. There are periodical pre-selections, audits, and evaluations for the component suppliers.

Cultural problems are avoided by means of contracts. *Experts* 6 and 7 use a systematic and comprehensive supplier selection and management process. Several types of simple and complex products and services are purchased. Language problems are avoided by means of bilingual suppliers. In general, cultural problems are avoided by means of know-how in the purchasing department or subsidiaries with own national staff. *Expert* 8 receives suggestions of recommended suppliers by the development department. Furthermore, supplier evaluation is applied. Several types of products and services are purchased.

3.2.2 Integration of Cultural Criteria in Supplier Evaluation

Regarding the company approaches we identified that cultural criteria to consider national and/or organizational cultural differences or similarities are not defined and integrated into supplier evaluation methods in the interviewed companies as suggested in the contributions of our literature review (Sect. 2.2). However, several cultural aspects are considered indirectly by other (main) criteria. For example, good or bad communication (*Experts* 6/7), common objectives (*Expert* 8), trust and same understanding (*Expert* 2), or public reputation (*Expert* 1) were mentioned.

Although the importance of national and organizational cultural criteria was ranked on average in the medium level by the experts, cultural criteria are not specifically considered by the companies in supplier evaluation. The main reason for the non-integration is that cultural issues play no or a minor role in contrast to the traditional evaluation criteria. Furthermore, we found that national culture aspects are not relevant in the situation where a company has to select a supplier of a determined country. We could also identify that there can be differences between the performance of suppliers of different countries. However, we could not identify that companies prefer suppliers of a country according to its national culture in international sourcing. This would be discrimination. Concerning organizational culture aspects there are no limitations for strategic partnership selection.

Furthermore, we asked at which steps in the supplier evaluation process (pre-selection, selection, controlling) cultural aspects should be included. All evaluation situations were mentioned. First, it is useful to include cultural criteria in pre-selection in order to avoid bad performance. Second, in supplier selection for an offer it was mentioned that, if other criteria are equally fulfilled, cultural criteria are useful. Third, in supplier controlling it is useful to consider problems regarding quality, communication, etc. due to different cultures in order to avoid future problems.

3.2.3 Cultural Problems with Suppliers

In order to identify cultural problems with suppliers we asked our interviewed experts about problems in relation to the different supplier evaluation dimensions to provide a holistic view. Janker (2008) differentiates the following dimensions: quality, information and communication, logistics, service, innovation, environmental, quantity, and payment performance. Furthermore, we added the social performance as dimension because social issues are also included in recent literature (e.g., Azadi et al. 2015).

First, regarding quality issues, problems due to different interpretations of requirements and different moral concepts with respect to quality can arise as reported by some of the interviewed experts. A mentioned action to avoid quality problems can be a clear definition of the requirements of the purchasing company. Second, in relation to the information and communication aspects, problems arise because of different languages, different systems (e.g., necessity of mandatory labelling), and different moral concepts regarding the communication of important information (e.g., feasibility, current status). Language problems can be avoided by means of interpreters or subsidiaries with foreign staff. Third, cultural problems in the logistical performance can arise due to different moral concepts. These can be avoided by means of contractual agreements. Fourth, there can be cultural problems with regard to the service performance according to different working times (holidays, time zones) and different values (e.g., reaction time). Fifth, regarding social issues, problems with foreign as well as domestic suppliers were mentioned. Such problems can be avoided by supplier selection or education. Furthermore, regarding environmental and innovation aspects, differences related to culture were mentioned; however, there were no problems. Last, cultural problems with respect to quantity and payment performance were not mentioned.

In summary, we could identify several problems related to different cultures. On the one hand, different moral concepts according to quality, communication, logistics, and service are a main reason for related problems. Such problems can arise because of both different national and organizational cultures. These problems can be avoided by a clear definition of requirements or rather with contractual agreements. On the other hand, there are problems because of different languages, systems, and working times due to different national cultures. Whereas language problems can be avoided by interpreters or foreign staff, requirements according to different systems have to be communicated. The problem of different working times cannot be avoided.

3.3 Discussion

This subsection discusses the importance of culture in supplier evaluation and gives recommendations for the integration of culture as evaluation criterion in the evaluation process. The importance of cultural criteria in supplier selection was rated on average in the medium level by the interviewed experts. However, most experts claimed that cultural issues play no or a minor role in comparison to traditional criteria in their supplier evaluation. Our results are in line with results of other papers. In literature we identified that the cultural match/fit or difference is less important. According to a study of Humphreys et al. (2003) company culture was ranked in the mid for collaborative relationships. In contrast to a study of Kannan and Tan (2002, 2003, 2004) the cultural match between the company and the supplier was regarded as one of the least important items in strategic supplier selection. Furthermore, Kar and Pani (2014) found that cultural differences between countries were not rated as being critical in the Indian manufacturing industries. Consequently, it seems that cultural differences are no ‘knock-out’ criteria in supplier evaluation and selection. However, we could identify a lot of problems with suppliers due to cultural differences so that the consideration of cultural differences is relevant in supplier evaluation. Due to the identified problems the aspects common values (regarding quality, etc.), different languages, systems, and working times should be included in supplier evaluation in practice to identify possible cultural problems. Afterwards, companies can use the mentioned approaches to overcome the identified cultural problems.

Last, we elaborated some recommendations for considering cultural differences in the supplier evaluation process which can be applied in practice. In pre-selection cultural criteria to avoid problems should be included. First, common moral concepts regarding quality, communication, logistics, and service should be recorded. In this context Briggs (1994) suggests the ranking of business subjects (e.g., manufacturing technology, quality improvement, cost reduction, etc.) regarding their importance (from 1 to 10) in order to identify the culture of the supplier. However, to assess different moral concepts regarding quality, communication, logistics, and service, we suggest the ranking or enquiry of aspects, for example, regarding communication (e.g., speed, accuracy, reliability), separately from the other dimensions in order to identify different values in each dimension. Moreover, supplier’s reputation in relation to ethical standards (e.g., treatment of employees) should be considered. Both, common values and supplier’s reputation are important for international sourcing and the selection of strategic partners. Second, in international sourcing communication barriers should be addressed. The existence of possible language barriers (e.g., possibility to communicate in English, own native speaking personnel in the country) should be identified. In the case of existing barriers, possible solutions should be worked out (e.g., use of an interpreter). Moreover, information about different systems, working times, and business customs should be gathered. Third, for the selection of strategic partners common objectives (e.g., management attitude/vision/outlook for the future), feeling of trust (e.g., honesty, professionalism, seriousness), the degree of (strategic) cooperation (e.g., cooperation understanding), and supplier-buyer-compatibility (e.g., top management, different functions) should be examined. Necessary information can be gathered by means of a supplier-self-assessment. If basic requirements are met, the supplier can be accepted in the supplier portfolio.

In supplier selection the mentioned cultural criteria can be included in order to select the best supplier for a certain order. For the cooperation it is advantageous to select the supplier with the least cultural differences. The weight of the cultural differences depends on the importance a company assigns to cultural differences. Because the focus is more on traditional criteria a low weight can be determined or the application as additional criteria is possible. In any case, it is important to try to avoid problems in relation to cultural differences by means of appropriate methods (e.g., definition of requirements, interpreter, etc.). Last, in supplier controlling bad performance regarding the different dimensions (e.g., communication, etc.) should be examined due to cultural aspects, especially if cultural differences were the cause of occurred problems. Evaluation results can be used for future supplier selections and actions to avoid cultural problems.

4 Conclusions

In this contribution we examined the integration of cultural criteria in supplier evaluation. On the basis of our literature review we distinguish between the national culture, relevant in international sourcing, and the organizational culture, relevant in the selection of strategic partners. In order to identify how companies integrate cultural issues in supplier evaluation and what problems arise due to cultural differences in practice we conducted an empirical study with purchasing experts. Main results are that the interviewed companies consider cultural issues indirectly but not as defined criteria in supplier evaluation, because the focus is more on traditional criteria. Furthermore, different moral concepts (regarding quality, communication, logistics, and service), different languages, different systems, and different working times are reported cultural problems in practice. On the basis of the empirical results and our literature review we recommend the inclusion of cultural criteria in pre-selection to avoid bad performance and in supplier selection to select the best supplier. Furthermore, in supplier controlling the performance of the supplier should be examined regarding cultural differences.

We contribute to the literature about supplier evaluation. In particular, we examine the possible application of cultural criteria in practice. Moreover, we contribute to literature about supplier relationship management by the identified actions to avoid cultural problems. Future research should address the application of cultural criteria in supplier evaluation on the basis of a quantitative survey and the question of how the inclusion of cultural criteria avoids problems due to national and organizational cultural differences.

References

Azadi M, Jafarian M, Saen RF, Mirhedayatian SM (2015) A new fuzzy DEA model for evaluation of efficiency and effectiveness of suppliers in sustainable supply chain management context. Comput Oper Res 54:274–285

Barkema HG, Bell JHJ, Pennings JM (1996) Foreign entry, cultural barriers, and learning. Strateg Manag J 17(2):151–166

Bayazit O, Karpak B, Yagci A (2006) A purchasing decision: selecting a supplier for a construction company. J Syst Sci Syst Eng 15(2):217–231

Beimborn D, Friedrich F, Blumenberg S (2009) The role of corporate cultural similarity for outsourcing relationship quality and outsourcing success. In: The 9th international conference on electronic business, pp 600–609

Bik OPG (2010) The behavior of assurance professionals: a cross-cultural perspective. Eburon, Delft

Briggs P (1994) Vendor assessment for partners in supply. Eur J Purch Supply Manag 1(1):49–59

Chan FTS, Chan HK, Ip RWL, Lau HCW (2007) A decision support system for supplier selection in the airline industry. Proc Inst Mech Eng B J Eng Manuf 221(4):741–758

Chen Y (2011) Structured methodology for supplier selection and evaluation in a supply chain. Inf Sci 181(9):1651–1670

Chou SY, Chang YH (2008) A decision support system for supplier selection based on a strategy-aligned fuzzy SMART approach. Expert Syst Appl 34(4):2241–2253

Cummins T, David M, Kawamoto K (2012) Contract and commercial management: the operational guide. Van Haren Publishing, Amersfoort

Das TK, Teng B-S (1998) Between trust and control: developing confidence in partner cooperation in alliances. Acad Manag Rev 23(3):491–512

Dogan I, Aydin N (2011) Combining bayesian networks and total cost of ownership method for supplier selection analysis. Comput Ind Eng 61(4):1072–1085

Eisenhardt KM (1989) Building theories from case study research. Acad Manag Rev 14(4): 532–550

Ellram LM (1990) The supplier selection decision in strategic partnerships. J Purch Mater Manag 26(4):8–14

Handfield RB (1994) US global sourcing: patterns of development. Int J Oper Prod Manag 14 (6):40–51

Hashemi SH, Karimi A, Tavana M (2015) An integrated green supplier selection approach with analytic network process and improved Grey relational analysis. Int J Prod Econ 159(1): 178–191

Heiman BA, Li W, Salvador GC, Aceves D (2008) Strategic, organizational, and cultural fit: effects on performance in China-U.S. joint ventures. J Asia Bus Stud 2(2):32–51

Hofstede G (2001) Culture's consequences: comparing values, behaviors, institutions and organizations across nations, 2nd edn. Sage Publications, Thousand Oaks

House RJ, Javidan M (2004) Overview of GLOBE. In: House RJ, Hanges PJ, Javidan M, Dorfman PW, Gupta V (eds) Culture, leadership, and organizations. Sage Publications, Thousand Oaks, pp 9–48

Huang SH, Keskar H (2007) Comprehensive and configurable metrics for supplier selection. Int J Prod Econ 105(2):510–523

Humphreys P, Shiu WK, Lo VHY (2003) Buyer-supplier relationship: perspectives between Hong Kong and the United Kingdom. J Mater Process Technol 138(1–3):236–242

Janker CG (2008) Multivariate Lieferantenbewertung: Empirisch gestützte Konzeption eines anforderungsgerechten Bewertungssystems, 2nd edn. Gabler, Wiesbaden

Kannan VR, Tan KC (2002) Supplier selection and assessment: their impact on business performance. J Supply Chain Manag 38(4):11–19

Kannan VR, Tan KC (2003) Attitudes of US and European managers to supplier selection and assessment and implications for business performance. Benchmarking Int J 10(5):472–489

Kannan VR, Tan KC (2004) Supplier alliances: differences in attitudes to supplier and quality management of adopters and non-adopters. Supply Chain Manag Int J 9(4):279–286

Kar AK, Pani AK (2014) Exploring the importance of different supplier selection criteria. Manag Res Rev 37(1):89–105

Kluckhohn C (1951) The study of culture. In: Lerner D, Lasswell HD (eds) The policy sciences. Stanford University Press, Stanford, pp 86–101

Kogut B, Singh H (1988) The effect of national culture on the choice of entry mode. J Int Bus Stud 19(3):411–432

Leung K, Bhagat RS, Buchan NR, Erez M, Gibson CB (2005) Culture and international business: recent advances and their implications for future research. J Int Bus Stud 36(4):357–378

Li JT, Lam K, Qian G (2001) Does culture affect behavior and performance of firms? the case of joint ventures in China. J Int Bus Stud 32(1):115–132

Min H (1994) International supplier selection: a multi-attribute utility approach. Int J Phys Distrib Logist Manag 24(5):24–33
Nukala S, Gupta SM (2007) A fuzzy mathematical programming approach for supplier selection in a closed-loop supply chain network. In: Proceedings of the 2007 POMS-Dallas meeting
Pothukuchi V, Damanpour F, Choi J, Chen CC, Park SH (2002) National and organizational culture differences and international joint venture performance. J Int Bus Stud 33(2):243–265
Rijamampianina R, Carmichel T (2005) A framework for effective cross-cultural co-opetition between organizations. Prob Perspect Manag 1(4):92–103
Sarkar MB, Echambadi R, Cavusgil ST, Aulakh PS (2001) The influence of complementarity, compatibility, and relationship capital on alliance performance. J Acad Mark Sci 29(4): 358–373
Sarkis J, Talluri S (2002) A model for strategic supplier selection. J Supply Chain Manag 38 (1):18–28
Stuart ID, McCutcheon D, Handfield R, McLachlin R, Samson D (2002) Effective case research in operations management: a process perspective. J Oper Manag 20(5):419–433
Thanaraksakul W, Phruksaphanrat B (2009) Supplier evaluation framework based on balanced scorecard with integrated corporate social responsibility perspective. In: Proceedings of the international multi-conference of engineers and computer scientists, vol II, pp 141–153
Wen UP, Chi JM (2010) Developing green supplier selection procedure: a DEA approach. In: International conference on IE&EM, pp 70–74
Yin RK (1989) Case study research: design and methods. Sage Publications, Newbury Park
Yin RK (1993) Applications of case study research. Sage Publications, Newbury Park
Yücenur GN, Vayvay Ö, Demirel NÇ (2011) Supplier selection problem in global supply chains by AHP and ANP approaches under fuzzy environment. Int J Adv Manuf Technol 56 (5–8):823–833

Part III
Supply Chain Coordination

RoRA-SCM—Review of the Integration of Flexibility Planning in Supply Chains

Immanuel Zitzmann

Abstract Planning in supply chains is characterized by uncertainties. These can be disruptive as well as operational risks, but also business opportunities. To deal with uncertainties, various approaches suggest supply chains that are robust, resilient, and agile (RoRA). To accomplish this, flexibility is necessary. Aspects of flexibility in supply chain management (SCM) have been analyzed from different angles. But still a holistic concept on how flexibility planning can be integrated in the process of supply chain planning does not exist. This paper gives an overview on optimization models used for flexibility planning in supply chains and identifies the characteristics of such a planning process. Which tasks of SCM the components of flexibility planning have to be integrated in is discussed along the structure of the supply chain planning matrix. An outlook on further research on flexibility planning in RoRA-SCM is also given.

1 Introduction

Competition takes place between supply chains not between single companies (Lambert and Cooper 2000). This knowledge leads to complex value creating networks that focus on efficiency and cost minimization. The drawback of this concentration is increased vulnerability (Craighead et al. 2007). Not only high risks but also low-probability/high-impact risks as well as operational uncertainties jeopardize supply chains (Klibi et al. 2010; Van Landeghem and Vanmaele 2002; Svensson 2000). Such dangers are increasing due to shorter product-life-cycles, market liberalization and new technologies. In order to address these issues, research suggests supply chains that are robust, resilient, and agile (RoRA). To create such networks, flexibility is necessary, a term and concept often mentioned in

I. Zitzmann (✉)
Lehrstuhl für BWL, insb. Produktion und Logistik, Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany
e-mail: immanuel.zitzmann@uni-bamberg.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_4

supply chain literature (Zitzmann 2014). This paper builds on existing literature on flexible supply chain management (SCM) and analyzes its contribution to creating RoRA supply chains. The principle of a model that integrates flexibility into supply chain planning is reviewed and the concept of planning flexibility is illustrated. To structure the different planning tasks, the supply chain planning matrix (SCPM) is used. The paper is ordered as follows.

In the second chapter, RoRA-SCM and flexibility planning are explained. Looking at a stochastic optimization model and analyzing its structure to plan flexibility is the first part of the third chapter. In the second part of the third chapter we look at the SCPM and identify the needed flexibility levels. The final chapter is the conclusion.

2 Theoretical Background

To handle uncertainties in value creating networks, RoRA-SCM is suggested. The core of such a management approach is to create flexibility within the network. Therefore, flexibility planning has to be integrated into the different planning tasks of SCM. Before flexibility planning is explained, RoRA-SCM is introduced as the foundation of this paper.

2.1 Robust, Resilient, Agile (RoRA) SCM

To explain RoRA-SCM, a definition of SCM is needed. This paper uses one from the Global Supply Chain Forum (Lambert and Cooper 2000): "*SCM is the integration of key business processes from end user through original suppliers that provides products, services, and information that add value for customers and other stakeholders.*" To create such uniform processes, the supply chain has to be planned on a strategic as well as on an operational level (Thomas and Griffin 1996). The planning tasks of SCM can be structured according to the SCPM which is illustrated in Fig. 2 (Stadtler 2005). It is important to note that these planning tasks are located within an organization. They have to be executed by each company within a supply chain. Additionally, process coordination between supply chain members is needed (Thomas and Griffin 1996). Fawcett and Magnan (2002) show how difficult the path to end-to-end integration is. In addition to these obstacles, supply chains as well as individual companies are operating in a dynamic environment that is characterized by uncertainties (Prater et al. 2001).

Uncertainties can be risks or chances for a supply chain (Simangunsong et al. 2012). Risks may have a disruptive nature or are operational (Tang 2006). To handle them, a supply chain should be robust and resilient (Klibi et al. 2010; Christopher and Peck 2004). To utilize opportunities the value creating network has to be agile (Naylor et al. 1999). Definitions of supply chains with these abilities and their intended goals are given in Table 1. The magnitude of each aspect depends on the individual circumstances of a supply chain (Cabral et al. 2012). Figure 1 shows that the core of all three concepts is flexibility (Zitzmann 2014). It is needed to create a RoRA supply chain. As well as other features of companies or value creation networks, flexibility has to be integrated into the management process. It can only be used when it is available. Also plans are needed on how the utilization is supposed to be executed. Therefore it is necessary to include thoughts about flexibility into supply chain planning which can be structured according to the SCPM.

The SCPM shown in Fig. 2 identifies the planning tasks of SCM (Stadtler 2005; Rohde et al. 2000). They are divided according to the functions of a firm, starting with procurement and followed by production, distribution, and sales. Corresponding to the time horizon, the tasks can also be structured from long-term to short-term. Literature suggested that *Strategic Network Planning* as well as *Master Planning* is executed by a supply chain coordinator. Different approaches and suggestions are made as to who this could be (Fawcett and Magnan 2002). So far such an institution does not exist. Therefore supply chain wide planning from source to sink is not possible (Li and Wang 2007; Pibernik and Sucky 2006). That does not mean that supply chain management is not happening. As there is no supply chain coordinator, every company plans its own value creating network. In certain points, these plans have to be coordinated with the supply chain partners (Arshinder et al. 2008; Thomas and Griffin 1996). An overview about supply chain collaboration is given by Min et al. (2005) or Barratt (2004). This paper focuses on the task of flexibility planning in the SCPM which is done by each company. This is shown in chapter "Integration of Cultural Aspects in Supplier Evaluation". Before that, the next paragraph looks at flexibility management in supply chains and flexibility planning.

Table 1 Definition and goal of robust, resilient and agile supply chains

Supply chain type	Definition	Goal
Robust supply chain	A supply chain that has the ability "to maintain its operational capabilities under different circumstances" (De Neufville 2004)	Match demand in an operationally uncertain environment
Resilient supply chain	Supply chain that has "the ability […] to return to its original state or move to a new, more desirable state after being disturbed" (Christopher and Peck 2004)	Handle internal and external disruptions of the supply chain and secure business continuity
Agile supply chain	"Using market knowledge and a virtual corporation to exploit profitable opportunities in a volatile market place" (Naylor et al. 1999)	Utilize chances in the marketplace to make profit

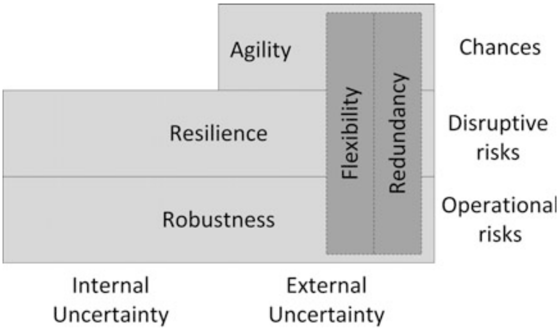


Fig. 1 Relationship between RoRA and flexibility in supply chains

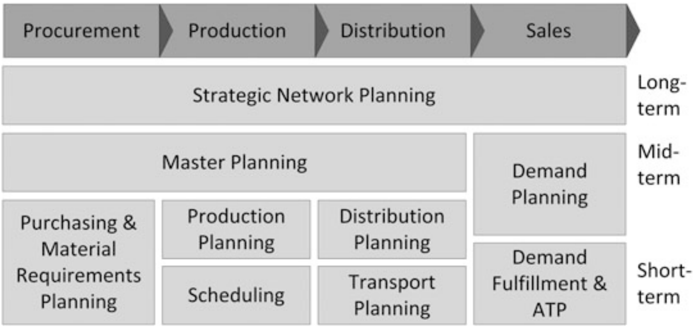


Fig. 2 The supply chain planning matrix (SCPM)

2.2 *Flexibility and Dimensions of Supply Chain Flexibility*

According to Garavelli (2003), "flexibility reflects the ability of a system to properly and rapidly respond to changes, coming from inside as well as outside the system". As mentioned in Sect. 2.1, this definition shows that flexibility is a suitable instrument to handle uncertainties which may be internal or external to a system and therefore to a supply chain (Tang 2006). Flexibility is a topic in economic and organizational literature as well as in the context of manufacturing (Sethi and Sethi 1990), but still there is no holistic flexibility theory. Sánchez and Pérez (2005), as well as Garavelli (2003), summarize the aspects of flexibility according to six dimensions. These include functional, hierarchical, measurement, and strategic aspects as well as the time horizon and the object of change. Different approaches exist on how flexibility may be achieved. The models of Vickery et al. (1999) and Duclos et al. (2003) are the most common concepts (Singer 2012). As the latter has

Table 2 Dimensions of supply chain flexibility

Dimension of supply chain flexibility	Explanation
Operations system flexibility	Ability of each supply chain node to adapt with its assets and operations as needed
Market flexibility	Designing and modifying products in close relationship with customers
Logistics flexibility	React cost-efficiently to changing locations of supplier and costumer
Supply flexibility	Ability to reconfigure the supply chain
Organizational flexibility	Flexibility of labor force skills to adjust to needed service/demand requirements.
Information system flexibility	Ability to change information system according to new circumstances

a more process orientated view, it is more suited when looking at supply chains. It has six dimensions of flexibility which are shown in Table 2. The dimensions of Duclos et al. (2003), as well as the attributes of other approaches, describe features of a flexible supply chain. They also explain what these dimensions are used for, yet no guideline on how to plan flexibility in supply chains is given.

Suggestions on how flexibility can be integrated in planning are made by Kaya et al. (2014), Pibernik (2001), Mulvey et al. (1995), Kouvelis et al. (1992) and others. Here, dynamic stochastic optimization models are used to handle uncertainties in the planning process. The integration of uncertainties by stochastic variables leads to configurations that are more flexible. Such a model and the integration of its principles into SCM are the focus of the next chapter.

3 Potential and Execution Flexibility in Supply Chain Models and the SCPM

In order to analyze how flexibility in supply chain planning works, the following chapter begins with an illustration of the model of Pibernik (2001), which is used as an example for planning flexibility with the help of dynamic stochastic optimization. The analysis shows the principle behind flexibility planning. Where this concept should be integrated into supply chain planning is considered in Sect. 3.2, which is structured according to the SCPM introduced in Chapter "Considering Small and Medium-Sized Suppliers in Public Procurement—The Case of the German Defence Sector".

3.1 Potential and Execution Flexibility in Supply Chain Flexibility Planning Models

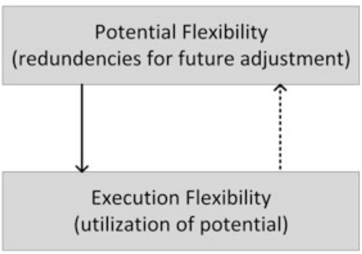
The model of Pibernik is only described informally here (for the mathematical depiction see Pibernik 2001). The purpose is to show its functional principle and to identify the components of planning flexibility.

As a hierarchical planning model the approach has two levels. The first, strategic one looks on the installation or modification of flexibility potential. The objective function maximizes the expected surplus of net payments max ENZ for all periods $t(t = 1, \ldots, T)$. In addition to the discount factor $(1 + k)^{-t}$, the function consists of the net payments of each period $\tilde{N}\tilde{Z}_{t+1}$ which depend on three components. The first one is the initial situation $\bar{s}_1$ in period $t = 1$ respectively the situation s_t in period $t = 2, \ldots, T$. Secondly, the decision of the management has to be taken into account. It is the decision which flexibility potential should be installed $fp_{j,1}^{install}$ in $t = 1$. In all following periods $t = 2, \ldots, T$, a decision has to be made about the modification $fp_{j,j(t)}^{\text{mod}}$ of the installed system. Finally, the stochastic manifestation of variables has to be considered. They are represented with $\tilde{r}_{\varpi,1}$ in $t = 1$ and $\tilde{r}_{\varpi,t}$ in $t = 2, \ldots, T$. This strategic level works with net payments which depend on gross cash flows BZ. They are generated by the second level of the hierarchical model.

In the operational model, a supply chain with the four process-levels supply, production, storage, and delivery is considered. Maximizing the gross cash flow max BZ_{t+1} in all periods after the installation of the flexibility potential $t = 2, \ldots, T$ is the goal of the objective function. The gross cash flows BZ consist of incomes and different kinds of costs (e.g. for transportation, production, storage, as well as exchange rates). The framework of income and costs is given by the three components (s_t, $fp_{j,t}$ and $r_{\bar{\pi}}^{m(t)}$) of the strategic level.

The described planning approach is a basic model which shows the principle of flexibility planning in mathematical models. It is structured according to the idea of hierarchical planning, shown in Fig. 3. Here, management decisions are divided into two interdependent tasks (Schneeweiß 1995). On the top-level long-term decisions are made. The model provides decision support about which flexibility potential should be installed. Information is incomplete and operational developments have to be anticipated. This is done on the base-level. Costs and profits of the supply chain performance are estimated here. The radius of action is limited by

Fig. 3 Hierarchical structure of flexibility planning



previous choices on the top-level. Using the hierarchical approach for managing flexibility means potential creating on the top-level which can be used on the base-level when needed. The available redundancies shall be called *potential flexibility* (*PF*). Planning *execution flexibility* (*EF*) means the development of emergency plans and guidelines to adjust processes in reaction to unexpected events. As this task considers available potential, it depends on the result of the top-level planning and is located on the base-level (Pibernik 2001).

Figure 3 summarizes the components of flexibility. PF as well as EF is needed. It also shows that a hierarchical dependency exists between these dimensions. In which planning task of the SCPM what kind of flexibility is needed is analyzed in the next section.

3.2 Potential and Execution Flexibility Requirements in the SCPM

For the creation of a RoRA supply chain that is able to handle uncertainties, it is necessary to integrate flexibility planning into the planning tasks of SCM. It has to be determined whether PF, EF, or both are needed. The analysis starts with the long-term planning tasks according to the SCPM (Meyr et al. 2008; Stadtler 2005; Rohde et al. 2000) and then continues with the more operational planning processes. As the tasks of *Production Planning* and *Scheduling* as well as *Distribution Planning* and *Transportation Planning* are often executed simultaneously, they are considered together (Stadtler 2005).

Strategic Network Planning defines the structure of the supply chain (Kauder and Meyr 2009). In contrast to what is often stated in literature on SCM this task does not automatically involve the whole process from suppliers of raw materials to end-user. There is no institution that designs the entire supply chain. Supply chains emerge from cooperation between individual companies. Therefore, *Strategic Network Planning* makes decisions about locations of warehouses, production plants, or possible suppliers within the value creating network of an organization. The consideration of changes in productivity, demand fluctuation, and emergencies in *Strategic Network Planning* lead to the creation of PF.

The planning process starts on the sales side with ***Demand Planning***. A variety of qualitative and quantitative methods exists to estimate future demand. The nature of forecasts is that they are not accurate. Differences between forecasts and real demand are the reason for demand uncertainty. Depending on the scale of the variance, more or less flexibility is needed. However, flexibility planning is not done in *Demand Planning*.

Balancing demand forecasts with available capacities is a major task of ***Master Planning***. It identifies bottlenecks and uses available flexibility. For this purpose EF within the organization is needed. In addition to capacity management, medium-range lot-sizing is the responsibility of *Master Planning* (Pibernik and Sucky 2007). This means, determining not only production quantities but also order sizes from suppliers. Lot-sizes specify the amount of raw-material and intermediate as well as finished products that are available for the next production and planning steps. This framework is again PF which can be used in the following planning tasks. Out of the perspective of flexibility, *Master Planning* can therefore be divided into a level where EF is used and a second level where PF is planned.

Detailed lot-sizing and order quantities are determined within the planning process of ***Purchasing and Material Requirements Planning***. It uses the medium-term quantity planning from *Master Planning* to plan the production of actual demand or detailed forecasts. EF is utilized to handle bottlenecks and reach service-levels with low costs. Flexible contracts were designed by *Master Planning*. By using EF this potential is exploited.

Again, depending on the production technology, ***Production Planning*** and ***Scheduling*** are executed simultaneously or sequentially. Both tasks are located on site or even on the production department level. The results of the planning process are production orders, detailed timetables, order sequences, and staff planning. All decisions are executed within the framework given by *Master Planning*. Concerning flexibility, only EF is applied here.

Distribution and ***Transportation Planning*** plans the flow of goods within the value creating network. Again, the results of *Master Planning* have to be considered. Given stock levels and lot-sizes are used to deliver goods to customers or warehouses. The detailed planning coordinates supply and demand and determines vehicle routing. This short-term planning uses PF created by *Strategic Network Planning* and *Master Planning* to fulfill customer demands. Therefore EF is planned and utilized here.

Order promising is the primary task of ***Demand Fulfillment and Available-to-Promise (ATP)***. It consists of matching customer orders with available inventory or production capability. The latter is called capable-to-promise. EF is used to make these promises.

Table 3 summarizes the need for flexibility in the tasks of the SCPM. Strategic and tactical decisions are concerned with PF. The example in Sect. 3.1 shows that models for planning flexibility in supply chains can help here. The characteristics of *Strategic Network Planning* correspond with the assumptions of the presented type of model. Forecasts about future developments must be taken into account and investments are significant and long-term. This justifies the high effort of such planning models. Before such a model is implemented, the selection of an appropriated objective function and the right model for the specific situation should be considered (Scholl 2001). For plans that have to be reviewed more often than the long-term investment decisions such an effort cannot be justified. The cost-benefit ratio would not be appropriate. Here, a heuristic that can be performed repeatedly with low effort is needed to support decisions about PF in *Master Planning*. Stochastic optimization models are only designed to give advice for planning PF. Though the operational level exists, no plans for EF are given.

Table 3 Flexibility planning in the SCPM

Planning task	Planning purpose	Flexibility planning	Flexibility level
Strategic network planning	Designing structure of supply chain	Yes	Potential
Master planning	Capacity leveling	Yes	Potential and execution
Purchasing and material requirements planning	Material supply and bottleneck management	Yes	Execution
Production planning and scheduling	Creating production plans	Yes	Execution
Distribution and transport planning	Material flow management	Yes	Execution
Demand planning	Planning expected demand	No	–
Demand fulfillment and ATP	Planning order execution	Yes	Execution

Such plans are necessary to react in a timely manner to unexpected events. With the help of supply chain event management systems such adjustments can be executed (Liu et al. 2007; Otto 2003; Nissen 2002). Here, decisions are made about how EF is best used. Therefore, event management systems need guidelines from the tactical and operative level of the SCPM were EF is planned. Today, *Purchasing and Material Requirements Planning* or *Distribution and Transport Planning* work with given capacities. They do not consider flexibility. Thoughts about the tradeoff between costs for the activation of EF and the costs resulting from unexpected events have to be integrated in the tasks of the SCPM. Also, emergency plans must be created as guideline for event management systems.

4 Conclusion and Outlook

Flexibility and its planning are key to establishing RoRA supply chains. By looking at a stochastic dynamic optimization model for flexibility planning in supply chains, two levels of flexibility could be identified: PF and EF which are related to each other in a hierarchical way. Only if PF exists are the operational systems able to use EF to handle uncertainties. By analyzing the SCPM, areas in which flexibility planning is needed could be determined. This is the case in all tasks but *Demand Planning*. Furthermore, the individual tasks of supply chain planning were analyzed to see whether PF or EF or perhaps both are relevant. Because of their structure, complexity, and implementation efforts optimization models are only useful for long-term planning of PF which is done in *Strategic Network Planning*. *Master Planning* is the second task of the SCPM where PF planning is needed. As the introduced model is too elaborate for this planning level, a heuristic is required that can easily be applied. Such a heuristic can be developed on the basis of existing

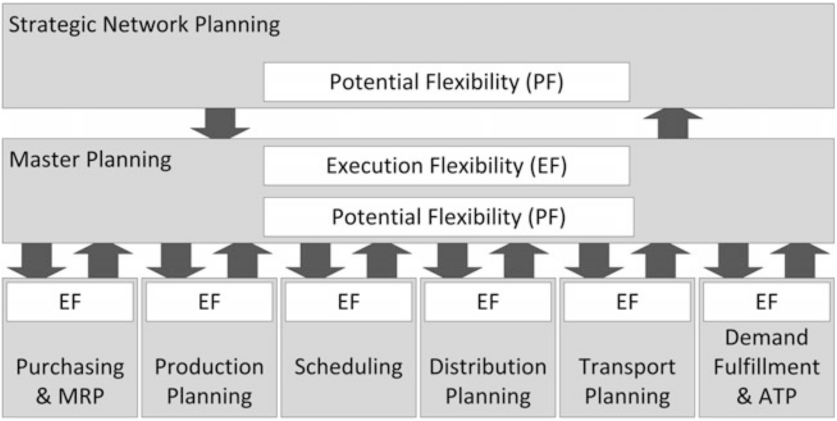


Fig. 4 Flexibility planning as part of SCM

models for flexibility planning. Unfortunately, these approaches give no guideline for planning EF which is used by supply chain event management to act on unexpected events. Additional research for integrating EF in the tactical and operational level of supply chain planning is necessary. It is important that these planning instruments do not stand for themselves. They have to be part of the normal SCM process as can be seen in Fig. 4. Here, PF and EF are part of the tasks in the SCPM. Resulting plans include flexibility aspects. If it were possible to develop a heuristic planning tool to establish PF and also a pragmatic approach on how EF is planned and integrated, huge steps in the direction of creating RoRA supply chains could be done.

References

Arshinder AK, Deshmukh SG (2008) Supply chain coordination: perspectives, empirical studies and research directions. Int J Prod Econ 115(2):316–335

Barratt M (2004) Understanding the meaning of collaboration in the supply chain. Supply Chain Manag Int J 9(1):30–42

Cabral I, Grilo A, Cruz-Machado V et al (2012) A decision-making model for lean, agile, resilient, and green supply chain management. Int J Prod Res 50(17):4830–4845

Christopher M, Peck H (2004) Building the resilient supply chain. Int J Logist Manag 15(2):1–14

Craighead CW, Blackhurst J, Rungtusanatham MJ, Handfield RB (2007) The severity of supply chain disruptions: design characteristics and mitigation capabilities. Decis Sci 38(1):131–156

De Neufville R (2004) Uncertainty management for engineering systems planning and design. Engineering Systems Monograph MITesd: March 29–31

Duclos LK, Vokurka RJ, Lummus RR (2003) A conceptual model of supply chain flexibility. Ind Manag Data Syst 103(6):446–456

Fawcett SE, Magnan GM (2002) The rhetoric and reality of supply chain integration. Int J Phys Distrib Logist Manag 32(5):229–361

Garavelli AC (2003) Flexibility configurations for the supply chain management. Int J Prod Econ 85:141–153

Kauder S, Meyr H (2009) Strategic network planning for an international automotive manufacturer —balancing flexibility and economic efficiency. OR Spectr 31:507–532

Kaya O, Bagci F, Turkay M et al (2014) Planning of capacity, production and inventory decisions in a generic reverse supply chain under uncertain demand and returns. Int J Prod Res 52 (1):270–282

Klibi W, Martel A, Guitouni A et al (2010) The design of robust value-creating supply chain networks: a critical review. Eur J Oper Res 203(2):283–293

Kouvelis P, Kurawarwala AA, Gutiérrez GJ (1992) Algorithms for robust single and multiple period layout planning for manufacturing systems. Eur J Oper Res 63:287–303

Lambert DM, Cooper MC (2000) Issues in supply chain management. Ind Mark Manag 29 (1):65–83

Li X, Wang Q (2007) Coordination mechanisms of supply chain systems. Eur J Oper Res 179 (1):1–16

Liu R, Kumar A, van der Aalst W (2007) A formal modeling approach for supply chain event management. Decis Support Syst 43(3):761–778

Meyr H, Wagner M, Rohde J (2008) Structure of advanced planning systems. In: Stadtler H, Kilger C (eds) Supply chain management and advanced planning. Springer, Berlin, pp 109–116

Min S, Roath AS, Daugherty PJ, Genchev SE, Chen H, Arndt AD, Richey RG (2005) Supply chain collaboration: what's happening? Int J Logist Manag 16(2):237–256

Mulvey JM, Vanderbei RJ, Zenios SA (1995) Robust optimization of large-scale systems. Oper Res 43(2):264–281

Naylor JB, Naim MM, Berry D (1999) Leagility: integrating the lean and agile manufacturing paradigms in the total supply chain. Int J Prod Econ 62(1/2):107–118

Nissen V (2002) Supply chain event management. Wirtschaftsinformatik 44(5):477–480

Otto A (2003) Supply chain event management: three perspectives. Int J Logist Manag 14(2):1–13

Pibernik R (2001) Flexibilitätsplanung in Wertschöpfungsnetzwerken. Gabler, Wiesbaden

Pibernik R, Sucky E (2006) Centralised and decentralised supply chain planning. Int J Integr Supply Manag 2(1/2):6–27

Pibernik R, Sucky E (2007) An approach to inter-domain master planning in supply chains. Int J Prod Econ 108(1/2): 200–212

Prater E, Biehl M, Smith MA (2001) International supply chain agility—tradeoffs between flexibility and uncertainty. Int J Oper Prod Manag 21(5/6):823–839

Rohde J, Meyr H, Wagner M (2000) Die supply chain planning matrix. PPS Manag 5:10–15

Sánchez AM, Pérez MP (2005) Supply chain flexibility and firm performance—a conceptual model and empirical study in the automotive industry. Int J Oper Prod Manag 25(7):681–700

Schneeweiß C (1995) Hierarchical structures in organisations: a conceptual framework. Euro J Oper Res 86:4–31

Scholl A (2001) Robuste planung und optimierung. Physica, Heidelberg

Sethi AK, Sethi SP (1990) Flexibility in manufacturing: a survey. Int J Flex Manuf Syst 2:289–328

Simangunsong E, Hendry LC, Stevenson M (2012) Supply-chain uncertainty: a review and theoretical foundation for future research. Int J Prod Res 50(16):4493–4523

Singer C (2012) Flexibilitätsmanagement zur Bewältigung von Unsicherheit in der supply chain. Eul, Köln

Stadtler H (2005) Supply chain management and advanced planning—basics, overview and challenges. Euro J Oper Res 163:575–588

Svensson O (2000) A conceptual framework for the analysis of vulnerability in supply chains. Int J Phys Distrib Logist Manag 30(9):737–749

Tang CS (2006) Perspectives in supply chain risk management. Int J Prod Econ 103(2):451–488

Thomas DJ, Griffin PM (1996) Coordinated supply chain management. Euro J Oper Res (94):1–15

Van Landeghem, Vanmaele H (2002) Robust planning: a new paradigm for demand chain planning. J Oper Manag 20:769–783

Vickery S, Calantone R, Dröge C (1999) Supply chain flexibility: an empirical study. J Supply Chain Manag 35(3):16–24

Zitzmann I (2014) How to cope with uncertainty in supply chains?—conceptual framework for agility, robustness, resilience, continuity and anti-fragility in supply chains. In: Kersten W, Blecker T, Ringle CM (eds) Next generation supply chains: trends and opportunities. Springer, Berlin, pp 361–377 (epubli)

Coordination in Multimodal Supply Chains: Drafting an Agent-Based Exchange

René Föhring and Stephan Zelewski

Abstract The more efficient configuration and coordination of multimodal transports is a topic consistently emerging in the last decades. Different projects, like e.g. LOGFOR, CODE24 and CENTRAL EUROPE, make efforts to achieve progress in this area. In this paper an already implemented and tested software prototype for the configuration of multimodal supply chains is presented. It is described how this prototype, which emerged from the CODE24 project, is able to facilitate contact between potential business partners. Subsequently, the paper shows how the implementation of the prototype and the research into freight exchanges led the authors of this paper to a new marketplace concept: Agent-based Freight Exchanges. These yet to be implemented, highly automated and interconnected marketplaces are designed to address problems commonly associated with existing intermediaries. They will provide support for decentralized and autonomous software agents to perform contractually binding auctions of multimodal freight transport services. It is also shown how the agents will utilize a double-sided combinatorial auction model to achieve this. Finally, an outlook on prospective concepts which support the negotiation of contracts for multimodal transport services using multi-agent systems is given.

1 A Brief Introduction to Online Freight Exchanges

Freight exchanges are marketplaces where offers for and demands for transport services find one another. Contrary to forwarders, which constitute the classic form of freight mediation, they themselves are no participants in the processing of transport services. They merely mediate transport services between shippers and carriers. The majority of the companies specializes in the mediation of truck

R. Föhring (✉) · S. Zelewski
Institute for Production and Industrial Information Management, University of Duisburg-Essen, Universitätsstraße 9, 45141 Essen, Germany
e-mail: rene.foehring@pim.uni-due.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_5

freights. By contrast, multimodal transports are being mediated fewest of all (Merkel and Kromer 2002). Since their origination in the 1970s and 1980s the freight exchanges conducted their business primarily via telephone, telefax and BTX. With the advent of the internet in the 1990s and 2000s the rise of e-commerce platforms opened up new sales channels and provided a more transparent and comprehensive offer for demanders.

2 Establishment of an Online Freight Exchange Within the Framework of the CODE24 Project

The multinational joint project CODE24 has been started in the year 2010. For an overview of the project within the framework of the INTERREG-IVB-NWE program of the European Union (EU), it is being referred to Brenienek (2014).

The primary goal of the joint project consists in the integration and advancement of the activities on the trans-European transport axis no. 24, the main railroad line through the Swiss Alps, which connects the harbors of Rotterdam and Genoa.

The challenges here are manifold: Comprehensive and publicly accessible information on how many freight trains will use the corridor is currently missing. It is also uncertain how much this capacity can be improved through a higher utilization of the existing infrastructure. Finally, a considerable market non-transparency exists for forwarders that take a transport by rail into consideration, especially regarding the connection possibilities to freight transports (Endemann et al. 2012).

As a result, a central component of the project is the conception and implementation of an online freight exchange (Endemann and Kaspar 2011). As a first step towards this goal, the Institute for Production and Industrial Information Management of the University Duisburg-Essen systematically ascertained the requirements of the essential logistical actors for an online rail freight exchange. This was achieved through the analysis of the relevant literature as well as interviews and workshops with industry experts (Bruns et al. 2010; Habib and Bruns 2012; Klippert et al. 2013). Further research results regarding user requirements were contributed by project partners of the institute (Dörr and Endemann 2014; Endemann et al. 2012). One of the most important conclusions was that a freight exchange which only supports rail freight traffic has no realistic market potential. A detailed market analysis consequently showed that no such online freight exchange could establish itself on the European transport market in the long term (Klippert et al. 2013). Especially the transport carrier road has to be involved in order to be able to exhaust the potential of multimodal transport chains.

The software prototype ORFE (“online rail freight exchange”) was implemented based on this research. It demonstrates functionalities for the facilitation of contact between potential business partners and the configuration of multimodal supply chains. Since the interviewed experts concluded that future users would hesitate to enter any monetary information, the prototype was build to support the pure mediation between potential business partners. As a consequence, it cannot guarantee contractually binding business transactions as these have to take place outside of the online platform after the contact initiation.

For detailed overviews on the concept and software development it is being referred to Bruns et al. (2012b), Föhring and Zelewski (2013).

After the conceptualization and implementation of the ORFE prototype, the project consortium agreed that the final version of the prototype would have to be reimplemented into a commercial software product. Additionally, a viable business model would have to be developed for its operation. For an early review of this work it is being referred to Dörr and Endemann (2014).

It was very important for all questioned project partners and also for other interviewed experts that the future operator of the online freight exchange behaves in an economically impartial way towards all exchange users. This demand can be attributed to the high intensity of competition and mutual distrust in the railway sector (Dörr and Endemann 2014; Klippert et al. 2013).

Currently two potential operators try to establish themselves on the market (Dörr and Endemann 2014): "Railcargo-Online" (http://www.railcargo-online.com) and "Freit-One" (http://www.freit-one.de). Both companies were given access to the ORFE prototype as a working basis and have started operations in late 2013.

3 Real Problems in Operating an Online Freight Exchange

The research around the CODE24 project revealed further obstacles to the successful establishment of an online freight exchange: If the virtual marketplace fails to reach the critical mass and provide a sufficient mediation rate, forwarders and transport carriers will keep settling their transactions the traditional way. Furthermore, freight exchanges are primarily suited for the mediation of transport services that are dealt with through spot markets, but many transports carried out within Europe are still bound to contracts. Therefore a potential exchange has to either control the existing spot market or strengthen the "spot character" of transport services in general (Merkel and Kromer 2002).

The requirement analysis for the ORFE prototype showed that the establishment of an online freight exchange in general meets four central real problems:

The first problem is the need for a business model that enables at least the loss-free operating of the marketplace and specifies a royalty for every user of the online freight exchange (Bruns et al. 2012a).

The second problem is the disclosure of competition-sensitive data to the future operator. All participants of a centrally organized marketplace are required to submit their data to the central operator in order for him to be able to perform his function as an intermediary. This requires a high confidence in the discretion of the operator.

The third problem is the demanded industry experience of the future operator. The role of the operator of an online freight exchange requires intimate knowledge of the respective transport sector. Yet at the same time the potential marketplace members will question his neutrality. It is therefore difficult to find an operator that has the necessary expertise but is not at the same time a participant of the market in any form (Bruns et al. 2012a).
The fourth problem is the consideration of multimodal transports. The ability to configure transports across different carriers is a requirement which can be found regularly in publications on the requirements for an online freight exchange (Endemann et al. 2012; Habib and Bruns 2012).

The challenge in solving the first three problems lies in the minimization of the costs of operation and participation and the believable guarantee of the neutrality, discretion and expertise of the operator. It becomes apparent that any future online freight exchange should support multimodal freight traffic by taking several traffic carriers into consideration for any given transport. Furthermore, it becomes clear that the first three problems can be attributed to the centralized nature of the marketplace. A single operator has to bear the costs for the provision of the infrastructure and will dispose of the data of all members. Moreover, he would have to reassure potential users about his expertise for the purpose of customer acquisition.

It should therefore be researched if an automated and *decentralized* approach would be an economically attractive alternative to the so far pursued *centralized* approaches. The basic premise of this idea is that a network of agents can form an interconnected marketplace in which they participate as equal trading partners. The agents are provisioning the computational infrastructure through the combination of their individual computing power where all agents share the same set of data amongst themselves. A single, central operator would not be needed, alleviating the first three real problems. The support of multimodal traffics would be easier to realize in an automated freight exchange than in an exchange organized in a central and purely contact mediating way, since the coordination could be left to the agents. Finally, an agent-based system could even strengthen the "spot character" of multimodal transport services.

In the following chapter chosen aspects and requirements for the development and implementation of such an agent-based freight exchange are presented.

4 A Concept for Agent-Based Freight Exchanges

4.1 State of Research

There are not many publications on the topic of "online freight exchange for transport services in the rail freight traffic". The majority of the publications on this topic were published within the project CODE24 at the Institute for Production and

Industrial Information Management of the University Duisburg-Essen (Föhring and Zelewski 2013; Klippert et al. 2013; Föhring et al. 2012; Habib and Bruns 2012; Bruns et al. 2010, 2012b; Bruns and Zelewski 2011). Beyond that, only few publications exist and from these many merely assert the need for such an exchange (Endemann and Kaspar 2011; Scheck and Wilske 2011). The usage of double-sided combinatorial auctions, on the other hand, is discussed elaborately in specialized literature for different markets (Ackermann et al. 2011; Parkes and Ungar 2001) just as is the usage of multi-agent systems (Davidsson et al. 2005; Fox et al. 2000; Jennings 2000).

The paper at hand suggests the merging of these findings on the requirements for an online rail freight exchange, on the usage of double-sided combinatorial auctions as well as on the organization of autonomous multi-agent systems in order to enable the conception and prototypical development of an agent-based freight exchange (or AFEX for short).

The proposed design is an automated exchange in the form of an electronic marketplace. It is organized as a decentralized system which is able to function without a central marketplace operator. The autonomous trade between equal actors is being enabled by the usage of agents that form a multi-agent system and employ double-sided combinatorial auctions in order to perform auctions of multimodal transport services. The subsequent prototypical implementation of AFEX will have a graphical user interface through which each human user can control his instance of the agent software.

4.2 Multi-Agent Systems as Decentralized Electronic Marketplaces

While traditional electronic marketplaces require a central operator, an AFEX-system has to be able to organize itself in a decentralized way. This means that, while in case of the central solution all market activity is coordinated by the marketplace operator, the configuration and coordination of the activities in the decentralized version happens by the actors themselves. The marketplace operator is no longer needed as an intermediary; a disintermediation of the trade chain occurs.

In order to develop a multi-agent system that is capable to coordinate itself without a central nodal point, the first requirement is that agents have to be able to locate trade partners. This is a nontrivial problem, as a central authority for mediating the contact between the agents is missing. This “contact problem” can, however, be solved if the agent software enables the manual entry of agent addresses. These describe the necessary information for making contact with another agent through the internet (i.e. an IP address and a port number).

Every time an agent contacts another agent they exchange all contact information known to them. Through this approach each agent gets to know the whole network known to the other agent. The agent software then has to save the gathered contact information in a way that enables it to contact the known agents again after a restart.

The contact problem can be solved substantially more user-friendly if other software agents can be discovered without requiring user interaction. For this purpose there should be one or more predefined agent instances on the internet whose fixed contact information is embedded in the agent software. These predefined agents have no trading preference but serve as a kind of beacon, i.e. their sole purpose is to answer contact requests. If a list of these “beacon-agents” is going to be embedded in all agents and stands at their disposal after installation, they can be contacted without of the human user.

Figure 1 illustrates the process: Agent A does not yet know other agents beside the beacon-agent B. He contacts agent B and gets further agent addresses from him, e.g. those of agent C and agent D. Agent A saves the received contacts and can recall them again at the next start and approach them without being dependent on the beacon-agent as a contact mediator.

The advantage of this method is that beacon-agents can be operated, communicated and used independently of each other. They support the decentralized organization of the AFEX marketplace since they solve the contact problem without requiring a user interaction. They are, however, not necessary for operating the decentralized network (as the human users could always build up their own “contact networks” with the manual entry method).

4.3 Capturing Trading Preferences

The agent software has to be usable by a human user. For this purpose an agent’s user can use input masks provided by the user interface to either capture his

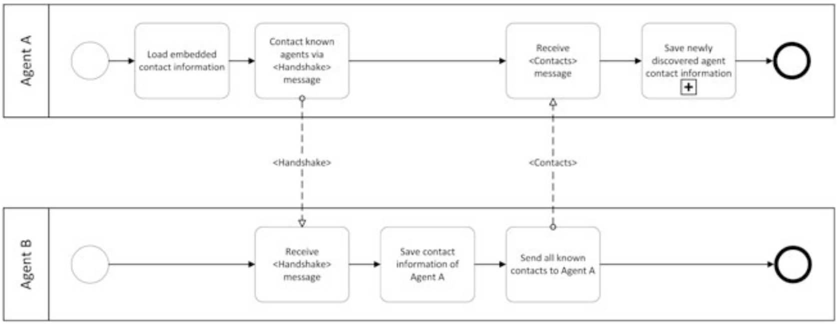


Fig. 1 Contact initiation between agents

preferences for an offer or demand for a transport service. This way he specifies similar transport-related preference data (for loading and unloading location, timeframe, etc.) that has also been captured in the ORFE prototype. Figure 2 illustrates this process schematically.

The difference between the ORFE prototype and AFEX becomes apparent afterwards: After the entry of their preferences the AFEX system does not need further user interaction. All agents that are in contact with each other exchange their preferences automatically and thus ensure complete market transparency. The preferences are hereby transmitted in a unified data format which all agents understand.

4.4 Coordinating Group Formation

Once started with a set of preferences, agents will always advertise the transport services their user offers and try to buy those transport services their user demands. For the sake of simplicity, agents that are offering transport services will be called "suppliers" in this paper (and agents that are demanding transport services "demanders").

All agents know the preferences of all other agents in their network and all preference data is exchanged in a unified format. Therefore it is possible for demanders to determine whether or not the transport services offered by a subset of suppliers can be combined in a way to accommodate at least one of their demands. If this is the case, the demander in question will contact the relevant suppliers and look for a "group" in which the demanded goods (transport services) are advertised. If no such group can be found, the demander will ask the suppliers to form one. In this group the agents will be able to auction the demanded goods. Other agents can find and join the group. Sometimes a demander will conclude that two groups would have to combine their auctioned goods to be able to accommodate one of his

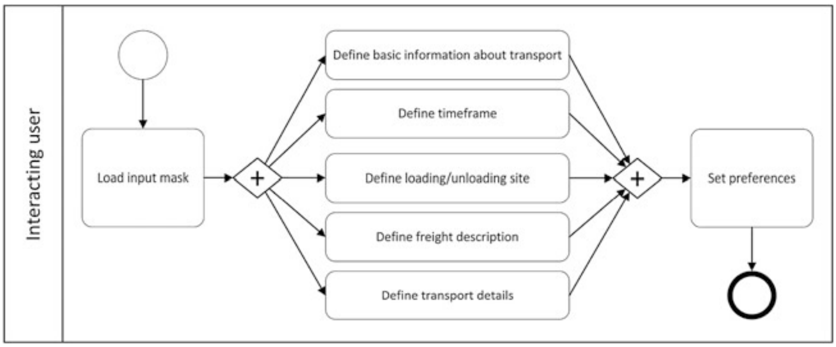


Fig. 2 Input process for a demand for a transport service

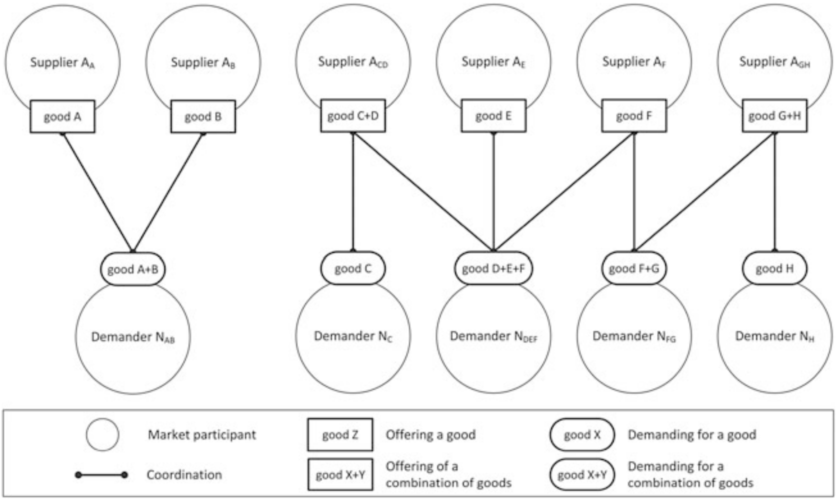


Fig. 3 Group formation

demands. In these cases he can ask both groups to merge in order to form a larger group which addresses more suppliers and demanders.

This way, the demanders in a network assist the suppliers in forming the right groups to ensure constant trading of the demanded goods in the network.

Figure 3 shows a schematic representation of a small AFEX system where two groups have formed: In group 1 two suppliers are offering two goods to one demander, in group 2 four suppliers are offering six goods to four demanders.

The group concept not only enables agents to frame the coordination of their efforts but also ensures that all participants in a group are interested in the offered goods. These aligned interests are a perquisite for the negotiation of prices.

4.5 *Using Double-Sided Combinatorial Auctions for Price Negotiation*

The pricing between buyer and seller is a challenge any marketplace faces. Three pricing models can be made out (Grieger 2003):

- The *bulletin board model* that serves primarily for the publication of advertisements and as a pure information and contact platform (this variation was implemented in the ORFE prototype).
- The *fixed price model* in which case the supplier and demander specify the final price for the service being in demand or offered.
- *Virtual exchanges* made possible by the internet that offer its members a dynamic pricing with the help of auctions.

For the design of an AFEX system, the chosen concept should ensure an efficient auction of the traded transport services. From the three mentioned alternatives, this requirement can be only met by the dynamic pricing through auction. The choice of auction form is crucial for the efficiency of the auction execution (Ausubel and Cramton 1998; Krishna and Perry 1998). There are two reasons why the employment of the double-sided auction form, in which case the auction participants can appear as buyer and as seller, is reasonable: Firstly, many exchanges and resource markets in the real world are organized as double-sided auctions (Yang 2002). Secondly, the participants are not assigned dedicated roles ("supplier" or "demander") but can act as both at any point according to their preferences.

In order to be able to depict multimodal transport services in an auction, other dimensions next to the price have to be taken into account when computing the optimal allocation of goods. Multidimensional auctions promise a high allocative efficiency despite the possibly complex preferences of the participants concerning the traded dimensions. Combinatorial auctions, sometimes called combinatorial exchanges, are very well researched multidimensional auctions that make it possible for participants to submit bids for indivisible combinations of goods and only win the bid if they receive exactly the desired combination (Bichler et al. 2005).

Resulting from these considerations it becomes clear that a double-sided combinatorial auction model meets the previously mentioned requirements. But while double-sided combinatorial auctions have major economic advantages, their computational complexity is a well-documented challenge that can be seen as a disadvantage (Sandholm et al. 2002). This complexity largely stems from the fact that each participant in a combinatorial auction has to submit bids for all relevant combinations, which means that the number of bids grows exponentially as the number of participants increases.

An AFEX system mitigates this issue by pre-selecting the participants of each auction through the previously described formation of groups and the concept of "ad hoc auctions".

4.6 Ad Hoc Auctions

The fact that an autonomously coordinated exchange without a central operator lacks the central figure of the auctioneer, who performs the auction and decides on the final allocation of goods, constitutes a design challenge. The agents do not only have to find each other and form groups based on their preferences but also have to coordinate the initiation and implementation of auctions by themselves.

After a group has formed and a sufficient number of suppliers and demanders have joined, the group is declared "complete" and the auction starts. For this purpose the agents carry out a spontaneous "ad hoc auction".

The difference between ad hoc auctions and "normal" auctions in centralized marketplaces is that the auctioneer is dynamically selected from the crowd of suppliers in a group. The role of the auctioneer falls to the supplier that tries to sell the highest number of goods or, if several suppliers make an equal number of offers, that supplier which entered the group first.

The auctioneer carries out a double-sided combinatorial auction according to the auction model and subsequently specifies the final allocation of goods within the group. After all participants agree to this new distribution the group dissolves.

This approach pairs well with the concept of loosely-coupled groups described before: Groups are not only a way to frame the context of an auction by ensuring that all participants are interested in the offered goods but also limit its complexity by limiting the number of participants. Handing the computationally expensive calculations needed to perform the auction to the supplier side is a design decision based on the assumption that suppliers have a natural interest in providing a solid technical foundation in order to enable auctions of the goods they offer.

4.7 Implications

The outlined marketplace concept AFEX should provide three implicit advantages in contrast to conventional approaches:

Equality—all members of the market are subject to the same rules of action. Although agents are started with individual preferences, they cannot deviate one-sidedly regarding their strategy, which is deposited in the software.
Efficiency—the usage of the double-sided combinatorial auctions allows for optimal solutions for pricing through the deployment of mathematical models. The efficiency criteria can be specified in a goal-oriented way during the design phase.
Transparency—from the point of view of the software agents the conditions of the market and the market activity are completely transparent: all agents make contact among themselves and exchange their trading preferences.

In addition, the described approach provides a realistic modeling of the roles played by the members of the marketplace. Agents do not only act explicitly as supplier or demander but also (analogously to operators in the real world) play either the role of a supplier or a demander dependent on the context.

5 Summary and Outlook

This paper addressed the efforts to establish an online freight exchange for the mediation of multimodal transport services using the example of the European rail freight traffic. The research on online freight exchanges, the development of the prototype ORFE and the challenges that any new freight exchange will face have been described. The investigation of these problems resulted in the draft of an innovative marketplace concept: AFEX, an online freight exchange that is based on autonomous software agents. Selected requirements for the development of these decentralized and autonomously trading agents have been outlined.

The next steps are the development of an adaptive agent behavior that is able to adjust to different situations, a generic traffic route notation for the description of transport routes and a description language for the offer and demand for transport services within auctions.

The last step will be the combination and implementation of all mentioned aspects into a prototypical agent software.

References

Ackermann H, Ewe H, Kopfer H, Küfer KH (2011) Combinatorial auctions in freight logistics. In: Computational logistics, second international conference, ICCL proceedings 2011, pp 1–17

Ausubel LM, Cramton PC (1998) The optimality of being efficient. World Bank, Private Sector Development Department, Private Participation in Infrastructure Division, Washington, DC

Bichler M, Pikovsky A, Setzer T (2005) Kombinatorische Auktionen in der betrieblichen Beschaffung – Eine Analyse grundlegender Probleme. Wirtschaftsinformatik 47(2):126–134

Brenienek J (unter Mitarbeit von Juchelka R et al) (2014) Logistikcluster entlang des Transportkorridors Rotterdam-Genua: Strukturen, Entwicklungen und Prozesse. Report, Lehrstuhl für Wirtschaftsgeographie, insb. Verkehr und Logistik, Universität Duisburg-Essen

Bruns AS, Zelewski S (2011) Übertragung von Projekterkenntnissen auf Kooperationen von Güterverkehrsunternehmen entlang der Alpentransversale "Rotterdam-Genua". In: Zelewski S, Jene S (Hrsg) Kooperationen zwischen Eisenbahnverkehrsunternehmen: Grundlagen – Konzepte – Praxisanwendungen. Berlin, pp 1071–1080

Bruns AS, Günes N, Zelewski S (2010) Online-Frachtenbörse für den transeuropäischen Schienengüterverkehr. Internationales Verkehrswesen 62(11):25–29

Bruns AS, Zelewski S, Lehr T (2012a) Analyse des Marktpotenzials für Online-Frachtenbörsen im Segment von Schienengüterverkehren und Kombinierten Güterverkehren. CODE24-Projektbericht Nr. 6, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen

Bruns AS, Föhring R, Zelewski S, Günes N (2012b) Eine Online-Frachtenbörse für Verlader und Güterbahnen. Güterbahnen 11(1):33–37

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-based approaches to transport logistics. Transp Res C Emerg Technol 13(4):255–271

Dörr H, Endemann P (2014) Moving forward freight mobility innovations. Internationales Verkehrswesen 66(2):60–63

Endemann P, Kaspar T (2011) Online-Transportbörse soll Schienengüterverkehr stärken – Was braucht die Praxis? Güterbahnen 10(4):31–36

Endemann P, Tracksdorf K, Kaspar T (2012) CODE 24 online rail freight exchange needs of potential users. In: European transport conference 2012, 08–10.10.2012 in Glasgow, pp 1–14. http://abstracts.aetransport.org/paper/download/id/3868. Accessed 9 Jan 2015

Föhring R, Zelewski S (2013) ORFE: online rail freight exchange—a software prototype for the configuration of multi-modal supply chains focused on rail freight traffic along the Alps Transversal Rotterdam-Genoa. OrGoLo-Projektbericht Nr. 20, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen

Föhring R, Kuhlmann AS, Zelewski S (2012) Vorstellung des finalen Softwareprototyps einer Online-Frachtenbörse, CODE24-Projektbericht Nr. 7, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen

Fox MS, Barbuceanu M, Teigen R (2000) Agent-oriented supply-chain management. Int J Flex Manuf Syst 12(2–3):165–188

Grieger M (2003) Electronic marketplaces: a literature review and a call for supply chain management research. Eur J Oper Res 144:280–294

Habib S, Bruns AS (2012) Analyse des kompetitiven Umfelds für die Etablierung einer Online-Frachtenbörse zur Vermittlung von Eisenbahngüterverkehren auf der Nord-Süd-Transversale "Rotterdam-Genua". CODE24-Projektbericht Nr. 5, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen

Jennings NR (2000) On agent-based software engineering. Artif Intell 117(2):277–296

Klippert SR, Zelewski S, Kuhlmann AS, Kowalski M (2013) Schienentransporte im KV über eine Online-Frachtenbörse organisieren – Das EU-Verbundprojekt Code24 gibt neue Impulse für den Kombinierten Verkehr (KV) auf der Schiene. EI – Der. Eisenbahningenieur 64(5):91–96

Krishna V, Perry M (1998). Efficient mechanism design. http://econ.la.psu.edu/~vkrishna/papers/vcg20.pdf. Accessed 9 Jan 2015

Merkel H, Kromer S (2002) Virtuelle Frachtbörsen - top oder flop? In: Hossner R (Hrsg) Jahrbuch der Logistik, pp 82–87

Parkes DC, Ungar LH (2001) An auction-based method for decentralized train scheduling. In: Proceedings of the fifth international conference on autonomous agents, pp 43–50

Sandholm T, Suri S, Gilpin A, Levine D (2002) Winner determination in combinatorial auction generalizations. In: Proceedings of the first international joint conference on autonomous agents and multi-agent systems: part 1, pp 69–76

Scheck C, Wilske S (2011) Infrastrukturentwicklung im europäischen Korridor Rotterdam-Genua: das INTERREG-IV-B-Projekt "Code 24". In: Hege H-P, Knapstein Y, Meng R, Ruppenthal K, Schmitz-Veltin A, Zakrzewski P (Hrsg) Schneller, öfter, weiter? Perspektiven der Raumentwicklung in der Mobilitätsgesellschaft. Hannover, pp 56–71

Yang J (2002) The efficiency of an artificial double auction stock market with neural learning agents. In: Chen S-H (Hrsg) Evolutionary computation in economics and finance, pp 85–105

Flexibility of 3PL Contracts: Practical Evidence and Propositions on the Design of Contract Flexibility in 3PL Relationships

Ute David

Abstract In 3PL relationships, complex, customised services are provided based on multi-year contracts. As practice reveals, contract design poses a major challenge to 3PL providers and clients. Contracts need to enable handling of changing requirements and contingencies that arise during contract terms. Contracts are designed flexible utilising specific flexibility mechanisms. These flexibility mechanisms reveal strengths and weaknesses depending on relationship characteristics. Inappropriate contracts provoke dissatisfaction, contract terminations, and transaction costs. The paper describes the use, strengths and weaknesses of flexibility mechanisms in 3PL contracts drawing on theoretical and practical analyses. Based on theory and practice, propositions and needs of further research are outlined.

1 Introduction

Outsourcing of complex logistics services plays a major role in commercial and industrial companies' business activities. The outsourcing companies aim at increasing their flexibility, saving costs, and concentrating on core competencies (Sanchis et al. 2012; Handfield et al. 2013). For these companies, logistics services are no core competency but a competitive factor as insufficient logistics performance leads to competitive disadvantage (Weber and Wallenburg 2010; Deepen 2007; Lusch and Vargo 2012).

Third-party logistics (3PL) involves complex, customised logistics services and specific investments. Clients depend on the providers as complex and customised services are not directly available from other firms. 3PL providers make specific investments that cannot completely or without additional costs be used for alternative business. This increases dependencies of 3PL providers on their clients (Hsuan and Prockl 2013; Williamson 2008). The multi-year duration implies

U. David (✉)
IPRI gGmbH, Königstr. 5, 70173 Stuttgart, Germany
e-mail: udavid@ipri-institute.com

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_6

dependencies and risks of cooperation for the contracting parties. Long-term relationships and investments lead to lock-in effects. As 3PL contracts are limited in their duration and applicability, renewals, adjustments, and extensions at specific conditions and points in time are necessary (Mühlencoert 2012).

Due to the characteristics of 3PL relationships, 3PL contracts are a specific challenge compared to other contracts. 3PL contracts need to be designed according to the particular requirements of service provision (Williamson 2000; Lumineau and Quélin 2012; Argyres et al. 2007; Lusch and Vargo 2014). Inadequate contracts hinder efficient collaboration and service provision, and reduce satisfaction. This increases the possibility of contract terminations that come with significant costs (Hofmann 2008; Wallenburg et al. 2010; Nyaga et al. 2010; Cabral et al. 2014).

In complex multi-year 3PL contracts, the requirements are not completely evident ex ante. During the term of contracts, contingencies arise and requirements change (Riordan and Williamson 1985). Foreseeing and planning of all contingencies is impossible or unreasonably costly (Lumineau and Quélin 2012; Argyres et al. 2007). Contracts cannot include all information ex ante (Yang et al. 2012). In order to handle incomplete information and contingencies, contracts are designed flexible to enable ex post adjustments as information and contingencies reveal (Yang et al. 2012). Depending on specific relationship characteristics, contract mechanisms come with strengths and weaknesses (Lumineau and Quélin 2012; Argyres et al. 2007; Williamson 2000).

In order to design contracts appropriately, relationships and contracts need to be analysed in detail. Such detailed analyses are not provided by research (Yang et al. 2012; Schepker et al. 2014). Empirical research on formal contracts concerns rather general influencing factors and dimensions (e.g. Van Hoek 2000; Wallenburg and Schäffler 2014). Dimensions of legal contracts that are addressed in research are formalisation in general (Schmoltzi and Wallenburg 2012), the extent to which contracts are detailed and fixed (Van Hoek 2000), and mechanisms like process control and formal controls (Wallenburg and Schäffler 2014). Relational governance is mainly analysed in terms of trust and commitment (e.g. Wallenburg et al. 2011; Kwon and Suh 2004).

This paper presents theoretical foundations and practical analyses of 3PL relationships and contracts. The use, strengths and weaknesses of flexibility mechanisms are examined in light of relationship characteristics. These analyses lead to propositions about the use of flexibility mechanisms.

The paper is structured as follows: The subsequent chapter provides theoretical foundations and the design of practical research. Then, 3PL success, drivers for adjustments and the use, strengths and weaknesses of flexibility mechanisms are discussed. Based on this, propositions and issues for further research are outlined.

2 Research Design

As theoretical foundations, contract theories, resource based view (RBV), principal agent theory (PA theory), and transaction cost economics (TCE) are described. Practical analyses draw on in-depth interviews and contract analysis.

2.1 Theoretical Foundations

Theoretical foundations of contract design are contract theories. Following the classical theory, contracts are complete and rigid without opportunities for adjustments. This requires that all information and contingencies are known ex ante when designing contracts (Narayanan and Narasimhan 2014). In 3PL, complete information on all contingencies during the contract term is not available or comes with high costs (Lumineau and Quélin 2012; Argyres et al. 2007). Multi-year 3PL contracts are necessarily incomplete (Williamson 1998). In order to govern long-term relationships, the legal perspective is not sufficient when designing and implementing 3PL contracts. This becomes clear as a 3PL client states:

> You can conclude a 3PL contract that is designed from a legal view, but you cannot 'live' it. (3)

An economic concept relevant for 3PL is the RBV. RBV describes resources as drivers of performance such as financial and logistics performance, and competitive advantage. By collaborating with external entities, firms gain access to additional resources as compared to the internal resources (Chen et al. 2010; Yeung et al. 2012). 3PL provider's resources and their suitability to the client's needs are particularly relevant for the performance in 3PL relationships (e.g. Liu and Lyons 2011; Lai 2004).

The PA theory considers the separation of ownership and control that is apparent in 3PL relationships between independent clients (principals) and providers (agents). In relationships between independent entities, agents have incentives to deviate and principals need to set appropriate incentives and monitor the agents' activities. PA theory focuses on single agents and considers contracts as incentive mechanisms from an ex ante perspective (Williamson 1996).

TCE considers contracts in the ex post implementation phase (Williamson 2000). Major determinants of relationship success are the transaction costs that arise during relationships. Transaction costs arise ex ante for planning, design and conclusion of contracts. Ex post transaction costs occur for implementation, monitoring and maladaptation. Governance mechanisms such as contracts are used to minimise transaction costs (Williamson 1979). TCE provides a framework to analyse governance mechanisms in outsourcing relationships (Halldórsson and Skjøtt-Larsen 2006). In order to ensure and coordinate efficient interaction and service provision, contracts need to be designed according to relationship requirements (Williamson 2000, 2008). This work draws on TCE as main theory foundation.

Referring to the assumptions of TCE, contract parties are subject to incomplete information, bounded rationality, and opportunism. In 3PL relationships, these assumptions apply. Due to the complex and uncertain conditions, not all information is known to the providers and clients. The contracting parties act boundedly rational and make arrangements and start exchanges under incomplete information (Williamson1981, 1985). Incomplete information and incomplete contracts give rise to opportunistic behaviour as contract parties exploit the relationships and contracting partners to their individual benefit (Williamson 1991).

Building on TCE, relationships differ with respect to the parameters uncertainty, frequency of interaction, and specific investments (Williamson 1981). These criteria influence the risks of opportunism (Carson et al. 2006). 3PL contracts are characterised by uncertainties and highly frequent interactions. Specific investments accrue as the contracting parties invest in specialised processes, know-how, and equipment. This raises dependencies and increases incentives to maintain and extend relationships (Hsuan and Prockl 2013; Williamson 2008; Drodofsky 2011; Nakos and Brouthers 2008; Large 2011). Specific investments are a major challenge when requirements are uncertain and changing. Dependencies raise risks of opportunism (Williamson 2008; Reeves et al. 2010). Appropriate contracts prevent opportunism and reduce transaction costs (Williamson 1991).

Further elements in multi-year relationships are relational factors (Macneil 1978; Parkhe 1993). These include trust, previous experience and expectations of long-term relationships as well as norms, customs and conventions that promise sanctions in case of deviations (Lumineau and Quélin 2012; Schepker et al. 2014; Knemeyer and Murphy 2005). In incomplete, complex contracts, relational factors decrease the risks of opportunism and support the long-term implementation (Leuschner et al. 2014; Reeves et al. 2010; Wallenburg and Lukassen 2011). Service-dominant logic points out that the value of service provision evolves over time. Experience and continued exchange are particular factors that create benefits of service provision (Lusch and Vargo 2014).

2.2 *Practical Analyses*

Foundations for practical analyses are ten semi-structured in-depth phone interviews that took between 30 and 60 min each. The interviews were conducted in a three months time period between August and November 2014. Table 1 provides characteristics of the companies and experts involved. In the wrap-up, the recordings were verified by the interviewees. The sample represents enterprises from different German industries including machinery and equipment manufacturing, high-tech, chemicals and logistics service providers. The interviewees provide logistics, supply chain, purchasing, management and legal backgrounds. The enterprises experience similar challenges regarding 3PL contract design.

Table 1 Sample

	Industry	Size	Expert position
1	Machinery and equipment manufacturer	Mid	Logistics manager
2	High-tech/consumer goods	Mid	Global head of logistics
3	High-tech/electronics	Mid	Head of logistics
4	High-tech/multimedia	Large	Legal
5	Automotive supplier	Large	Head of logistics
6	Automotive supplier	Large	Head of logistics purchasing
7	Chemicals	Mid	Procurement
8	3PL provider	Large	Legal
9	3PL provider	Mid	Managing director
10	3PL provider	Large	Key account manager entertainment

The logistics service providers focus 3PL services and carry out complex and customised warehousing, additional and value-added services. The client firms include large and medium-sized enterprises with limited as well as broad international orientation. Some purchase complex warehousing services including inventory and order management. Others also obtain customisation, logistics planning and IT services from 3PL providers. The services are partially or fully customised.

The use of flexibility mechanisms under different circumstances is analysed in a sequence of contracts. Contract adjustments enable insight into the development of flexibility mechanisms as requirements change. By means of the adjustments, the contracts are supposed to better match new requirements, and improve applicability and efficiency within the 3PL relationship.

The results of the in-depth interviews and contract analyses are structured and compared. The findings apply to other enterprises in the sample and supposedly a large number of further enterprises in 3PL relationships. The text refers to the interviews utilising the numbering in Table 1.

3 Use, Strengths, and Weaknesses of Flexible 3PL Contracts

In the following, the use, strengths, and weaknesses of flexible 3PL contracts are discussed drawing on extant literature and statements from practice.

3.1 Success of 3PL Relationships

Contracts minimise transaction costs and take functions of control and safeguarding, coordination and adaptation (Schepker et al. 2014; Malhotra and Lumineau

2011; Ryall and Sampson 2009; Lumineau and Quélin 2012). Successful 3PL contracts ensure effective and efficient collaboration between providers and clients. These contracts last long, avoid early terminations, and improve the possibility of contract renewals and extensions.

During long-term relationships, collaboration and service provision improve as the contracting parties get accustomed and enhance processes and coordination. Long-term relationships enable efficient utilisation of investments so that price conditions improve (9). If contracts are renewed and not terminated, no switching costs accrue. Switching costs include transaction costs for contract terminations and phase-out, tenders, negotiations for new contracts, and set-up of relationships. Switching also comes with performance losses and delays in service provision (1, 2, 4, 5, 6). If new contracting partners do not have previous experiences, switching costs also arise due to uncertainties about the new partners (Hofmann 2008; Wallenburg et al. 2010; Nyaga et al. 2010; Cabral et al. 2014). These transaction costs play a major role in decision making (4, 6, 9).

A crucial factor for long-term success of 3PL relationships is client satisfaction regarding the interaction and collaboration in the relationship (Large 2011; Cahill et al. 2010; Stank et al. 1999; interviews 3, 6). The clients can determine the value of service provision (Lusch 2011; interview 9) and satisfaction is the direct result of an efficient relationship (Poppo and Zenger 2002).

3.2 *Drivers of Ex Post Adjustments*

In the course of long-term 3PL relationships, changing requirements give rise to ex post adjustments of contracts. Volumes deviate from the plans so that the services cannot be fulfilled under the conditions initially agreed upon. In case of repeated increases and decreases of volumes, service providers need to have additional capacities available and handle idle capacities (9, 10).

Contracts allow for varying volumes. Volumes that are covered by a contract are typically limited within specific boundaries. If volume exceeds these boundaries, appropriate service provision cannot be ensured. In these cases, pricing and further conditions of service provision are renegotiated (6, 8, 9). Changing customer requirements and processes initiate ex post contract adjustments. Improving internal processes that influence service provision induces decreasing costs that are passed on to the contracting partners. This is particularly the case the longer 3PL relationships exist (2, 9).

One driver of volume, requirement and process changes are adjustments of business models. If 3PL clients change their ways of distribution and structure of customers, requirements for 3PL service provision alter significantly. A change of the distribution concept from indirect B2B to direct B2C distribution implies that large amounts of small orders need to be handled instead of small amounts of large orders. This demands considerably different processes and IT. Such changes even imply that contracts are terminated and other 3PL providers are employed (2).

3.3 Flexibility Options

Contracts are designed flexible by including flexibility options. These include renegotiation options, flexible pricing, incentive contracting, short contract duration and early termination clauses (Harris et al. 1998). Renegotiations during the contract term offer opportunities to react to foreseen and unforeseen changes. In the course of renegotiations, selected parts or all elements of contracts are changed. Renegotiations are allowed at specific points in time or when specific conditions apply (Harris et al. 1998). Renegotiations enable the contracting parties to postpone decisions that are difficult to agree upon during the initial contract negotiations. This is the case if necessary information is missing or the contracting parties differ in their positions (7, 8, 9).

Renegotiations, however, come at considerable costs. They induce transaction costs for administration, contract design, and technical issues (Argyres et al. 2007; Schepker et al. 2014). Renegotiations imply uncertainties and increase the risks of opportunism (Reuer and Ariño 2002; Argyres and Mayer 2007). Renegotiations are regarded beneficial as uncertainties are high and unforeseen changes occur. The risks of opportunism are particularly relevant as dependencies increase (Williamson 1991). This is considered in P-I:

P-I: (1) The higher the uncertainties or (2) the lower the dependencies between the contracting parties are, the more favourable are renegotiation options.

Dependencies between the contracting parties are particularly relevant in 3PL relationships that involve complex and customised services. 3PL providers experience dependencies as they set up specialised service provision. They cannot employ the capacities in other contracts without incurring additional costs (8, 9). Complexity and customisation imply dependencies for clients as they cannot obtain the services from other providers at the same conditions. Changing service providers implies switching costs which tie clients to 3PL providers (1, 2, 4, 5, 6).

As renegotiations take considerable effort and time, contracting parties consider them as barriers to operational performance. Practice focuses on precise adjustment mechanisms to concentrate on fulfilling the operational tasks (9). One precise adjustment mechanism is flexible pricing. Remuneration for 3PL providers then comprises of fixed and variable components that depend on volumes handled. Price adjustments further build on increasing wages and operating expenses. Pricing adjustments enable the supplier to keep up with service provision even when volumes or other conditions change (3, 9).

Another element of pricing is incentive contracting. Within bonus malus systems, high performance is rewarded with additional compensation and low performance is punished using penalty fees (Sols et al. 2007; Harris et al. 1998). Incentive contracting enables 3PL clients to remunerate the actual achievement and stimulate high performance. At the same time, incentive contracting comes with high transaction costs for setting performance indicators, reward and penalty schemes, and enforcement and monitoring. Incentive contracting is particularly relevant as services become more customised and essential for the client's business (5). This leads to proposition P-II:

P-II: The more (1) customised or (2) critical the services are for the client, the more efficient is incentive contracting.

Short contract durations increase flexibility as contracting parties are able to conclude new contracts and switch partners after a short period of time. This is also implied by early termination clauses that allow for contract terminations prior to the agreed expiration date if predefined conditions occur. Contract terminations induce significant switching costs (Hofmann 2008; Wallenburg et al. 2010; Nyaga et al. 2010; Cabral et al. 2014). Considering short contract durations or early terminations, 3PL providers compare switching costs with potential savings or profit increases in alternative business (9). 3PL clients consider the switching costs, but neglect the potential savings that might come with new contracts and different 3PL providers (3). If specific investments are involved, short contract durations worsen contract conditions as providers' average costs and therefore prices increase. Early termination clauses imply that contracting parties are uncertain if their investments can be utilised during a sufficiently long time (9). This implies P-III:

P-III: The higher the specific investments are, the less beneficial are (1) short contract durations or (2) early termination clauses.

3.4 Development of 3PL Contracts

The flexibility options are combined within the scope of 3PL contracts. They are applied differently as relationship characteristics change. This is illustrated on the basis of a sequence of contract adjustments.

The considered relationship is characterised by high uncertainty, low trust and no experience between the contracting parties. Remuneration is set as a fixed price. A bonus malus system allows for limited additional payments and penalties. These arrangements do not provide considerable flexibility. Limited flexibility is induced by including early termination clauses and short periods of notice. The contract is intended to last for five years and termination clauses enable regular contract terminations. These mechanisms provide limited flexibility to account for highly uncertain circumstances, low trust and no experience between the companies. The fixed price is set relatively high as the termination options induce the risk that the investments cannot be earned during the life of the contract (3).

These arrangements are not applicable as additional specific investments are necessary and volumes vary over time. In light of specific investments, the contract duration and the period of notice are extended significantly. For handling of volume changes, variable compensation is introduced. If volume exceeds the planned level, additional payments apply. These contract arrangements provide flexibility taking into consideration specific investments and volume changes. Flexibility shifted from termination to pricing flexibility and is still limited. Limited flexibility is beneficial in particular as specific investments are involved (3).

The benefits of limited flexibility are of decreasing relevance as the contract parties build up trust and experience in collaboration. As these circumstances emerge, fixed remuneration is reduced, and the variable compensation option and the bonus malus system are strengthened. Contract duration is shortened. These arrangements offer increased flexibility and enable contract execution under the conditions of high uncertainty, specific investments, increasing trust and experience between provider and client (3). This implies proposition P-IV:

P-IV: The stronger the trust between the contracting parties, the more beneficial are (1) short contracts or (2) early termination clauses even though specific investments are involved.

4 Conclusion

This paper presents the use, strengths and weaknesses of flexibility mechanisms in 3PL contracts. Flexible contracting is particularly important in 3PL relationships as not all information and contingencies that arise during the life of contracts are known ex ante. Flexibility mechanisms enable contracts to be adjusted as contingencies occur and requirements change. Appropriate contracts minimise transaction costs, and ensure adequate service provision and collaboration. This improves client satisfaction and ensures long lasting relationships (Large 2011; Cahill et al. 2010; Stank et al. 1999).

The use, strengths and weaknesses of flexibility mechanisms need to be understood in light of relationship requirements (Williamson 2000; Lumineau and Quélin 2012; Argyres et al. 2007; Lusch and Vargo 2014). In order to design appropriate contracts, detailed analyses of relationship requirements and contracts are necessary. This paper provides practical evidence and propositions for the design of efficient 3PL contracts. The analyses are based on theoretical foundations and incorporate practical analyses from a variety of industries and areas of expertise. However, the analyses need to be extended to a more comprehensive sample and tested in quantitative analyses. The propositions span areas for further research.

References

Argyres NS, Mayer KJ (2007) Contract design as a firm capability: an integration of learning and transaction cost perspectives. Acad Manag Rev 32(4):1060–1077

Argyres NS, Bercovith J, Mayer KJ (2007) Complementarity and evolution of contractual provisions: an empirical study of IT services contracts. Organ Sci 18:3–19

Cabral S, Quélin BV, Maia W (2014) Outsourcing failure and reintegration: the influence of contractual and external factors. Long Range Plan 47(6):365–378

Cahill DL, Goldsby TJ, Knemeyer AM, Wallenburg CM (2010) Customer loyalty in logistics outsourcing relationships: an examination of the moderating effects of conflict frequency. J Bus Logist 31(2):253–277

Carson SJ, Madhok A, Wu T (2006) Uncertainty, opportunism, and governance: the effects of volatility and ambiguity on formal and relational contracting. Acad Manag J 49:1058–1077

Chen H, Tian Y, Ellinger AE, Daugherty PJ (2010) Managing logistics outsourcing relationships: an empirical investigation in china. J Bus Logist 31:279–299

Deepen JM (2007) Logistics outsourcing relationships: measurement, antecedents, and effects of logistics outsourcing performance. Springer, Heidelberg

Drodofsky M (2011) Beendigung von Kontraktlogistikbeziehungen - Dienstleisterperspektive. In: Sucky E, Asdecker B, Dobhan A, Haas S, Wiese J (eds) Logistikmanagement, Logistik und supply chain management, Bamberg, pp 233–252

Halldórsson A, Skjøtt-Larsen T (2006) Dynamics of relationship governance in TPL arrangements: a dyadic perspective. Int J Phys Distrib Logist Manag 36(7):490–506

Handfield R, Straube F, Pfohl H-C, Wieland A (2013) Trends and strategies in logistics and supply chain management: embracing global logistics complexity to drive market advantage, Hamburg

Harris A, Giunipero LC, Hult GTM (1998) Impact of organizational and contract flexibility on outsourcing contracts. Ind Mark Manag 27:373–384

Hofmann E (2008) Vom Outsourcing zum Insourcing: Organisatorische Alternativen und Wege zur Beendigung von Geschäftsbeziehungen. Zeitschrift für Organ 77(2):104–113

Hsuan J, Prockl G (2013) Measuring service individuality in third party. In: Arnäs PO, Arvidsson N, Bergqvist R, Johansson M, Pahlén P-O (eds) Proceedings of the 25th NOFOMA conference, Göteborg

Knemeyer AM, Murphy PR (2005) Exploring the potential impact of relationship characteristics and customer attributes on the outcomes of third-party logistics arrangements. Transp J 44 (1):5–19

Kwon I-WG, Suh T (2004) Factors affecting the level of trust and commitment in supply chain relationships. J Supply Chain Manag 40:4–14

Lai K-H (2004) Service capability and performance of logistics service providers. Transp Res E Logist Transp Rev 40(5):385–399

Large RO (2011) Partner-specific adaptations, performance, satisfaction and loyalty in third-party logistics relationships. Logist Res 3:37–47

Leuschner R, Carter CR, Goldsby TJ, Rogers ZS (2014) Third-party logistics: a meta-analytic review and investigation of its impact on performance. J Supply Chain Manag 50(1):21–43

Liu C-L, Lyons AC (2011) An analysis of third-party logistics performance and service provision. Transp Res E Logist Transp Rev 47:547–570

Lumineau F, Quélin BV (2012) An empirical investigation of interorganizational opportunism and contracting mechanisms. Strateg Organ 10(1):55–84

Lusch RF (2011) Reframing supply-chain management: a service-dominant logic perspective. J Supply Chain Manag 47(1):14–18

Lusch RF, Vargo SL (2012) Gaining competitive advantage with service-dominant logic. In: Lilien GL, Grewal R (eds) Handbook on business to business marketing, Cheltenham, pp 109–124

Lusch RF, Vargo SL (2014) Service-dominant logic: premises, perspectives, possibilities. Cambridge University Press, Cambridge

Macneil IR (1978) Contracts: adjustment of long-term economic relations under classical, neoclassical, and relational contract law. Northwest Univ Law Rev 72:854–905

Malhotra D, Lumineau F (2011) Trust and collaboration in the aftermath of conflict. Acad Manag J 54(5):981–998

Mühlencoert T (2012) Kontraktlogistik-Management: Grundlagen - Beispiele - Checklisten. Wiesbaden

Nakos G, Brouthers KD (2008) International alliance commitment and performance of small and medium-size enterprises: the mediating role of process control. J Int Manag 14:124–137
Narayanan S, Narasimhan R (2014) Governance choice, sourcing relationship characteristics, and relationship performance. Decis Sci J 45(4):717–751
Nyaga GN, Whipple JM, Lynch DF (2010) Examining supply chain relationships: do buyer and supplier perspectives on collaborative relationships differ? J Oper Manag 28:101–114
Parkhe A (1993) Strategic alliance structuring: a game theoretic and transaction cost examination of interfirm cooperation. Acad Manag J 36:794–829
Poppo L, Zenger T (2002) Do formal contracts and relational governance function as substitutes or complements. Strateg Manag J 23:707–725
Reeves KAJ, Caliskan F, Ozcan O (2010) Outsourcing distribution and logistics services within the automotive supplier industry. Transp Res E 46:459–468
Reuer JJ, Ariño A (2002) Contractual renegotiations in strategic alliances. J Manag Gov 28:47–68
Riordan MH, Williamson OE (1985) Asset specificity and economic organization. Int J Ind Organ 3:365–378
Ryall MD, Sampson RC (2009) Repeated interaction and contract structure: evidence from technology development contracts. Manag Sci 55:906–925
Sanchís-Pedregosa C, del Mar González-Zamora M, Machuca JAD (2012) Identifying the objectives of outsourcing logistics services as a basis for measuring its financial and operational performance. Eng Technol 69:909–915
Schepker DJ, Oh W-Y, Martynov A, Poppo L (2014) The many futures of contracts: moving beyond structure and safeguarding to coordination and adaptation. J Manag Gov 40(1):193–225
Schmoltzi C, Wallenburg CM (2012) Operational governance in horizontal cooperations of logistics service providers: performance effects and the moderating role of cooperation complexity. J Supply Chain Manag 48:53–74
Sols A, Nowicki D, Verma D (2007) Defining the fundamental framework of an effective performance-based logistics contract. Eng Manag J 19(2):40–50
Stank TP, Goldsby TJ, Vickery SK (1999) Effect of supplier performance on satisfaction and loyalty of store managers in the fast food industry. J Oper Manag 17:429–447
Van Hoek RI (2000) The purchasing and control of supplementary third-party logistics services. J Supply Chain Manag 36:14–26
Wallenburg CM, Lukassen P (2011) Proactive improvement of logistics service providers as driver of customer loyalty. Eur J Mark 45(3):438–454
Wallenburg CM, Schäffler T (2014) The interplay of relational governance and formal control in horizontal alliances: a social contract perspective. J Supply Chain Manag 50:41–58
Wallenburg CM, Cahill DL, Goldsby TJ, Knemeyer AM (2010) Logistics outsourcing performance and loyalty behaviour. Int J Phys Distrib Logist Manag 40(7):579–602
Wallenburg CM, Cahill DL, Knemeyer AM, Goldsby TJ (2011) Commitment and trust as drivers of loyalty in logistics outsourcing relationships: cultural differences between the united states and Germany. J Bus Logist 32:83–98
Weber J, Wallenburg CM (2010) Logistik- und supply chain-controlling. Stuttgart
Williamson OE (1979) Transaction-cost economics: the governance of contractual relations. J Law Econ 22(2):233–261
Williamson OE (1981) The economics of organization: the transaction cost approach. Am J Sociol 87(3):548–577
Williamson OE (1985) The economic institutions of capitalism: firms, markets, relational contracting. Springer, New York
Williamson OE (1991) Comparative economic organization: the analysis of discrete structural alternatives. Adm Sci Q 36:269–296
Williamson OE (1996) The mechanisms of governance. Springer, New York
Williamson OE (1998) TCE: How It Works. Where It is Headed. De Economist 146(1):23–58
Williamson OE (2000) The new institutional economics: taking stock, looking ahead. J Econ Lit 38(3):595–613

Williamson OE (2008) Outsouring: TCE and supply chain management. J Supply Chain Manag 44 (2):5–16

Yang C, Wacker JG, Sheu C (2012) What makes outsourcing effective? a transaction-cost economics analysis. Int J Prod Res 50(16):4462–4476

Yeung K, Zhou H, Yeung ACL, Cheng TCE (2012) The impact of third-party logistics providers' capabilities on exporters' performance. Int J Prod Econ 135(2):741–753

Part IV
Resource Management

Extended Model Formulation of the Facility Layout Problem with Aisle Structure

Armin Klausnitzer and Rainer Lasch

Abstract An efficient layout arrangement and its aisle structure are particularly important due to material handling costs, which directly affect product costs. Traditionally this is a procedure of sequential steps. In recent literature there exist a few approaches to handle layout arrangement and aisle structure simultaneously using heuristics. Only one approach could be found to solve it exactly, but with a lack of exactness due to its discrete representation. In this article, a mixed integer programing model (MIP) is proposed to achieve an arrangement of equipment with fixed shapes, input/output points, and aisle structure with predefined widths simultaneously and exactly. Furthermore, a continuous representation is considered. For that purpose the MIP model of Kim and Kim (2000) is modified to reduce computing time and is enhanced later on to consider aisles. The applicability is demonstrated by a small computational test.

1 Introduction

One of the most researched and still challenging issues is the facility layout problem (FLP). It deals with the arrangement of equipment like machinery, cells, and departments, as well as human resources, usually to achieve a minimum of material flow (Emami and Nookabadi 2013). In the further description the expression "department" is used as a synonym for machine, cell, workstation, office, and miscellaneous components of a factory.

For practitioners, FLP is of considerable interest considering that in Germany about 11 % of all gross investment was committed to factories in the year 2004 (Dombrowski et al. 2007). Furthermore, up to 50 % of total operating expenses are generated by material handling within manufacturing. In addition, it is estimated

A. Klausnitzer (✉) · R. Lasch
Lehrstuhl für Betriebswirtschaftslehre, insb. Logistik, Technische Universität Dresden, Münchner Platz 1/3, 01062 Dresden, Germany
e-mail: armin.klausnitzer@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_7

that at least between 10 and 30 % of these costs result from an inferior layout, which can be avoided by effective planning (Tompkins et al. 2010).

FLP has been of interest for more than five decades. It belongs to the class of NP-hard combinatorial problems, which was treated as a quadratic assignment problem (QAP) in early research (Singh 2009). This approach is addressed to determine the relative locations of departments without consideration of their dimensions. In case of equally sized departments, QAP may be suitable for FLP. In real world applications, the dimensions of departments usually differ and should be taken into account to achieve more precise planning. This requirement led to enhanced models by breaking departments into small parts which are assigned to several elements of a grid-based floor (Ripon et al. 2013). The main drawback of the discrete representations is that these are only able to determine rough positions on the facility floor. To handle exactness, continuous representations of locations and dimensions provide an alternative and are suited to consider further layout constraints (Drira et al. 2007). Further constraints for MIP formulations to enforce planning exactness can be defined as follows:

- There is the possibility of obligatory neighborhoods of departments motivated by large material flow between them (Tavares et al. 2002).
- Constrained distances between departments or departments from fixed points can be considered, motivated for instance by regarding temperature controlled areas (Tavares et al. 2002).
- There is the obligation or prohibition to place, respectively not to place departments within a defined area in order to extend the available floor space to a not perfect rectangle (Tavares et al. 2002).
- Considering the dependency of a department's rotation and its relative position, i.e. 'at front/back of' or 'at right/left of', may be necessary due to technological correlations (Tavares et al. 2002).
- The usage of material input and output points within departments may improve distance measurement of material flow (Kim and Kim 2000).
- By the implementation of a combined rotation and mirror image of departments, a broader set of representations of departments is offered (Meller et al. 2010).
- To allow departments to differ from an ordinary rectangular shape, there are also approaches to combine these to rectangular polygons, whereby more realistic models can be achieved (Bukchin et al. 2006).
- It is possible to implement a clearance factor to provide distances between borders of departments (Heragu and Kusiak 1991).
- FLP also addresses multi-floor problems (Patsiatzis and Papageorgiou 2002) as well as three-dimensional problems (Barbosa-Póvoa et al. 2002).

Traditionally, FLP is treated as a sequential procedure. Firstly, departments are allocated within the floor space to achieve minimal material handling effort. This is calculated using material handling distances weighted by interdepartmental flow intensity and, if necessary, weighted by unit transportation costs. The result of this step is known as a block layout and will be executed to a detailed layout later on (Bukchin and Tzur 2014). The current ambition is to consider details such as the above enhancements which are most likely not complete. However, they exemplify the development from early continual representations (Heragu and Kusiak 1990), with the outcome of a simple block layout, to the attempt of generating a detailed layout in one step.

Traditionally, space requirements enabling material transport are considered in further planning steps, as well as the detailed constraints mentioned before, which have recently been integrated to achieve a detailed layout in one step. Disregarding aisle structures in respective problem formulations leads to a significant gap in the aim to improve further precision and can lead to poor planning solutions (Hu et al. 2007). To the best of our knowledge there exists only one MIP-formulation to model FLP with aisle structure, which was presented by Bock and Hoberg (2007). The drawback of this approach is its lack of exactness due to the use of a discrete representation. In recent years a few heuristic approaches have been proposed to consider aisle structure in layout planning (Leno et al. 2012; Hu et al. 2007), but these cannot guarantee optimality. Hence, the intention of this paper is to present a MIP-formulation for continuous representation to close this gap and to broaden the above refinements with this meaningful component.

The remainder of this paper is organized as follows: After the introduction in Sect. 1, a well-known MIP-formulation which is suited to arranging departments with input and output points is presented and enhanced in Sect. 2. Furthermore, a model extension to integrate aisle structure in order to enforce simultaneous planning is proposed. After presenting the new approach, its applicability is illustrated with a set of numerical examples. Conclusions and further research are contained in Sect. 3.

2 Problem Description

Rectangular departments with fixed material input and output locations considering a suited orientation are to be placed in a restricted rectangular facility area. The main challenge is the integration of aisles with defined widths between input and output points whenever there is a material flow between the respective departments. In addition to the width of aisles, the following data sets have to be predefined (Meller et al. 2010):

- dimension of available floor space
- width of aisles and dimension of departments to be placed
- locations of input and output points within departments in their initial orientation
- estimated intensity of material flow measured in unit loads per time unit between departments

Although the model's intention is to converge to a more realistic and exact representation, there is a need to make assumptions. It is allowed to place departments and aisle structure in parallel to floor space borders. Furthermore, aisles between two departments are restricted to obtaining at most one bend with a 90° angle. The latter is caused by the usage of the rectilinear distance metric to define aisles. This restricting assumption is not seriously unrealistic because there is an effort to minimize bends for the material flow path to provide smooth and fast material transport (Leno et al. 2012).

2.1 Model Description

The used MIP-approach to allocate departments with input and output points was originally formulated by Kim and Kim (2000). Slight changes to the linear programing model are proposed to improve it. To shorten the explanation, terms with similar structures are presented by the use of brackets. The following notation is assumed:

Indices:

i, j, m, n department indices range from 1 to N

Parameters:

f_{ij}	material flow in unit loads per time unit between dep. i and j
c_{ij}	=1, if material flow between dep. i and j exists, 0 otherwise (c_{ij} relates to f_{ij} through $c_{ij} \cdot M \geq f_{ij}$ ^ $f_{ij} \geq c_{ij}$)
(P_i^x, P_i^y)	horizontal and vertical distances between the input point and the left bottom corner of dep. i in its initial orientation
(D_i^x, D_i^y)	horizontal and vertical distances between the output point and the left bottom corner of dep. i in its initial orientation
w_i, h_i	width and height of dep. i
W, H	horizontal and vertical dimension of available floor space
M	arbitrarily large positive number

Decision variables:

e_{ij}	horizontal distance between input and output point of dep. i and j
d_{ij}	vertical distance between input and output point of dep. i and j
$e(d)_{ij}^{+}$	$\max(0, x(y)_i^O - x(y)_j^I)$
$e(d)_{ij}^{-}$	$\max(0, x(y)_j^I - x(y)_i^O)$
$ee(dd)_{ij}$	=1, if $e(d)_{ij}^{+}$ is positive and $e(d)_{ij}^{-}$ is zero, 0 otherwise
(x_i^O, y_i^O)	spatial coordinate of output point of dep. i
(x_i^I, y_i^I)	spatial coordinate of input point of dep. i
(x_i', y_i')	spatial coordinate of right upper corner of dep. i
(x_i, y_i)	spatial coordinate of left bottom corner of dep. i
l_{ij}	=1, if dep. i is placed to the left of dep. j, 0 otherwise
b_{ij}	=1, if dep. i is placed below dep. j, 0 otherwise

u_i =1, if dep. i is placed in 90° or 270° rotation, 0 otherwise
v_i =1, if dep. i is placed in 180° or 270° rotation, 0 otherwise
s_i =1, if dep. i is placed in 270° rotation, 0 otherwise

MIP model:

$$\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} f_{ij}(e_{ij}+d_{ij}) \rightarrow \min \tag{1}$$

$$e(d)_{ij} = c_{ij}(e(d)_{ij}^{+} + e(d)_{ij}^{-}) \quad \forall i,j \tag{2}$$

$$x(y)_j^I - x(y)_i^O = e(d)_{ij}^{+} - e(d)_{ij}^{-} \quad \forall i,j \tag{3}$$

$$e(d)_{ij}^{+} \leq ee(dd)_{ij}M \quad \forall i,j \tag{4}$$

$$e(d)_{ij}^{-} \leq (1-ee(dd)_{ij})M \quad \forall i,j \tag{5}$$

$$x_i^{'} = x_i + (1-u_i)w_i + u_ih_i \quad \forall i \tag{6}$$

$$y_i^{'} = y_i + (1-u_i)h_i + u_iw_i \quad \forall i \tag{7}$$

$$\begin{aligned} x_i^{I(O)} =& x_i + \overbrace{(1-u_i-v_i+s_i)}^{0^\circ} P(D)_i^x + \overbrace{(u_i-s_i)}^{90^\circ} P(D)_i^y \\ &+ \underbrace{(v_i-s_i)}_{180^\circ}(w_i-P(D)_i^x) + \underbrace{s_i}_{270^\circ}(h_i-P(D)_i^y) \quad \forall i \end{aligned} \tag{8}$$

$$\begin{aligned} y_i^{I(O)} =& y_i + (1-u_i-v_i+s_i)P(D)_i^y + (u_i-s_i)(w_i-P(D)_i^x) \\ &+ (v_i-s_i)(h_i-P(D)_i^y) + s_iP(D)_i^x \quad \forall i \end{aligned} \tag{9}$$

$$s_i \geq u_i + v_i - 1 \quad \forall i \tag{10}$$

$$s_i \leq u_i, \;\; s_i \leq v_i \quad \forall i \tag{11}$$

$$l_{ij} + l_{ji} + b_{ij} + b_{ji} \geq 1 \quad \forall i<j \tag{12}$$

$$x_i^{'} \leq x_j + W(1-l_{ij}) \quad \forall i \neq j \tag{13}$$

$$y_i^{'} \leq y_j + H(1-b_{ij}) \quad \forall i \neq j \tag{14}$$

$$x_m + 0.5((1-u_m)w_m + u_mh_m) \leq x_n + 0.5((1-u_n)w_n + u_nh_n) \tag{15}$$

$$y_m + 0.5((1 - u_m)h_m + u_m w_m) \leq y_n + 0.5((1 - u_n)h_n + u_n w_n) \tag{16}$$

$$x_i' \leq W, \quad y_i' \leq H \quad \forall i \tag{17}$$

$$x_i, y_i, x_i', y_i', x_i^{I(O)}, y_i^{I(O)} \geq 0 \quad \forall i \tag{18}$$

$$u_i, v_i \in \{0, 1\} \quad \forall i \tag{19}$$

$$l_{ij}, b_{ij}, ee_{ij}, dd_{ij} \in \{0, 1\} \quad \forall i, j \tag{20}$$

The objective function (1) minimizes the sum of rectilinear distances between input and output points weighted by their material flow intensity. Thus, departments with high material interaction will be placed as close as possible to each other. Rectilinear distances between input points are defined by constraints (2)–(5), using a linearization scheme of absolute values (Kallrath 2002). Equations (6) and (7) define coordinates of the upper right corner of department i, depending on the department's horizontal or vertical placement.

Spatial coordinates of input and output points are defined by Eqs. (8) and (9). Figure 1 illustrates the above terms. However, the original model contains four binary variables, each for a 90° rotation step. Hu et al. (2007) proposed a non-linear alternative by the multiplication of two binary variables to represent 0°, 90°, 180°, and 270° cases. To reduce complexity, we propose a linearized alternative with only two binary variables u_i, v_i, and one continual variable s_i. The latter operates like a binary variable due to inequalities (10) and (11) (Kallrath 2002). Table 1 shows an overview of terms usable for determining the department's orientation.

To prevent departments from overlapping, inequalities (13) and (14) demand that the left (bottom) border of machine j has a bigger coordinate than the right (upper) border of machine i. Inequality (12) assures that the separation for every pair of departments i and j occurs in at least one dimension. An unnecessary but advantageous constraint proposed by Sherali et al. (2003) is placed with the inequalities (15) and (16). By determining lower centroid coordinates of the department m than centroid coordinates of department n, the model's run time showed promising improvement. The intention is to eliminate symmetric solutions. Furthermore, (17) and (18) assure that all departments will be placed within floor dimensions, while (19) and (20) are definitions of binary variables.

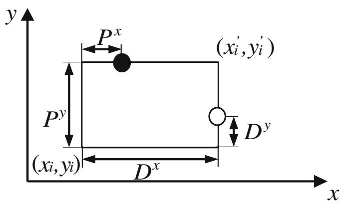
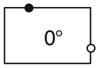
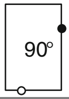
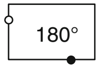
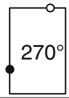


Fig. 1 Department's parameter

Table 1 Terms causing the department's orientation

Rotation				
Kim and Kim (2000)	u_i^1	u_i^2	u_i^3	u_i^4
	$u_i^1=1$	$u_i^1=0$	$u_i^1=0$	$u_i^1=0$
	$u_i^2=0$	$u_i^2=1$	$u_i^2=0$	$u_i^2=0$
	$u_i^3=0$	$u_i^3=0$	$u_i^3=1$	$u_i^3=0$
	$u_i^4=0$	$u_i^4=0$	$u_i^4=0$	$u_i^4=1$
Hu et al. (2007)	$(1-u_i)(1-v_i)$	$u_i(1-v_i)$	$(1-u_i)v_i$	u_iv_i
	$u_i=0$	$u_i=1$	$u_i=0$	$u_i=1$
	$v_i=0$	$v_i=0$	$v_i=1$	$v_i=1$
Proposal	$1-u_i-v_i+s_i$	u_i-s_i	v_i-s_i	s_i
	$u_i=0$	$u_i=1$	$u_i=0$	$u_i=1$
	$v_i=0$	$v_i=0$	$v_i=1$	$v_i=1$
	$s_i=0$	$s_i=0$	$s_i=0$	$s_i=1$

2.2 *Model Extension*

In the following we present a non-linear MIP approach as an extension to implement aisles simultaneously. Thus, there is need to introduce further notations:

Indices

z department index ranges from 1 to N

k aisle index equals 1 in case of horizontal, 2 in case of vertical segment

Parameter

w_{ij}^a width of aisle between dep. i and j

Decision variables

(x_{ijk}^a, y_{ijk}^a) spatial coordinate of diagonally opposite point of $(x_i^{I(O)}, y_i^{I(O)})$

t_{ij} =1, if segment k = 1(2) is spanned by $(x_i^{I(O)}, y_i^{I(O)})$, 0 otherwise

p_{ijk} =1, if $x_{ij1}^a \geq x_i^{I(O)}$, also if $y_{ij2}^a \geq y_i^{I(O)}$, 0 otherwise

q_{ijk} =1, if $y_{ij1}^a \geq y_i^{I(O)}$, also if $x_{ij2}^a \geq x_i^{I(O)}$, 0 otherwise

r_{ijzk}^1 =1, if segment k is placed to the left of dep. z, 0 otherwise

r_{ijzk}^2 =1, if segment k is placed to the right of dep. z, 0 otherwise

r_{ijzk}^3 =1, if segment k is placed below dep. z, 0 otherwise

r_{ijzk}^4 =1, if segment k is placed above dep. z, 0 otherwise

Additional constraints:

$$x^a_{ij1} = c_{ij}(x^o_i(1 - t_{ij}) + x^I_j t_{ij} - e_{ij} + 2e_{ij}p_{ij1}) \quad \forall i \neq j \tag{21}$$

$$y^a_{ij1} = c_{ij}(y^o_i(1 - t_{ij}) + y^I_j t_{ij} - w^a_{ij} + 2w^a_{ij}q_{ij1}) \quad \forall i \neq j \tag{22}$$

$$x^a_{ij2} = c_{ij}(x^o_i t_{ij} + x^I_j(1 - t_{ij}) - w^a_{ij} + 2w^a_{ij}q_{ij2}) \quad \forall i \neq j \tag{23}$$

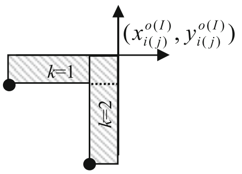
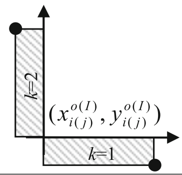
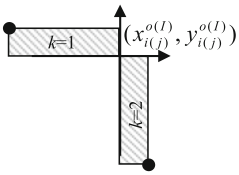
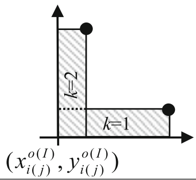
$$y^a_{ij2} = c_{ij}(y^o_i t_{ij} + y^I_j(1 - t_{ij}) - d_{ij} + 2d_{ij}p_{ij2}) \quad \forall i \neq j \tag{24}$$

Due to the assumption of at most one bend per aisle, each aisle consists of one horizontal ($k = 1$) and one vertical ($k = 2$) segment for every pair of departments. Each segment is spanned by two diagonally opposite points. One point of each segment is either an input or output point, depending on the binary variable t_{ij}. Spatial coordinates of the diagonally opposite points are defined by the Eqs. (21)–(24). These consider aisle width and aisle length. The latter uses vertical (d_{ij}) and horizontal (e_{ij}) distances between the material flow points previously defined. Binary variables p_{ijk} and q_{ijk} control the span direction of aisle segments. Allowable configurations are shown in Table 2. To reduce computing time, binary variable c_{ij} puts spatial coordinates of the aisle point to zero, if there is no flow and no need of an aisle.

$$(x^a_{ij1} - x^O_i t_{ij} - x^I_j(1 - t_{ij}))c_{ij} = 0 \quad \forall i \neq j \tag{25}$$

$$(y^a_{ij2} - y^O_i(1 - t_{ij}) - y^I_j t_{ij})c_{ij} = 0 \quad \forall i \neq j \tag{26}$$

Table 2 Available aisle segment orientations

	$p_{ijk} = 0$	$p_{ijk} = 1$
$q_{ijk} = 0$		
$q_{ijk} = 1$		

To ensure an aisle is continuous, Eqs. (25) and (26) require that segments between two positioning departments need to touch each other. Thus, the horizontal segment has to start at an input point of department j and has to reach up to the output point of department i regarding the x direction. Furthermore, the corresponding vertical segment needs to start in the output point of department i and has to reach vertically up to the input point of department j and vice versa.

Figure 2 illustrates this requirement and shows two cases. The two segments can be connected like a continuous aisle (case 1) or with a gap in the bend (case 2). With further extensions, gaps can be avoided. But during the testing phase such cases do not matter, because further aisle configurations are included and fill up such gaps to use available floor space efficiently.

Aisles and departments are not allowed to overlap, which is managed in the inequalities (27) until (37). The overlap mechanism is analogous to the initial model. Assuming department z is positioned to the left (lower) side of aisle segment k between departments i and j, binary variable $r^{1(3)}_{ijzk}$ returns 0 to fulfill inequalities (27), (29), (31), and (33). Without an effective M, these would require that the left (lower) border of department z is to the right (upper) of the right (upper) border of aisle segment k between departments i and j. In contrast, case $r^{2(4)}_{ijzk}$ returns 0 in (28), (30), (32), and (34), if department z is placed to the right (upper) side of aisle segment k. If there is material flow between departments i and j, inequality (35) needs not less than one $r^{n}_{ijzk} = 1$ to differ positioning coordinates in at least one direction. Finally, (36) assures that aisles are placed within the floor dimensions and (37) defines binary variables.

$$x_z + M(1 - r^1_{ijz1}) \geq x^a_{ij1} + e_{ij}(1 - p_{ij1}) \quad \forall z, i \neq j \tag{27}$$

$$x'_z \leq x^a_{ij1} - e_{ij}p_{ij1} + M(1 - r^2_{ijz1}) \quad \forall z, i \neq j \tag{28}$$

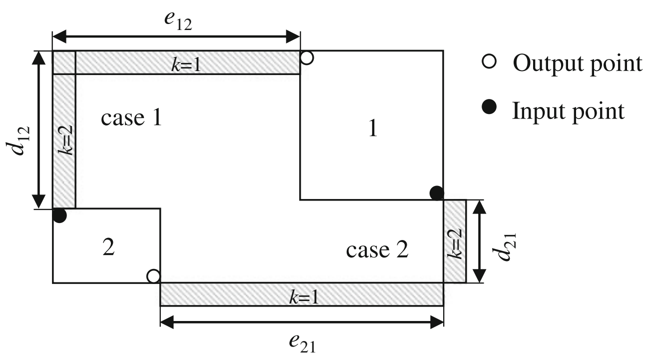


Fig. 2 Available aisle segment connections

$$y_z + M(1 - r^3_{ijz1}) \geq y^a_{ij1} + w^a_{ij}(1 - q_{ij1}) \quad \forall z, i \neq j \tag{29}$$

$$y^{'}_z \leq y^a_{ij1} - w^a_{ij}q_{ij1} + M(1 - r^4_{ijz1}) \quad \forall z, i \neq j \tag{30}$$

$$x_z + M(1 - r^1_{ijz2}) \geq x^a_{ij2} + w^a_{ij}(1 - q_{ij2}) \quad \forall z, i \neq j \tag{31}$$

$$x^{'}_z \leq x^a_{ij2} - w^a_{ij}q_{ij2} + M(1 - r^2_{ijz2}) \quad \forall z, i \neq j \tag{32}$$

$$y_z + M(1 - r^3_{ijz2}) \geq y^a_{ij2} + d_{ij}(1 - p_{ij2}) \quad \forall z, i \neq j \tag{33}$$

$$y^{'}_z \leq y^a_{ij2} - d_{ij}p_{ij2} + M(1 - r^4_{ijz2}) \quad \forall z, i \neq j \tag{34}$$

$$r^1_{ijzk} + r^2_{ijzk} + r^3_{ijzk} + r^4_{ijzk} \geq c_{ij} \quad \forall z, k, i \neq j \tag{35}$$

$$0 \leq x(y)^a_{ijk} \leq W(H) \quad \forall k, i \neq j \tag{36}$$

$$t_{ij}, c_{ij}, q_{ij}, p_{ij}, r^1_{ijzk}, r^2_{ijzk}, r^3_{ijzk}, r^4_{ijzk} \in \{0, 1\} \quad \forall i, j, z, k \tag{37}$$

2.3 Test Case

To demonstrate feasibility, four test instances with four and six departments are presented in the following. The above model was implemented in the optimization software Lingo 10.0, which employs a branch-and-bound algorithm. It supports a feature to linearize non-linear models, which we use to find the optimal solution and to overcome local optima. Note that the automatic linearization of Lingo results in an exceptionally long computing time. Thus, an extensive performance analysis is reasonable by testing a linearized version of this MIP, which would, however, exceed the length of this article.

The departments' data of the four test instances are presented in Tables 3and4. Input and output points are assumed to be in the corners of departments. Aisle widths are defined to be 1. The dimension of floor space is 30 × 30. Because there is high material flow intensity between departments 5 and 6, the centroid coordinates of department 6 are determined to be less than or equal to the centroid coordinates of department 5, considering (15) and (16). Table 4 provides an overview of the four test instances, the resulting optima, and computing times. The first two instances consist of departments 3–6 (grey-shaded area in Table 3) and the remaining instances consider all six departments. To show effects of the quantity of aisles, test instances with an equal number of machines differ in needed material flow from departments 4–6 (slash in Table 3).

The results of the test instances show that complexity increases the more departments and aisles that are considered. Furthermore, Table 4 indicates that computing time is more affected by the number of departments than the number of

Table 3 Test instances

	Material flow f_{ij} from/to						Resulting need of aisles c_{ij} from/to						Department dimensions and coordinates of input/output points					
dep i	1	2	3	4	5	6	1	2	3	4	5	6	h_i	w_i	P_i^x	D_i^x	P_i^y	D_i^y
1	0	1	2	0	0	0	0	1	1	0	0	0	2	2	0	2	2	0
2	0	0	0	0	3	0	0	0	0	0	1	0	4	2	0	2	4	0
3	0	0	0	3	0	2	0	0	0	1	0	1	6	3	0	3	6	0
4	4	0	0	0	0	0/2	1	0	0	0	0	0/1	5	4	0	4	5	0
5	2	0	0	0	0	1	1	0	0	0	0	1	7	3	0	3	7	0
6	0	0	0	0	10	0	0	0	0	0	1	0	4	3	0	3	4	0

Table 4 Optimal solutions

No. Instance	# departments	$\sum c_{ij}$	Optimal solution	Computing time [hh:mm:ss]
1	4	4	11	00:00:22
2	4	5	29	00:00:35
3	6	9	65	04:36:45
4	6	10	72	05:32:50

needed aisles. On the other hand, further experimentation showed that computing times are strongly influenced by the choice of departmental pairs which require material flow. For example, considering an additional material flow from department 6 to 4 in test instance no. 4 was not solvable within 24 h.

As an illustrative example, the optimal solution of 65 total travel distances of test instance no. 3 is given in Fig. 3. It shows that if output and input points are placed side by side, no aisles are needed. Hence, needed aisles are efficiently placed around the departments. Furthermore, floor dimensions are chosen as loose and do not constrain the relative position of departments.

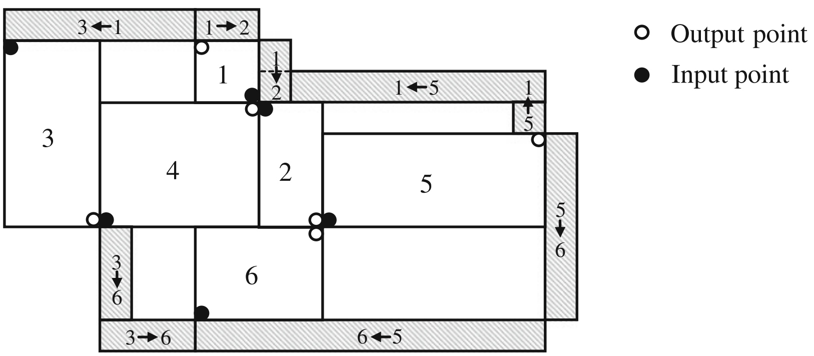


Fig. 3 Optimal layout of the test instance

3 Conclusions

In this article we proposed a new optimal approach for the FLP and provided a two-fold contribution. Firstly, to realize the departments' 90° step orientation we reduced responsible variables to two binary variables and one continuous variable. This leads to shorter computing times. Secondly, a MIP-approach was offered to handle aisle structure and the placement of departments simultaneously.

Due to the NP-hard nature of the FLP, only small problems are solvable. Earlier studies dedicated the allocation of departments to the approaches of Group Technology and Cellular Manufacturing (Tam 1992). Thus, the allocation within cells seems to be less complex due to less equipment being handled at once. The study of Wemmerlöv and Hyer (1989) showed that most companies composed cells of at most six machines. This indicates applicability of exact approaches. Furthermore, prior research showed huge potential through logical improvement and linearization (Sherali et al. 2003). Further research considering the proposed model could be profitable. Besides, to the best of our knowledge there is no MIP-approach to handle contour distances. Considering this, it would be of interest to deal with more than one bend per aisle.

References

Barbosa-Póvoa AP, Mateus R, Novais AQ (2002) Optimal 3D layout of industrial facilities. Int J Prod Res 40(7):1669–1698

Bock S, Hoberg K (2007) Detailed layout planning for irregularly-shaped machines with transportation path design. Eur J Oper Res 177(2):693–718

Bukchin Y, Meller RD, Liu Q (2006) Assembly system facility design. IIE Trans 38(1):67–79

Bukchin Y, Tzur M (2014) A new MIP approach for the facility layout design problem with rectangular and L/T shaped departments. Int J Prod Res 52(24):7339–7359

Dombrowski U, Hennersdorf S, Palluck M (2007) Die wirtschaftliche Bedeutung der Fabrikplanung am Standort Deutschland. Zeitschrift für wirtschaftlichen Fabrikbetrieb 102 (1/2):14–18

Drira A, Pierreval H, Hajri-Gabouj S (2007) Facility layout problems. Ann Rev Control 31 (2):255–267

Emami S, Nookabadi AS (2013) Managing a new multi-objective model for the dynamic facility layout problem. Int J Adv Manuf Technol 68(9–12):2215–2228

Heragu SS, Kusiak A (1990) Machine layout: an optimization and knowledge based approach. Int J Prod Res 28(4):615–635

Heragu SS, Kusiak A (1991) Efficient models for the facility layout problem. Eur J Oper Res 53 (1):1–13

Hu GH, Chen YP, Fang HC (2007) A genetic algorithm for the inter-cell layout and material handling system design. Int J Adv Manuf Technol 34(11/12):1153–1163

Kallrath J (2002) Gemischt-ganzzahlige optimierung: modellierung in der Praxis. Vieweg, Braunschweig

Kim J-G, Kim Y-D (2000) Layout planning for facilities with fixed shapes and input and output points. Int J Prod Res 38(18):4635–4653

Leno IJ, Saravanasankar S, Ponnambalam SG (2012) Layout design for efficient material flow path. Procedia Eng 38:872–879

Meller RD, Kirkizoglu Z, Chen W (2010) A new optimization model to support a bottom-up approach to facility design. Comput Oper Res 37(1):42–49

Patsiatzis DI, Papageorgiou LG (2002) Optimal multi-floor process plant layout. Comput Chem Eng 44(3):575–583

Ripon KSN, Glette K, Khan KN, Hovin M, Torresen J (2013) Adaptive variable neighborhood search for solving multi-objective facility layout problems with unequal area facilities. Swarm Evol Comput 8:1–12

Singh SP (2009) Solving facility layout problem. Three-level tabu search metaheuristic approach. Int J Recent Trends Eng 1(1):73–77

Sherali HD, Fraticelli BMP, Meller RD (2003) Enhanced model formulations for optimal facility layout. Oper Res 51(1):629–644

Tam KY (1992) A simulated annealing algorithm for allocating space to manufacturing cells. Int J Prod Res 30(1):63–87

Tavares J, Ramos C, Neves J (2002) Addressing the facility layout design problem through constraint logic programming. Int J Manuf Sci Prod 4(3):149–167

Tompkins JA, White JA, Bozer YA, Tanchoco JMA (2010) Facilities Planning, 4th edn. Wiley, New York

Wemmerlöv U, Hyer NL (1989) Cellular manufacturing in the U.S. industry: a survey of users. Int J Prod Res 27(9):1511–1530

Integrated Facility Location, Capacity and Production Planning in a Multi-Commodity Closed Supply Chain Network

Leena Steinke and Kathrin Fischer

Abstract In this paper the network design, the capacity and production planning of a closed-loop supply chain (CLSC) with multiple make-to-order (MTO) products consisting of common components is studied. The components can be remanufactured or can be purchased new from suppliers. It is assumed that products are returned by customers to the same facilities where the final products are sold. Transportation between facilities of the same type, i.e. redistribution, is allowed. The network is explored over multiple time periods. The objective of this work is to extend the strategic CLSC management by integrating decisions regarding production planning on an aggregate level to achieve an integrated optimization of facility locations, capacity equipment and production quantities at the facilities and forward and reverse material flows. For this purpose a mixed-integer linear optimization model is developed and solved for an example set of data.

1 Introduction

A supply chain can be closed by collecting non-desirable or not functioning products which are returned by customers after usage, at the end of the leasing period or of product life. There are different product recovery options, as repair, refurbishing, recycling, remanufacturing and cannibalization (Thierry et al. 1995) which enable the recovery of the products, their components or materials in the same or another supply chain. A narrow definition of a CLSC, in contrast to an open loop supply chain, requires the original equipment manufacturer (OEM) to recover his products and to reintegrate them, thus the recovered products enter the original supply chain again. Remanufacturing can be seen as the "foundation" of the CLSC (Guide 2000). Especially OEMs are qualified for remanufacturing

L. Steinke (✉) · K. Fischer
Institute of Operations Research and Information Systems, Technische Universität Hamburg-Harburg, Schwarzenbergstr 95, 21073 Hamburg, Germany
e-mail: l.steinke@tuhh.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_8

because of their specific product knowledge, e.g. the copier manufacturer Xerox uses remanufactured parts in the assembly of new copiers (King et al. 2006). The reasons for producing companies to close their supply chain with product recovery can be either the fulfillment of legal requirements, gaining an economical benefit due to lower costs, or a sense of responsibility to decrease the social or environmental impact (Dekker et al. 2004). The savings by remanufacturing can be higher when common components are used in different products than without component commonality (Walther et al. 2009).

In the next section, selected relevant literature is presented and the modelling approach taken in this work is compared to existing models. In Sect. 3 the network and problem structure which is studied here is described. Afterwards an optimization model for a CLSC with multiple product types, component commonality and MTO over multiple time periods is presented. The model is solved for a specific data set and the solution is discussed. In the last section, conclusions and further research perspectives are given.

2 Literature Review

Souza (2013) classifies the CLSC management literature depending on the impact level: there are operational, tactical and strategic issues of a CLSC. Decisions on disassembly planning, material requirement plans, scheduling and routing are made every day and are operational. The decisions regarding production planning and inventory management are made on a weekly or monthly basis and hence, are tactical. Decisions concerning strategic issues as network design, collection strategy, take-back arrangements and supply chain coordination have effects for years and hence are strategic. However, decisions made on one impact level are not independent of the other levels, but influence the respective decisions. Here, the focus is on strategic planning, i.e. the network design of a CLSC, but by incorporating aspects of production planning, implications of other planning levels are taken into account.

Following Fleischmann et al. (2000) there are three different types of networks for product recovery: Networks for recycling, for reusable items and for recovery of components or subassemblies. Products with a higher value, as e.g. copiers and automobiles, are suitable for recovery on component level which is the area studied here. In such a network remanufacturable products are collected, suitable products are disassembled into parts and repaired or refurbished, if necessary. Afterwards the parts are reassembled and reintegrated into the original supply chain as as-new items or sold in secondary markets as remanufactured items. Hence, the structure of a CLSC network is complex and has several levels. To design such a network optimally, the processing has to be considered in detail.

Mathematical optimization models used for network design with product recovery are extensions of the Warehouse Location Problem (WLP) and are based mostly on Mixed Integer Programming (MIP) and Mixed Integer Linear Programming (MILP). The models differ regarding number of product types, consideration of capacities and time horizon. Moreover, there are different approaches to model reverse flows and the reintegration of the recovered products into the supply chain. Demand and the quantities and quality of returned products are assumed as deterministic or stochastic. E.g. Fleischmann et al. (2004) model a network with three echelons and a single product over one time period under capacity limitations for product flows and deterministic demand. Furthermore, they present extensions, which consider uncertainty of demand and of the returned product quantities. Schultmann et al. (2003) study an existing network for multiple product types, which is to be expanded to handle an increased amount of returned products. Capacity is measured in terms of product units, which can be processed at the inspection and sorting facilities. Capacity is product-specific and induces costs.

Salema et al. (2007) extend the model of Fleischmann et al. (2004) by maximum and minimum capacity limits for the facilities and multiple product types. The described networks include separate distribution and collection centers (DCCs), while in this work, combined DCCs are considered. Furthermore, they assume that products are manufactured or remanufactured at the same, hybrid facility, whereas in this paper the production system is modelled in two stages: On the first stage, components can be newly purchased or can be remanufactured, and on the second stage, different types of products are assembled. Each product type is a specific combination of components, which can be used commonly among the different product types. Moreover, facility closings and transport between locations of the same type are possible in the planning problem studied here.

All models discussed above are Capacitated Facility Location Problems (CFLP). A more detailed literature review on network design of supply chains with product recovery can be found in Akcali et al. (2009).

The work by Pishvaee and Torabi (2010) is one of the few contributions which studies a CLSC network over multiple time periods, to consider in addition to the strategic issues, as facility location and allocating transportation quantities, tactical planning issues, as delivery dates at the facilities. A single product is produced and production consists of one stage. They use a multi-objective approach to combine cost minimization of the network and minimization of delivery tardiness. In this work we follow Pishvaee and Torabi's (2010) idea and extend the network design by production planning. Aggregate Production Planning (APP) is used to model the production, as all decisions are made on a strategic level. APP determines the production and inventory quantities which are needed to fulfill demand under a cost minimization objective (Nam and Logendran 1992). The aggregate plan is developed usually for a planning horizon of 6–24 months (Akinc and Roodman 1986). Jayaraman (2006) develops an APP model for a recovering company, the

Remanufacturing APP (RAPP) model. However, in this paper an integrated planning of remanufacturing and manufacturing is pursued. Furthermore, Jayaraman (2006) considers capacities in terms of labour hours as fixed for the total planning horizon, whereas in this work capacities can be adjusted over time. Moreover, the production planning aspect is taken into account in the network design, and capacities are also modeled in volume units. This combination of APP and CFLP allows to study the effect of facility opening and capacity equipment on the production planning, inventory holding and distribution system of the CLSCN, and vice versa.

3 Problem Definition

In this paper, a multi-echelon and multi-commodity network of a CLSC with remanufacturing on component level and bidirectional facilities for DCCs is studied over multiple time periods, which can be interpreted as years. The planning horizon T can be defined as the product lifecycle. In this section the network and its production and distribution system are introduced.

As shown in Fig. 1, there are three levels of facilities in the network: The remanufacturing centers R and suppliers Z which deliver the components, the plants F and the DCCs V. It is possible to hold remanufactured components and returned products on stock at the remanufacturing centers. Moreover, there is a component inventory at the plants, but there are no final product inventories at the plants or at the DCCs. The final products P are specific combinations of components C and are assembled whenever it is optimal to satisfy an order (MTO), otherwise the demand is lost. Customers demand different quantities of final products in each period. Demand is assumed as deterministic and known. Customer locations are defined by the set K.

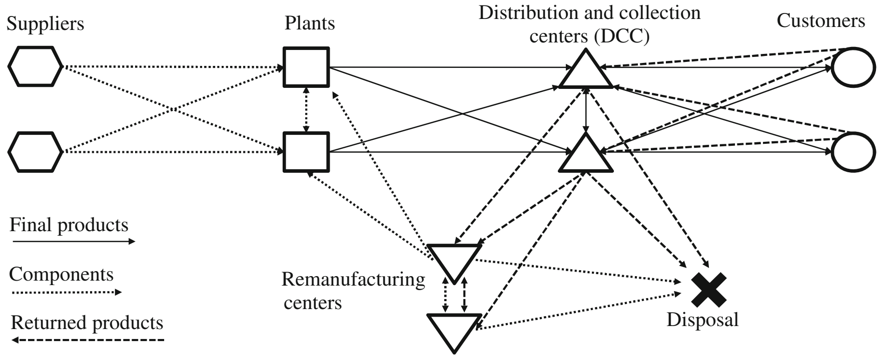


Fig. 1 Closed-loop supply chain network

Besides the distribution of new products from the DCCs to the customers, used products are returned by customers and collected in DCCs. The components for product assembly are either purchased from suppliers or shipped from remanufacturing centers to the plants, assuming that components from remanufacturing centers have the same quality as the components from suppliers. While production planning approaches on a tactical planning level mostly assume independence of demand and product returns, and study the impact of uncertainty on the remanufacturing decision, e.g. Inderfurth (2005), it is assumed here that returned products are a known fraction of the products shipped to customers in the previous period. As we study the network and the product flows in the long run, the assumption of related demand and return quantities is valid, especially for the case of returned products at the end of leasing period, as, e.g., in the copier industry (King et al. 2006). The returned products which are collected in the DCCs are shipped to the remanufacturing centers or are disposed, i.e. shipped to the disposal unit D. There are no lead-times considered, as production is planned on an aggregate level.

Given the demand and return product quantities for each period, the corresponding CLSC has to be designed such that total cost is minimized. Decisions to be made consist in the determination of facilities to be opened and closed and their capacity levels in working hours and volume units in the different periods, as well as products to be assembled, components to be purchased, returned products to be remanufactured and the amount of returned products and components to be stored at open facilities. The capacity levels at open facilities can be increased or decreased in fixed steps in every period. Furthermore, the model includes the calculation of the quantities shipped between facilities of different echelons and between facilities of the same echelon. The transport between facilities of one echelon can, for example, be optimal in the case of different capacity limits at facilities of the same type.

4 Model Formulation

In this section the MILP model for the CLSC network (CLSCN) design problem described in Sect. 3, the Capacitated Facility Location and Aggregate Production Planning with Remanufacturing (CFLAPPR) model, is presented. The following notation is used in the formulation of the model (Tables 1 and 2).

Using the above notation, the multi-period and multi-commodity CFLAPPR problem can be formulated as follows. The objective function looks as follows; explanations of the different components are given below.

Table 1 Definition of relevant parameters

Parameter	Definition
a_{cp}	Number of component c yielded by the remanufacturing of one product unit of $p, c \in C$ and $p \in P$
b_{cp}	Number of component c needed for producing one unit of product $p, c \in C$ and $p \in P$
Cap_{yt}^{O}	Maximum capacity at facility y in period t (in m^3), $\forall y \in F, V$ or $R, t \in T$
Cap_{y}^{U}	Minimum capacity at facility y (in m^3), $\forall y \in F, V$ or R
c_z^c	Unit cost for procuring component c from supplier $z, \forall c \in C$ and $z \in Z$
c^{DEnt}	Disposal cost (per unit)
c_k^U	Unit penalty cost for unmet demand $k, \forall k \in K$
c_{yw}^x	Cost for transportation of a unit x from facility y to facility w (per km), $\forall y, w \in F, V$ and $R : y \neq w, x \in C$ or P
c_y^x	Unit cost for processing at facility $y, \forall y \in F, V$ or R
dy_x	Time required for processing a unit of x at facility $y, \forall y \in F, V$ or $R,\ x \in C$ or P
e	Size of capacity step by which the locations can be extended within one period (in m^3)
$fcap_y$	Cost or revenue for capacity expansion or reduction at facility $y, \forall y \in F, V$ or R
f_y	Cost for opening facility $y, \forall y \in F, V$ or R
f_t^y	Cost for open facility y in period $t, \forall y \in F, V$ or $R, t \in T$
g_x	Volume of one unit of x (in m^3), $\forall x \in C$ or P
h_f^c	Cost per period for holding a unit of c at plant f in inventory, $\forall f \in F$ and $c \in C$
h_r^x	Cost per period for holding a unit of x in inventory at remanufacturing center $r, \forall r \in R, x \in C$ or P
$LabCap_{yt}^{O}$	Maximum labour hours available at facility y in period $t, \forall y \in F, V$ or $R, t \in T$
$LabCap_{y}^{U}$	Minimum labour hours at facility $y, \forall y \in F, V$ or R
$labcc_y$	Hourly cost or revenue for labour capacity expansion or reduction at facility $y, \forall y \in F, V$ or R
le	Labour hours per worker (in hours) in one period
LT	Last planning time period, $LT \in T$
M	Sufficiently large number
md_t	Minimum proportion of returned products that has to be disposed after visual inspection at the DCCs in period $t, \forall t \in T$
md_t^c	Minimum proportion of component c that has to be disposed after disassembly, repair and testing in period $t, \forall t \in T$ and $c \in C$
N_{kpt}	Demand of customer k for product p in period $t, \forall k \in K, p \in P$ and $t \in T$
q_{kpt}	Return rate of customer k for product p in period $t, \forall k \in K, p \in P$ and $t \in T$
sf_y	Cost for closing facility $y, \forall y \in F, V$ or R

(continued)

Table 1 (continued)

Parameter	Definition
sh_r^x	Cost for disposing a stored unit of x at remanufacturing center r at the last planning period $LT, \forall r \in R, x \in C$ or P
sh_f^c	Cost for disposing a stored unit of c at plant f at the last planning period $LT, \forall f \in F$ and $c \in C$
t_{yw}^x	Distance of facility y to another facility w (in km), $\forall y, w \in F, V$ and $R : y \neq w, x \in C$ or P

Table 2 Definition of relevant variables

Variable	Definition
Cap_t^y	Number of capacity steps at open facility y in period t (in m^3), $\forall y \in F, V$ or $R, t \in T$
$CCap_t^y$	Expansion or reduction of capacity steps at facility y in period t (in m^3), $\forall y \in F, V$ or $R, t \in T$
$CLCap_t^y$	Expansion or reduction of the workforce at facility y in period $t, \forall y \in F, V$ or $R, t \in T$
EI_c^f	Quantity of c remaining in the inventory of plant f at the end of the last planning period $LT, \forall f \in F$ and $c \in C$
EI_x^r	Quantity of x remaining in inventory of remanufacturing center r at the end of the last planning period $LT, \forall r \in R, x \in C$ or P
EXI_{xt}^{yw}	Quantity of x transported from facility y to another facility w of the same echelon in period $t, \forall y, w \in F, V$ and $R : y \neq w, x \in C$ or $P, t \in T$
H_t^y	$\begin{cases} 1, & \text{if facility } y \text{ is closed} \\ & \text{in period } t, \forall t \in T \text{ and } y \in F, V \text{ or } R \\ 0, & \text{otherwise} \end{cases}$
I_{ct}^f	Quantity of c remaining at plant f at the end of period $t, \forall f \in F, c \in C, t \in T$
I_{xt}^r	Quantity of x remaining at remanufacturing center r at the end of period $t, \forall r \in R, x \in C$ or $P, t \in T$
$LCap_t^y$	Workforce available at facility y in period $t, \forall y \in F, V$ or $R, t \in T$
U_{pt}^k	Number of unmet demand for product p of customer k in period $t, \forall p \in P,\ k \in K, t \in T$
X_{xt}^y	Quantity of x processed in facility y in period $t, \forall y \in F$ or $R, x \in C$ or $P, t \in T$
X_{xt}^{yw}	Quantity of x transported from facility y to another facility w or to the disposal unit D in period $t, \forall y, w \in FD, V, RD$ and $Z : y \neq w, x \in C$ or $P, t \in T$
Y_t^y	$\begin{cases} 1, & \text{if facility } y \text{ is open} \\ & \text{in period } t, \forall t \in T \text{ and } y \in F, V \text{ or } R \\ 0, & \text{otherwise} \end{cases}$
Y^y	$\begin{cases} 1, & \text{if facility } y \text{ is opened in the planning horizon,} \\ & \forall y \in F, V \text{ or } R \\ 0, & \text{otherwise} \end{cases}$

$$
\begin{aligned}
\min \sum_{r \in R} f_r \cdot Y^r &+ \sum_{v \in V} f_v \cdot Y^v + \sum_{f \in F} f_f \cdot Y^f \\
&+ \sum_{t \in T} \Bigg(\sum_{r \in R} f_t^r \cdot Y_t^r + \sum_{v \in V} f_t^v \cdot Y_t^v + \sum_{f \in F} f_t^f \cdot Y_t^f \\
&+ \sum_{r \in R} sf_r \cdot H_t^r + \sum_{v \in V} sf_v \cdot H_t^v + \sum_{f \in F} sf_f \cdot H_t^f \\
&+ \sum_{r \in R} fcap_r \cdot CCap_t^r + \sum_{v \in V} fcap_v \cdot CCap_t^v \\
&+ \sum_{f \in F} fcap_f \cdot CCap_t^f + le \cdot \Bigg(\sum_{r \in R} labcc_r \cdot CLCap_t^r \\
&+ \sum_{v \in V} labcc_v \cdot CLCap_t^v + \sum_{f \in F} labcc_f \cdot CLCap_t^f \Bigg) \\
&+ \sum_{z \in Z, c \in C, f \in F} (c_{zf}^c \cdot t_{zf}^c + c_z^c) \cdot X_{ct}^{zf} + \sum_{r \in R, c \in C, f \in F} c_{rf}^c \cdot t_{rf}^c \cdot X_{ct}^{rf} \\
&+ \sum_{r \in R, c \in C} c_r \cdot X_{ct}^r + \sum_{f \in F, p \in P} c_f \cdot X_{pt}^f \\
&+ \sum_{f \in F, v \in V, p \in P} c_{fv}^p \cdot t_{fv}^p \cdot X_{pt}^{fv} \\
&+ \sum_{v \in V, k \in K, p \in P} c_{vk}^p \cdot t_{vk}^p \cdot X_{pt}^{vk} + \sum_{k \in K, p \in P} c_k^U \cdot U_{pt}^k \\
&+ \sum_{v \in V, k \in K, p \in P} c_{kv}^p \cdot t_{vk}^p \cdot X_{pt}^{kv} + \sum_{v \in V, r \in R, p \in P} c_{vr}^p \cdot t_{vr}^p \cdot X_{pt}^{vr} \\
&+ c^{DEnt} \cdot \Bigg(\sum_{v \in V, p \in P} X_{pt}^{vD} + \sum_{r \in R, c \in C} X_{ct}^{rD} \Bigg) \\
&+ \sum_{c \in C, (r,s) \in R : r \neq s} c_{rs}^c \cdot t_{rs}^c \cdot EXI_{ct}^{rs} + \sum_{p \in P, (r,s) \in R : r \neq s} c_{rs}^p \cdot t_{rs}^p \cdot EXI_{pt}^{rs} \\
&+ \sum_{c \in C, (f,i) \in F : f \neq i} c_{fi}^c \cdot t_{fi}^c \cdot EXI_{ct}^{fi} + \sum_{p \in P, (v,j) \in V : v \neq j} c_{vj}^p \cdot t_{vj}^p \cdot EXI_{pt}^{vj} \\
&+ \sum_{r \in R, p \in P} h_r^p \cdot I_{pt}^r + \sum_{r \in R, c \in C} h_r^c \cdot I_{ct}^r + \sum_{f \in F, c \in C} h_f^c \cdot I_{ct}^f \Bigg) \\
&+ \sum_{r \in R, p \in P} sh_r^p \cdot EI_p^r + \sum_{r \in R, c \in C} sh_r^c \cdot EI_c^r + \sum_{f \in F, c \in C} sh_f^c \cdot EI_c^f
\end{aligned} \tag{1}
$$

The objective of the CFLAPPR model is to minimize the total costs of the CLSCN over multiple time periods (1). The first line represents the costs for opening facilities, as remanufacturing centers, plants and DCCs. Whenever a facility is in use, fixed costs occur, as shown in line two of the objective function. Moreover, it is possible to close facilities; in doing so, costs for closing facilities are incurred which are stated in the third line. The costs in the following lines are the costs or revenues of capacity expansion or reduction, respectively, at the existing facilities in the network. We differentiate between capacity in terms of storage space in m^3, see line four and five, and labour hours of the workforce, see line five and six. The seventh line contains the costs of procuring new components from suppliers and the transportation costs for moving components between suppliers and plants. The costs for shipping components from remanufacturing centers to plants are listed in the seventh line, too. The costs of remanufacturing components and of assembling products are captured in the eighth line. The following line defines the costs for shipping products from plants to DCCs. In the next line, the shipping costs from DCCs to customers are listed. The penalty costs induced by unmet demand are stated in this line, too. In line eleven, the costs for shipping used products from the customers to the DCCs and further in the network onwards to the remanufacturing centers are listed. The costs for shipping returned products and repaired components from the DCC and the remanufacturing centers to the disposal unit are stated in line twelve. The costs for shipping components or products, respectively, between facilities of the same kind are listed in line thirteen and fourteen. Line fifteen describes the inventory holding costs at the remanufacturing centers and the plants, since every stored item occupies space in terms of volume units. The costs captured in line two to fifteen occur in every period and hence, these costs are added up over the planning horizon. Finally, in the last line costs are listed which are caused by the items remaining in the inventories at the end of planning horizon. They may be interpreted as disposal costs.

Below the constraints are presented, followed by their respective explanation.

$$\sum_{t \in T} Y_t^y \leq (LT + 1) \cdot Y^y \quad \forall y \in F, R \text{ } or \text{ } V \tag{2}$$

$$Y_{t-1}^y - Y_t^y \leq H_t^y \quad \forall y \in F, R \text{ } or \text{ } V, t \in T : t > 0 \tag{3}$$

$$\sum_{t \in T} H_t^y \leq 1 \quad \forall y \in F, R \text{ } or \text{ } V \tag{4}$$

$$\sum_{t \in T : t \geq s} Y_t^y \leq M \cdot \left(1 - H_s^y\right) \quad \forall y \in F, R \text{ } or \text{ } V, s \in T \tag{5}$$

$$\sum_{v \in V} X_{pt}^{vk} + U_{pt}^k = N_{kpt} \quad \forall k \in K, p \in P, t \in T \tag{6}$$

$$\sum_{v \in V} X_{pt}^{kv} = q_{kpt} \cdot \sum_{v \in V} X_{pt-1}^{vk} \quad \forall k \in K, p \in P, t \in T \tag{7}$$

$$I_{ct}^f = I_{ct-1}^f + \sum_{z \in Z} X_{ct}^{zf} + \sum_{r \in R} X_{ct}^{rf} + \sum_{i \in F : i \neq f} EXI_{ct}^{if} - \sum_{i \in F : i \neq f} EXI_{ct}^{fi} - X_{ct}^f$$
$$\forall f \in F, c \in C, t \in T : t > 0 \tag{8}$$

$$I_{pt}^{r} = I_{pt-1}^{r} + \sum_{v \in V} X_{pt}^{vr} + \sum_{s \in R: s \neq r} EXI_{pt}^{sr} - \sum_{s \in R: s \neq r} EXI_{pt}^{rs} - X_{pt}^{r} \quad \forall r \in R, p \in P, t \in T : t > 0 \tag{9}$$

$$I_{ct}^{r} = I_{ct-1}^{r} + X_{ct}^{r} + \sum_{s \in R: s \neq r} EXI_{ct}^{sr} - \sum_{s \in R: s \neq r} EXI_{ct}^{rs} - \sum_{f \in F} X_{ct}^{rf} \quad \forall r \in R, c \in C, t \in T : t > 0 \tag{10}$$

$$I_{c0}^{f} = 0 \quad \forall f \in F, c \in C \tag{11}$$

$$I_{x0}^{r} = 0 \quad \forall r \in R, x \in C \text{ or } P \tag{12}$$

$$I_{cLT}^{f} = EI_{c}^{f} \quad \forall f \in F, c \in C \tag{13}$$

$$I_{xLT}^{r} = EI_{x}^{r} \quad \forall r \in R, x \in C \text{ or } P \tag{14}$$

$$\sum_{f \in F} X_{pt}^{fv} + \sum_{j \in V: j \neq v} EXI_{pt}^{jv} = \sum_{k \in K} X_{pt}^{vk} + \sum_{j \in V: j \neq v} EXI_{pt}^{vj} \quad \forall v \in V, p \in P, t \in T \tag{15}$$

$$\sum_{v \in V} X_{pt}^{fv} = X_{pt}^{f} \quad \forall f \in F, p \in P, t \in T \tag{16}$$

$$X_{ct}^{f} = \sum_{p \in P} b_{cp} \cdot X_{pt}^{f} \quad \forall f \in F, c \in C, t \in T \tag{17}$$

$$\sum_{r \in RD} X_{pt}^{vr} = \sum_{k \in K} X_{pt}^{kv} \quad \forall v \in V, p \in P, t \in T \tag{18}$$

$$X_{pt}^{vD} \geq md_{t} \cdot \sum_{k \in K} X_{pt}^{kv} \quad \forall v \in V, p \in P, t \in T \tag{19}$$

$$X_{ct}^{r} = \sum_{p \in P} a_{cp} \cdot X_{pt}^{r} \quad \forall r \in R, c \in C, t \in T \tag{20}$$

$$md_{t}^{c} \cdot X_{ct}^{r} \leq X_{ct}^{rD} \quad \forall r \in R, c \in C, t \in T \tag{21}$$

$$\sum_{p \in P} dV_{p} \cdot \left(\sum_{k \in K} \left(X_{pt}^{kv} + X_{pt}^{vk} \right) \right) \leq le \cdot LCap_{t}^{v} \quad \forall v \in \mathrm{V}, t \in T \tag{22}$$

$$\sum_{c \in C} dR_{c} \cdot X_{ct}^{r} \leq le \cdot LCap_{t}^{r} \quad \forall r \in R, t \in T \tag{23}$$

$$\sum_{p \in P} dF_p \cdot X_{pt}^{f} \leq le \cdot LCap_t^{f} \quad \forall f \in F, t \in T \tag{24}$$

$$LabCap_y^{U} \cdot Y_t^{y} \leq le \cdot LCap_t^{y} \leq LabCap_{yt}^{O} \cdot Y_t^{y} \quad \forall y \in F, R \text{ or } V, t \in T \tag{25}$$

$$LCap_t^{y} - LCap_{t-1}^{y} = CLCap_t^{y} \quad \forall y \in F, R \text{ or } V, t \in T : t > 0 \tag{26}$$

$$LCap_0^{y} = CLCap_0^{y} \quad \forall y \in F, R \text{ or } V \tag{27}$$

$$\sum_{c \in C} g_c \cdot I_{ct}^{f} \leq e \cdot Cap_t^{f} \quad \forall f \in F, t \in T \tag{28}$$

$$\sum_{c \in C} g_c \cdot \left(\sum_{z \in Z} X_{ct}^{zf} + \sum_{r \in R} X_{ct}^{rf} + \sum_{i \in F: i \neq f} EXI_{ct}^{if} \right) \leq e \cdot Cap_t^{f} \quad \forall f \in F, t \in T \tag{29}$$

$$\sum_{p \in P} g_p \cdot X_{pt}^{fv} \leq e \cdot Cap_t^{f} \quad \forall f \in F, t \in T \tag{30}$$

$$\sum_{c \in C} g_c \cdot I_{ct}^{r} + \sum_{p \in P} g_p \cdot I_{pt}^{r} \leq e \cdot Cap_t^{r} \quad \forall r \in R, t \in T \tag{31}$$

$$\sum_{c \in C} g_c \cdot \left(X_{ct}^{r} + \sum_{s \in R: s \neq r} EXI_{ct}^{sr} \right) + \sum_{p \in P} g_p \cdot \left(\sum_{v \in V} X_{pt}^{vr} + \sum_{s \in R: s \neq r} EXI_{pt}^{sr} \right) \leq e \cdot Cap_t^{r} \quad \forall r \in R, t \in T \tag{32}$$

$$\sum_{p \in P} g_p \cdot \left(\sum_{f \in F} X_{pt}^{fv} + \sum_{k \in K} X_{pt}^{kv} + \sum_{j \in V: j \neq v} EXI_{pt}^{jv} \right) \leq e \cdot Cap_t^{v} \quad \forall v \in \mathrm{V}, t \in T \tag{33}$$

$$Cap_y^{U} \cdot Y_t^{y} \leq e \cdot Cap_t^{y} \leq Cap_{yt}^{O} \cdot Y_t^{y} \quad \forall y \in F, R \text{ or } V, t \in T \tag{34}$$

$$Cap_t^{y} - Cap_{t-1}^{y} = CCap_t^{y} \quad \forall y \in F, R \text{ or } V, t \in T : t > 0 \tag{35}$$

$$Cap_0^{y} = CCap_0^{y} \quad \forall y \in F, R \text{ or } V \tag{36}$$

$$\begin{aligned}&U^k_{pt}, X^{kv}_{pt}, X^{vk}_{pt}, X^{vr}_{pt}, X^{zf}_{ct}, X^r_{ct}, X^r_{pt}, X^{rf}_{ct}, X^{fv}_{pt}, X^f_{pt}, X^f_{ct},\\ &EXI^{rs}_{ct}, EXI^{rs}_{pt}, EXI^{fi}_{ct}, EXI^{vj}_{pt}, I^f_{ct}, I^r_{ct}, I^r_{pt}, EI^f_c, EI^r_c, EI^r_p\\ &Cap^r_t, Cap^f_t, Cap^v_t, LCap^r_t, LCap^f_t, LCap^v_t \in \mathbb{Z}^+\\ &\forall p \in P, c \in C, r \in RD, k \in K, v \in V, f \in FD, z \in Z, t \in T\end{aligned} \tag{37}$$

$$CCap^y_t, CLCap^y_t \in \mathbb{Z} \forall y \in R, V \text{ or } F, t \in T \tag{38}$$

$$Y^y, Y^y_t, H^y_t \in \{0, 1\} \forall y \in F, R \text{ or } V, t \in T \tag{39}$$

Constraints (2) ensure that opening costs for facilities are considered if a facility is opened. If a facility is open in one period, but not in the next, it must be closed, see Constraints (3). Facilities can be closed just once, as shown in Constraints (4). Constraints (5) ensure that there is no opening in the periods following the closing. Demand for a product in period t can either be satisfied or not, see Constraints (6). Constraints (7) determine that every returned product is collected from the customers. Constraints (8), (9) and (10) are the inventory balance equations of the component inventory at the plants and of inventories for components and returned products at the remanufacturing centers. Every inventory is initially set to zero, see (11) and (12). At the end of the planning horizon the remaining items on stock at the facilities are determined by (13) and (14). We assume that these remaining items have to be disposed. DCCs are inventory-free, therefore the product inflows have to meet the amounts of products leaving DCCs in every period, as shown in Constraints (15). Constraints (16) ensure that the total amount of products assembled at plant f in period t is shipped to the DCCs. Constraints (17) determine the amount of components, which is needed for assembling the products in every period. The returned products are shipped from the DCCs to the remanufacturing centers and the disposal unit in the same period, see (18). There is a fraction md_t of returned products that fails the inspection at the DCCs and has to be disposed. This minimum disposal fraction is based on historical experience regarding the quality of returned products. However, it is possible to dispose more returned products than this fraction, see (19). Constraints (20) define the quantity of component c, which can be generated by disassembling and repairing returned products in period t. There is a fraction of components that does not meet the required quality level for as-new components and has to be disposed, but also more components than the defect ones can be disposed, (21).

At the open facilities workers are needed for the respective operations. Constraints (22), (23) and (24) determine the required working hours at the DCCs, the remanufacturing centers and the plants. They can be increased or decreased in every period by steps of le. However, if a facility is opened, its workforce level should be between the minimum and maximum quantity, which is specified by the operations and the availability of matching workers, (25). The changes in workforce levels at the facilities are defined in (26); (27) defines the starting level.

The inventory in period t at the plants and the remanufacturing centers cannot exceed the volume capacity available in period t, (28) and (31). Constraints (29), (30) and (32) are necessary to ensure the adherence to the capacity restrictions and to limit the flows to the volume capacity. Constraints (33) ensure that the volume of inflows to a DCC does not exceed its capacity. The volume capacity at the facilities can be increased and decreased by steps of e, but there is a minimum and maximum capacity level at open facilities, see (34). The changes in the volume capacity at the facilities are determined in (35) and (36).

The variables describing unsatisfied demand, the flows of products or components between echelons and between facilities of one echelon, produced units, units on stock and the capacities of the facilities have to be integers and positive, see (37). The change of capacities, expansion or reduction, is represented by variables which are integers and can be positive or negative, see (38). The variables indicating the opening and closing of facilities are binary, see (39).

5 Example and Solution

Based on data for the copier industry described in Fleischmann et al. (2001) for the network design and in van der Laan and Salomon (1997) for the production planning and inventory management, a network for Germany is designed. At the five biggest cities of Germany, based on number of inhabitants, see StatAmt (2015), there are suppliers and possible plant locations. The DCCs and remanufacturing centres can be located in the fifteen biggest German cities, where customer demand is predicted as 10 units per 1000 inhabitants, (Fleischmann et al. 2001).

In extension to Fleischmann et al. (2001), who only consider one year, the planning horizon is extended to three years and later on expanded to ten years. We study a network with two final products and three components. These two different final products consist of one common component and one specific component each. Disassembly is assumed as the reverse operation of the assembly operation. The return rate is independent of customer and product type, and all parameters are time-independent. There are no minimum capacity limits for the facilities. The distances between the locations are taken from Google Maps (2015). All other parameter values are shown in Table 3.

By using Gurobi 6.0.0 on a computer with two 3.10 GHz Intel® Xeon® Processors E5-2687 W and 128 GB RAM, the following solution is found for the above stated model within 6.7995 min: Plants and DCCs are open in Berlin and Cologne over the total planning horizon. A remanufacturing centre is open for the second and third period in Berlin. There is no closing of facilities, no unsatisfied demand and no transport between facilities of the same type. The capacities at the DCCs and the plants are at the upper bounds for the first two periods and lower in the last period. In the second and third period returned and recoverable products are brought to the remanufacturing centre in Berlin. In the second period products are remanufactured and the remanufactured components are shipped to the Berlin plant,

Table 3 Parameter values

Parameter	Definition
Cap_{yt}^{O}	500,000 m^3, $\forall y \in F, V$ or $R, t \in T$
c_z^c	10 monetary units (MU) per component, $\forall c \in C$ and $z \in Z$
c^{DEnt}	100 MU per unit
c_k^U	1,000 MU per unit, $\forall k \in K$
c_{yw}^c	0.0030 MU per km, $\forall c \in C, y \in Z$ or $R, w \in F$
c_{vk}^p	0.01 MU per km, $\forall p \in P, v \in V, k \in K$
c_{kv}^p	0.0050 MU per km, $\forall p \in P, v \in V, k \in K$
c_{vr}^p	0.0030 MU per km, $\forall p \in P, v \in V, r \in R$
c_{fv}^p	0.0045 MU per km, $\forall p \in P, f \in F, v \in V$
c_f	1 MU per unit, $\forall f \in F$
c_r	5 MU per unit, $\forall r \in R$
dV_p	0.5 h, $\forall p \in P$
dR_c	2 h, $\forall c \in C$
dF_p	1 h, $\forall p \in P$
e	1,000 m^3
$fcap_y$	1,000 MU, $\forall y \in F, V$ or R
f_y	50,000 MU, $\forall y \in F, V$ or R
f_t^y	10,000 MU, $\forall y \in F, V$ or $R, t \in T$
g_p	10 m^3, $\forall p \in P$
g_c	2 m^3, $\forall c \in C$
h_r^x	0.5 MU per unit, $\forall x \in C$ or $P, r \in R$
h_f^c	1 MU per unit, $\forall c \in C$ and $f \in F$
$LabCab_{yt}^{O}$	1,000,000 h, $\forall y \in F, V$ or $R, t \in T$
$labcc_y$	15 MU per hour, $\forall y \in F, V$ or R
le	1,610 h
md	0.3
md^c	0.1, $\forall c \in C$
(q_0, q_1, q_2)	(0, 0.3, 0.3)
sf_y	50,000 MU, $\forall y \in F, V$ or R
sh_y^x	2 MU per unit, $\forall y \in F$ or $R, x \in C$ or P

whereas in the last period the returned products are being stored at the remanufacturing centre. The volume and operational capacity is adapted to these flows: In the second period they are at their maximum levels, but in the last period there is only storage capacity at the remanufacturing centre, because it is not cost optimal to remanufacture. The corresponding costs of the network over three years are 9,827,628.28 [in monetary units (MU)].

Solving the problem with any return rate between 0.1 and 1 (varying in 0.1 steps) results in the same facility openings and remanufacturing decision, i.e. the return rate does not have an influence on the facility selection, and there never is remanufacturing in the last period. When the planning horizon is expanded to 10 periods, the solution time increases to 3.2 h. In the solution, products are remanufactured in period 2–9 and again returned products are stored in the last period. Hence, the trade-off between production and storage units is exploited. In the case of 10 periods, an additional plant and DCC in Munich are installed, i.e. the network is extended. This shows the impact of the planning horizon on the facility location decisions.

6 Conclusions

In this paper a model for a CLSCN with multiple MTO products and component commonality is presented and solved for an example case. To our knowledge this is the first approach to combine a WLP with APP. To study the effect of facility location and capacity planning on production planning and inventory holding, the CLSC is modelled over multiple periods. Experiments show that the length of the planning horizon and the capacities at the facilities influence the solution, i.e. have an impact on the network design and the decision to remanufacture. Further numerical studies are necessary to study the effect of an integrated planning of facility location, production and capacity planning under different conditions, e.g. under different costs or capacity limits. However, due to space limitations such a detailed sensitivity analysis has to be omitted here.

With the model presented above it is possible to study a CLSCN with multiple product and component types with different product-life cycles and the effect of component commonality on the remanufacturing decision. Moreover, more complex product structures than in the example presented in this paper, as described e.g. by Taleb et al. (1997) can be considered.

As described in Guide (2000), there are often uncertainties associated with product recovery. When products are returned at the end of the leasing period, timing and quantity of returned products are known. In contrast to this, for the case of returns at the end of the product life and at the end of use, the timing and quantities are difficult to predict. Also the quality of the returned products is unknown and therefore the remanufacturing costs, remanufacturing time and scrap rate can vary a lot among the returned products and components. The effect of uncertain demand and returned product quantities on the network can be studied by a scenario-based approach; this is left for future research.

References

Akcali E, Cetinkaya S, Uster H (2009) Network design for reverse and closed-loop supply chains: an annotated bibliography of models and solution approaches. Networks 53:231–248

Akinc U, Roodman GMA (1986) New approach to aggregate production planning. IIE Trans 18:88–94

Dekker R, Fleischmann M, Inderfurth K, Van Wassenhove LN (2004) Quantitative models for reverse logistics decision making. In: Dekker R, Fleischmann M, Inderfurth K, Van Wassenhove LN (eds) Reverse logistics: quantitative models for closed-loop supply chains, pp 29–41

Fleischmann M, Krikke HR, Dekker R, Flapper SDP (2000) A characterisation of logistics networks for product recovery. Omega 28:653–666

Fleischmann M, Beullens P, Bloemhof-Ruwaard JM, Van Wassenhove LN (2001) The impact of product recovery on logistics network design. Prod Oper Manag 10:156–173

Fleischmann M, Bloemhof-Ruwaard JM, Beullens P, Dekker R (2004) Reverse logistics network design. In: Dekker R, Fleischmann M, Inderfurth K, Van Wassenhove LN (eds) Reverse logistics: quantitative models for closed-loop supply chains, pp 65–94

Google Maps (2015) Google Maps. https://maps.google.com. Version: checked on: 12 Jan 2015

Guide VDR Jr (2000) Production planning and control for remanufacturing: industry practice and research needs. J Oper Manag 18:467–483

Inderfurth K (2005) Impact of uncertainties on recovery behavior in a remanufacturing environment: a numerical analysis. Int J Phys Distrib Logistics Manag 35:318–336

Jayaraman V (2006) Production planning for closed-loop supply chains with product recovery and reuse: an analytical approach. Int J Prod Res 44:981–998

King A, Miemczyk J, Bufton D (2006) Photocopier remanufacturing at Xerox UK. A description of the process and consideration of future policy issues. In: Brissaud D, Tichkiewitch S, Zwolinski P (eds) Innovation in life cycle engineering and sustainable development, pp 173–186

Nam S, Logendran R (1992) Aggregate production planning—A survey of models and methodologies. Eur J Oper Res 61:255–272

Pishvaee MS, Torabi SA (2010) A possibilistic programming approach for closed-loop supply chain network design under uncertainty. Fuzzy Sets Syst 161:2668–2683

Salema MIG, Barbosa-Povoa AP, Novais AQ (2007) An optimization model for the design of a capacitated multi-product reverse logistics network with uncertainty. Eur J Oper Res 179:1063–1077

Schultmann F, Engels B, Rentz O (2003) Closed-loop supply chains for spent batteries. Interfaces 33:57–71

Souza GC (2013) Closed-loop supply chains: a critical review, and future research. Decis Sci 44:7–38

Statistische Ämter des Bundes und der Länder: Gemeindeverzeichnis-Sonderveröffentlichung Gebietsstand: 31.12.2011 (Jahr): Großstädte (mit mindestens 100 000 Einwohnerinnen und Einwohnern) in Deutschland nach Bevölkerung am 31.12.2011 auf Grundlage des Zensus 2011 und früherer Zählungen. https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/Administrativ/GrosstaedteEinwohner.html. Version: 2013. Last checked on 15 Jan 2015

Taleb KN, Gupta SM, Brennan L (1997) Disassembly of complex product structures with parts and materials commonality. Prod Plann Control 8:255–269

Thierry MC, Salomon M, van Nunen JAEE, van Wassenhove LN (1995) Strategic issues in product recovery management. Calif Manag Rev 37:114–135 van der Laan E, Salomon M (1997) Production planning and inventory control with remanufacturing and disposal. Eur J Oper Res 102:264–278

Walther G, Steinborn J, Spengler TS (2009) Variantenvielfalt und Lebenszyklusbetrachtungen im Remanufacturing. In: Heyde F, Löhne A, Tammer C (eds) Methods of multicriteria decision theory and applications, pp 179–204

Integrated Make-or-Buy and Facility Layout Problem with Multiple Products

Bernd Hillebrand and Grigory Pishchulov

Abstract Layout decisions represent an important step in production planning as they significantly affect the efficiency of a facility. They typically follow after make-or-buy decisions and capacity planning of a firm, and belong to the lowest level in the hierarchy of tactical production decisions. Sequential decision making neglects however the potential interaction between consecutive planning levels and may lead to inefficiencies. We propose an approach to a simultaneous make-or-buy decision making and facility layout planning in the presence of multiple products. To this end, we introduce a generalization of the classical quadratic assignment problem that takes into account (i) the costs of outsourcing, and (ii) the impact of layout decisions on the production costs of individual products. We further propose a solution approach based on the linearization of the resulting model and derive basic properties of the optimal planning decisions.

1 Introduction

The tactical production planning can be divided into three different levels: The first level involves decisions about the production range, specifically about which products from the product line of the firm should be made in-house, and which should be sourced from external suppliers. The second level determines the production process and the necessary resources. The third level aims to locate resources within the firm's premises while seeking to minimize unproductive movement of material and employees in the production process—with the general

B. Hillebrand (✉) · G. Pishchulov
Supply Chain Management, Technische Universität Dortmund,
Martin-Schmeißer-Weg 12, 44227 Dortmund, Germany
e-mail: bernd.hillebrand@tu-dortmund.de

B. Hillebrand
Lehrstuhl für Produktion und Logistik, Technische Universität Dortmund,
Martin-Schmeißer-Weg 12, 44227 Dortmund, Germany

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_9

objective of cutting throughput times and saving costs (cf. Tompkins et al. 2010). A production site of the firm is typically referred to as facility (Reid and Sanders 2007). In general, the decision about the assignment of different objects to locations within a facility is called a facility layout problem (FLP) (cf. Drira et al. 2007).

The process of tactical production planning is organized hierarchically and therefore may neglect certain interactions between the different planning steps. For example, including an additional product into the production range typically increases the material flow within the production process. The additional material flow costs must not necessarily be attributed to the new product exclusively—as they may be determined to a certain extent by the existing facility layout, which has been tailored to accommodate the existing material flow. Attempting to minimize the material flow costs associated with each individual product is however likely to lead to conflicting choices with regard to the facility layout. Therefore, re-designing the existing layout to accommodate the entire production range would require seeking a trade-off between the material flow costs related with all of the products. The material flow costs of the resulting optimal layout that go beyond the total of the material flow costs which would have been generated by the production of each individual product in isolation from the others on the respective optimal layout, can be defined as complexity costs of facility layout (Gössinger 2008).

In the present work we propose an approach to a simultaneous make-or-buy decision making and facility layout planning in the presence of multiple products. It simultaneously takes into account (i) the costs of product outsourcing, and (ii) the impact of facility layout on the production costs of individual products, leading to a managerial decision which is jointly optimal from both perspectives: choosing a production range and a facility layout. To focus on the main effects of such integrated decision making, we assume that the make-or-buy decision has no influence on the process and resource planning, so that the second decision level of the tactical production planning shall not be part of our analysis. However, in Sect. 6 we explain how this decision level can be integrated into the proposed decision-making approach. The rest of the paper is structured as follows. In Sect. 2 we briefly review the relevant literature. Section 3 describes the problem under study and the modeling approach. In Sect. 4 we study basic properties of the optimal planning decisions and illustrate them on a numerical example. Section 5 presents a solution approach based on the linearization of the underlying optimization problem. In Sect. 6 we conclude with a summary and an outlook.

2 Literature Review

Facility layout planning comprises a broad range of planning tasks (cf. Tompkins et al. 2010) of which we restrict our attention in the present work to those of them that are represented by the FLP. The Quadratic Assignment Problem (QAP) of Koopmans and Beckmann (1957) can be considered as one of the earliest models which suits several common variants of facility layout problems. The QAP is an integer quadratic optimization problem, which models the task of optimally assigning a number of interconnected objects to a number of given locations by minimizing the distance-weighted connection intensity (manifested e.g. in the resulting material flow) between the objects.

Subsequent research introduced several additional aspects to this problem setting, such as specific layout configurations and special handling devices (e.g. Afentakis 1989; Kouvelis and Chiang 1992; Braglia 1996), and demonstrated a range of applications of the QAP in diverse facility layout planning contexts, e.g. in universities (Dickey and Hopkins 1972), hospitals (Elshafei 1977) and forest zoning (Bos 1993). The QAP is considered to be one of the most popular models of the FLP (Loiola et al. 2007).

The QAP is often solved by its reduction to an (mixed-)integer linear program ((M)ILP). The most often cited reason for this is to obtain better interim solutions within problem-specific solution procedures to increase their efficiency or to form the basis for new solution approaches. The linearization of Lawler (1963) is one of the first approaches to obtaining lower bounds for this purpose. Kaufman and Broeckx (1978) provide another linearization for obtaining lower bounds when applying Benders' decomposition (Benders 1962). This linearization was improved by Xia and Yuan (2006) as well as Zhang et al. (2013). Also Bazaraa and Sherali (1980) use their linearization with Benders' decomposition. Christofides et al. (1980) provide a linearization to reduce complexity in the solution process of large-scale instances by means of a decomposition approach. Frieze and Yadegar (1983) apply a Lagrange relaxation to their linearization in order to tighten the bounds. Adams and Johnson (1994) provide another linearization to further improve the bounds. Helber et al. (2014) apply a fix-and-optimize heuristic (Helber and Sahling 2010) to their linearization of the QAP.

In comparison to the different variants of FLPs, where the intensity of the material flow between the objects is pre-defined, only a few contributions mention variable material flow between the objects. A multi-period setting was investigated for the first time by Rosenblatt (1986). Such problem settings are called dynamic facility layout problems, as the material flow is specified for each individual period. A fuzzy representation of a variable material flow (e.g. Grobelny 1987; Dweiri and Meier 1996; Enea et al. 2005) and a scenario-based robustness analysis (Rosenblatt and Lee 1987) are examples of including a restricted knowledge about the material flow intensity into the facility layout planning. Still, the decision maker cannot influence the actual realization of the flow intensity in these problem settings.

Note that the decision maker can however influence the flow intensity by varying the production range—e.g., by outsourcing the production of certain products (see e.g. Gunasekaran et al. 2015, for a comprehensive discussion of outsourcing decision making). In order to evaluate such opportunity, the material flow needs to be considered at the product level. We are aware of an only study that adopts this perspective in the context of facility layout planning—that by Bazaraa and Sherali (1980), based on the earlier work by Graves and Whinston (1970). However in their problem setting, the production range is not subject to change. To our best knowledge, the present work is the first one to address make-or-buy decisions and facility layout planning within an integrated approach.

For a more profound insight to the different problems setting in FLPs please see Singh and Sharma (2006) as well as Drira et al. (2007).

3 The Integrated Make-or-Buy and Facility Layout Problem

3.1 Problem Description and Modeling Assumptions

Consider a company which has to satisfy the given demand for a set P of products within a specific period. The length of this period corresponds usually to the range of tactical production planning which depends on the industry and spans typically a number of years. The entire demand for each single product can either be produced in-house or purchased from an outside supplier. As in most of FLPs studied in the literature, we assume that the demand is deterministically known (cf. Drira et al. 2007). The respective make-or-buy decision with regard to product $p \in P$ is represented by the variable $q_p \in \{0, 1\}$. To make the products in-house, it is necessary to arrange a set I of objects (e.g. equipment units or machines) on the set K of locations within the facility. For brevity, we will refer to the objects as machines. Every machine can be placed on any location; one location can accommodate at most one machine. For this reason, $|I| \leq |K|$ is required. The decision about the assignment of machine i to location k is represented by the variable $u_{ik} \in \{0, 1\}$. Determining their values for all $i,\ k$ defines therefore the layout of the facility. We assume that set I of machines to be located is not affected by the make-or-buy decision. As an example, consider a joinery that ships furniture products to a retail chain.

If $q_p = 1$, the direct material flow from machine i to machine j causes costs λ_{pij} per distance unit, where $\lambda_{pij} \geq 0$ for every $i, j \in I$ is determined by the production process of product p. The distance between the locations k and l is represented by d_{kl}. If a specific product p is not made in-house $(q_p = 0)$, the entire demand for this product has to be met via outsourcing at the cost c_p. In general, c_p should represent the total cost of procuring the entire demand volume of product p from an external supplier, less any savings realized due to not producing product p in-house—so that c_p can also be negative. To simplify the exposition, we further assume that the production costs of any product, except for that part of theirs which results from the material flow within the facility layout, remain unaffected by the make-or-buy decision.

It should be noted that the total procurement costs are determined in the industry practice by a broad range of factors, both financial and non-financial, e.g. the purchase price, supplier service level, monitoring costs, and alliances' risk perspectives, to name a few (see Gunasekaran et al. 2015). For the purposes of the present study we assume that the outsourcing cost c_p, $p \in P$, accounts in our model for all relevant procurement cost factors. Alternatively, it can be interpreted as a penalty cost for not producing the respective product at all.

The objective is to minimize the total costs of sourcing by choosing between production and outsourcing for each product p, as represented by variables $q_p,\ p \in P$, and a choosing a facility layout represented by variables $u_{ik}, i \in I, k \in K$. This defines the integrated make-or-buy and facility layout problem.

3.2 Model Statement

The following optimization problem seeks to determine the optimal solution of the integrated make-or-buy and facility layout problem:

$$TC^* := \min \sum_{p \in P} \sum_{i \in I} \sum_{j \in I} \sum_{k \in K} \sum_{l \in K} q_p \cdot \lambda_{pij} \cdot d_{kl} \cdot u_{ik} \cdot u_{jl} + \sum_{p \in P} (1 - q_p) \cdot c_p \tag{1}$$

$$\sum_{k \in K} u_{ik} = 1 \quad \forall i \in I \tag{2}$$

$$\sum_{i \in I} u_{ik} \leq 1 \quad \forall k \in K \tag{3}$$

$$u_{ik} \in \{0, 1\} \quad \forall i \in I; k \in K \tag{4}$$

$$q_p \in \{0, 1\} \quad \forall p \in P \tag{5}$$

Problem (1)–(5) is an extension of the QAP. The first part of (1) represents the total material flow costs of in-house production. Specifically, if machine i is assigned to location k and machine j is assigned to location l, then $u_{ik} = u_{jl} = 1$. Should product p be made in-house $(q_p = 1)$, the cost of material flow between machines i and j associated with that product must be taken with the factor d_{kl}, which accordingly represents the distance between the machines. The second part of (1) expresses the total costs of outsourcing the products, which are not produced in-house. The total of material flow costs and outsourcing costs has to be minimized. Constraints (2) and (3) express that each machine is assigned to exactly one location and that at most one machine is assigned to each location. Constraints (4) and (5) define the feasible range of decision variables. Note that (1)–(5) is a non-convex integer optimization problem with a cubic objective function, $|I| \cdot |K| + |P|$ binary variables and $|I| + |K|$ linear constraints.[1]

[1]Note that integrality of make-or-buy decision variables expressed by constraint (5) can be relaxed, meaning essentially that any product demand can be shared between in-house production and outsourcing. This would render the problem under study to an integrated make-and-buy and facility layout problem. It can however be shown that such relaxation does not improve the optimal objective value in (1). We omit the detailed proof here for reasons of space.

In the next section we derive basic properties of optimal solutions of problem (1)–(5) and illustrate them by a numerical example.

4 Properties of Optimal Make-or-Buy and Facility Layout Decisions

4.1 Superadditivity of Minimum Flow Costs

In this section we establish a basic property of material flow costs induced by a subset of products in their respectively optimal facility layout: This property states that merging two subsets of products for in-house production cannot reduce the material flow costs—compared to the total of these costs which would have arisen if each subset of products would have been produced in isolation on the respectively optimal layout. This property can be described as superadditivity of the material flow costs in the respectively optimal layout as a function of the set of products made in-house. Let $S \subseteq P$ be a non-empty set of products and Λ_S the optimization problem (1)–(4) with $q_p = 1$ for every $p \in S$ and $q_p = 0$ as well as $c_p = 0$ for every $p \in P \backslash S$. Consequently, the value of q_p for $p \in P$ is not variable anymore and represents a parameter of problem Λ_S. Further, let Z_S^* be the optimal objective function value of Λ_S and $L_S^* = \{u_{ik}^* | i \in I, k \in K\}$ its optimal solution. Thus Z_S^* displays the minimal material flow costs resulting from in-house production of product set S. There holds the following

Proposition 1 $Z_S^* + Z_{P\backslash S}^* \leq Z_P^* \quad \forall S \subset P, S \neq \emptyset.$

Proof It is easy to see that the objective function of Λ_P is the sum of the objective functions of problems Λ_S and $\Lambda_{P\backslash S}$. Note further that all three problems Λ_P, Λ_S and $\Lambda_{P\backslash S}$ have an identical feasible region. Furthermore, L_P^* cannot deliver an objective value to problem Λ_S lower than Z_S^* as well as it cannot deliver an objective value to $\Lambda_{P\backslash S}$ lower than $Z_{P\backslash S}^*$. This leads to the assertion of the proposition. □

This property implies that, compared to the exclusive in-house production of the entire set P, it can be more economically to restrict the set of products made in-house to S in order to achieve savings in material flow costs of the magnitude $K = Z_P^* - \left(Z_S^* + Z_{P\backslash S}^*\right) + Z_{P\backslash S}^*$. However this will lead to outsourcing costs of $\sum_{p \in P\backslash S} c_p$. In the case of $|P| = 2$ the amount $Z_P^* - \left(Z_S^* + Z_{P\backslash S}^*\right)$ represents the complexity costs as defined in Sect. 1.

The above exposition makes clear that optimal make-or-buy decisions depend on the trade-off between complexity and material movement costs on the one hand and outsourcing costs on the other. The following section analyses accordingly the structure of optimal make-or-buy decisions.

4.2 Structure of Optimal Make-or-Buy Decisions

Below we investigate the dependence of optimal make-or-buy decisions on the costs of outsourcing. We will describe this dependence as optimal make-or-buy decision policy. To analyze its structural properties we will focus on the case $P = \{1, 2\}$ and vary the outsourcing costs c_1 and c_2 on the interval $[0; \infty)$ respectively. Let further the notation of Sect. 4.1 hold. For the sake of brevity, let $Z^*_p \equiv Z^*_{\{p\}}$ for $p \in P$. There holds the following

Proposition 2 *Depending on the outsourcing costs c_1 and c_2, it is optimal to decide between in-house production and outsourcing as follows:*

(a) $c_1 \le Z^*_1 \wedge c_2 \le Z^*_2 \Rightarrow (q^*_1, q^*_2) = (0, 0)$;
(b) $c_1 \ge Z^*_1 \wedge c_2 \le \min\{c_1 + Z^*_2 - Z^*_1; Z^*_P - Z^*_1\} \Rightarrow (q^*_1, q^*_2) = (1, 0)$;
(c) $c_1 \le \min\{c_2 + Z^*_1 - Z^*_2; Z^*_P - Z^*_2\} \wedge c_2 \ge Z^*_2 \Rightarrow (q^*_1, q^*_2) = (0, 1)$;
(d) $c_1 \ge Z^*_P - Z^*_2 \wedge c_2 \ge Z^*_P - Z^*_1 \Rightarrow (q^*_1, q^*_2) = (1, 1)$.

Proof If the condition of case (a) holds then outsourcing of each product is cheaper than its exclusive in-house manufacturing on the respectively optimal layout. Note that there are four possible make-or-buy decisions: $\{(q_1, q_2) | q_1, q_2 \in \{0, 1\}\}$. Solving the problem (1)–(5) while fixing q_1, q_2 to particular values will result in four problem instances. A direct comparison of their objective values implies that $(q_1, q_2) = (0, 0)$ results in the lowest objective value. This proves case (a). The proof of the subsequent cases can be carried out by a similar approach. □

The proof of Proposition 2 yields the following result:

Corollary 1 *The optimal objective value of problem* (1)–(5) *in cases* (*a*)–(*d*) *of Proposition 2 are respectively* $c_1 + c_2$, $Z^*_1 + c_2$, $c_1 + Z^*_2$ *and* Z^*_P.

Optimal make-or-buy decisions arising in cases (a)–(d) for the given combinations of outsourcing costs c_1 and c_2 comprise four respective areas on the coordinate plane Oc_1c_2 which are represented by areas A–D in the illustration provided on the right of Fig. 1.

As the figure illustrates, case (a) applies when neither the exclusive in-house production of product 1, nor that of product 2 is cheaper than outsourcing the respective product. Therefore both products will be purchased from the external supplier, which results in costs $c_1 + c_2$. If $c_1 \ge Z^*_1$, case (b) applies. In this case, outsourcing is more expensive than exclusive in-house production of product 1. Therefore product 1 will be produced and product 2 will be purchased if $c_2 + Z^*_1 \le c_1 + Z^*_2$—i.e., it is not cheaper to do the opposite and, further, $c_2 \le Z^*_P - Z^*_1$—i.e., the outsourcing costs of product 2 are not higher than the additional production costs; the latter include the complexity costs if they emerge. Case (c) can be interpreted in a similar way. If $c_1 \ge Z^*_P - Z^*_2$ and $c_2 \ge Z^*_P - Z^*_1$ then case (d) applies. In this case, production of both products on a shared optimal layout

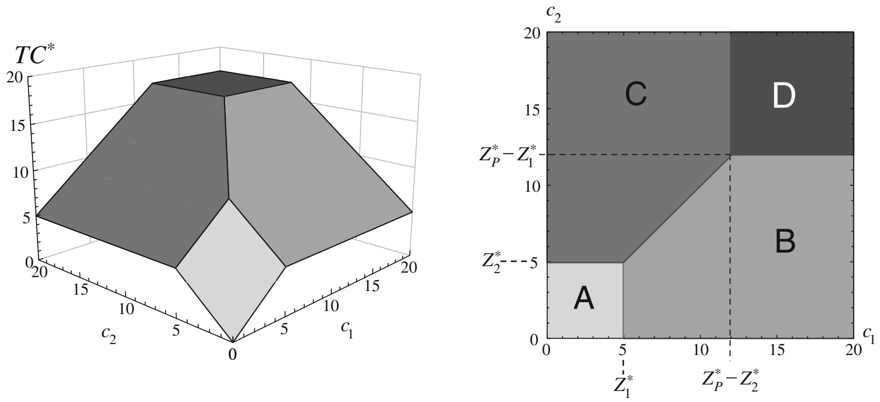


Fig. 1 Optimal make-or-buy decisions and resulting objective values as a function of c_1, c_2

is cheaper than outsourcing any of the products—despite of the existence of possible complexity costs. So both products will be produced at the cost Z_P^*.

4.3 Numerical Example

In the following example we consider a facility with six locations arranged in a row. Between two adjacent locations, the distance amounts to one distance unit. If there are n locations between any two given locations, then the distance between the given two amounts to $n + 1$ distance units. Six machines have to be located. The product set is given by $P = \{1,\ 2\}$. Product 1 passes the machines in the following order: 1–2–3–4–5–6, for the second product this order is respectively 5–2–3–4–1–6. The direct transfer of a product from one machine to another results in material flow costs of one currency unit per distance unit.

Firstly we compute the objective values Z_1^*, Z_2^* and Z_P^* using CPLEX v. 12.5; these have been found to be 5, 5 and 17 respectively. We then let the outsourcing costs c_1 and c_2 vary in the range from 0 to 20 and determine the optimal make-or-buy decisions by means of Proposition 2. The right-hand part of Fig. 1 illustrates cases (a)–(d) from Proposition 2. Its left-hand part shows the respective optimal objective values of problem (1)–(5) as per Corollary 1. The four colored faces of the graph correspond to one of cases (a)–(d). It can be seen that the projection of the graph on the horizontal coordinate plane yields the right-hand part of Fig. 1.

If a layout exists, which is optimal for both products in exclusive in-house production, then the boundary between areas B and C collapses to a point; there are no complexity costs in this case.

Note that other FLPs investigated in the literature are not equipped with the possibility to outsource products. But as shown on the left-hand part of Fig. 1, it might be an economically attractive option because of the possible presence of complexity costs.

It may be impractical or impossible to derive optimal make-or-buy and facility layout decisions for larger instances of problem (1)–(5) using a general-purpose solver as it represents an integer non-convex optimization problem. An established solution approach which guarantees an exact solution or at least a guaranteed quality of approximate solutions consists in the reduction of the problem at hand to a mixed integer linear program and solving it with respective state-of-the-art methods. This approach will be pursued in the following section.

5 Solution Approach via Linearization

On the first sight, it appears logical to use one of the existing linearizations of the QAP to solve problem (1)–(5) by transforming it to a MILP. A closer examination however reveals that the available linearizations cannot serve this purpose, or do so only to a limited extent.

With regard to the additional variables and constraints, the linearization of Kaufman and Broeckx (1978) is known as the most compact, thus as the most efficient formulation. However adopting their linearization technique for the purpose of linearizing problem (1)–(5) would lead to quadratic constraints in the resulting optimization problem. Their linearization is further only suitable for QAPs with a symmetric distance matrix, what cannot be assumed for facilities with special handling devices, e.g. a directed loop layout (Drira et al. 2007). The other known linearization approaches will lead to a quadratic objective function with a relatively high number of additional variables and constraints.

For the above reasons, we propose a new linearization approach which reduces (1)–(5) from a problem with a cubic objective function to the following mixed-integer linear program:

$$TC^* := \min \sum_{k \in K} \sum_{l \in K} d_{kl} \cdot v_{kl} + \sum_{p \in P} (1 - q_p) \cdot c_p \tag{6}$$

$$v_{kl} \geq \sum_{p \in P} q_p \cdot \lambda_{pij} - M_{ij} \cdot (2 - u_{ik} - u_{jl}) \quad \forall i, j \in I; k, l \in K \tag{7}$$

$$v_{kl} \geq 0 \quad \forall k, l \in K \tag{8}$$

and (2), (3), (4) and (5).

The new decision variables v_{kl} represent in the optimal solution the resulting material flow between the locations k, l in currency units per distance unit. To determine the total material flow costs in the facility, they become multiplied in objective function (6) with the distance d_{kl} between the respective locations.

If machines i,j are placed on locations k,l respectively, then the parenthetical expression in (7) turns to zero. Accordingly, constraint (7) implies $v_{kl} \geq \sum_{p \in P} q_p \cdot \lambda_{pij} \cdot u_{ik} \cdot u_{jl}$ for these i,j. Therefore, v_{kl} has to be at least as large as the resulting material flow between the machines in question. For all other combinations of i,j, a sufficiently large number M_{ij} just assures that v_{kl} is bounded from below by a non-positive value. For a tight linearization, we suggest to choose $M_{ij} = \sum_{p \in P} \lambda_{pij}$. Further, constraint (8) assures the non-negativity of v_{kl} in every case. Because the objective function (6) is to be minimized, v_{kl} will take on the smallest possible value for every k,l, thus representing the material flow between k and l. As a result, problem (2)–(8) involves $|I|^2 \cdot |K|^2$ additional constraints and $|K|^2$ continuous variables in comparison to problem (1)–(5).

Note that the above linearization approach is not restricted to any specific product range P. Note further that it can also be applied to the classical QAP by setting $q_p = 1$ for each $p \in P$.

6 Conclusion and Outlook

The present study proposes an approach to an integrated make-or-buy and facility layout planning which takes into account interdependencies between these planning levels. We formulate a non-linear integer programming problem which underlies the suggested planning approach. Basic properties of its optimal solutions are investigated and illustrated on a numerical example. Furthermore, we propose a linearization of the said problem to facilitate its solution with conventional tools. In addition to the above, the present study provides an operationalization of the notion of complexity costs in the context of facility layout decisions.

It should be noted that the second level of tactical production planning—namely that of production process and resource planning—can be integrated into the proposed planning approach by taking into account the decision with regard to the acquisition of machines comprising set I. This is achieved by incorporating the following constraint into problem (1)–(5) as well as (2)–(8):

$$\sum_{p \in P} \left[q_p \cdot \sum_{j \in J} \left(\lambda_{pij} + \lambda_{pji} \right) \right] \leq \sum_{j \in J} \left(M_{ij} + M_{ji} \right) \cdot \sum_{k \in K} u_{ik} \quad \forall i \in I \tag{9}$$

Furthermore, the expression of the objective function in (1) as well as (6) has to involve the additional term $\sum_{i \in I} \sum_{k \in K} a_{ik} \cdot u_{ik}$, where a_{ik} is the cost of ownership of machine i when it is employed at location k, and constraint (2) has to be changed to an inequality with the less-or-equal sign.

Future research should attend this model extension, as well as focus on deriving solution properties for a larger number of products and gaining insights into the magnitude of complexity costs by conducting a numerical study in a variety of settings. Furthermore, it would be of interest to address the following problem settings in the future research: Division of labor between multiple facilities of the company, planning over product life-cycles, as well as problem settings with supply- or demand-side uncertainties. Last but not least, it is essential to develop efficient heuristic solution methods for the problem under study.

Acknowledgments We are thankful to Rolf Wendt of TU Dortmund University for the useful insights with regard to relaxation of the problem referred to in footnote 1.

References

Adams WP, Johnson TA (1994) Improved linear programming-based lower bounds for the quadratic assignment problem. In: Pardalos PM, Wolkowicz H (eds) *Quadratic* assignment and related problems, vol 16, DIMACS series in discrete mathematics and theoretical computer science, pp 43–75

Afentakis PA (1989) Loop layout design problem for flexible manufacturing systems. Int J Flex Manuf Syst 1:143–175

Bazaraa MS, Sherali HD (1980) Benders' partitioning scheme applied to a new formulation of the quadratic assignment problem. Naval Res Logistics Q 27:29–41

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numer Math 4:238–252

Bos J (1993) Zoning in forest management: a quadratic assignment problem solved by simulated annealing. J Environ Manag 37:127–145

Braglia M (1996) Optimization of a simulated-annealing-based heuristic for single row machine layout problem by genetic algorithm. Int Trans Oper Res 3:37–49

Christofides N, Mingozzi A, Toth P (1980) Contributions to the quadratic assignment problem. Eur J Oper Res 4:243–247

Dickey JW, Hopkins JW (1972) Campus building arrangement using topaz. Transp Res 6:59–68

Drira A, Pierreval H, Hajiri-Gabouj S (2007) Facility layout problems: a survey. Ann Rev Control 31:255–267

Dweiri F, Meier FA (1996) Application of fuzzy decision-making in facilities layout planning. Int J Prod Res 34:3207–3225

Elshafei AN (1977) Hospital layout as a quadratic assignment problem. Oper Res Q 28:167–179

Enea M, Galante G, Panascia E (2005) The facility layout problem approached using a fuzzy model and a genetic search. J Intell Manuf 16:303–316

Frieze AM, Yadegar J (1983) On the quadratic assignment problem. Discrete Applied Mathematics 5:89–98

Gössinger R (2008) Produktion und Logistik. In: Corsten H, Reiß M (eds) Betriebswirtscha-ftslehre, 4th edn, vol 1. Chapter E, pp 443–539

Gunasekaran A, Irani Z, Choy K-L, Filippi L, Papadopoulos T (2015) Performance measures and metrics in outsourcing decisions: a review for research and applications. Int J Prod Econ 161:153–166

Graves GW, Whinston AB (1970) An algorithm for the quadratic assignment problem. Manag Sci 17:453–471

Grobelny J (1987) The fuzzy approach to facility layout problems. Fuzzy Sets Syst 23:175–190

Helber S, Sahling F (2010) A fix-and-optimize approach for the multi-level capacitated lot sizing problem. Int J Prod Econ 123:247–256

Helber S, Böhme D, Oucherif F, Lagershausen S, Kasper S (2014) A hierarchical facility layout planning approach for large and complex hospitals. Diskussionspapiere Hannover Econ Pap 527:1–27

Kaufman L, Broeckx F (1978) An algorithm for the quadratic assignment problem using Benders' decomposition. Eur J Oper Res 2:204–211

Koopmans TC, Beckmann M (1957) Assignment problems and the location of economic activities. Econometrica 25:53–76

Kouvelis P, Chiang WC (1992) A simulated annealing procedure for the single row layout problems in flexible manufacturing systems. Int J Prod Res 30:717–732

Lawler EL (1963) The quadratic assignment problem. Manag Sci 9:586–599

Loiola EM, de Abreu NMM, Boaventura-Netto PO, Hahn P, Querido T (2007) A survey for the quadratic assignment problem. Eur J Oper Res 176:657–690

Reid RD, Sanders NR (2007) Operations management, 3rd edn. Wiley, Hoboken

Rosenblatt MJ (1986) The dynamics of plant layout. Manag Sci 32:76–86

Rosenblatt MJ, Lee HL (1987) A robustness approach to facilities design. Int J Prod Res 25:479–486

Singh SP, Sharma RRK (2006) A review of different approaches to the facility layout problems. Int J Adv Manuf Technol 30:425–433

Tompkins JA, White JA, Bozer YA, Tanchoco MA (2010) Facilities planning, 4th edn. Wiley, Hoboken

Xia Y, Yuan Y-X (2006) A new linearization method for quadratic assignment problems. Optim Methods Softw 21:803–816

Zhang H, Beltran-Royo C, Ma L (2013) Solving the quadratic assignment problem by means of general purpose mixed integer linear programming solvers. Ann Oper Res 207:261–278

Qualification and Competence Related Misfits in Logistics Companies: Identification and Measurement

Sebastian Wünsche

Abstract Qualification and competence requirements in logistics companies are influenced by internal and external processes of change. To stay competitive and reactive to the market, those companies have to face the problem of qualification and competence related misfits. This paper presents a mixed method approach to identify and measure qualification and competence related misfits in different occupational groups and closes a gap in the literature. The use of the tool could provide valuable information to the HR development function of a logistics company.

1 Introduction

Depending on the general development of the economy, logistics can be seen as a growth industry with a huge demand for employees, in particular for logistics/supply chain managers and qualified personnel (see e.g. BVL 2008a, b). Despite these good predictions, logistics companies face qualification and competence related problems that result from influences of internal and external processes of change. For example due to social and economical megatrends, as well as changes in business activities, job requirements rise quickly. This development is propelled by the characteristics of the sector, which can be described as highly dynamic, networked and service orientated (Jahns and Darkow 2008; BVL 2008a; Klumpp 2009). Existing qualification and competence profiles are not able to keep up with, which ends up in a qualitative skill shortage (PwC 2012; Carter and Carter 2007). This shortage can be observed in many industries but due to high dynamics in logistics processes there is a distinct tendency in this special sector (Oxford Research 2010). In this kind of situation, a logistics company is reliant to attract, develop and keep qualifications and competencies (Carter and Carter 2007).

S. Wünsche (✉)
University Bremen, Wilhelm-Herbst-Str. 12, 28359 Bremen, Germany
e-mail: swuensch@uni-bremen.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_10

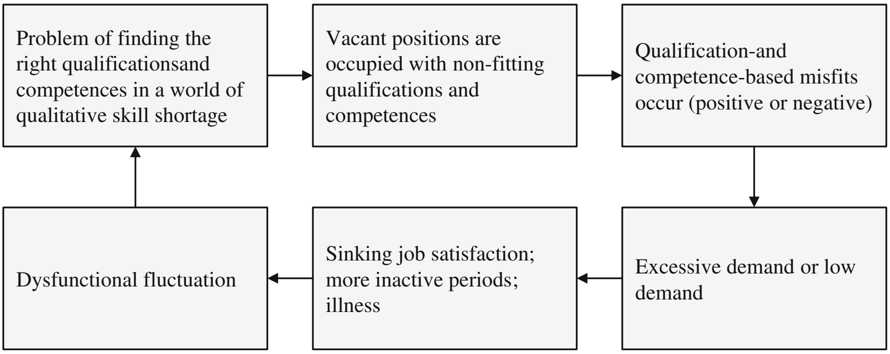


Fig. 1 Dysfunctional fluctuation as a result of qualification and competence based misfits

Instead, companies often suffer from undesired retirement of their employees (Felfe 2008). Dysfunctional fluctuation is the last element in a vicious circle beginning with a qualitative skill shortage, which leads to qualification and competence based misfits in a company, as shown in Fig. 1. A major problem for logistics companies is the fact that qualification and competence based misfits constitute extra costs. Based on a negative misfit, an employee can be underchallenged, dissatisfied and/or show inactive periods at work (Brauweiler 2010). In this case opportunity cost for the company appear, as the potential of the employee is not fully practiced. If dysfunctional fluctuation occurs, a company faces the extra costs of finding, recruiting and instructing new personnel, as well as the cost of losing expertise.

Logistics companies are not always able to match their spectrum of business activities with the employees' qualifications and competences. In this "era of turbulence" they have to stay competitive and reactive to the market (Christopher and Holweg 2011). Just as little a single logistic company is able to influence megatrends as situations which can last decades (Froehlich 2010). How can this vicious circle be interrupted? A task of the HR-development function is to close qualification and competence related gaps (Holtbrügge 2010) to avoid follow up problems. But first, those gaps or so called misfits have to be visible. The on hand research starts at this point and asks:

How can qualification and competence related misfits be identified and measured in logistics companies?

2 Qualifications and Competences in the Literature

2.1 Literature Review

Investigating 109 scientific articles from 1998 to 2014, Hohenstein et al. (2014) identify seven different HRM/SCM research streams: Skills, knowledge and abilities; training and development; HRM impact on performance; education and teaching; hiring and recruiting; compensation and pay as well as global mindset. In the first stream, Ellinger and Ellinger (2014) point out skills for managing supply chains. Murphy and Poist (2007) undertake long-term studies to investigate skills for senior level logistics and supply chain managers. They differ between general business skills, managerial attributes and specific logistics skills. In contrast to other researchers, Wu et al. (2013) discuss skills like rank communication, financial analysis and customer relationship management. Wünsche et al. (2014) emphasize professional experience, intercultural experience, foreign language skills, computer skills and different vehicle driving licenses as relevant for logistics professionals. The body of literature contains different tools for measuring competences (e.g. Hathaway et al. 2000; Bambeck 2003; Balzer et al. 2004; Frey and Balzer 2007). However, there is no tool for identifying and measuring qualification and competence based misfits in logistics. Those misfits in logistics have not even been recognized in the SCM literature. When it comes to skills, authors often investigate single occupational groups, e.g. supply chain managers (Murphy and Poist 2007). A closer examination of different groups is very seldom (e.g. Wünsche et al. 2014).

The skill related research in SCM and logistics came to its peak in 2013. Not only for this reason Hohenstein et al. (2014) conclude that there is a high relevance in this issue. Regarding further research, there is a call for mixed method approaches (Golicic and Davis 2012).

2.2 *Defining the Components*

To be more precise when it comes to "skills", it can be differed between qualifications and competences.

Tippelt and Schmidt (2009) understand a qualification as a certificated and standardized skill. In order to gain a qualification, one has to pass an evaluated exam (Preissing 2001). This ensures that a person is able to execute relevant actions within a special profession on a given standard (Tippelt and Schmidt 2009). Qualifications are used to document knowledge and are very important for starting a career (Müller 2008).

Competence refers to an individual personality aspect, which means that they are dependent from action and can only be obtained by application (North 2007). Even though competences can be trained and taught, they remain individual and cannot be imitated (Müller 2008). Competences enable the effective performance of a job (see Bartram and Roe 2005). A typical differentiation of competences refers to professional, methodological, social and self- and personal competences (Fölsch 2010; Erpenbeck and Heyse 2007). Misleadingly, often competences are put on a level with capabilities. Ulrich and Smallwood (2004) make a distinction between both terms. While an organization has capabilities, an individual has competences.

If there is a gap between job requirements and the qualifications and competencies owned by an employee, a qualification and competence based misfit occurs (Kristof 1996; Gericke et al. 2009). They can be positive or negative and end up in over- or under-qualification as shown in Fig. 1 (Erdogan et al. 2011; Allen and van der Velden 2001; Rumberger 1981).

To strengthen international competitive ability and the comparability of qualifications and competences, different tools exist (Müller 2008). The European Qualifications Framework, as well as the German Qualifications Framework and the Classification of Professions are such systems (EC 2008; BfBF 2014; GFLMA 2011). An assignment of qualifications, competences, organizational hierarchy levels and occupational groups becomes possible (Kotzab and Wünsche 2013).

3 Developing a Strategy to Identify and Measure Qualification and Competence Based Misfits

3.1 Propositions and Approach

To develop a tool for identifying and measuring misfits, propositions for orientation and later assessment have to be set up first.

As shown in Fig. 1, it is not easy for a logistics company to find the right qualifications and competencies in a world of qualitative skill shortage (PwC 2012; Carter and Carter 2007). It can be assumed, that

P1: Not every employee in a logistics company is skilled adequately for his/her position.

Then a qualification and competence related misfit should occur. If it occurs it must be recognizable or visible, which leads to

P2: If P1 is true, qualification and competence related misfits can be shown and/or measured.

If P2 is true, there must be an instrument that is suitable for such an undertaking, which leads to

P3: Qualification and competence related misfits can be identified and/or measured by a suitable instrument.

When it comes to qualification and competence based variations, the HR development function talks about a deviation in target/actual (Oechsler 2000), which gives a first hint for developing a tool. Finance knows a tool for measuring target and actual business results, called target-performance comparison (Jung 2003). First, c-level management and controlling define financial targets in measurable figures. By means of a variation analysis the controlling identifies variations in target and actual figures. Afterwards it is possible to plan corrective actions (Preissler 2014).

With some adjustments it is possible to use this approach for identifying and measuring variations in target and actual qualifications and competences. Figure 2 shows the used approach in three steps.

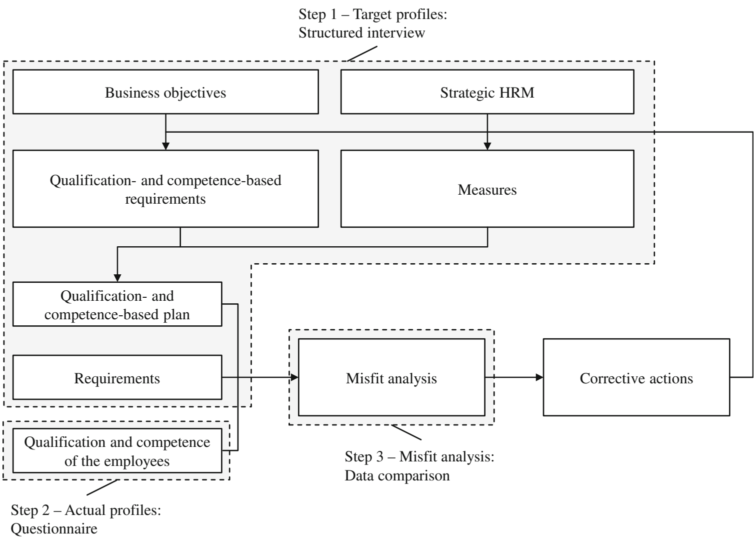


Fig. 2 Adjusted target-performance comparison (based on Jung 2003)

In a first step, qualification and competence based requirements have to be elaborated. The strategic function of HRM is responsible for implementing the company's business strategy into HRM practices (Ackermann et al. 1989; Evans 1986). For this reason strategic HR managers are in charge of assessing requirements for each hierarchy level and occupational group. This can be done by a structured interview (see Sect. 3.2).

In a second step the strategic HRM has to prepare qualification and competence related employee profiles for a comparison. The data collection can be done by an employee survey (see Sect. 3.2).

In a last step, strategic HRM can compare both data sets for identifying and measuring qualifications and competence related misfits (Sect. 3.2).

The field "corrective actions" is not part of this research. But the results from former measurement provide valuable data for the HR development function regarding the continuous planning of qualification and competence related development measures. Dynamical development within the company becomes possible.

3.2 Configuration of the Research Steps

As mentioned above, strategic HR managers are in charge of assessing requirements und can be seen as experts to ask. This assessment process can be carried out by a semi-structured and guided expert interview according to Flick (1999). A semi-structured and guided interview results in precise statements (Flick 1999), avoids disorientation (Mayer 2008) and allows room for discussing critical issues

(Atteslander 1984). The interviewer should be a neutral and trained person (Maccoby and Maccoby 1965).

For setting up requirements, first qualifications and competencies for assessment have to be chosen. The German Framework for University Degrees and the German Institute for International Pedagogic Research differ between 8 qualifications that will be used for the interview (DIPF 2012; BMBF 2005). Table 1 gives an overview.

Additionally, Table 1 shows the competences used for research. The catalogues by Frey and Balzer (2007), as well as Wünsche et al. (2014) have been chosen. Wünsche et al. (2014) extracted competences from logistic job advertisements of different occupational groups and provide logistics specific competences. Frey and Balzer (2007) discuss general competences for several industries, which also are typical for the logistics sector.

The German Federal Agency of Labour differs between four occupational groups with different sets of qualifications and competences needed (GFLMA 2011). Each group can be linked to different hierarchy levels and tasks. The system by Hildebrandt and Roth (2008) can help to establish a connection. The different groups used for research are shown in Fig. 3.

Subsequently all statements regarding qualifications and competences have to be interpreted and prepared for measurement. The question of the importance of a qualification can easily be answered with yes or no. This process is more difficult for competences as they are not simple certificates, but individual personality aspects. For this reason the competence measurement literature uses rating scales (e.g. Frey and Balzer 2007). This interpretation and rating process subsequent to the

Table 1 Set of qualifications and competences used for research

Qualification according to DIPF (2012), BMBF (2005)	Competences according to Frey and Balzer (2007)	Logistics competences according to Wünsche et al. (2014)
Promotion	Self-dependence	Professional experience
Diploma/master-degree	Cooperation competence	Simple computer literacy
Bachelor-degree	Social competence	Special computer literacy
Professional education	Conflict-dealing competence	Cultural experience
General qualification for university entrance	Communication competence	Relevant languages for German logistics companies
General qualification for university of applied science entrance	Leadership ability	Different driving licences
General certificate of secondary education	Situation-related behaviour	
Certificate of general education	Reflexivity	
	Analytical competence	
	Flexibility	
	Target-oriented acting	
	Work techniques	

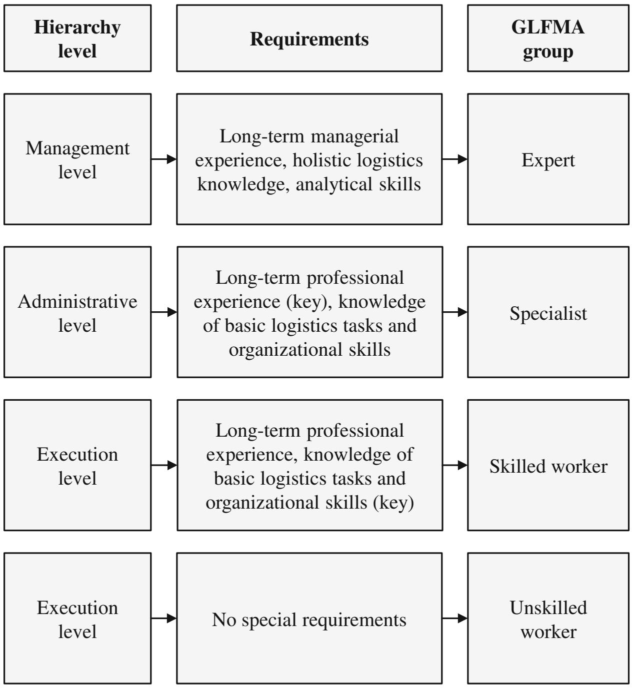


Fig. 3 Assignment of hierarchy levels, requirements and occupational categories [based on Hildebrandt and Roth (2008), 72 and GFLMA (2011), 27–28]

interview can be done by content analysis according to Mayring (2010). This kind of analysis allows each statement regarding a qualification or competence to be placed on a scale, if configured accordingly towards a classification (Mayring 2010). In step three, the SMk questionnaire by Frey and Balzer (2007) is used to collect data regarding the qualifications and competences of the employees. Frey and Balzer (2007) use a six-step Likert scale to rate assessments. For the purpose of comparability in step three, the rating of assessments in connection with the interview is bound to the same scale.

To avoid a static assessment, a tolerance interval for each scale value from −0.5 to 0.5 is used for the interview results.

For the investigation of qualifications and competences owned by the employees, the competence measurement tool SMk by Frey and Balzer (2007) can be used. Inherently it includes all general competences shown in Table 1. According to the authors the tool is easy to modify, thus competences for logisticians according to Wünsche et al. (2014) have been added. Furthermore the set of qualifications from Table 1 has been used for setting up the questionnaire. The dimensions of the instrument have been validated statistically and by experts (Frey and Balzer 2007). The authors appraise the scale’s stability as satisfying to good. Objectivity is given. The tool copes with the requirements of classical testing theory.

In a final step, both profiles have to be compared with each other to identify whether there are misfits or not. The Likert scale allows an estimation of the "size" of a misfit. For easier interpretation and a more precise look, the scale can be transformed to a percentage scale (Schön 2012).

4 Identifying and Measuring Qualification and Competence Related Misfits

4.1 Testing Environment

The instrument has been tested by carrying out a case study in cooperation with a German logistics provider that kindly agreed to be available for the study.

In a first step, two strategic HR managers from the company have been interviewed in a structured and guided way as described above. By analyzing the statements, requirement profiles for each occupational group (see Fig. 3) have been built. The interview took place on June 16th, 2013 in Bremen. According to Mayring (2002), a summary has been sent to the managers for approval.

In a second step, employees from all positions and hierarchy levels of the company have been asked for their qualification and for a self-assessment of their competences. Due to the situation of the company, the questionnaires have been handed out online, by hand and postal. In the end, 106 out of 332 possible persons participated. 105 questionnaires were suitable for use. From 105 participants, 38 could be classified as experts, 25 as specialists, 18 as skilled workers from administrative job, 21 as skilled workers from operative jobs and 3 as unskilled workers.

In a last step, both profiles have been compared with each other in order to identify and measure qualifications and competence related misfits.

4.2 Presentation of Key Results

Regarding the qualification no misfit could be identified in any of the groups. Each employee had the required qualification at the moment of data collection. The image changes on closer examination of competences according to Frey and Balzer (2007) and Wünsche et al. (2014). Misfits can be shown for every occupational group. Table 2 shows the most striking results with misfits of more than 10 % for each occupational group. It contains no values for unskilled workers due to the very low number of participants.

Table 2 Identified competence related misfits for each occupational group

Expert	Actual employee self-assessment on Likert scale	Target area on Likert scale (tolerance included)	Misfit expressed in per cent
Conflict-dealing competence	0.79	0.92–1	−13
Communication competence	0.81	0.92–1	−10
Leadership ability	0.71	0.92–1	−21
Analytical competence	0.78	0.92–1	−14
Special computer literacy	0.8	0.25–0.41	39
Specialist			
Analytical competence	0.81	0.92–1	−11
Special computer literacy	0.75	0.92–1	−17
Skilled worker (administrative)			
Self-dependence	0.86	0.58–0.75	11
Target-oriented acting	0.79	0.42–0.58	21
Special computer literacy	0.62	0.75–0.92	−13

Negative values show a tendency to under-qualification while positive values show a tendency to over-qualification

4.3 Discussion and Meaning for the Approach

Identified misfits depend on hierarchy level and typical tasks of the job.

Within the occupational group of experts, the most and strongest negative misfits can be identified. Here employees with authority to give directives can be found. However, the biggest discrepancies in this occupational group can be spotted at leadership- and interaction-related competences. Experts of the company seem to be a little over-qualified when it comes to special computer literacy.

Some misfits can be identified on the specialist level. Here the differences are more task-related. The values for analytical competence and special computer literacy are interesting. According to the interviewed strategic HR managers, these are typical competencies for the specialist-level.

On the skilled worker levels negative and positive misfits can be identified. Interestingly the employees from administration self-assessed better in self-dependence and target-oriented acting than required. According to the interviewed strategic HR managers, especially self-dependence does not play a crucial role at this hierarchy level.

Interestingly competence related misfits could be shown, while there were not any qualification related misfits. Although the employees passed exams to qualify in general for their job it can be questioned whether they work in the right position according to their abilities or if they are trained properly. Gaining a qualification does not seem to result in gaining the right competences.

All results have been discussed with the strategic HR managers subsequent to the misfit analysis. In some instances the experts were able to recognize an occupational group-dependent behavior, which supports the results from misfit analysis.

At this point, coming back to the propositions from Sect. 3.1 can help to review the used approach.

Proposition *P*1 can be confirmed for the case study. Not every employee was equipped with the desired competences. Therefore proposition *P*2 can be confirmed as well. It was possible to show misfits regarding the employee's qualifications and competences. By means of a Likert scale used in the SMk questionnaire by Frey and Balzer (2007) and a transformation to percentage values, it is possible to give an impression about the "size" of a misfit. Like assumed in *P*3, generally it can be said that the developed tool is suitable for such an operation. The discussion subsequent to the misfit analysis supports this. However the application contains a variety of restrictions.

First, the results come from a case study with a German logistics service provider. As this is a very limited testing environment, generalization is not possible. Every company features a different culture, different employees or different business activities.

Second, the used set of qualifications and competences is limited. Murphy and Poist (2007) as well as Gammelgaard and Larson (2001) provide bigger sets of logistic specific competences compared to Wünsche et al. (2014). The selection has been made against the background of timeliness. Furthermore the distinction of GFLMA (2011) is not tailored to logistics. The application of a logistics specific system could provide a comparison.

Third, it can be questioned whether a strategic HR manager is capable of judging about qualifications and competences needed in specialist departments away from their own department. This way of imposing profiles has been chosen from the perspective of HRM. Asking specialists in line function could generate more reliable data.

Furthermore the comparability of assessments from interviews and questionnaires can be questioned. The methodological choice is based on a call for mixed method approaches in HRM related SCM/logistics research (Golicic and Davis 2012).

Another restriction comes with the use of self-assessments. It can be questioned, whether the employees reflect their own competences correctly (Frey and Balzer 2007). A third party assessment could provide a comparison for more resilient results.

5 Conclusion and Outlook

Logistics companies find themselves in an "era of turbulence" (Christopher and Holweg 2011). To stay competitive and reactive to the market those companies have to face the problem of qualification and competence related misfits.

The purpose of this paper was to find a way to identify and measure qualification and competence related misfits in logistics companies. The results from Sect. 4 show that this kind of misfits indeed exists. Through using an adjusted approach from finance it is possible to identify and measure qualification and competence related misfits in three steps. Some restrictions have been discussed above. A closer look on those can lead the way for further research.

Comparative studies would be interesting for more evidence regarding applicability in logistics practice.

In practice, results from misfit measurement could provide valuable information to the HR-development function. Targeted training and development can eliminate or minimize misfits and prevent follow up problems as shown in Fig. 1 subsequently.

References

Ackermann K-F, Danert G, Horváth P (1989) Personalmanagement im Wandel. Schäffer-Poeschel, Stuttgart

Allen J, van der Velden R (2001) Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. Oxford Econ Pap 53(3):434–452

Atteslander P (1984) Methoden der empirischen Sozialforschung. de Gruyter, Berlin

Balzer L, Bodensohn R, Frey A (2004) Diagnose und Rückmeldung von Handlungskompetenzen von Studierenden im Blockpraktikum. J für LeherInnenbildung 4:30–36

Bambeck J (2003) Bambeck-Competence-Instrument (BCI). In: Erpenbeck J, Rosenstil L (eds) Handbuch Kompetenzmessung. Schäffer-Pöschl, Stuttgart, pp 19–30

Bartram D, Roe R (2005) Definition and assessment of competences in the context of the European Diploma in Psychology. Eur Psychol 10(2):93–102

BfBF (2014) Bundesministerium für Bildung und Forschung, Der DQR. Online access: www.dqr.de. Checked on 10.10.2014

BMBF (2005) Bundesministerium für Bildung und Forschung, Qualifikationsrahmen für Deutsche Hochschulabschlüsse

Brauweiler J (2010) Retention Management. Erfolgreiches Personalmanagement im demografischen Wandel. In: Preißig D (ed) Erfolgreiches Personalmanagement im demografischen Wandel. Munich, Oldenbourg, pp 77–105

Bundesvereinigung Logistik (2008) Studie Trends und Strategien in der Logistik 2008: Die Kernaussagen

Bundesvereinigung Logistik (2008) Trends und Strategien in der Logistik - Globale Netzwerke im Wandel, Bremen

Carter PL, Carter JR (2007) Organization+talent. Supply Chain Manage Rev 11(8):37–43

Christopher M, Holweg M (2011) “Supply Chain 2.0”: managing supply chains in the era of turbulence. Int J Phys Distrib Logistics Manage 41(1):63–82

Deutsches Institut für Internationale Pädagogische Forschung (2012) Bildung in Deutschland 2012: Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf, Bildung in Deutschland, vol 2012. Bertelsmann, Frankfurt am Main, Bielefeld

European Commission (2008) Der Europäische Qualifikationsrahmen für lebenslanges Lernen, Luxembourg

Ellinger A, Ellinger A (2014) Leveraging human resource development expertise to improve supply chain managers’ skills and competencies. Eur J Training Dev 38(1/2):118–135

Erdogan B, Bauer TN, Peiró JM, Truxillo DM (2011) Overqualified employees: making the best of a potentially bad situation for individuals and organizations. Ind Organ Psychol 4(2):215–232
Erpenbeck J, Heyse V (2007) Die Kompetenzbiographie: Wege der Kompetenzentwicklung. Waxmann Verlag GmbH
Evans P (1986) The strategic outcomes of human resource management. Hum Resour Manage 25(1):149–167
Felfe (2008) Mitarbeiterbindung. Hogrefe, Göttingen
Flick U (1999) Qualitative Forschung: Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften, Rowohlts Enzyklopädie, vol 546. Rowohlt Taschenbuch-Verlag, Reinbek bei Hamburg
Fölsch T (2010) Kompetenzentwicklung und Demografie, Schriftenreihe Personal- und Organisationsentwicklung, vol 9. Kassel University Press, Kassel
Frey A, Balzer L (2007) Beurteilung zu sozialen und methodischen Kompetenzen. Die SMK-Reihe. Available at: www.kompetenzscreening.de/Beurteilungsbogen_smk_FreyBalzer_Beschreibung2007.pdf. Checked on 11.05.2015
Froehlich R (2010) Investment-Megatrends. profitieren Sie von den großen Veränderungen unserer Zeit. Börsenmedien-Verlag, Kulmbach
Gammelgaard B, Larson P (2001) Logistic skills and competencies for supply chain management. J Bus Logistics 22(2):27–50
Gericke N, Krupp T, Troltsch K (2009) Unbesetzte Ausbildungsplätze–warum Betriebe erfolglos bleiben. Ergebnisse des BIBB-Ausbildungsmonitors. BIBB Report, vol 10(09)
German Federal Agency of Labour (2011) Klassifikation der Berufe 2010: Band I: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg
Golicic S, Davis D (2012) Implementing mixed methods research in supply chain management. Int J Phys Distrib Logistics Manage 42(8/9):726–741
Hathaway SR, McKinley JC, Engel R (2000) Minnesota multiphasic personality inventory 2 (MMPI-2). Hogrefe, Göttingen
Hildebrand W-C, Roth A (2008) Führungskräfte für die Logistik - Akademische Ausbildung in Deutschland. In: Baumgarten H (ed) Das Beste der Logistik. Springer, Berlin, pp 69–79
Hohenstein N, Feisel E, Hartmann E (2014) Human resource management issues in supply chain management research: a systematic literature review from 1998 to 2014. Int J Phys Distrib Logistics Manage 44(6):434–463
Holtbrügge D (2010) Personal management, 4th edn. Springer, Berlin
Jahns C, Darkow I-L (2008) Die besten Köpfe für die Logistik gewinnen. In: Baumgarten H (ed) Das Beste der Logistik Innovationen, Strategien, Umsetzungen. Springer, Berlin, pp 81–87
Jung H (2003) Controlling. Oldenbourg, München
Klumpp M (2009) Logistiktrends- und Logistikausbildung 2020. ILD, Essen
Kotzab H, Wünsche S (2013) Competences and qualifications of logistics managers—an empirical analysis of job postings in Germany. In: Dethloff J, Haasis H-D, Kopfer H, Kotzab H, Schönberger J (eds) Logistics management. Lecture Notes in Logistics. Springer International Publishing, pp 517–527
Kristof A (1996) Person-organization fit: an integrative review of its conceptualizations, measurements, and implications. Pers Psychol 49:1–49
Maccoby E, Maccoby N (1965) Das Interview: Ein Werkzeug der Sozialforschung. In: König R (ed) Das Interview, Praktische Sozialforschung I, 4th edn. Köln und Berlin
Mayer O (2008) Interview und schriftliche Befragung, 4th edn. Oldenbourg, München
Mayring P (2002) Einführung in die Qualitative Sozialforschung, 5th edn. Beltz, Weinheim und Basel
Mayring P (2010) Qualitative Inhaltsanalyse, 11th edn. Beltz Verlag, Weinheim und Basel
Müller K (2008) Europäischer und deutscher Qualifikationsrahmen. Wirtschaft Berufserziehung 60(5):13–18
Murphy P, Poist R (2007) Skill requirements of senior-level logisticians. Int J Phys Distrib Logistics Manage 28(4):284–301

North K (2007) Kompetenzrad und Kompetenzmatrix. In: Erpenbeck J, von Rosenstiel L (eds) Handbuch Kompetenzmessung - Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis. Schäffer Poeschel, Stuttgart, pp 175–186

Oechsler WA (2000) Personal und Arbeit: Grundlagen des Human Resource Management und der Arbeitgeber-Arbeitnehmer-Beziehungen, 7th edn. Oldenbourg, München und Wien

Oxford Research (2010) Querschnittsstudie zur Entwicklung der Qualifikationsanforderungen in 19 Wirtschaftssektoren. Online address: http://ec.europa.eu/social/BlobServlet?docId=4687&langId=de. Checked on 21.01.2013

Preissing D (2001) Erfolgreiches Personalmanmagement im demografischen Wandel. Oldenbourg, München

Preissler PR (2014) Controlling: Lehrbuch und Intensivkurs, 14th edn. Oldenbourg

PricewaterhouseCoopers (2012) Transportation and logistics 2030. Winning the talent race, vol 5

Rumberger R (1981) The rising incidence of overeducation in the U.S. labor market. Econ Educ Rev 1(3):75–86

Schön D (2012) Planung und Reporting im Mittelstand. Gabler, Wiesbaden

Tippelt R, Schmidt B (2009) Handbuch Bildungsforschung: Arbeit, Qualifikation und Kompetenzen. Verlag für Sozialwissenschaften, Wiesbaden

Ulrich D, Smallwood N (2004) Organizational capabilities. Harvard Business manager, June, pp 119–127

Wu Y-C, Huang S, Goh M, Hsieh Y (2013) Global logistics management curriculum: perspective from practitioners in Taiwan. Supply Chain Manage: Int J 18(4):376–388

Wünsche S, Kotzab H, Teller C (2014) In search of excellence: what is the competence and qualification profile of logistics managers? In: Gammelgaard B, Prockl G, Kinra A, Aastrup J, Holm Andreasen P, Schramm H-J, Hsuan J, Malouf M, Wieland A (eds) Proceedings 26th conference of the nordic logistics research network—NOFOMA 2014; Competitiveness through supply chain management and global logistics, 11–13 June, pp 235–248

Part V
Flexible Production Management

A Lot Streaming Model for a Re-entrant Flow Shop Scheduling Problem with Missing Operations

Richard Hinze

Abstract Re-entrant scheduling problems are characterized by a multiple processing of the jobs on one or more machines. Such problems occur in wafer manufacturing, paint shops and rework operations. This paper proposes a lot streaming formulation for a re-entrant flow shop scheduling problem with missing operations. The examined objective values are makespan and the sum of completion times. The model is compared to an existing approach. In addition the impact of an increasing number of sublots is examined.

1 Introduction

This article describes a mathematical model for a re-entrant flow shop scheduling problem. In machine scheduling problems a set of jobs needs to be processed on one or more machines. Each job in flow shop problems consists of multiple operations. The operations are performed on several machines. The machine sequence is predetermined and identical for each job. A re-entrant flow shop scheduling problem additionally requires the jobs to be processed more than once on at least one machine. Re-entrant product flows occur for instance in wafer fabrication (Kumar et al. 2004), LCD panel production (Choi et al. 2011) and air plane engine manufacturing (Hekmatfar et al. 2011).

A re-entrant flow shop was first described in 1983 (Graves et al. 1983). Since then different models and solution techniques for the re-entrant flow shop as well as re-entrant job shop scheduling problems are discussed (Uzsoy et al. 1992, 1994; Danping and Lee 2011). Different types of re-entrants have been discussed in literature (Emmons and Vairaktarakis 2012). This paper considers cyclic re-entrant, which requires the jobs to be processed in the production system in loops. These

R. Hinze (✉)
Department of Business Administration, Logistics and Production Management, University of Applied Sciences Merseburg, Eberhard-Leibnitz-Str. 2, 06217 Merseburg, Germany
e-mail: richard.hinze@hsmerseburg.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_11

loops are called levels subsequently. The objective functions examined in this paper are makespan and the sum of completion times. Makespan is a common measure of evaluating schedules. A low makespan implies high machine utilization. The total completion times can be used to represent the work-in-process inventory (Conway et al. 1967) as a performance measure in production logistics.

2 Re-entrant Flow Shop

The model in Sect. 3 is based on two prior formulations for re-entrant flow shops (Pan and Chen 2003; Hinze et al. 2013), which includes the properties described in Sects. 2.1 and 2.2. The multiple runs of jobs in a production system are formulated as levels. A job enters a new level $l + 1$, when it returns to a machine to be processed again. Each level of a job is assigned to one sequence position for the flow shop. The total number of sequence positions is calculated as the product of the number of jobs and the number of levels per job. Re-entrant flow shop scheduling problems are NP-hard (Wang et al. 1997).

The lot splitting approach described in Sect. 2.3 is based on a job shop formulation with lot streaming (Buscher and Shen 2011).

2.1 *Mixed Levels*

The mixed levels property allows a job to start its level $l + 1$ after it finished its level l preceding the levels l of the other jobs. That leads to a possible sequence where a job in level l is preceded and followed by jobs in a level equal or lower than l. Mixed levels allow jobs to be finished earlier and therefore reduce completion time. This effect is illustrated by the example in Sect. 2.2.

2.2 *Missing Operations*

Missing operations represented by processing times equal to zero are possible, for example if a job does not re-enter a production system at the first machine but maybe at the second or third machine. The following example (Tables 1 and 2) for two jobs ($i = 1, 2$) being processed on three machines ($k = 1, 2, 3$) over two levels ($l = 1, 2$) illustrates such a behavior.

Both jobs re-enter the system on the second machine. Therefore the second processing time ($l = 2$) on the first machine ($k = 1$) is zero.

Solving the problem with the model from Pan and Chen (2003) leads to a makespan of 13 time units represented in Fig. 1. A model that deals effectively with missing operations and mixed levels (Hinze et al. 2013) leads to the schedule with

Table 1 Example processing times job $i = 1$

$i = 1$	$k = 1$	$k = 2$	$k = 3$
$l = 1$	2	1	1
$l = 2$	0	1	3

Table 2 Example processing times job $i = 2$

$i = 2$	$k = 1$	$k = 2$	$k = 3$
$l = 1$	3	3	2
$l = 2$	0	1	1

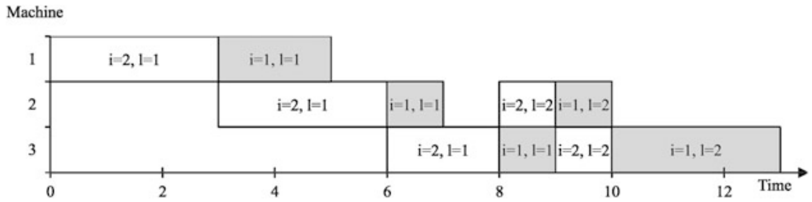


Fig. 1 Example for two jobs on three machines with a re-entry on machine two without mixed levels and proper consideration of missing operations

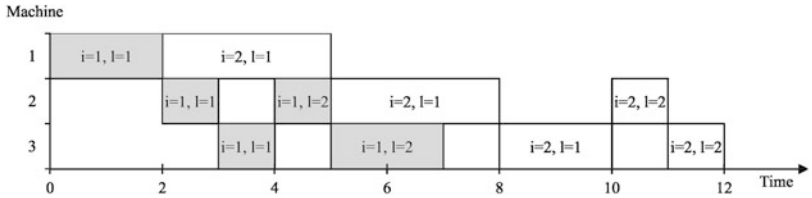


Fig. 2 Example for two jobs on three machines with a re-entry on machine two with mixed levels and proper consideration of missing operations

reduced makespan of 12 time units shown in Fig. 2. The schedule shown in Fig. 2 is not possible in the former formulation, because the operation of job $i = 1$ in level $l = 2$ on machine $k = 2$ needs to wait for machine $k = 1$ to process job $i = 1$ zero time units before it can change to machine $k = 2$.

2.3 Lot Streaming

It is possible to split the operations to process a job, if the job can be assumed as a lot consisting of multiple parts. Dividing the job into two sublots makes it possible to transfer the first sublot earlier to the next processing step.

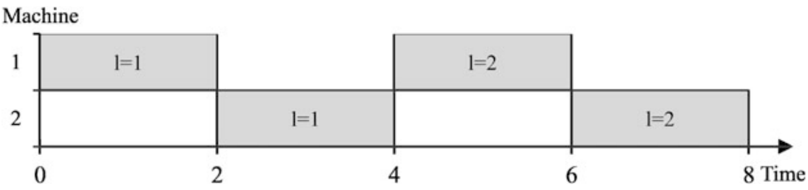


Fig. 3 Example for one job on two machines without lot streaming

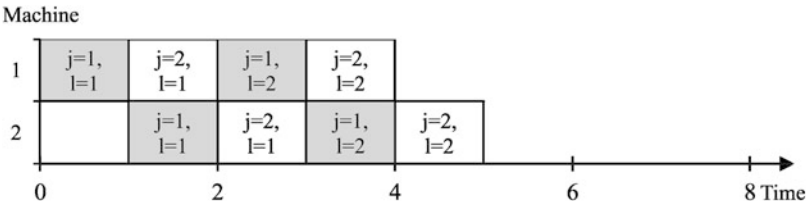


Fig. 4 Example for one job on two machines with lot streaming

Figure 3 shows the processing of a job, which needs to visit two machines twice. The job is not split into sublots. The throughput time is longer than in a schedule of a job divided into two sublots (Fig. 4).

Lot streaming makes it possible to schedule the sublots of a job. The model in Sect. 3 is able to split jobs into J sublots in order reduce the objective values makespan and completion times.

3 Re-Entrant Flow Shop Model

The proposed model uses the notations given in Table 3 to introduce lot streaming to a re-entrant flow shop with missing operations (Sect. 2.2).

3.1 Objective Functions

The first objective is the minimization of makespan (1). Makespan is the time between the start of the first operation of the whole set of jobs and the end of the last operation of all jobs.

$$C_{\max} \rightarrow \min \tag{1}$$

Table 3 Notations

n	Number of jobs
J	Number of sublots
L	Number of levels
m	Number of machines
i, i'	Job indices
j, j'	Sublot index
l, l'	Levels indices
k, k'	Machine indices
T	Sufficiently large number
p^i_{lk}	Unit processing time of job i on level l on machine k $\forall i = 1, \ldots, n;\ j = 1, \ldots, J;\ l = 1, \ldots, L$
D^i	Size of job i $\forall i = 1, \ldots, n$
X^{ij}	Size of the jth sublot of job i $\forall i = 1, \ldots, n;\ j = 1, \ldots, J$
$y^{il}_{i'l'}$	Binary variable, 1 if job i on level l precedes job i' on level l'; 0 otherwise $\forall i, i' = 1, \ldots, n;$ $l, l' = 1, \ldots, L$
s^{ij}_{lk}	Start time of sublot j of job i on level l on machine k $\forall i = 1, \ldots, n;\ j = 1, \ldots, J;\ l = 1, \ldots, L;$ $k = 1, \ldots, m$
$C_{\max}$	Makespan
C_i	Completion time of job i $\forall i = 1, \ldots, n$

The second objective function (2) in this paper is the sum of completion times. The completion time of a job is the time when a job is finished.

$$\sum_{i=1}^{n} C_i \rightarrow \min \tag{2}$$

3.2 Constraints

Equation (3) ensure that the set of sublots contains all units of the complete job.

$$\sum_{j=1}^{J} X^{ij} = D^i \quad \forall i, j \tag{3}$$

The multiplication with the processing times in the constraint sets (4)–(6) enables the ability to skip missing operations. Both sides of the constraints equal zero, if one of the considered operations is missing.

Constraints (4) keep the sequence of the jobs operations, starting at $k = 1$.

$$p^i_{lk'} \cdot \left(s^{ij}_{lk} + p^i_{lk} \cdot X^{ij}\right) \cdot p^i_{lk} \le p^i_{lk'} \cdot s^{ij}_{lk'} \cdot p^i_{lk} \quad \forall i, j, l, k < k' \tag{4}$$

The sublots of a job i should be processed in order, starting with sublot $j = 1$ and ending with sublot $j = J$ in (5).

$$p^i_{lk'} \cdot \left(s^{ij}_{lk} + p^i_{lk} \cdot X^{ij}\right) \cdot p^i_{lk} \le p^i_{lk'} \cdot s^{i,j+1}_{lk'} \cdot p^i_{lk} \quad \forall i, j, l < L, k \le k' \tag{5}$$

The level transition of a job is regulated in (6). A level l + 1 is not allowed to be processed before level l is finished.

$$p^i_{l+1,k'} \cdot \left(s^{ij}_{lk} + p^i_{lk} \cdot X^{ij}\right) \cdot p^i_{lk} \le p^i_{l+1,k'} \cdot s^{i,j}_{l+1,k'} \cdot p^i_{lk} \quad \forall i, j, l < L-1, k, k' \tag{6}$$

The job levels are kept in a basic sequence by constraint set (7). The job sequence within each level is the same. These constraints are used to give the formulation a stronger structure.

$$y^{ijl}_{i'j'l} = y^{ijL}_{i'j'L} \quad \forall i, i', j, j', l < L \tag{7}$$

Equation (8) ensure the unambiguousness of sequence variables.

$$y^{ijl}_{i'j'l'} + y^{i'j'l'}_{ijl} = 1 \quad \forall i, i', j, j', l, l' \tag{8}$$

The model includes consistent sublots, so the sublots of the same job follow each other subsequently (9).

$$y^{ijl}_{i'j'l'} + y^{i'j'l'}_{ijl} = 1 \quad \forall i \ne i', j < j', l, l' \tag{9}$$

The constraint sets (10) and (11) prevent jobs to be scheduled on a machine k, when there is already another job at the same time on k.

$$T \cdot \left(1 - y^{ijl}_{i'j'l'}\right) + s^{i'j'}_{l'k} - s^{ij}_{lk} \ge p^i_{lk} \cdot X^{ij} \quad \forall i \ge i', j, j', l, l', k,\ if(i = i')\{j \ne j\} \tag{10}$$

$$T \cdot y^{ijl}_{i'j'l'} + s^{ij}_{lk} - s^{i'j'}_{l'k} \ge p^{i'}_{l'k} \cdot X^{i'j'} \quad \forall i \ge i', j, j', l, l', k,\ if(i = i')\{j \ne j\} \tag{11}$$

The makespan of the schedule is calculated by (12). It is necessary to check the end of all operations to compute the makespan, since it is not known whether the operation on the last machine has a processing time equal zero.

$$s^{ij}_{lk} + p^i_{lk} \cdot X^{ij} \le C_{\max} \quad \forall i, j, l, k \tag{12}$$

Equation (13) measure the completion time of each job. It is only necessary, if the sum of completion times is the objective.

$$s^{ij}_{lk} + p^i_{lk} \cdot X^{ij} \le C_i \quad \forall i, j, l, k \tag{13}$$

The constraints (14) are the binary constraints to the sequence variables. The non-negativity constraints apply to the sublot size (15) and (16) to the starting times of the sublots.

$$y^{ijl}_{i'j'l'} \in \{0;1\} \quad \forall i, i', j, j', l, l' \tag{14}$$

$$X^{ij} \geq 0 \quad \forall i, j \tag{15}$$

$$s^{ij}_{lk} \geq 0 \quad \forall i, j, l, k \tag{16}$$

4 Computational Experiments

The computational experiments are performed with IBM CPLEX ILOG Optimization Studio 12.4 on a 4 GB RAM, Intel Core 2 Duo 2 GHz, Windows 7 system.

4.1 Test Instances

384 test instances have been created for 24 different problem sizes. Each of the instances contains randomly uniform distributed processing times between 0 and 20. The number of jobs ranges from 2 to 4, the number of levels from 2 to 3 and the number of machines from 2 to 5.

The test problems have been solved using the re-entrant flow shop model from Pan and Chen (2003) and the model proposed in this article. The test with the lot streaming model has been performed for each test instance, assuming 1, 2, 3 or 4 sublots.

4.2 Results

The values in Tables 4 and 5 are calculated using the mean reduction over all 16 instances for each problem size.

The mean reductions of objective values by using the model able to deal with zero processing times are shown in the column “Reduction 1 sublot”. In that calibration the proposed model is set to $J = 1$ sublot per job, so the jobs are not split into sublots. The effect of lot streaming is examined by increasing the number of sublots for each job to $J = 2$ and $J = 3$ in the columns “Reduction 2 sublots” and “Reduction 3 sublots”. The results in these columns are compared to the objective values obtained with a configuration of $J - 1$.

The makespan reductions with $J = 1$ range from 2.71 to 6.29 %. The largest relative reduction of makespan in the experiments is achieved by introducing the lot streaming ability with a number of sublots $J = 2$. The mean reduction of makespan is between 2.90 and 37.85 % compared to the configuration of the new model with

Table 4 Mean reductions of makespan

Jobs	Levels	Machines	Reduction[a] 1 sublot (%)	Reduction[b] 2 sublots (%)	Reduction[b] 3 sublots (%)
2	2	2	3.13	9.90	2.05
		3	3.03	19.81	5.09
		4	2.71	25.25	8.48
		5	2.98	34.49	11.30
	3	2	3.65	13.11	3.42
		3	2.87	23.13	6.06
		4	2.84	32.88	10.01
		5	2.73	37.85	10.55
3	2	2	3.21	6.15	1.17
		3	3.54	13.59	4.30
		4	3.28	22.38	6.36
		5	2.99	26.84	7.39
	3	2	4.19	4.45	0.65
		3	2.95	13.49	4.59
		4	3.03	21.04	5.27
		5	3.00	28.25	6.35
4	2	2	3.29	2.90	0.65
		3	4.53	10.79	3.17
		4	2.98	13.12	3.50
		5	2.98	20.14	5.46
	3	2	6.29	3.58	0.63
		3	4.86	8.38	2.15
		4	3.79	13.45	3.42
		5	2.86	20.56	5.33

[a]The reduction is calculated between the makespan calculated with the model of Pan and Chen (P&C) and the lot streaming model considering zero processing times and mixed levels for sublots $J = 1$: $(C_{\max,P\&C} - C_{\max,J=1})/C_{\max,P\&C}$

[b]The reduction is calculated by comparing the makespan for calculations with a given number of sublots J and $J - 1$: $(C_{\max,J-1} - C_{\max,J})/C_{\max,J-1}$

$J = 1$. The mean relative reduction of makespan by increasing the number of sublots from $J = 2$ to $J = 3$ is lower for each of the test problem sizes. Makespan reductions are achieved for all test instances.

The mean reduction of total completion time by skipping missing operations and $J = 1$ sublots per job are between 3.01 and 19.63 %. The largest relative reduction of total completion time is also achieved by increasing the number of sublots from $J = 1$ to $J = 2$. This is the same effect as for makespan minimization. The range of reduction is from 12.82 to 36.08 %. The effect of an increasing number of sublots per job is also lower for the minimization of completion time, when the number of sublots switches from $J = 2$ to $J = 3$.

Table 5 Mean reductions of the sum of completion times

Jobs	Levels	Machines	Reduction[a]	Reduction[b]	Reduction[b]
			1 sublot (%)	2 sublots (%)	3 sublots (%)
2	2	2	5.43	16.86	5.28
		3	4.22	24.04	10.72
		4	3.01	28.73	12.02
		5	3.05	33.18	15.83
	3	2	8.12	20.90	7.52
		3	4.58	26.58	11.66
		4	3.75	33.26	14.03
		5	3.40	36.08	15.92
3	2	2	11.34	15.73	4.60
		3	5.86	22.61	9.12
		4	5.24	24.82	11.25
		5	3.80	28.96	12.69
	3	2	13.20	16.12	5.64
		3	8.37	22.75	9.71
		4	4.53	25.43	11.78
		5	4.30	30.24	13.34
4	2	2	14.56	12.82	4.30
		3	10.91	19.23	7.02
		4	5.92	22.29	9.40
		5	3.78	24.90	10.75
	3	2	19.63	14.43	4.43
		3	10.71	18.20	7.96
		4	7.01	22.42	9.85
		5	4.93	26.06	11.51

[a]The reduction is calculated between the total completion times calculated with the model of Pan and Chen and the lot streaming model considering zero processing times and mixed levels for sublots $J = 1$: $(\sum C_{i,\ P\&C} - \sum C_{i,\ J=1})/\sum C_{i,\ P\&C}$

[b]The reduction is calculated by comparing the total completion times for calculations with a given number of sublots J and $J - 1$: $(\sum C_{i,\ J-1} - \sum C_{i,\ J})/\sum C_{i,\ J-1}$

5 Conclusions

A formulation for a re-entrant flow shop scheduling problem is proposed, that is able to avoid waiting times for missing operations and can divide jobs into sublots. The formulation leads to significantly reduced objective values. Makespan and the sum of completion times have been used as objective function in two separate experiments.

The target values could be reduced compared to a formulation from literature by avoiding waiting times for missing operations. Further reductions are achieved by splitting the jobs into sublots. The highest reduction of target values is achieved by splitting each job into two sublots. The reduced total sum of completion time implies a lower average completion time, which means a reduced work-in-process. Future models should examine the possibilities to vary sublot sizes for the different levels of a job. Further research should be spent on applying heuristic solution approaches on re-entrant flow shop problems with lot streaming, as the model requires large computational resources to solve bigger problem classes.

References

Buscher U, Shen L (2011) An integer programming formulation for the lot streaming problem in a job shop environment with setups. In: International multi-conference of engineers and computer scientists (IMECS 2011), pp 1343–1348

Choi HS, Kim JS, Lee DH (2011) Real-time scheduling for reentrant hybrid flow shops: a decision tree based mechanism and its application to a TFT-LCD line. Expert Syst Appl 38:3514–3521

Conway RW, Maxwell WL, Miller LW (1967) Theory of scheduling

Danping L, Lee CK (2011) A review of the research methodology for the re-entrant scheduling problem. Int J Prod Res 49:2221–2242

Emmons H, Vairaktarakis G (2012) Flow shop scheduling: theoretical results, algorithms, and applications

Graves SC, Meal HC, Stefek D, Zeghmi AH (1983) Scheduling of re-entrant flow shops. J Oper Manage 3:197–207

Hekmatfar M, Fatemi Ghomi S, Karimi B (2011) Two stage reentrant hybrid flow shop with setup times and the criterion of minimizing makespan. Appl Soft Comput 11:4530–4539

Hinze R, Sackmann D, Buscher U, Aust G (2013) A contribution to the reentrant flow-shop scheduling problem. In: Proceedings of IFAC conference on manufacturing modelling, management and control, pp 718–723

Kumar R, Tiwari M, Allada V (2004) Modelling and rescheduling of a re-entrant wafer fabrication line involving machine unreliability. Int J Prod Res 42:4431–4455

Pan J, Chen J (2003) Minimizing makespan in re-entrant permutation flow-shops. J Oper Res Soc 54:642–653

Uzsoy R, Lee CY, Martin-Vega LA (1992) A review of production planning and scheduling models in the semiconductor industry part I: system characteristics, performance evaluation and production planning. IIE Trans 24:47–60

Uzsoy R, Lee CY, Martin-Vega LA (1994) A review of production planning and scheduling models in the semiconductor industry part II: shop-floor control. IIE Trans 26:44–55

Wang M, Sethi S, van de Velde S (1997) Minimizing makespan in a class of reentrant shops. Oper Res 45:702–712

Identifying Complexity-Inducing Variety: Adapting ClustalW for Semiconductor Industry

Jan Müller, André Wenzel and Rainer Lasch

Abstract Identifying complexity-inducing factors is an essential precondition for the successful management of manufacturing complexity. However, established methods do not fully address the characteristics of semiconductor front-end manufacturing. This paper introduces Multiple Workflow Alignment (MWA) as a new approach to identify differences between products that increase semiconductor manufacturing complexity—so called complexity-inducing variety. MWA is an adaption of the biological sequence alignment algorithm ClustalW and is characterized by high accuracy and reliability. The resulting alignments can be used for logistic-oriented workflow-design and complexity measurement.

1 Introduction

Semiconductor front-end manufacturing facilities probably represent the most capital-intensive and complex manufacturing plants today (Ignizio 2011). The main complexity drivers of this specific type of manufacturing are (Aelker et al. 2013; Chien et al. 2011; Keil 2012; Mönch and Yugma 2012):

- Highly heterogeneous customers demanding many different products
- Up to 1.000 very complex process steps in one workflow
- Up to 800 machines of 100 different types in one facility, high capacity costs
- Non-linear workflows characterized by re-entrant flows where the same sequence is passed several times
- High variability and demand volatility, fierce competition.

Neshati (2013) states that complexity in semiconductor manufacturing has increased during the last decades and will continuously climb. Thus, complexity

J. Müller (✉) · A. Wenzel · R. Lasch
Lehrstuhl Für Betriebswirtschaftslehre, Insb. Logistik, Technische Universität Dresden, Münchner Platz 1/3, 01069 Dresden, Germany
e-mail: jan.mueller1@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_12

management has become a vital competitive factor. According to Meier and Hanenkamp (2004), the first step in complexity management has to be the **identification** of complexity-inducing factors. In this paper we focus on **complexity-inducing variety** in semiconductor front-end manufacturing. A large amount of manufacturing complexity is caused by the variety of manufacturing processes and products, which can be described by the amount and the structure of applied workflows. Workflows are defined as sequences of process steps that have to be conducted to create the desired functionality of a semiconductor device.

However, due to the characteristics of this industry, many approaches of identifying complexity-inducing variety cannot be applied properly (see Sect. 2). Therefore, we propose using **Multiple Workflow Alignments (MWA)** for the identification of such variety in a front-end facility. Aligning workflows of different products means to arrange them in such a way that relatively similar process steps are placed next to each other. During the alignment, the order of the process sequences is maintained. Dissimilar steps that cannot be arranged with another step are instead assigned to a gap. An exemplary alignment of three different workflows A, B and C is depicted in Fig. 1. For illustration purposes also a simplified version of the alignment is displayed (also used in Fig. 5).

The alignments reveal structural differences between the considered workflows, which increase complexity by causing personnel training, sophisticated dispatching rules, ramified material flows, etc. In combination with information on the technological complexity of single process steps, those differences are the key to describe variety-induced complexity. In a subsequent step, the complexity can be reduced by standardizing the regarded workflows (Frizelle and Woodcock 1995).

For calculating the alignments we adapted the biological sequence alignment algorithm ClustalW (see Sect. 3). The information gained from the MWA-algorithm serves as a profound basis for detailed workflow analyses in order to manage variety-induced complexity (see Sects. 4 and 5). Finally, a conclusion with a summary of the results and future research directions is given.

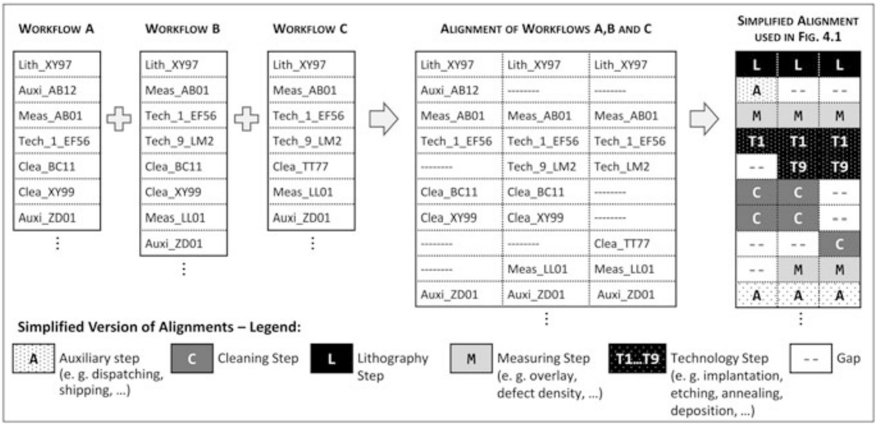


Fig. 1 Basic concept of multiple workflow alignments

2 Identifying Complexity-Inducing Variety: Requirements and Existing Approaches

To capture complexity-inducing variety in semiconductor front-end manufacturing, an underlying method should fulfil several requirements (e.g. Ignizio 2009; Keil et al. 2014):

- *Consideration of product variety*: In addition to describing the complexity of single products the method should also allow to consider **interrelations** that might occur due to the differences between various products.
- *Detailed depiction of manufacturing processes*: To ensure an accurate description of the differences between products a suitable method should analyze workflows on **single process level**. Additionally, the method should retain the order of process steps, to enable the consideration of interrelations between subsequent technological processes.
- *Automation*: Front-end workflows contain up to 1.000 process steps. Moreover, semiconductor manufacturing processes and product portfolios are continuously changing. Therefore, the method has to be fully automatable without losing accuracy or reliability.
- *Consideration of semiconductor manufacturing characteristics*: The method should consider typical characteristics like the layer-oriented production process and re-entrant flows. For detailed information on the semiconductor manufacturing process, please refer i. a. to van Zant (2014).

In literature, there are many approaches on identifying and visualizing complexity-inducing variety. However, they cannot be applied on semiconductor manufacturing due to various reasons.

Schuh (1989) introduced variety-and feature trees. **Variety trees** visualize the different components and assembly orders of variants, **feature trees** show product variety on the basis of different features (Schuh 2005). A related method is the **classification of parts** depending on geometry and raw material (Pawellek 2014). These approaches are very often used for variant management and product design. Unfortunately, they are not suitable for semiconductor production where product variants do not emerge from the assembly of different components but from different process sequences that are conducted on one substrate, the wafer.

Braglia et al. (2006) and Lee (2014) use **machine-part matrices** to link products and processes by displaying how often a machine/process is used. This way, products can be grouped depending on their similarity of production. However, this approach only considers the number of shared process steps and not their order. The result is a similarity score of different products that does not provide a detailed overview of complexity-inducing variety.

Keil (2012) eliminates the deficit of the missing consideration of process sequences and uses **cluster analysis** in semiconductor industry to build sequence families as a basis for the implementation of flow production. It is shown that cluster analysis is a good approach for grouping similar sequences. However, besides the identification of similar sequences we also intend to detect complexity-inducing differences between multiple semiconductor workflows. Nevertheless, cluster analysis could be a good supplement to our approach (see Sect. 5).

Several approaches directly address the complexity of manufacturing systems: Based on **graph theory**, Espinoza et al. (2012) develop six different indices to measure the complexity of the manufacturing layout. ElMaraghy et al. (2005) develop a **complexity code** to depict the inherent structural complexity of manufacturing systems. Both methods are not really suitable for an application in semiconductor industry due to the high number of process steps and re-entrant-flows. Moreover, both methods are mainly designed for assembly production.

None of the described methods meets the stated criteria satisfactorily. Ignizio (2009) suggests the creation of a process-step-centric flowchart for identifying complexity reduction potential. Workflow alignments are quite comparable to this idea (see Sect. 1), however, contrary to Ignizio's approach they allow the joint consideration of multiple product variants: Alignments can be composed of two (pairwise alignment) or more workflows (multiple alignment). By using multiple workflow alignments, one gains several advantages (see Sect. 3):

- Detailed depiction of complexity-inducing variety down to single process level, both for pairs of workflows and across all workflows
- Depiction of similarity relationships in a technology tree
- Adjustable, fully automatable model, short runtime, high accuracy
- Consideration of semiconductor characteristics: Retention of sequence order enables consideration of re-entrant flows; two-step approach reflects layer-oriented production.

3 Adapting ClustalW for Multiple Workflow Alignment

Sequence alignment is extensively used in bioinformatics (Thompson et al. 2011) but is also applied in social sciences (Gauthier et al. 2014), linguistics (List 2012) and other fields (Xing et al. 2010). In bioinformatics there are several approaches to compute the optimal alignment of two sequences; best known are dynamic programming algorithms like that of Needleman and Wunsch (1970). In contrast, the calculation of optimal multiple sequence alignments (MSA) is incomputable for more than a few sequences of average length (Thompson et al. 1994). For that reason, several heuristics have been developed.

For a long time **progressive approaches** were of high importance. They are built on the premise that similar sequences should be aligned at an early stage of the multiple alignment because their close relation reduces the probability of false alignments. In the last twenty years, increasing requirements regarding speed and accuracy of MSA-algorithms led to the development of more efficient approaches (Thompson et al. 2011). Unfortunately, the achieved improvements rest mainly on specific characteristics of protein sequences and are therefore hardly applicable on the alignment of workflows. Considering the fact that workflow alignments are not as complex as biological benchmark alignments (cf. Thompson et al. 2011), the adaption of a less specific algorithm stood to reason.

With this in mind, the progressive alignment algorithm ClustalW of Thompson et al. (1994) appears to be a good choice for an efficient adaption. It has been extensively used in bioinformatics, but has also been adapted by researchers of other scientific fields (e. g. Steiner et al. 2011; Wilson 2008). Like ClustalW, the MWA-algorithm runs through three consecutive stages: Pairwise Alignment → Generation of Technology Tree → Actual Progressive Alignment (cf. Fig. 2).

In sum, the pairwise alignment determines the similarity relations between all pairs of workflows. Based on this information, the neighbor joining algorithm is used to calculate a hierarchical technology tree. The order, in which the workflows are joining the tree, is used to guide the progressive alignment: At the beginning very similar workflows are aligned, the most dissimilar ones come last (cf. Fig. 2—please refer to Thompson et al. (1994) for more information on the basic functionality of sequence alignment algorithms). Despite the obvious similarities, there are significant differences between the adapted algorithm and the original ClustalW. Those are described below:

Differences during pairwise alignment (stages 0 and 1 in Fig. 2):

Two-step approach: Based on the lithography steps, the process sequence of a semiconductor device can be divided into self-contained sub-sequences, so called layers. Thus, a workflow can also be regarded as a sequence of layers, which in turn consist of process steps. To reflect this characteristic (and contrary to ClustalW) the pairwise alignment in MWA is calculated in two steps:

During the first step a process step alignment is calculated for all pairs of layers of the concerned workflows. Afterwards, the resulting alignment is evaluated, leading to a score s_L that describes the degree of similarity between both layers (cf. (2)). After the completion of the first step, the scores of all layer pairings are collected in the layer similarity matrix S_L. This matrix is used in step 2, where the layer sequences of both workflows are aligned so that similar layers are assigned to each other. The division into layers prevents misalignments of common process steps and eases the analysis of the alignment. It provides information on the similarity relationships between the considered layers. Coincidentally, the accuracy of the alignment increases with shorter sequences.

Gotoh's algorithm: For the pairwise alignment in MWA we use Gotoh's dynamic programming algorithm (Gotoh 1982). It calculates global alignments and enables affine gap costs, i.e. gap costs that vary in dependence of the gap length. To calculate the optimal alignment, Gotoh's algorithm assigns scores to occurring matches (pairs of identical elements) and penalties to mismatches (pairs of non-identical elements with more or less similarity) and gaps (single elements corresponding to a space in the other sequence). In general, the optimal alignment is equal to the alignment with maximum total score.

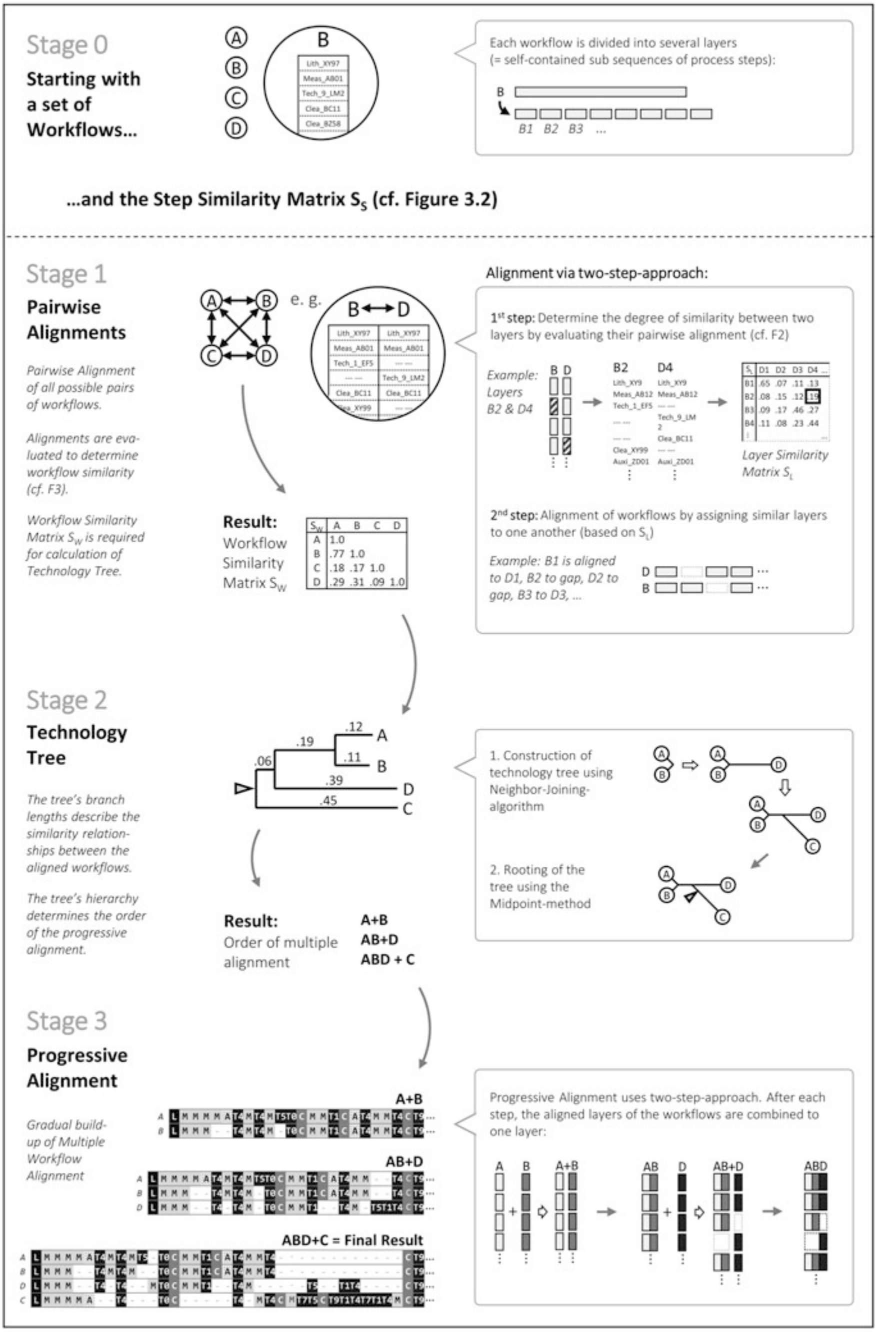


Fig. 2 Basic structure of MWA-algorithm

Similarity Matrix: In case of application, most algorithms are implemented with similarity matrices (resp. score matrices) instead of general match and mismatch scores (Thompson et al. 1994). A similarity matrix S assigns individual scores to all possible pairs of sequence elements. Generally, a match of two elements becomes more likely with increasing score. The exact value, based on which a specific pair is matched, depends on the one hand on the gap penalties, on the other hand on the scores of the adjoining alignment segment.

In MWA, the step similarity matrix S_S rates the similarity and relative importance of process steps: major technology steps like lithography or implantation define the characteristics of a specific layer to a large extend. Therefore, matches of such steps are rated higher than those of cleaning or measuring steps. To build a suitable matrix S_S, during application we classified all steps into 53 categories of process technologies (e.g. technology_1 1–4, cleaning 1–5, measuring 1–6) to which we assigned individual match scores based on the categories' technological relevance. Mismatch penalties were set to $-\infty$ since not all process steps can be assigned to one another (cf. Fig. 3).

Gap Penalties: Another important point is choosing suitable gap penalties. It was shown for sequence alignments that affine gap penalties deliver good results. Since they are also considerably easier to handle and more efficient than other penalties (Cartwright 2006; Liu et al. 2009), we use them for MWA:

$$G = g_O + g_E \cdot k \quad (1)$$

G = gap penalty; g_O = gap opening penalty (GOP); g_E = gap extension penalty (GEP); k = gap length.

It is not possible to define "typical" or "optimal" values for g_O or g_E because they are greatly depending on the alignment's purpose and the structure of the treated sequences. In general, g_E should be lower than g_O. In this case, using a long

Fig. 3 Exemplary step similarity matrix S_S

S_S	Tech_1_IM01	Tech_1_IM02	Tech_1_IM99	…	Tech_2_ET01	…	Clea_CW01	Clea_CW08	…	Meas_TH01	Meas_AL01	…	Auxi_ID01	…
Tech_1_IM01	10													
Tech_1_IM02	7	10												
Tech_1_IM03	5	5	10											
⋮				⋱										
Tech_2_ET01	-∞	-∞	-∞		10									
⋮						⋱								
Clea_CW01	-∞	-∞	-∞		-∞		4							
Clea_CW02	-∞	-∞	-∞		-∞		2	4						
⋮									⋱					
Meas_TH01	-∞	-∞	-∞		-∞		-∞	-∞		3				
Meas_AL01	-∞	-∞	-∞		-∞		-∞	-∞		2	3			
⋮												⋱		
Auxi_ID01	-∞	-∞	-∞		-∞		-∞	-∞		-∞	-∞		2	
⋮														⋱

gap is more favorable than dividing it into two or even more gaps. Consequently, the alignment contains rather long gaps and long match sequences.

We tested various gap penalty functions; of these $G_1 = 2 + k$ for step 1 and $G_2 = 0,2 + 0,1 \cdot k$ for step 2 delivered the best results. Nevertheless, they should be regarded as a mere guideline, which, when MWA is applied, should be adjusted to the characteristics of the specific workflows.

Calculation of similarity scores s_L and s_W: After each alignment during the first step, the similarity score s_L of the two layers is calculated (2). It describes the relation between the actual alignment and the best possible one, which would contain a match for every process step. The aligned sequence is composed of n elements, which in turn are composed of two sub-elements (a and b). Either both sub-elements represent a process step or one sub-element represents a gap.

$$s_L = \sum_{j=1}^{n} \frac{2 \cdot S_S(a_j, b_j)}{S_S(a_j, a_j) + S_S(b_j, b_j)} \tag{2}$$

For a and b the value of the symmetrical similarity matrix is $S_S(a, b)$. The best possible score is the sum of the diagonal values of the score matrix for a and b: $S_S(a, a)$ and $S_S(b, b)$. Penalties for gap opening or extension are not included in the equation since they can distort the score by shifting it into negative numbers—especially for rather different sequences.

In the second step, the similarity score of both workflows s_W is calculated. It is equal to the average layer similarity score of the m aligned layers (3).

$$s_W = \frac{1}{m} \cdot \sum_{i=1}^{m} s_{Li} \tag{3}$$

Technology tree instead of phylogenetic tree (**stage 2 in** Fig. 2):

As mentioned, biological progressive alignments use rooted phylogenetic trees to determine the order in which the single sequences join the multiple alignment. Basically, such a tree describes the degree of relationship between different species. In MWA, the technology tree describes the similarity of semiconductor products. Consequently, it is based on the workflow similarity matrix S_W, which was calculated during the pairwise alignment. The tree is grown from an unresolved, star-shaped tree. Initially, the two most similar sequences are joined to a cluster. The cluster replaces the sequences in S_W, which has to be recalculated afterwards. This procedure is repeated until no sequence is left. Usually clustering algorithms like neighbor-joining (Saitou and Nei 1987) or UPGMA (Sokal 1958) are used for this task (Thompson et al. 1994; Higgins and Sharp 1988).

Contrary to UPGMA, the neighbor joining (NJ) algorithm requires no ultrametric workflow similarity matrix. Moreover, the unrooted tree resulting from NJ has definite branch lengths, which provide additional information on the similarity relations within workflow families. After NJ, the root of the tree is set between the most dissimilar workflows using the midpoint method. The hierarchy of the resulting tree determines the order of the progressive alignment. For a more detailed description of the NJ-algorithm, please refer to Studier and Keppler (1988).

Differences during progressive alignment (stage 3 in Fig. 2):

Two-step approach: During the progressive alignment the workflows are joined to a multiple alignment. This is done by a series of pairwise alignments. Again, the alignments follow the introduced two-step approach.

Scoring: As the alignment progresses, the single workflows are merged to larger and larger multiple alignments that find their equivalent in the clusters of the technology tree. Accordingly, for the progressive alignment the evaluation process of Gotoh's algorithm has to be adapted to these conditions. Like in ClustalW, in MWA the score for a pair of cluster elements is the average score of all possible pairs of sub-elements. For instance, in a case of two cluster-alignments, which are composed of three resp. two workflows, the total score is aggregated of six scores taken from the step similarity matrix S_S (cf. Fig. 4).

No variable gap penalties: ClustalW provides several possibilities for the implementation of variable gap costs. Some are not applicable in MWA, like reduced gap-penalties in hydrophilic stretches; others represent interesting concepts, whose effects on the alignment of workflows need to be further evaluated.

No sequence weighting: In ClustalW sequence weights are calculated based on the phylogenetic tree. Those attach more weight to dissimilar sub-elements when the score of a pair of elements is calculated. The idea is that similar sequences contain duplicate information, which is less interesting for the exploration of evolutionary relations than the information of dissimilar sequences. This specific aspect of the biological sequence alignment cannot be directly transferred to MWA, though it is possible that the general thought of weighting sequences can be put to use. Currently, sequence weighting is not implemented in the algorithm.

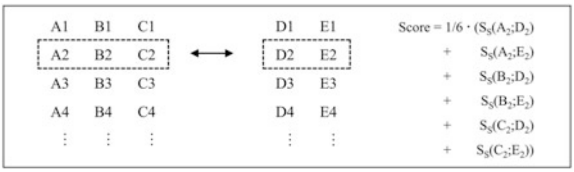


Fig. 4 Scoring during progressive alignment

4 Application

The MWA-algorithm was tested on four workflows of a front-end facility, each standing for a different product family. The workflows were assembled to represent typical challenges that may occur in workflow alignment: (1) Mixture of similar and dissimilar workflows, (2) differences in structure, length, and number of layers and (3) similar layers within one workflow.

Figure 5 shows the multiple alignment of those four semiconductor workflows A, B, C and D in a simplified visualization, in which only general process types are displayed (cf. Fig. 1 for legend, the actual alignments were done for workflows with more than 200 different types of process steps). The product families A and B show only few differences, which mostly concern less complex measure or auxiliary steps (boxes M and A). In contrast, especially D contains several, typically more complex, main technology steps that do not occur in the layers of A or B (boxes T0-T9).

In a hypothetical system that solely produces technology A the introduction of D would add more complexity than the introduction of B. This is caused by D's different material flow, process technologies, interfaces, etc., which could require additional measures like personnel training, process qualification, equipment or dispatching rules.

Altering the product mix share of a relatively divergent product can have similar effects. Thus, products should be standardized as much as possible by reducing the identified disparities (see Sect. 5).

For the set parameters, the MWA-algorithm reaches an accuracy of 99 % for pairwise and 98 % for the multiple alignment compared to the optimal alignment. Apparently, the current version of the algorithm is able to deal with the first two challenges. However, regarding the issue of similar layers within one workflow currently a layer can only be assigned to one other layer, even if there is a whole group of comparably similar layers. Instead of assigning the unconsidered layers to gaps, they should be included into the alignment as a group.

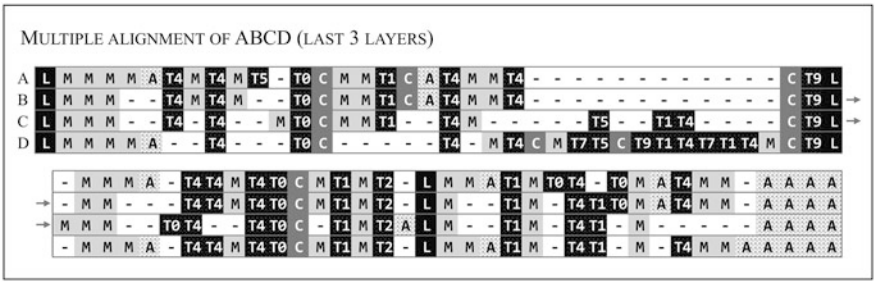


Fig. 5 Multiple alignments

5 Conclusions

In this paper we introduced MWA as an innovative method for the identification of complexity-inducing variety emerging from a specific product mix in semiconductor manufacturing. MWA is an adaption of the biological sequence alignment algorithm ClustalW and enables the fully automated alignment of workflows with very high accuracy. Due to the retention of sequence orders and the presented two-step approach, MWA allows the consideration of semiconductor manufacturing characteristics. The information provided by MWA serves as a basis for further investigations in order to reduce, avoid, and control variety-induced complexity. Therefore, besides using MWA as a complexity identification tool, we can think of several other applications:

Logistic oriented workflow-design: Ignizio (2009) states that "10 to 40 % of the process steps employed by a typical firm may be eliminated or refined." Partly this is due to the common praxis of developing new workflows based on older ones. In this context, divergences that were identified by the MWA-algorithm could indicate improvement potential (cf. Keil et al. 2013, 2014):

- Similar sequences with insertions/deletions of single steps (esp. inspection steps) as indicator for potential dispensable/mergeable process steps
- Swapped steps as indicator for flexibilization potential
- Small differences between aligned process steps/blocks of similar steps as indicator for standardization potential in general.

Complexity measurement: Following the identification of complexity-inducing variety, the workflow alignment could be used for complexity measurement. Because of the retention of sequence order, related complexity drivers like *degree of reentrancy* or *interacting process steps* can be measured more accurately—esp. in combination with process-and material flow data. Furthermore, the alignment enables the differentiation of workflows into common and uncommon segments, making it easier to evaluate the impact of new workflows or product mix changes on complexity.

Of course the current version of MWA leaves room for improvement. The implementation of unused features of ClustalW (e.g. variable gap costs, sequence weighting) could increase the algorithm's performance. To enable the consideration of dynamic aspects of variety-induced complexity, production data (e. g. machine data or key figures) should be linked to the alignment. Furthermore, approaches like those of Keil (2012) or Espinoza et al. (2012) could be a useful supplement.

Additionally, it should be evaluated whether the general idea of workflow alignments could be transferred to other industries that have to deal with very complex manufacturing processes.

Acknowledgments The authors would like to thank G. Schneider from Infineon Technologies Dresden GmbH and S. Keil from TU Dresden for supporting this paper. The work presented in this paper is performed in the project EPPL, financially supported by grants from Austria, Germany, The Netherlands, France, Italy, Portugal and the ENIAC Joint Undertaking (ENIAC JU Grant Agreement no. 325608).

References

Aelker J, Bauernhansl T, Ehm H (2013) Managing complexity in supply chains: a discussion of current approaches on the example of the semiconductor industry. Procedia CIRP 7:79–84

Braglia M, Carmignani G, Zammori F (2006) A new value stream mapping approach for complex production systems. Int J Prod Res 44:3929–3952

Cartwright RA (2006) Logarithmic gap costs decrease alignment accuracy. BMC Bioinf 7:527–538

Chien CF, Dauzère-Pérès S, Ehm H, Fowler JW, Jiang Z, Krishnaswamy S, Mönch L, Uzsoy R (2011) Modelling and analysis of semiconductor manufacturing in a shrinking world: challenges and successes. Eur J Ind Eng 5:254–271

ElMaraghy HA, Kuzgunkaya O, Urbanic RJ (2005) Manufacturing systems configuration complexity. CIRP Ann-Manuf Technol 54:445–450

Espinoza V, ElMaraghy H, AlGeddawy T, Samy SN (2012) Assessing the structural complexity of manufacturing systems layout. In: Proceedings of the 4th CIRP conference on assembly technologies and systems, pp 65–69

Frizelle G, Woodcock E (1995) Measuring complexity as an aid to developing operational strategy. Int J Oper Prod Manag 15:26–39

Gauthier JA, Bühlmann F, Blanchard P (2014) Introduction: sequence analysis in 2014. In: Blanchard P, Bühlmann F, Gauthier JA (eds) Advances in sequence analysis: theory, method, applications. Springer International Publishing, pp 1–17

Gotoh O (1982) An improved algorithm for matching biological sequences. J Mol Biol 162:705–708

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Ignizio JP (2009) Optimizing factory performance. McGraw-Hill

Ignizio JP (2011) Estimating the sustainable capacity of semiconductor fab workstations. Int J Prod Res 49:5121–5131

Keil S (2012) Flussorientierte Gestaltung von Produktionssystemen. Springer, Berlin

Keil S, Eberts D, Lasch R (2013) Designing product workflow for logistics a hidden potential for cycle time and cost reduction as well as quality improvement in high-tech-factories. In: Advanced semiconductor manufacturing conference 2013, pp 27–34

Keil S, Eberts D, Lasch R (2014) Re-designing product workflow for excellence. IEEE Trans Semicond Manuf 27(4):539–550

Lee Y (2014) Wertstromdesign als Instrument der wertorientierten Unternehmensführung. Diss

List JM (2012) Multiple sequence alignment in historical linguistics. In: Proceedings of ConSOLE XIX, pp 241–260

Liu K, Nelesen S, Raghavan S, Linder CR, Warnow T (2009) Barking up the wrong treelength: the impact of gap penalty on alignment and tree accuracy. IEEE/ACM Trans Comput Biol Bioinf 6:7–21

Meier H, Hanenkamp N (2004) Komplexitätsmanagement in der supply chain. In: Busch A, Dangelmaier W (eds) Integriertes supply chain management. Gabler Verlag, pp 111–130

Mönch L, Yugma C (2012) Planning, scheduling, and coordination issues in semiconductor manufacturing. IEEE CASE 2012

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Neshati R (2013) A statistical model to analyze the complexity of semiconductor manufacturing. In: Technology management in the IT-driven services (PICMET), 2013 proceedings of PICMET'13, pp 2757–2764

Pawellek G (2014) Ganzheitliche Fabrikplanung: Grundlagen, Vorgehensweise. Springer, EDV-Unterstützung

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schuh G (1989) Gestaltung und Bewertung von Produktvarianten: ein Beitrag zur systematischen Planung von Serienprodukten. VDI Verlag
Schuh G (2005) Produktkomplexität managen. Hanser Verlag
Sokal RR (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438
Steiner L, Stadler PF, Cysouw M (2011) A pipeline for computational historical linguistics. Lang Dyn Change 1:89–127
Studier JA, Keppler KJ (1988) A note on the neighbor-joining algorithm of Saitou and Nei. Mol Biol Evol 5:729–731
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4676–4680
Thompson JD, Linard B, Lecompte O, Poch O (2011) A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. PLoS ONE 6:1–14
Van Zant P (2014) Microchip fabrication. McGraw-Hill
Wilson C (2008) Activity patterns in space and time: calculating representative Hagerstrand trajectories. Transportation 35:485–499
Xing Z, Pei J, Keogh E (2010) A brief survey on sequence classification. ACM SIGKDD Explor Newslett 12:40–48

Consideration of Redundancies in the Configuration of Automated Flow Lines

Christoph Müller, Christian Weckenborg, Martin Grunewald and Thomas S. Spengler

Abstract Highly automated manufacturing systems offer several advantages and are widely introduced as a strategy to improve the performance of manufacturing organizations. Various forms of automated equipment, such as industrial robots or flexible automatic machines, are used extensively in high-volume industrial production. Especially for flow production systems automation has advanced considerably. This type of production system can, for instance, be found in the automotive, home appliance, or electronics industry where highly automated flow lines are mainly implemented for safety, quality, and productivity reasons. A significant challenge in the operation of highly automated assembly lines is the occurrence of equipment failures which impair the throughput rate. Therefore, buffer space is allocated between the stations of an assembly line in order to achieve a desired throughput rate in spite of equipment failures. However, the installation of buffer space requires considerable investments and also leads to an increase of the average work-in-process inventory in the line. A different approach to achieve a desired throughput rate despite equipment failures is a redundant configuration, in which downstream stations automatically take over the operations of failed stations in the event of failure. The throughput loss in these situations mainly depends on the level of redundancy designed into the system. We present an assembly line balancing model for automated assembly lines which maximizes the lines' level of redundancy for a given number of stations. In a numerical analysis we demonstrate

C. Müller (✉) · C. Weckenborg
Institute of Automotive Management and Industrial Production, Technische Universität Braunschweig, Mühlenpfordtstraße 23, 38106 Brunswick, Germany
e-mail: christ.mueller@tu-braunschweig.de

M. Grunewald · T.S. Spengler
Production and Logistics Research Group, Institute of Automotive Management and Industrial Production, University of Braunschweig, Mühlenpfordtstraße 23, 38106 Braunschweig, Germany
e-mail: m.grunewald@tu-bs.de

T.S. Spengler
e-mail: t.spengler@tu-bs.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_13

the effectiveness of the approach and show that a redundant configuration allows for significant reductions of required buffer sizes in flow lines with unreliable equipment for a given throughput rate.

1 Introduction

Over the past decades, the role of automation technology in modern manufacturing enterprises has increased significantly. In this new environment, several forms of automated equipment, such as industrial robots or flexible automatic machines are available for inclusion in a production system (Graves and Redfield 1988) and are used extensively in high-volume industrial production. Especially for flow production systems automation has advanced considerably. This type of production system can, for instance, be found in the automotive, home appliance, or the electronics industry where highly automated flow lines are mainly implemented for safety, quality, and productivity reasons. In 2013, for instance, worldwide robot sales increased to 178.132 units, the highest level ever recorded for one year. The automotive and electronics industry account for 60 % of the total sales (International Federation of Robotics 2015).

Despite continuous improvement of automated equipment, the occurrence of equipment failures which reduce the throughput remains one major challenge in the operation of highly automated flow lines. As a consequence, the consideration of equipment failures in the configuration of highly automated assembly lines is necessary to limit the resulting throughput loss, especially with regard to the high investments involved. These failures occur randomly over time and can arise for a number of different reasons, e.g., program errors or defects of parts (Moon et al. 2006). In general, the repair times of automated stations, measured as mean time to repair (MTTR), are short. Along with the mean time between failures (MTBF), the technical availability of a robot or machine can be determined.

Within mid-term configuration of automated flow lines two planning problems, namely the assembly line balancing problem and the buffer allocation problem, have to be addressed. The assembly line balancing problem, first defined by Salveson (1955), is set up to determine the required number of workstations in the line and to allocate the total workload for manufacturing any unit of the product to be assembled to the workstations along an assembly line. This problem has received considerable attention in the literature and numerous exact and heuristic methods have been developed for a wide range of ALB problems (for an overview see, e.g., Baybars 1986; Scholl and Becker 2006; Boysen et al. 2007, 2008). However, despite increasing automation, stochastic equipment failures have not been addressed within assembly line balancing so far. Instead, previous research has solely focused on a respective placement and dimensioning of buffers between workstations to maintain production in the event of failures (cf. e.g., Gershwin and Schor 2000). Based on the number of workstations and the distribution of the workload as a result of the deterministic ALB problem, buffer space is allocated within the line to achieve a prespecified target system throughput rate in a consecutive planning step. The installation of buffers requires considerable investments and scarce factory space. Additionally, buffers cause substantial holding costs due to the increase of the average work-in-process (WIP) inventory as well as maintenance costs. Therefore, the minimization of the total number of buffers is the primary objective within the buffer allocation problem (Spieckermann et al. 2000; Tempelmeier 2003).

Another approach to achieve a target system throughput rate despite equipment failures is a redundant configuration. This means that several stations in the line are capable of performing a certain operation. Such a configuration allows for downstream stations to perform the operations of failed stations in addition to those operations originally assigned to them in the event of failure. Therefore, the achievable throughput rate mainly depends on the number of downstream stations capable of performing each operation, which can be considered as a measure for the lines' level of redundancy. Clearly, the lines' level of redundancy could easily be increased by installing additional (backup) stations at the end of the line. However, these additional stations would require substantial investments. A higher level of redundancy can already be achieved for a given number of stations, provided that the configuration of the flow line is carefully planned. The ability of a downstream station to take over the operations of a failed station mainly depends on the selected equipment type, its position within the line, and precedence relations between the operations. Therefore, the problem of including redundancies into an automated flow line is a variation of an assembly line balancing problem where the decisions regarding the equipment types to be installed and operation assignments to stations are made under consideration of the resulting redundancy. As the degree of overlap between the capabilities of the different stations in the line increases, the level of redundancy inherent to the system also increases. Against this background, we present a mathematical formulation to maximize a lines' level of redundancy with respect to a given number of stations.

The remainder of the paper is organized as follows. In the next section, the related literature in the field of assembly line balancing is reviewed. In Sect. 3, a model formulation which maximizes the level of redundancy of an automated flow line is presented. Section 4 deals with a numerical analysis to investigate the effect of redundancies on required buffer sizes for a given throughput rate. The paper closes with a conclusion and an outlook for further research in Sect. 5.

2 Literature Review

The widely studied assembly line balancing problem, where a single product is considered, is the Simple Assembly Line Balancing Problem (SALBP). This problem describes assembly line balancing in its simplest form, where a set of assembly operations is assigned to workstations while the assignment is only restricted by precedence relations between assembly operations and cycle time constraints. Due to these simplifying assumptions the use of SALBP methods is restricted in industrial practice since the definition of the problem ignores many crucial aspects of the real-world problem. Therefore, SALBPs have been extended to a variety of generalized assembly line balancing problems (GALBP) to close the gap between research and real-world. An overview of GALBP has been presented by Becker and Scholl (2006).

While assembly line balancing problems for manual lines have received considerable attention, there is significantly less research devoted to the configuration of automated flow lines. Works for the design of automated flow lines need to consider the equipment required to perform the operations explicitly. Since different equipment alternatives have different capabilities and in some cases also different specializations and processing times for each operation, the assignment of equipment types to the stations restricts the assignment of operations to stations. When equipment decisions are considered the term assembly system design is usually used.

One of the first works in this field has been presented by Graves and Lamar (1983). The authors propose a formulation for the single-product design problem of an automated assembly system where each operation can be performed by one or more alternative types of equipment. The problem consists of the simultaneous assignment of a fixed sequence of operations and the equipment selection with the objective of minimizing total system costs while maintaining a pre-defined production rate. Graves and Redfield (1988) extend the formulation and present a procedure, where multiple similar products with a fixed sequence of assembly operations are considered. The system costs include the annualized fixed costs for the equipment as well as the variable workstation operating costs.

Rubinovitz et al. (1993) present a heuristic algorithm for a robotic assembly line balancing problem (RALBP) where several types of robots with different capabilities are available. The proposed algorithm aims at allocating equal amounts of work to the stations within the line while assigning the most efficient robot type to each station with the objective of minimizing the total number of robots used for a given cycle time. Purchasing costs of the robots are not considered in this approach. Bukchin and Tzur (2000) consider a robotic assembly line balancing problem with the objective of minimizing total equipment costs for a given cycle time. This formulation is extended by Bukchin and Rubinovitz (2003) to consider station paralleling.

Common to all approaches mentioned above is that equipment failures are not considered within assembly line planning. Therefore, the systems' productivity in a stochastic environment can differ significantly. Kahan et al. (2009) are the first to consider equipment failures in a variation of the assembly line balancing problem. They propose a backup strategy for robotic assembly lines where working robots take over the tasks of failed robots during repair time, using the example of an automotive body shop system. Therefore, a mixed-integer formulation, which minimizes the throughput loss after a robot fails by utilizing the redundancy inherent to the system, is developed. The reallocation of operations is subject to precedence constraints as well as assignment restrictions. However, the approach is restricted to the case where the equipment types for the robots are already selected and therefore their capabilities, in terms of the operations the robots can perform, are given. Based on the idea of backup strategies for robotic assembly lines, Müller et al. (2014) presented an approach to maximize the level of redundancy for an automotive body shop line. In a numerical analysis they compare the performance of their approach with existing RALBP methods and can show that the average cycle time can be decreased by maximizing the lines' level of redundancy. However, the approach does not consider precedence relations when determining the level of redundancy. Furthermore, buffers are not considered in the numerical analysis although buffers are generally used in industrial practice to cope with failures.

From this literature analysis we conclude that a line balancing approach which accounts for equipment failures in the configuration of automated flow lines is missing. In the following we adopt and extend the idea of utilizing a lines' inherent redundancy to develop an approach for a redundant configuration of automated flow lines.

3 An Approach for Flow Line Design Under Consideration of Redundancies

3.1 Problem Description

In this section, we address the design problem for automated flow lines with the objective of maximizing the lines' level of redundancy. Therefore, we develop an approach which maximizes the number of feasible downstream stations for each operation, and consider this to be a measure for the lines' level of redundancy. As the number of stations that are capable of performing a certain operation increases, the throughput loss during equipment failures is reduced. This is due to the fact that the operations assigned to the failed station(s) can be reallocated to a larger number of downstream stations.

The ability of a station to perform a certain operation mainly depends on three factors: First, an equipment type which can perform the operation has to be assigned to the station. Second, a backup station should be located downstream with respect to the failed station to which the operation is originally assigned to. Third, precedence relations between operations restrict the number of backup stations for each operation. The example depicted in Fig. 1 illustrates such a case.

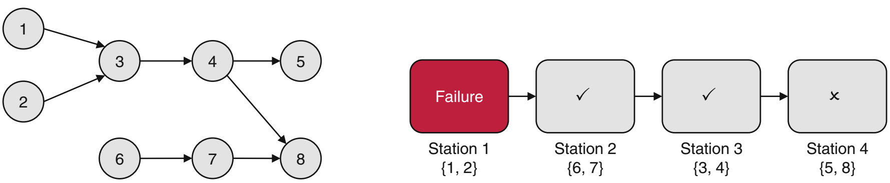


Fig. 1 Problem illustration

Assuming the first station in the example fails, its operations 1 and 2 could only be reallocated to the second or third station. A reallocation to the fourth station, however, is prevented since operation 3 which is an immediate successor of operations 1 and 2 is performed at the third station.

The modeling approach is based on the following four assumptions: First, it is assumed that each equipment type can be assigned to each station in the line. Second, we assume that at most one equipment type can be assigned to each station due to space constraints on the shop floor. Third, the processing times for the operations are constant, regardless of the selected equipment type. Fourth, in case assembly operations are reallocated to downstream stations, it is assumed that the material required to perform the operations is made available to the station or that no material is required. The latter case can, for instance, be observed in automotive body shops where respot welding operations which do not require any material account for a large portion of all assembly operations (Kahan et al. 2009). Other examples are transfer or machining lines for mechanical parts manufacturing such as cylinder heads or engine blocks. Therefore, the capabilities of the stations with regard to the operations they can perform depend only on the three factors discussed above.

3.2 Model Formulation

To precisely describe the decision problem presented above, we develop a model formulation in this section. In the following we give a detailed description of the used notation and the optimization model.

The assembly line consists of a given number of installed stations K which are connected with a material handling system. A set of operations I has to be assigned to the stations and each operation requires a certain processing time t_i. A pre-defined cycle time C indicates how long a product can stay in each station, i.e., the maximum processing time per station. Whether an operation is assigned to a station is indicated by the binary decision variables x_{ik}. The assignment of operations requires the installation of automated equipment to the stations. If an equipment type j is installed at station k is modeled by the binary decision variable y_{jk}. To model the capabilities of the different equipment types we use a capability matrix CM_{ij} which indicates if operation i can be performed by equipment type j. The capabilities of the stations with regard to the operations they can perform are captured by the auxiliary variables z_{ik}. The used notation is summarized in Table 1. With this notation we can state the problem as follows:

$$\text{Maximize} \sum_{i \in I} \sum_{k \in K} z_{ik} \tag{1}$$

Table 1 Notation

Sets and parameters	
I	Set of operations (index h, i)
J	Set of equipment types (index j)
K	Set of stations (index k, l)
IP_i	Set of immediate predecessors of operation i
C	Cycle time (maximum processing time per station)
t_i	Processing time of operation i
M	Sufficiently large number
CM_{ij}	$= \begin{cases} 1, & \text{if operation } i \text{ can be performed by equipment type } j \\ 0, & \text{else} \end{cases}$
Decision and derived variables	
x_{ik}	$= \begin{cases} 1, & \text{if operation } i \text{ is assigned to station } k \\ 0, & \text{else} \end{cases}$
y_{jk}	$= \begin{cases} 1, & \text{if equipment type } j \text{ is assigned to station } k \\ 0, & \text{else} \end{cases}$
z_{ik}	$= \begin{cases} 1, & \text{if station } k \text{ can perform operation } i \\ 0, & \text{else} \end{cases}$

Subject to:

$$\sum_{k \in K} x_{ik} = 1 \quad \forall i \in I \tag{2}$$

$$\sum_{j \in J} y_{jk} \leq 1 \quad \forall k \in K \tag{3}$$

$$x_{ik} \leq \sum_{j \in J} \left(y_{jk} \cdot CM_{ij} \right) \quad \forall i \in I,\ k \in K \tag{4}$$

$$\sum_{k \in K} k \cdot x_{hk} \leq \sum_{k \in K} k \cdot x_{ik} \quad \forall i \in I,\ h \in IP_i \tag{5}$$

$$\sum_{i \in I} x_{ik} \cdot t_i \leq C \quad \forall k \in K \tag{6}$$

$$z_{ik} \leq \sum_{j \in J} \left(y_{jk} \cdot CM_{ij} \right) \quad \forall i \in I,\ k \in K \tag{7}$$

$$\sum_{k \in K} x_{ik} \cdot k \leq l + (1 - z_{il}) \cdot M \quad \forall i \in I,\ l \in K \tag{8}$$

$$z_{hl} \cdot l \leq \sum_{k \in K} x_{ik} \cdot k \quad \forall i \in I,\ h \in IP_i,\ l \in K \tag{9}$$

$$x_{ik}, z_{ik} \in \{0, 1\} \quad \forall i \in I, k \in K \tag{10}$$

$$y_{jk} \in \{0, 1\} \quad \forall j \in J, k \in K \tag{11}$$

The objective is to maximize the lines' level of redundancy, which is calculated in (1) as the sum over all entries of the auxiliary variables z_{ik}. Constraints (2) assure that each operation i is assigned to exactly one station using the binary decision variables x_{ik}. Constraints (3) assure that each station has at most one equipment type. The suitability of each station to perform a certain operation, based on the selected equipment type, is verified in constraints (4). Precedence relations restrict the assignment of operations to stations. Before operation i can be executed, all immediate predecessors in the set IP_i have to be completed (5). Constraints (6) further restrict the assignment of operations to stations since the pre-defined cycle time C must not be exceeded.

Constraints (7)–(9) define the entries of the auxiliary variables z_{ik} which capture the redundancies within the line. Constraints (7) determine whether a certain station k can perform an operation i based on the selected equipment type. Constraints (8) assure that only downstream stations are feasible backup stations. An entry z_{il} can only be equal to one if the station index l is greater than or equal to the index k of the station to which operation i is initially assigned to. Similarly to the initial assignment of operations to stations constraints (9) assure that precedence relations are also considered when determining possible backup stations. If operation $h \in IP_i$ has to be reallocated due to an equipment failure, it can only be assigned to a station which has an index l which is less than or equal to the index k of the station to which operation i is assigned to.

4 Numerical Analysis

4.1 Instance Generation and Benchmark

To illustrate the functionality of the presented approach, a small numerical analysis is carried out in this section. Since no established test datasets for the problem considered in this paper are available in the literature, we first describe the generation of test instances. We randomly choose 20 different instances from the small-sized SALBP dataset of Otto et al. (2013) with 20 operations. These instances include precedence graphs with properties that can be observed in real-world, e.g., bimodal time distribution of processing times. However, for the problem considered in this paper these SALBP instances have to be extended to consider equipment types based on typical values that can be observed in industrial practice.

An operation can be assigned to one or more of the available equipment types. In our numerical analysis we consider 10 different equipment types. Feasible assignments of operations to equipment types are captured by the capability matrix

CM. As the number of equipment alternatives for each operation increases, the potential to increase the lines' level of redundancy also increases due to this flexibility. Therefore, two different levels of equipment flexibility (low and high) are considered. The entries of the capability matrices are randomly drawn with a probability of each 30 % (low level of flexibility) or 60 % (high level of flexibility) that a given equipment type can perform a certain operation. In case none of the existing equipment types is capable of performing a certain operation, we randomly choose one equipment type for this operation with equal probability.

In a real-world setting, equipment failures occur randomly over time and the repair time is also a random variable. Furthermore, the number of stations failing simultaneously may change. To consider these stochastic effects, a discrete-event simulation (implemented in Plant Simulation 10.1) is utilized for the performance evaluation throughout the numerical analysis. In our numerical analysis, we consider an assembly line with 10 stations which are connected with a material handling system and a manual repair station at the end of the line. This production environment is, for instance, typical for automotive body shops (Kahan et al. 2009). We assume exponentially distributed times between failure and exponentially distributed repair times for all stations in the line. Based on an empirical evaluation from an automotive body shop, Inman (1999) concludes that this assumption is reasonable in most cases. The values of the availability and MTTR of the stations are set to 98 % and 120 s, respectively. In the event of failure we assume that the station fails completely, i.e., no operation can be performed at the station during repair time.

To compare the performance of our approach for a redundant configuration of automated flow lines, we use a standard approach for the design of automated assembly lines with the objective of minimizing the cycle time for a given number of stations. This approach uses information on the precedence relations and the capabilities of the different equipment types to find a line configuration, i.e., an assignment of equipment types and operations to stations. In order to compare the two different approaches, the minimum cycle time obtained by using the standard approach has been used as the maximum processing time C for each station in our formulation. Based on the line configurations obtained by using the different approaches the backup plans have been generated. Therefore, a backup formulation which is based on the approach presented by Kahan et al. (2009) is implemented. The backup plan contains the assignment of failed operations to downstream stations. For each failure scenario two options arise: If a feasible solution exists, the operations originally allocated to the failed station are reallocated to one or more downstream stations. If no feasible solution exists, the operations are reallocated to the manual repair station at the end of the line. As in the simulation study of Kahan et al. (2009), we assume that the allocation to the manual repair station is always possible as a last resort when no backup exists, regardless of the precedence constraints. However, if an operation is assigned to the manual repair station, the processing time is assumed to be three times higher compared to the processing time of the automated equipment.

For the buffer allocation a genetic algorithm (GA) is combined with the simulation model. Based on the workload distribution obtained by applying the two different approaches buffer space is allocated in a second step in order to meet a prespecified target system throughput rate. The fitness function of the algorithm comprises the number of buffers allocated in the line in order to punish growing buffer sizes. For each solution candidate, the genetic algorithm calls the simulation model for the performance evaluation. The computation times required for the GA/simulation approach range from 45 to 90 min for each scenario. Each configuration was simulated with the heuristic solution obtained for the buffer allocation by running 1.000 replications with a length of three eight-hour shifts of which the first shift has been used as warm-up period, i.e., no statistics were collected during this period. The model formulations have been implemented using a mathematical programming language (AIMMS) and solved using CPLEX 12.5. All instances could be solved to optimality on a standard computer in less than 10 min.

4.2 Results

To evaluate our planning approach, we calculate the relative reduction of required buffer sizes compared to the benchmark for all instances. In Fig. 2 the boxplots for the relative reductions of buffer sizes are illustrated. It should be noted that our approach strictly outperforms its benchmark with regard to the required buffer sizes for all considered test instances. This result is to be expected, since our approach generally yields line configurations with a higher level of redundancy. Yet, it is possible that the benchmark accidentally finds line configurations which also

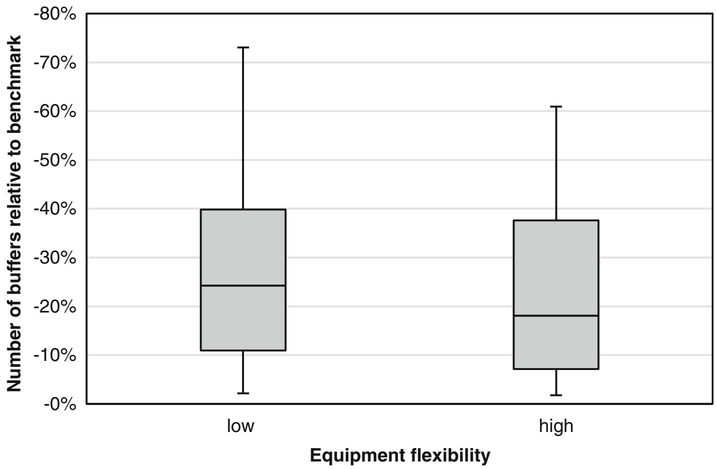


Fig. 2 Boxplot for the reduction of buffer sizes compared to the benchmark

feature a high level of redundancy. Therefore, the relative reductions of required buffer sizes vary widely with values between 1.8 and 73 %. Compared to the benchmark, buffer sizes can be reduced by 26.8 % on average for low equipment flexibility. With an increase in the equipment flexibility the relative reduction of required buffer sizes compared to the benchmark decreases. This result is quite intuitive: The number of operations each equipment type can perform increases in the values of the equipment flexibility. Hence, the effect of the assignment of equipment types to stations on the lines' level of redundancy diminishes. For instance, in the case of full flexibility (i.e., each equipment type can perform each operation) the assignment of equipment types to stations has no effect on the lines' level of redundancy at all. While this case is generally not of practical importance, it illustrates why a higher level of equipment flexibility leads to a decline of the average reduction of buffer sizes of our approach compared to the benchmark. Still, even for the case of high equipment flexibility our approach allows for an average reduction of required buffer sizes by 21.6 %.

5 Conclusion and Outlook

Given the trend towards increasing automation of production processes, we consider the novel planning problem of configuring automated flow lines under consideration of redundancies. This means that several stations in the line are capable of performing a certain operation. Such a configuration allows for downstream stations to perform the operations of failed stations in addition to those operations originally assigned to them in the event of failure. Therefore, the achievable throughput rate depends on the number of downstream stations capable of performing each operation, which can be considered as a measure for the lines' level of redundancy. We present an assembly line balancing model for automated flow lines which maximizes the lines' level of redundancy for a given number of stations. First numerical results demonstrate the effectiveness of the approach and show that a redundant configuration allows for significant reductions of required buffer sizes in flow lines with unreliable equipment for a given throughput rate. For a flow line with ten stations average reductions of required buffer sizes range from 21.6 to 26.8 %.

Further research is especially needed in developing the objective function. With the approach presented in this paper, the total number of redundancies in the line is maximized. This can result in solutions where multiple redundancies are included for a few operations while other operations are not covered at all or where the majority of redundancies are included for stations which have a high technical availability while stations with low technical availability are neglected. Therefore, objective functions which aim at a leveled allocation of redundancies in the line or that use weights for the redundancies, for instance, by taking the technical availabilities into consideration, should be investigated.

Another promising direction for further research is the development of a specialized algorithm for the allocation of buffer space within the line. Furthermore, an analytical algorithm could be used for the performance evaluation of the system instead of simulation. Thus, the speed of optimization could be increased substantially allowing for the evaluation of a large number of line configurations with different allocations of redundancies within short time. This would also allow for extending the approach to explore trade-offs between an increased level of redundancy obtained by installing additional stations and required buffer sizes. For instance, if the investments required for the installation of an additional station are lower than the savings due to reduced buffer sizes, the line configuration with the lowest number of stations is not necessarily the most efficient one.

The presented approach can further be extended to consider different processing times for a certain operation, depending on the selected equipment type. In real-world, for instance, the processing time between a special-purpose machine and a flexible robot would differ significantly. This tradeoff between equipment flexibility and performance should be taken into account. Additionally, the hierarchical level could be refined, i.e., the assignment of equipment types to stations could be complemented by the assignment of tools to the equipment types.

References

Baybars I (1986) A survey of exact algorithms for the simple assembly line balancing problem. Manag Sci 32 (8):909–932

Becker c, Scholl A (2006) A survey on problems and methods in generalized assembly line balancing. Eur J Oper Res 168(3):694–715

Boysen N, Fliedner M, Scholl A (2007) A classification of assembly line balancing problems. Eur J Oper Res 183(2):674–693

Boysen N, Fliedner M, Scholl A (2008) Assembly line balancing: Which model to use when? Int J Prod Econ 111(2):509–528

Bukchin J, Rubinovitz J (2003) A weighted approach for assembly line design with station paralleling and equipment selection. IIE Trans 35(1):73–85

Bukchin J, Tzur M (2000) Design of flexible assembly line to minimize equipment cost. IIE Trans 32(7):585–598

Gershwin SB, Schor JE (2000) Efficient algorithms for buffer space allocation. Ann Oper Res 93 (1–4):117–144

Graves SC, Lamar BW (1983) Integer programming procedure for assembly system design problems. Oper Res 31(3):522–545

Graves SC, Redfield CH (1988) Equipment selection and task assignment for multiproduct assembly system design. Int J Flex Manuf Syst 1(1):31–50

Inman R (1999) Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. Prod Oper Manag 8(4):409–432

International Federation of Robotics (2015) Industrial robots statistics. http://www.ifr.org/industrial-robots/statistics/

Kahan T, Bukchin Y, Menassa R, Ben-Gal I (2009) Backup strategy for robots' failures in an automotive assembly system. Int J Prod Econ 120(2):315–326

Moon D, Cho H, Kim H, Sunwoo H, Jung J (2006) A case study of the body shop design in an automotive factory using 3D simulation. Int J Prod Res 44(18–19):4121–4135

Müller C, Spengler TS, Sodhi MS (2014) Robust flowline design for automotive body shops. In: Guan Y, Liao H (eds) Proceedings of the 2014 industrial and systems engineering research conference, pp 3077–3085

Otto A, Otto C, Scholl A (2013) Systematic data generation and test design for solution algorithms on the example of SALBPGen for assembly line balancing. Eur J Oper Res 228(1):33–45

Rubinovitz J, Bukchin J, Lenz E (1993) RALB—a heuristic algorithm for design and balancing of robotic assembly lines. CIRP Ann—Manuf Technol 42(1):497–500

Salveson ME (1955) The assembly line balancing problem. J Ind Eng 6(3):18–25

Scholl A, Becker C (2006) State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. Eur J Oper Res 168(3):666–693

Spieckermann S, Gutenschwager K, Heinzel H, Voß S (2000) Simulation-based optimization in the automotive industry—a case study on body shop design. Simulation 75(5–6):276–286

Tempelmeier H (2003) Practical considerations in the optimization of flow production systems. Int J Prod Res 41(1):149–170

Part VI
Distribution Management

Prepositioning of Relief Items Under Uncertainty: A Classification of Modeling and Solution Approaches for Disaster Management

Emilia Graß and Kathrin Fischer

Abstract Natural and man-made disasters often have devastating effects on the affected population and the economy. Optimization models for determining the locations of relief facilities and the amount of first-aid items to be stored before the occurrence of an emergency can help decisively to mitigate the impact of a disaster. As research interest in disaster management has grown significantly over the last decade, this paper reviews the state-of-the-art literature concerning prepositioning of relief items in conjunction with facility location decisions, taking the uncertain nature of a disaster into account. A classification of the respective modeling approaches and solution methods is provided to facilitate the identification of relevant research gaps. One of the main findings is the lack of efficient solution methods especially for large scale problems, although generating optimal solutions in a reasonable timeframe is crucial for the success of a relief operation. Based on the detected gaps, future research directions are suggested.

1 Introduction

An increasing number of natural disasters like flooding, storms and earthquakes causes countless fatalities, ecological devastation and economic damage running into billions of US-dollars every year (IFRC 2010). The Centre for Research on the Epidemiology of Disasters (CRED) defines a disaster as: "a situation or event which overwhelms local capacity, necessitating a request to a national or international level for external assistance; an unforeseen and often sudden event that causes great damage, destruction and human suffering" (Guha-Sapir et al. 2012). In general, disaster management seeks to reduce such consequences of a disaster by designing plans according to three phases: preparedness-, response- and recovery-phase.

E. Graß (✉) · K. Fischer
Institute of Operations Research and Information Systems,
Schwarzenbergstr. 95, 28359 Hamburg, Germany
e-mail: e.grass@tuhh.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_14

The two latter phases are essential in the aftermath of a disaster, e.g. aid distribution and reconstruction of buildings and roads, whereas the preparedness phase contains systematic planning of the relief supply chain *before* the occurrence of a disaster. Prepositioning of relief items, i.e. the storage of these items in a disaster-prone area, is of special importance in the pre-disaster context as it reduces the time needed to react to an emergency situation. Moreover, relief organizations who purchase in advance can benefit from price negotiations with suppliers and discounts due to large quantity orders (Rawls and Turnquist 2012). On the contrary, if no emergency supplies are stored in advance, large scale disasters lead to a sudden increase in demand for relief items and relief organizations have to purchase these items from local or global vendors whose inventory levels are usually not sufficient. Such ad hoc purchase decisions result in higher costs, longer delivery times due to shortages and therefore to delayed disaster response and unsatisfied demand, e.g. Balcik and Beamon (2008). As stated in Rawls and Turnquist (2012), the first 72 h in the aftermath of a disaster are crucial for the success of a relief operation. Indeed, prepositioning decisions are of vital importance with regard to response time and at the same time especially challenging due to uncertainties about where, when and to what extent a disaster may occur. Researchers have realized the necessity of incorporating the stochastic nature of disasters into planning approaches and therefore deterministic models are only used as reference models to show their inadequateness, e.g. Verma and Gaukler (2014).

Mainly the work of Rawls and Turnquist (2010) has encouraged researchers to address prepositioning decisions for emergency supplies resulting in an increased number of publications. However, research in this area is still scarce and needs further development in order to enhance responsiveness in the aftermath of a disaster. This paper reviews literature concerning decisions about where to locate emergency facilities, such as shelters and distribution centers, in combination with prepositioning issues under uncertainties. These decisions have to take place in advance, i.e. before the severity of a disaster becomes apparent. Research considering solely inventory decisions in an existing network or isolated location decisions is not considered here. Recent developments, i.e. of the last decade, in the relevant literature are covered in this review, but no claim of completeness is made. In particular, working papers, theses, conference proceedings, books and book sections are excluded from this review. Mainly, published journal articles are considered and keywords like “disaster”, “humanitarian”, “emergency” in combination with “prepositioning” were used. Unlike other reviews, e.g. the one by Caunhye et al. (2012), this review focuses not only on optimization models, but also on solution procedures. Moreover, a specific aspect within the preparation phase is considered allowing for a more detailed analysis of the state-of-the-art modelling and solution approaches. The key contribution of this survey is to classify the corresponding literature currently available, to identify trends and to determine needs for future progress with an emphasis on solution methods.

The remainder of this paper is structured as follows: Sect. 2 classifies the relevant publications based on the modeling approach that is used. In particular, Sect. 2.1 considers stochastic models with scenarios and Sect. 2.2 is dedicated to probability based approaches. Due to space limitations, no models can be presented here; the reader is referred to the respective publications. In Sect. 3 solution techniques, i.e. heuristics (Sect. 3.1) and decomposition methods (Sect. 3.2), for the proposed models are discussed. Both sections close with future research directions. Final conclusions are presented in Sect. 4.

2 Features of Surveyed Optimization Models

All the papers reviewed in this survey are listed in Table 1 and are classified according to model type, area in which uncertainty is assumed, assumptions made, the respective approach capturing uncertainty and the suggested solution method.

In terms of model type, integer programming (IP) and mixed integer programming (MIP) are distinguished. Two-stage stochastic programming models refer to problems where decisions at the first stage are made before the realization of the stochastic parameters and second stage variables are determined afterwards. Based on the classification given in Table 1 two major streams of approaches, namely with and without the use of scenarios, can be identified. As shown in the table, the preferred way to capture uncertainties is scenario-based whereas alternative approaches, e.g. chance constraints, are less frequently used.

2.1 Scenario-Based Approaches

In the literature discussed here, each scenario contains information on demand for relief goods and, optionally, on potential damages to inventories and/or transportation links, i.e. a scenario is a data set containing the relevant information on these aspects, often in combination with a probability of its occurrence. Stochastic models using the scenario-based approach are shown in Table 2 with the corresponding objectives, constraints and assumptions, additionally to those given in Table 1.

Balcik and Beamon (2008) propose a maximal covering location model under uncertain demand. The objective is to decide where to locate a distribution facility and how much of different relief items have to be stored at these capacitated centers such that the expected demand coverage over all scenarios is maximized, i.e. prepositioning issues are effectively addressed. In their model, budgets available before and after the occurrence of a disaster are considered, limiting the costs, e.g. for inventory holding and transportation. Although Balcik and Beamon (2008) consider demand uncertainties they assume undamaged distribution centers such that the prepositioned amounts are completely available in the aftermath of a disaster. The same holds for the models of Salmeron and Apte (2012); Mete and Zabinsky (2010); Duran et al. (2011), additionally assuming undamaged transportation links, see Table 1. Since the probability of network destruction is very

Table 1 Classification of publications according to model type, assumed parameter uncertainties, corresponding approaches, assumptions and solution methods

Author(s)	Year	Model type	Uncertainty	Assumptions	Approach	Solution method
Balcik and Beamon	2008	MIP	Demand; time to satisfy demand	No damages to facilities	Scenario-based	Solver
Mete and Zabinsky	2010	Two-stage MIP	Demand; transportation links	No damages to facilities	Scenario-based	Solver
Rawls and Turnquist	2010	Two-stage MIP	Demand; inventory; transportation links		Scenario-based	Heuristic
Salmerón and Apte	2010	Two-stage MIP	Demand; transportation time	No damages to facilities	Scenario-based	Solver
Campbell and Jones	2011	MIP	Demand (normal distribution)	No damages to transportation links	Newsvendor	Heuristic
Duran et al.	2011	MIP	Demand	No damages to facilities and transportation links	Scenario-based	Solver
Rawls and Turnquist	2011	Two-stage MIP	Demand, inventory, transportation links		Scenario-based	Solver
Bozkurt and Duran	2012	MIP	Demand	No damages to facilities and transportation links	Scenario-based	Solver
Döyen et al.	2012	Two-stage MIP	Demand; capacity; transportation time		Scenario-based	Heuristic
Murali et al	2012	MIP	Demand (log-normal distribution)	No damages to facilities	Chance-constrained	Heuristic
Noyan	2012	Two-stage MIP	Demand; inventory; transportation links		Scenario-based	Decomposition

(continued)

Table 1 (continued)

Author(s)	Year	Model type	Uncertainty	Assumptions	Approach	Solution method
Paul and Hariharan	2012	MIP	Demand; capacity; transportation time and time an individual can survive without health care		Scenario-based	Solver
Rawls and Turnquist	2012	Two-stage MIP	Demand; inventory; transportation links		Scenario-based	Solver
Galindo and Batta	2013	IP	Demand; inventory (uniform integer distribution)	No damages to transportation links	Expected value	heuristic
Klibi et al	2013	Two-stage MIP	Demand; facility and vendor`s inventory damages; transportation links		Scenario-based	solver
Renkli and Duran	2014	MIP	Demand, transportation links	No damages to facilities	Chance-constrained	solver
Verma and Gaukler	2014	Two-stage MIP	Demand; inventory (uniform and normal distributions)	No damages to transportation links	Distance-damage function	decomposition

Table 2 Scenario-based approaches: objectives, constraints and additional assumptions

Author(s)	Year	Objective	Constraints	Additional assumptions
Balcik and Beamon	2008	Max demand covered	Pre-/post-disaster budget	Single demand point
Rawls and Turnquist	2010	Min costs, e.g. Penalties	Facility and transportation link capacities	
Mete and Zabinsky	2010	Min warehouse operation costs, unmet demand and transportation time	Facility and vehicle capacities	Sufficient number of vehicles available
Salmerón and Apte	2010	Min casualties and unmet demand	Budget, facility and vehicle capacities; maximum travel time	
Duran et al	2011	Min response time	Number of open facilities; total inventory	Replenishment lead time of 2 weeks
Rawls and Turnquist	2011	Min costs, e.g. Penalties	Service quality; average shipment distance	
Bozkurt and Duran	2012	Min response time	Number of open facilities; total inventory	
Döyen et al	2012	Min costs	Maximum transportation time; facility capacities	Each demand point served by one facility
Noyan	2012	Min costs and cvar	Facility capacities	Perfect information
Paul and Hariharan	2012	Min costs, e.g. Fatality costs	Facility capacities; budget	Each demand point served by one facility
Rawls and Turnquist	2012	Min costs, e.g. Penalties	Service quality; facility, dispatch and link capacitites	Four time periods
Klibi et al	2013	Min penalties (produced by unsatisfied demand)	Inventory of warehouses	Multi disasters; multi-period

high especially for large scale disasters these assumptions are rather unrealistic. Moreover, in the model proposed by Balcik and Beamon (2008) demand can only arise at one point in the network.

In order to develop a more realistic model, Rawls and Turnquist (2010) present a two stage MIP taking uncertainties concerning demand, network disruptions and inventory damages into account. First stage decisions involve locating relief facilities and storing emergency supplies before uncertainties are realized. After a disaster has occurred, the conditions of transportation links and storage facilities are known and second stage decisions are made, i.e. how much of what type of relief item has to be transported to which demand point. Besides common costs associated with acquisition and transportation, Rawls and Turnquist (2010) consider penalty costs for unsatisfied demand and unused commodities. Their objective is to minimize expected total costs subject to capacity restrictions of transportation links and facilities. This minimum is determined over all scenarios. After the publication by Rawls and Turnquist (2010) researchers have realized the importance of prepositioning decisions in order to enhance emergency preparedness. Their approaches differ either in the way of dealing with uncertainties as stated in Sect. 2.1 and/or in the choice of objectives and restrictions, e.g. minimization of response time (Duran et al. 2011; Bozkurt and Duran 2012), minimization of casualties (Salmerón and Apte 2010) and minimization of warehouse operating costs, unsatisfied demand and transportation time (Mete and Zabinsky 2010), minimization of different categories of costs (Döyen et al. 2012; Paul and Hariharan 2012) and/or in providing extensions. For example, Rawls and Turnquist (2011) extend their previous prepositioning model by introducing service quality requirements such that demand is fulfilled with a predefined probability. Noyan (2012) augments the model of Rawls and Turnquist (2010) by adding the risk averse measure "conditional Value at Risk" (cVaR) to their objective function such that worst case scenarios are weighted higher in the decision-making process. Other extensions are presented by Rawls and Turnquist (2012) where multi-period allocation of emergency items within the first 72 h following a disaster is planned in advance. The dynamic character is also adopted by Klibi et al. (2013), taking multi disasters over several periods into account. In their publication, disasters may hit vendor`s inventory, apart from possible facility damages, such that the total available quantities of items may be reduced significantly.

2.2 *Scenario-Free Approaches*

In the contributions reviewed in this section it is argued that approaches based on scenarios limit the set of future realizations. If historical data are not readily available to formulate a sufficient number of scenarios that adequately represent possible outcomes, alternative approaches have to be found. Table 3 summarizes objectives, constraints and assumptions, in addition to those already given in Table 1, of the relevant scenario-free approaches.

The first who have refrained from using scenarios in the context of aid prepositioning were Campbell and Jones (2011) who assume normally distributed demand. Their objective is to decide where to open facilities and how much of certain relief items to store such that total costs, including e.g. restocking and disposal costs, are minimized. In their model demand only occurs at one specified

Table 3 Scenario-free approaches: objectives, constraints and additional assumptions

Author(s)	Year	Objective	Constraints	Additional assumptions
Campbell and Jones	2011	Min costs, e.g. for	Solely demand fulfillment	Single demand point
Murali et al	2012	Max demand covered	Facility capacities	Optimal assignment of demand to facilities
Galindo and Batta	2013	Min costs	Budget; facility capacities	Short-term planning (2 days before disaster occurence)
Renkli and Duran	2014	Min weighted distance	Maximal service distance of relief items	
Verma and Gaukler	2014	Min transportation costs	Facility capacities	

location, given that supply facilities and hence inventories can be destroyed by a disaster with known and independent probabilities. In the situation studied by Galindo and Batta (2013), emergency items are stored two days before disaster occurrence using expected values of demand based on available disaster forecasts. However, long-term planning requires more advanced approaches to model uncertainty. Recently, Verma and Gaukler (2014) introduce a random variable which is distance-dependent and represents the fraction of total facility capacity which is no longer available in the aftermath of a disaster. Here, the destruction level of facilities decreases with increasing distance to the disaster location. However, transportation disruptions are not considered in their model.

Uncertainty can also be captured by the so-called chance-constrained method as in Murali et al. (2012). Their optimization model is a variant of the maximal covering location problem, in which the amount of satisfied demand depends on the distance to a service facility. Here, constraints dealing with the assignment of relief items to demand points should hold with a predefined probability; these are chance-constraints. Due to the fact that demand is non-negative Murali et al. (2012), unlike Campbell and Jones (2011), assume log-normally distributed demand. A second paper using probability constraints within the disaster preparedness phase has been published by Renkli and Duran (2014). Deliveries can only take place if the distance is not too large and if the corresponding transportation link is not damaged.

The main disadvantage of continuous probability distribution approaches is that their applicability is limited; see e.g. the model of Campbell and Jones (2011) where only one demand point is served. In order to cope with multi-facility and multi-demand networks a common methodology is to define a finite number of scenarios as presented in the previous section.

2.3 Future Research

Sections 2.1 and 2.2 reveal that only the minority of contributions uses alternatives to scenario-based approaches to capture the uncertainty related to disastrous events. It is worth investigating more advanced methods in stochastic programming to overcome the disadvantages caused by the use of the presented approaches, e.g. a limited set of outcomes in case of scenario-based approaches and high complexity due to continuous probability distributions in scenario-free approaches. Dynamic models as well as relief chains with several stages and actors could help to gain further insights into more realistic settings. Within the framework of dynamic models, the transfer of relief items between storage facilities can overcome possible capacity shortages and is a potential future research direction, as identified by Caunhye et al. (2012).

As shown in Tables 2 and 3, minimization of costs is a frequently used objective, which is reasonable since donations are generally given only after the occurrence of a disaster; therefore, cost aspects play an important role in the preparedness phase. Moreover, maximization of covered demand is a crucial objective within disaster management and many authors, e.g. Rawls and Turnquist (2010), use penalty costs for unsatisfied demand to incorporate this objective. Other multi-objective approaches may also be valuable for a holistic view of disaster relief problems, see e.g. Rottkemper and Fischer (2013).

3 Solution Methods

The literature described above proposes different modeling approaches to address prepositioning and location decision problems under uncertainties within disaster management. However, none of them focus on the development of efficient solution methods. As stated in Table 1, either standard solvers like CPLEX are used for relatively small problems, or known heuristics are adapted to the specific model structure to tackle larger problems. The main reason for using a commercial solver is that the focus is rather on the model presentation than on solution techniques. Only seven out of 17 articles reviewed in this paper suggest a specific solution method, but still with a clear emphasis on the model formulation. The different heuristics and decomposition methods used in these papers are presented in the following.

3.1 Heuristics

The model presented by Campbell and Jones (2011), see Table 3, can be expanded to the case with multiple demand and facility locations. They solve the problem where every demand point is serviced by only one facility by adapting a heuristic, originally invented for solving p-median problems, proposed by Cooper (1964). Here, every supply node has to be allocated to an arbitrary demand node at the beginning of the solution process; otherwise this heuristic is not able to solve the underlying problem.

In order to obtain a lower bound for the minimization problem, Döyen et al. (2012) use the Langrangean relaxation, i.e. specific constraints, e.g. the flow balance equations, associated with the corresponding Lagrange multipliers, are integrated into the objective function. The multipliers are updated in every iteration by the subgradient method such that the lower bound is successively improved. In combination with a local search algorithm this Langrangean heuristic determines solutions for problems where up to 25 scenarios are considered in Döyen et al. (2012).

Murali et al. (2012) replace the uncertain demand with the corresponding quantile of the log-normal distribution such that a deterministic mixed integer programming model is obtained. They adapt the locate-allocate heuristic (e.g. Taillard 2003) to their problem formulation, since in Jia et al. (2007) this heuristic outperforms a genetic algorithm as well as the Langrangean relaxation concerning computing time for the uncapacitated maximal covering location problem. However, the study of Jia et al. (2007) reveals high sensitivity to the initial solution since convergence is slow and solutions far away from optimality might be obtained in case of an unfavorable starting solution.

Reduction approaches as proposed in Galindo and Batta (2013) aim to delete unnecessary constraints resulting in less computation time. For example, if the capacities of supply points are sufficiently high and the probability of their destruction is relatively low, the constraint that restricts the expected amount of damaged items can be omitted. Besides removing variables and constraints, they aggregate several demand points such that problem size is decreased further. The resulting problem with a smaller number of constraints, variables and demand points is solved subject to a limitation of the error occurrence due to aggregation.

All of the above mentioned papers select relatively small problem sizes to compare the performance of their proposed heuristic with the solution computed by a commercial solver like CPLEX.

3.2 Decomposition Methods

In general, decomposition techniques exploit the specific structure of a problem in such a way that smaller sub-problems have to be solved in order to find the exact solution of the original problem. This is of special interest for large scale two-stage stochastic programs where the first stage involves decisions before and the second stage considers decisions after the realization of uncertainty. In particular, for models with a two-stage mixed integer programming formulation, (see Table 1), the L-shaped method is applicable which yields outer approximations, so called optimality cuts, of the second-stage function (Van Slyke and Wets 1969). These optimality cuts are inserted into the first stage problem, the so-called master problem, which has to be solved to obtain the corresponding values. These first stage variables are taken to solve the second stage sub-problems which in turn are used to update the optimality cuts in each iteration. When the master problem and the corresponding sub-problems are solved to optimality, the L-shaped algorithm finds the exact solution of the original problem. However, only three papers within this review adapt this decomposition technique to their underlying model.

Rawls and Turnquist (2010) solve their scenario-based problem by means of the integer L-shaped method in combination with the Lagrangian relaxation, i.e. the Lagrangian L-shaped method (LLSM). Since the first stage problem consists of binary variables concerning the location decisions, the integer version of the L-shaped method, proposed by Laporte and Louveaux (1993), is suitable. Rawls and Turnquist (2010) assume unlimited link capacities such that the multi-item network can be reduced to problems with a single-item each. Based on this assumption the second-stage problems, i.e. distribution of relief items to demand points using specific links, are simplified. It should be noted that the proposed solution strategy LLSM is not straightforwardly applicable for a generalized network where link capacity constraints and multi-item flows are considered. Due to the application of the Langrangean relaxation heuristic, the optimality of the solution cannot be guaranteed. In their study, Rawls and Turnquist (2010) demonstrate the beneficial feature of their heuristic by comparing solutions found by LLSM and the standard solver CPLEX. For relatively small problem instances LLSM finds a solution which is near the optimal value found by CPLEX, but with less computation time. However, it has not been shown that the algorithm proposed by Rawls and Turnquist (2010) yields near optimal solutions also for large scale problems.

Noyan (2012) compares two versions of the L-shaped method for the two-stage model of Rawls and Turquist (2010) extended by the conditional VaR. The first variant of the L-shaped method adds variables to the second-stage problem whereas the second L-shaped method makes use of the subgradient algorithm. Then Noyan (2012) augments these single-cut approaches to multi-cut algorithms. Numerical experiments are performed for the same case study for hurricane occurrence in the southeastern US as presented in Rawls and Turnquist (2010), i.e. 51 scenarios each with 30 potential facility locations, 58 transportation links and three types of relief items. Noyan (2012) uses the commercial software CPLEX to solve the master problem, i.e. the first stage problem plus the optimality cuts, and the second-stage sub-problems. Due to the existence of binary variables within the first stage, solving the master problem to optimality is computationally expensive. Hence, Noyan (2012) terminates the computations of CPLEX after a maximal time period of 2 h. His results show that the performance of the algorithms, in terms of optimality gaps, is highly problem-dependent and hence not reliable.

As mentioned before, solving the master problem at each iteration is time consuming and prohibitive especially for large scale problems. As a remedy, Verma and Gaukler (2014) employ a greedy heuristic in combination with local search which determines a solution of the master problem until this solution does not change between two iterations. Only in this case the master problem is solved exactly and provides a lower bound for the original minimization problem. This modified L-shaped procedure is repeated until the optimal objective value of the overall problem is obtained. A case study for earthquakes in California is presented by Verma and Gaukler (2014) containing 20 demand points, 58 potential facility locations of three different sizes and up to three facilities to be opened. Note that despite this relatively small problem size, the corresponding CPU times exceed several hours, i.e. the procedure is very time-consuming. While this is not problematic in purely strategic planning, computation time can be crucial for models with a short-term time horizon. For instance, in Galindo and Batta (2013) decisions are made two days before the landfall of a hurricane and hence results are required within a rather limited time-frame. Indeed, location decisions within disaster management often have a provisional character since in many cases existing facilities, like schools or other institutions, are used to store relief items temporarily. In contrast, location planning within a commercial supply chain is concerned with facilities which should retain their functionality over a long period of time.

3.3 Future Research

Research gaps concerning the use of adequate solution methods are much more obvious than those identified within modeling frameworks. Several drawbacks of the algorithms, as stated in Sects. 3.1 and 3.2, make it necessary to devote more attention to the development of efficient solution methods in the future.

Approaches based on the use of scenarios often limit the computational effort by limiting the number of possible future developments. However, taking only a subset into account bears the risk of excluding the relevant disaster scenarios and hence being ill-prepared for the actual development. In order to solve more realistic problems, i.e. with a large number of possible scenarios, facility locations, different types of relief items etc., large scale models are required for which exact methods are computationally prohibitive. Heuristics are a reasonable alternative for researchers who prefer fast solutions at relatively low computational costs. However, heuristics may be highly problem-dependent, i.e. applying the same heuristic to another large problem instance does not necessarily result in a near-optimal solution again. Consequently, efficient techniques generating benchmark solutions are useful to assess the quality of heuristics for large scale problems. One possibility is to use an iterative algorithm, e.g. the Krylov subspace method (Saad 2003), which can be terminated at any point if a prescribed accuracy level is reached. Iterative solvers provide approximate solutions with measurable distance to optimality but with less storage requirements than an exact method; hence they have significant advantages over the methods proposed in the literature.

4 Conclusions

This paper reviews the state-of-the-art research relating to preparedness enhancement in disaster management. Thanks to the work of Rawls and Turnquist (2010), there has been a shift from pure facility location models to a combination of prepositioning issues and facility location decisions. However, research in this field is still in its infancy.

Although recent review papers, e.g. Caunhye et al. (2012), mention the need of efficient solution methods, the focus of the relevant literature is mostly on modeling aspects. To the few authors who propose, in addition to a model, a solution algorithm, heuristics are the preferred method of solving the underlying problem. The reasons are twofold: First, solutions for large scale problems may be determined within a reasonable timeframe whereas exact methods take too long. Second, the concept of a heuristic is often relatively simple and intuitive, e.g. the greedy heuristic and genetic algorithms. However, heuristics possess several disadvantages putting their application into question: If the heuristic fails to find a solution, it cannot be determined if this is due to the applied heuristic procedure or if a solution does not exist for the underlying problem. It is also not possible to verify the optimality of the solution in general, i.e. it is well-known that heuristics do not necessarily yield optimal or near-optimal solutions and might lead to a good solution in one case and to a bad one in another. Therefore, the most serious drawback when applying heuristics is the inability to measure the deviation from the optimum. But especially in the context of pre-disaster activities where strategic decisions may be crucial for the success of relief operations, evaluating the quality of the solution is essential. Non-optimal solutions may produce unfavorable positions of relief locations and corresponding inventory levels resulting in high financial burden, unsatisfied demand and, more importantly, in fatalities. Such consequences are even intensified when the available data is scarce and of poor quality. Hence, the development of exact solution procedures is an important area for future research and iterative methods like Krylov subspace algorithms are promising alternatives.

References

Balcik B, Beamon BM (2008) Facility location in humanitarian relief. Int J Logistics Res Appl: A Lead J Supply Chain Manag 11:101–121

Bozkurt M, Duran S (2012) Effects of natural disaster trends: a case study for expanding the pre-positioning network of care international. Int J Environ Res Publ Health 9:2863–2874

Campbell AM, Jones PC (2011) Prepositioning supplies in preparation for disasters. Eur J Oper Res 209:156–165

Caunhye AM, Nie X, Pokharel S (2012) Optimization models in emergency logistics: a literature review. Socio-Econ Plann Sci 46:4–13

Cooper L (1964) Heuristic methods for location-allocation problems. SIAM Rev 6:37–53

Döyen A, Aras N, Barbarosoğlu G (2012) A two-echelon stochastic facility location model for humanitarian relief logistics. Optim Lett 6:1123–1145
Duran S, Gutierrez MA, Keskinocak P (2011) Pre-positioning of emergency items for CARE international. Interfaces 41:223–237
Galindo G, Batta R (2013) Prepositioning of supplies in preparation for a hurricane under potential destruction of prepositioned supplies. Socio-Econ Plann Sci 47:20–37
Guha-Sapir D, Vos F, Below R (2012) Annual disaster statistical review 2011: the numbers and trends. Center for Research on the Epidemiology of Disasters, Brussels
IFRC (2010) World disasters reports 2010: focus on urban risk
Jia H, Ordónez F, Dessouky MM (2007) Solution approaches for facility location of medical supplies for large-scale emergencies. Comput Ind Eng 52:257–276
Klibi W, Ichoua S, Martel A (2013) Prepositioning emergency supplies to support disaster relief: a stochastic programming approach. Technical Publication CIRRELT-2013-19. Université de Montréal. Montréal. Université Laval. Québec
Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. Oper Res Lett 13:133–142
Mete HO, Zabinsky ZB (2010) Stochastic optimization of medical supply location and distribution in disaster management. Int J Prod Econ 126:76–84
Murali P, Ordóñez F, Dessouky MM (2012) Facility location under demand uncertainty: Response to a large-scale bio-terror attack. Socio-Econ Plann Sci 46:78–87
Noyan N (2012) Risk-averse two-stage stochastic programming with an application to disaster management. Comput Oper Res 39:541–559
Paul JA, Hariharan G (2012) Location-allocation planning of stockpiles for effective disaster mitigation. Ann Oper Res 196:469–490
Rawls CG, Turnquist MA (2010) Pre-positioning of emergency supplies for disaster response. Trans Res Part B: Methodol 44:521–534
Rawls CG, Turnquist MA (2011) Pre-positioning planning for emergency response with service quality constraints. OR Spectrum 33:481–498
Rawls CG, Turnquist MA (2012) Pre-positioning and dynamic delivery planning for short-term response following a natural disaster. Socio-Econ Plann Sci 46:46–54
Renkli, Ç, Duran S (2014) Pre-positioning disaster response facilities and relief items. Hum Ecol Risk Assess: An Int J (to appear)
Rottkemper B, Fischer K (2013) Decision making in humanitarian logistics–A multi-objective optimization model for relocating relief goods during disaster recovery operations. Proceedings of the 10th international conference on information systems for crisis response management, Germany, pp 647–657
Saad Y., 2003: Iterative methods for sparse linear systems. *Siam*, Philadelphia, second edition
Salmerón J, Apte A (2010) Stochastic Optimization for Natural Disaster Asset Prepositioning. Prod Oper Manag 19:561–574
Taillard ED (2003) Heuristic methods for large centroid clustering problems. J Heuristics 9:51–73
Van Slyke RMV, Wets R (1969) L-shaped linear programs with applications to optimal control and stochastic programming. J Appl Math 17:638–663
Verma A, Gaukler GM (2014) Pre-positioning disaster response facilities at safe locations: An evaluation of deterministic and stochastic modeling approaches. Comput Oper Res (to appear)

Forecasting Misused E-Commerce Consumer Returns

Stefan Drechsler and Rainer Lasch

Abstract High return rates are a problem particularly concerning e-tailers. Often, these returns are considered to come back as good as new which enables an immediate resale. This assumption does not take into account that some consumers show a dishonest buying intention by returning items they have already used. For reasons of an improved operational planning it could be helpful to predict this kind of product returns. Therefore, we adapt and enhance already existing forecasting approaches to situations where the e-tailer matches consumer returns with their date of sale and furthermore, where the consumer has to register the return first. The so generated product-individual data are covered by developing three forecasting methods. It is examined under which circumstances the models show a satisfying accuracy and correspondent requirements that have to be met in this context are inferred. The overall premise is the practicability in order to ensure an uncomplicated application.

1 Introduction

Due to reasons like lenient return policies (Hsiao and Chen 2012), return rates for goods sold on the internet are correspondingly high (Mostard and Teunter 2006). Fashion products in particular have to record the highest shares whereby a study from Pur et al. (2013) revealed that over 50 % of e-tailers belonging to the fashion industry reported rates of over 25 %. A great challenge lies in the efficient and effective handling and integration of returns into forward logistics processes (Brandl 2014). Literature particularly concentrating on consumer returns originally purchased online is scarce and frequently deals with product- or process-related influencing factors (e.g. Rao et al. 2014). Furthermore, there is the problem of already used and still returned products which causes low-quality consumer returns

S. Drechsler (✉) · R. Lasch
TU Dresden, Münchner Platz 1/3, 01062 Dresden, Germany
e-mail: stefan.drechsler@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_15

(Asdecker and Weigel 2013). Forecasting in this context could be very helpful to support and simplify planning actions, for example the refurbishment of misused items as well as optimizing inventory and replenishment (Mostard et al. 2005). Besides, the gathering of respective return data could be used to set adequate return policies (Hess and Meyhew 1997). Even if in the last years, the amount of approaches focusing on return forecasting increased, so far there are no elaborations dealing with the problem of misused items by exclusively concentrating on observing return rates. It is an interesting question to which extent the accuracy of forecasting procedures is raised if the information level increases, too.

The remainder of this paper is structured as follows: The adjoining "Considering Small and Medium-Sized Suppliers in Public Procurement—The Case of the German Defence Sector" deals with literature reviews on return quality concerns and already existing forecasting methods. Here, we only consider those papers dealing with return forecasting as such. In "Integration of Cultural Aspects in Supplier Evaluation" selected models are adapted to a utilization of product-individual data representing a situation where a return can be matched directly with the date of sale. To preserve practicability, a high (programming) effort should be avoided which encouraged the application of techniques based on the returns' time lag. By means of simulated data it is shown that forecasting results do not need to be satisfying as long as return rates fluctuate distinctly. Beyond, an advancement particularly interesting for seasonal occurring misused returns is introduced by considering a preceding return merchandize authorization. A conclusion, limitations, and future research possibilities are presented in "RoRA-SCM—Review of the Integration of Flexibility Planning in Supply Chains".

2 Literature

2.1 Quality Concerns Regarding Consumer Returns

Quality concerns are more often brought into connection with end-of-life returns, because most of these products have been used already (e.g. Jin et al. 2013). Consumer returns by contrast often are regarded as reusable from the outset which is why this uncertainty is mostly excluded from respective expositions (e.g. van Nunen and Zuidwijk 2004). However, return quality problems are not unknown for consumer returns, too. Besides the issue that goods are already delivered or bought in a bad condition (Hsiao and Chen 2012) there is the problem of illegal item's consumption which generates fraudulent respectively misused returns (Harris 2010). Rosenbaum and Kuntze (2005) linked this behavior to a compulsion to buy in order to derive prestige. Schmidt et al. (1999) carried out a survey among British consumers discovering that men rather concentrate their fraudulent behavior on tools or CDs, whereas women focus on fashion products which Speights and Hilinski (2005) call wardrobing. Asdecker and Weigel (2013) stated that seasonal products, for instance carnival costumes or flat screens returned after certain incidents like the

FIFA World Cup, are particularly affected by misuse. Furthermore, fraudulent returns show an average loss in value of about 37 % and cause increasing process costs. The authors emphasized the dishonest buying intention which they especially exposed as a problem of the German online-trading. Pur et al. (2013) confirmed this assertion in their study by revealing that a planned misuse is more often seen as a frequent return cause among e-tailers, particularly within the clothing industry. About 12 % of fashion returns cannot be sold as new goods anymore. This is less than in other sectors, but compared to other products more items (22 %) have to be put under more or less comprehensive refurbishment. The focus of this paper should lie solely on fraudulent returns. Clothing serves as base for the simulated case examples in Sect. 3. Nevertheless, it should be noted that all explained methods are generic and can be easily applied to other product and quality categories, too.

2.2 Literature on Forecasting Product Returns

So far, there are no approaches concentrating on forecasting abusive consumer returns, as they mostly focus on the prediction of end-of-life returns. Some of them are working with periodical data, so that the forecasts are solely based on historical information about the total number of sales and returns per period and not on observed rates. Goh and Varaprasad (1986) used a transfer function model, whereas Kelle and Silver (1989) introduced 4 forecasting methods based on varied types of data, but assuming known distributions and parameters for their recommended probability distributions. Toktay et al. (2000) and Clottey et al. (2012) utilized distributed lag models with discrete respectively continuous distributions to represent return rates. Krapp et al. (2013a, b) stated that the infinite planning horizon and the normally distributed error term assumed by Toktay et al. (2000) could be a problem during the parameter estimation. They create a more generic model without these limiting assumptions. Still they worked with predetermined probability distributions and applied Bayesian techniques for parameter estimation which cannot be regarded as simple practical procedure. Product-individual data find integration to different extent. The expectation minimization algorithm applied by Toktay et al. (2000) is based on right-censored information and requires a certain distribution assumption, too. The application of fuzzy techniques and soft-computing represents another possibility (Marx-Gómez et al. 2002) which enables the integration of various factors that impact the number of returns (Temur et al. 2014). These models generally serve for the creation of certain networks to predict the number of scrapped product returns for the purpose of reuse and recycling (Hanafi et al. 2008; Agrawal et al. 2014). The approaches are not discussed in detail here, because their adaption to a consumer return network of an e-tailer as well as the adjoining forecasting imply a high programming and simulation effort. This contradicts against the required practicability.

Looking at consumer returns, the often required distribution assumptions that should describe return rates are rather difficult to make, especially for certain categories like misused returns. If such an assumption is required, it should be combined with corresponding researches concerning the customer's post-purchase behavior. Hess and Mayhew (1997) dealt with this topic by using an explorative split adjusted hazard rate model adapted from comparatively extensive tracking data. Potdar and Rogers (2012) addressed the problem of forecasting categorized product returns since they stated that distributions of the products' time lag depend on the reason why the items are returned. The authors used moving averages to estimate future return rates and also introduced a data envelopment analysis based on competitive products. The authors desisted from forecasting total return amounts which makes it difficult to assess the quality of their proposed methods. Summarizing it can be said that there are no approaches directly usable for the prediction of fraudulent returns by similarly dealing with estimated return rates acquired with the help of product-individual information.

3 Forecasting Approaches

All following approaches concentrate on forecasting the amount of fraudulent returns expected in a period τ, represented by the variable $\mu_{f,\tau}$. The two approaches explained in Sect. 3.1 refer to simple product-individual information where a return is matched with its date of sale. The so observed return rates afterwards are used to estimate rates for the following periods respectively sales lots in order to determine the forecasts $\mu_{f,\tau}$. An option to advance the information level is the implementation of an effective return merchandize authorization which means the consumer needs to receive an allowance before the parcel is brought on its way (Norek 2002). Consequently, the overall quantity of returns is known in advance. An example could be a printable return label as it is done by several e-tailers. Utilizing the information stemming from this registration leads to an enhancement introduced in Sect. 3.2. By means of simulated data and exemplarily conducted forecasts it should be clarified if and under which circumstances the approaches are suitable for return forecasting.

3.1 Forecasting Fraudulent Returns Based on Sales Lots

As explained above, the estimated return rates required for the calculation of $\mu_{f,\tau}$ are gained with the help of observations stemming from previous periods and applying suitable forecasting approaches to the time series of observed return rates. Besides these estimations, the fraudulent return forecasts are also based on the (known or predicted) sales lots v sold in periods $\tau - k$. A weekly time basis should be considered. The time lag k denotes the respective week after the week of sale with the maximum z (Krapp et al. 2013a). The earliest period for an item to return should be one period after the week of sale which applies $k \in \{1, 2,\ldots, z\}$. This declaration is no obstacle for stating a periodical basis where sales could return within the same period, too. Both elements together lead to the premise that each estimated rate should be clearly allocated to a certain sales lot $v_{\tau - k}$ for all k.

Two elaboration options are distinguished. Method 1 contains the observation of simple fraudulent return rates s_f whereby it is recorded which share of v returned in each following week k as misused. Thus, it is possible to determine the estimated return rates $\tilde{s}_f$ for each k and sales lot. This method is motivated by one of the approaches introduced by Potdar and Rogers (2012) for the first two of their assumed reason codes. It can be expressed by:

$$\mu_{f,\tau} = \sum_{k=1}^{z} v_{\tau-k} \tilde{s}_{f,\tau-k}^{(k)}. \tag{1}$$

The so observed and estimated time-lag-specific return rates are—depending on the return's extent—likely to be quite low and therefore hard to interpret. This could be particularly disadvantageous if return rates should be analyzed in order to for instance adapt return policies (Davis et al. 1998). That is why the additional scaling of return rates also used by Toktay et al. (2000) and Clottey et al. (2012) might be useful. In doing so, the simple return rate s_f is split into several components. First of all, there is the share of v that actually comes back within z periods, expressed by p_R. The variable q_f indicates the proportion of the actually returned products that come back as misused items. Finally, the time-lag-specific ratios a_f for each k refer to the share of misused products returned after k periods. All these time series of observed return rates are used again to determine the according estimations. As it is the straight implication that for each sales lot the time-lag-specific actual return rates a_f sum up to one over all weeks k, their estimations should be adapted afterwards, too, as they also pertain for a whole sales lot. Subsequently, method 2 can be expressed by:

$$\mu_{f,\tau} = \sum_{k=1}^{z} v_{\tau-k} \tilde{p}_{R,\tau-k} \tilde{q}_{f,\tau-k} \tilde{a}_{f,\tau-k}^{(k)}. \tag{2}$$

In order to evaluate which of these two methods shows the better forecasting results, both were applied exemplarily on simulated data, like it was done for instance by Krapp et al. (2013a, b). As already mentioned above, an e-tailer for fashion products with a quite high average overall return rate p_R serves as foundation. With respect to the survey referred to in Sect. 2.1 the average of q_f is set to approximately 15 %. This includes the cases where the product cannot be sold as new and by supposing that some overused items regain their as good as new condition after a more comprehensive refurbishment. The maximum time lag z should be 4 periods with the data basis for a_f stemming from the assumption that customers plan a misuse and therefore may use the product a certain period of time before they return it (Asdecker and Weigel 2013). This circumstance coincides with the third reason code described by Potdar and Rogers (2012) representing a more or less comprehensive test of the good. For simulation purposes each fraudulent return

Table 1 Mean return rates and standard deviations for simulation

	p_R	q_f	$a_f^{(1)}$	$a_f^{(2)}$	$a_f^{(3)}$	$a_f^{(4)}$
μ	0.35	0.15	0.2	0.25	0.3	0.25
σ_A	0.01	0.01	0.01	0.01	0.01	0.01
σ_B	0.02	0.02	0.02	0.02	0.02	0.02
σ_C	0.04	0.04	0.02	0.02	0.02	0.02
σ_D	0.07	0.05	0.04	0.04	0.04	0.04

rate should be normally distributed with different standard deviations to create noise (Krapp et al. 2013b) and as consequence generating different scopes of variation. The time-lag-specific return rates a_f were adjusted afterwards to ensure a sum of 1 over k for each sales lot. Thus, the four scenarios A, B, C and D—shown in Table 1—were produced. It should be noted that the comparably low standard deviations were approved to depict reality by practitioners.

For method 1 all four time series of estimated return rates were gained by utilizing a 4-period moving average (MA 4) recommended by Potdar and Rogers (2012). For method 2 the shares were estimated additionally with a 10-period moving average (MA 10) and exponential smoothing (ES) which are a natural choice if values are fluctuating around a rather constant level (Brown 1963). The smoothing factor required for ES was dynamically adjusted after each new observation by minimizing the mean absolute deviation (MAD), Theil's inequality coefficient (TIC) (Theil 1971) and mean absolute percentage error (MAPE). The simulation comprised 60 periods with the sales lots derived from random, uniformly distributed numbers. The two assumed intervals of 15,000–20,000 as well as 20,000–40,000 sold items should represent a different degree of sales tips and lows. With a consistent forerun of 14 periods due to MA 10 the amounts $\mu_{f,\tau}$ in 46 periods were calculated and afterwards compared to the simulated actual return numbers. The so obtained MAD, TIC, and MAPE are displayed in Table 2. The respective first values originate in the first mentioned, rather narrow interval of sales whereas the parenthesized measures stem from the second interval. It is apparent that only the forecasts in situation A can be seen as consistently satisfying. Values for MAD and MAPE clearly imply a decreasing accuracy with increasing variation of return rates and sales, except for situation C where the MAPE improves with increasing fluctuations in sale. The variation of quotas has the greater impact, because the amount of returns increase together with the sales, thus a higher MAD can be seen as consequential. Values for the TIC improve for both methods, when there are higher sales tips and lows, because the naïve forecasts lose accuracy in that case.

Especially under situations C and D both forecasting procedures cannot be described as valuable anymore, as the errors are too high to base a profound planning on the forecasts. Despite the generally unsatisfying outcomes, it is recognizable that method 1 is slightly superior to method 2 by showing better values for MAD, MAPE, and TIC in nearly all situations which is why there is a mild tendency to this procedure. In case that the additional scaling shall still be used because of the already mentioned disadvantages of simple return rates, MA 4

Table 2 Error measures under different scopes of return rate variation

		Method 1	Method 2		
			MA 4	MA 10	ES
MAD	Sit. A	26.8530 (46.0304)	31.1717 (57.3735)	29.7420 (53.9067)	28.4433 (52.0289)
	Sit. B	63.9643 (112.3630)	73.2578 (125.6450)	76.8863 (133.3300)	71.3457 (124.3383)
	Sit. C	106.6850 (184.3028)	112.9593 (187.8767)	109.6187 (172.5041)	107.3272 (167.5137)
	Sit. D	139.7617 (263.3654)	186.4785 (331.2141)	136.8487 (256.4215)	152.2957 (276.1172)
MAPE	Sit. A	0.0299 (0.0296)	0.0349 (0.0370)	0.0334 (0.0351)	0.0320 (0.0336)
	Sit. B	0.0707 (0.0719)	0.0818 (0.0818)	0.0850 (0.0860)	0.0793 (0.0803)
	Sit. C	0.1443 (0.1412)	0.1527 (0.1466)	0.1514 (0.1395)	0.1518 (0.1394)
	Sit. D	0.1674 (0.1839)	0.2237 (0.2297)	0.1657 (0.1784)	0.1797 (0.1883)
TIC	Sit. A	0.6818 (0.4256)	0.7735 (0.5013)	0.7600 (0.4860)	0.7428 (0.4763)
	Sit. B	1.1618 (0.9497)	1.5317 (1.2293)	1.4708 (1.1874)	1.3882 (1.1131)
	Sit. C	1.2020 (1.0533)	1.1978 (0.9798)	1.1902 (0.9607)	1.1966 (0.9467)
	Sit. D	1.0168 (0.9215)	1.1977 (1.0664)	0.9391 (0.8776)	1.0181 (0.9246)

should not be applied. As the results for MA 10 and ES are close, none of them can be ruled out. More importantly, the examples clearly show the crucial fact that even if estimated return rates are not based on distribution assumptions, forecasts do not have to be helpful under any circumstances. Consequently, the return forecasting based on time lag deliberations cannot be regarded as robust. Our results imply that high deviations within actual return rates should be avoided, because forecasting might miss the target if the sales highly fluctuate, too. However, in case the company has to face significant increases or downturns and nevertheless records low fluctuating return rates the introduced methods yet could be of avail. Here, a reasonable summary of product types is inevitable. Another option could be a more appropriate return categorization, for example on the base of certain quality levels instead of only considering misuse. Since our simulations were merely exemplary more comprehensive tests should be performed.

3.2 Forecasting Fraudulent Returns Based on the Return Period

For the purpose of accurate forecasts an integration of additional accessible information should be contemplated. Because of their inclusion of sales lots the so far explained methods ignore the effect of returns that are dependent on the specific period of return. As already explained there are certain events encouraging the emergence of misuse. These might be more numerous in periods after the incidents, so it could be helpful to take this fact into account, too. The essential premise lies in the preceding return registration by the consumer, thus the e-tailer knows about it in advance. The first step is the determination of d, which defines the time interval between the return's registration and arrival, in order to predict the period $\hat{\tau}$ of return. With $\tilde{\tau}$ denoting the period of approval it applies:

$$\hat{\tau} = \tilde{\tau} + d. \tag{3}$$

For the calculation of d some auxiliary variables are necessary. First, there is the weekday of the return's registration t within the respective week $\tilde{\tau}$ whereby only workdays should be considered by treating all Sundays as Mondays. Every day is assigned to a certain number. Accordingly, the variable $t \in \{1, 2,\ldots, w)$ is introduced by matching Monday with 1, Tuesday with 2 etc. This implies $w = 6$ if only Sundays are not recognized. Variable $c \in \{1, 2,\ldots\}$ should denote the number of working days between applying the return and actually bringing it on the way which could ideally depend on relevant influencing factors. Thinkable are, for instance, the customer's past behavior, the weekday of approval, or simply the number of granted periods already passed. The delivery time $l \in \{1, 2,\ldots\}$ represents the number of working days a parcel delivery provider needs to submit the good. Thus, the variable m is calculated by

$$m = t + c + l. \tag{4}$$

This enables the determination of the actual delay d given by Eq. (5), assuring that d appears as non-negative integer:

$$d = \begin{cases} \lceil m/w - 1 \rceil & \forall m \geq w \\ \lfloor m/w \rfloor & \forall m < w. \end{cases} \tag{5}$$

At the same time, we suppose the customers need to add a cause respectively reason code for their return (Potdar and Rogers 2012; Krapp et al. 2013a). Consequently, there have to be wrongly stated reasons, because no one will admit to send back a misused item (Asdecker and Weigel 2013). Since not each reason code j may be relevant in this context, the set J only includes all j that were ever wrongly stated. The variable $\mu_{j,\tau}$ should denote the estimated number of products with reason code j assumed to return in τ determined based on the number of approvals and Eqs. (3)–(5) for each of this potential return. Next, the number of fraudulent returns is forecasted via estimates for the return rate q_f again. The estimation for q_f could be specified either for each $j \in J$ or treated as general ratio like it is done in Eq. (6):

$$\mu_{f,\tau} = \tilde{q}_{f,\tau} \sum_{j \in J} \mu_{j,\tau}. \tag{6}$$

The following exemplary simulation illustrates the forecasting quality of the method. A German e-tailer primarily known for carnival articles is expected. The e-tailer records high fraudulent return rates particularly in February with the highest shares after Carnival Monday (Paletta 2014). The relevant time frame consists of seven periods starting two weeks before Carnival Monday with respect to earlier events before that day. The amount of sold products and returns slightly increases over the years (IW Köln 2010; Dpa-Afx 2012). The e-tailer summarizes carnival products for joint forecasting. The registered returns—already summed and filtered by relevant reason codes—were simulated using uniformly distributed random numbers in increasing intervals. Each year, seven weeks are regarded as particularly affected. In order to simulate clear seasonal patterns the first seven values of a discrete negative binomial distribution with the parameters n = 8 and p = 0.7 served as background for the observable return rates. These probabilities were rounded and an adaption of the third and fourth value was conducted to create a clear peak. This leads to the vector $\boldsymbol{q_f}$*: = (0.06; 0.14; 0.17; 0.2; 0.15; 0.1; 0.07). The quotas $\boldsymbol{q_f}$* were randomized assuming a normal distribution with a consistent standard deviation of $\sigma = 0.01$ and afterwards they were arranged in ascending order to generate a growing seriousness of the problem. The return rate estimation was carried out by means of the additive Holt-Winters-approach (HWA) (Holt 1957; Winters 1960) with a forerun of two years respectively 14 periods. Beyond, two scenarios are distinguished. In the first, we presumed a fixed value for c and l. With $w = 6$, $c = 4$ for $t \in \{1,\ldots, 4\}$, $c = 5$ for $t \in \{5, 6\}$ and $l = 3$ for all t it applies $d = 1$ for registrations on Monday to Thursday and accordingly $d = 2$ for Friday and Saturday valid for all returns. Subsequently, the amount of products assumed to return in τ is known. The second scenario regards varying values for c and l by supposing that both are uniformly distributed in intervals of $c \in \{2,\ldots, 5\}$ and $l \in \{2, 3, 4\}$. For each weekday this causes different shares of products that will return the same week (d = 0), after one week (d = 1) and after two weeks (d = 2). Our observations indicated a fluctuation around constant levels for all of these shares. For forecasting, we formed simple averages by utilizing the observations of the respective previous years for d in order to calculate estimations for each of these rather constant quotas. By multiplying the estimates with the known amount of registered returns a day we could determine the number of products μ_τ expected to return in τ. For both scenarios Eq. (6) was applied to determine $\mu_{f,\tau}$. The straight forecasts of the misused returns without estimated rates—realized by using HWA, too—serve as benchmark. The values for MAD, MAPE, and TIC are displayed in Table 3. The TIC was calculated with an adapted naïve forecast using the return amounts of the respective previous year (Makridakis and Hibon 2000).

The results clearly show the superiority of our method. The variation of c and l did not negatively affect the forecasting accuracy. Rather the contrary, the TIC and MAD show even better results, whereas the error measures for the direct forecasting of returns deteriorate. This means our method is still valuable if c and l cannot be determined for sure. The reason can be traced back to the balancing effect of the simultaneous under- respectively overestimation of μ_τ in opposite to $\tilde{q}_{f,\tau}$. The procedure is particularly advantageous if deviations within the return rates are low. It should be clear that a high fluctuation will impact the forecasting accuracy distinctly again. For seasonal products this means the patterns of these quotas show a clear recurring structure of comparable cycle durations. But since there are extensions dealing with this kind of problems it does not mean the approach is not applicable to time series which do not fulfill this requirement (e.g. Goodwin 2010). The forecasting accuracy might also decrease if the interval of c and l increases. Another disadvantage is the expense with which data has to be analyzed. For all impacted reason codes this involves the continuous gathering of corresponding return rates. Beyond, further investigations are necessary to improve the estimation of $\mu_{j,\tau}$ by reliably define c and l to accurately determine d.

Summarizing it can be said that all tested methods lead to reliable results if particular conditions are fulfilled. The accuracy particularly depends on deviations within the quotas. Forecasts are more reliable in case these are low and accordingly if potential seasonal patterns recur in a comparable manner. However, both models

Table 3 Error measures for seasonal return forecasting

	Fixed c and l		Varying c and l	
	Only returns	With est. rates	Only returns	With est. rates
MAD	11.9989	5.0532	13.1368	4.8321
MAPE	0.1086	0.0524	0.1232	0.0528
TIC	0.9872	0.4717	1.0416	0.4212

can be regarded as superior to the mere assumption of probability distributions to represent return rates, as misspecifications together with additional variation of the actual quotas would cause even higher errors. That is why the smart summary of products and product categories for joint prediction as well as a reasonable categorization are inevitable for a target-aimed forecasting with observed return rates.

4 Conclusion

In this paper, we adapted and further developed existing forecasting approaches by underlying e-commerce consumer returns affected by misuse. Furthermore, it was expected that the e-tailer uses product-individual data, so the estimated return rates are not based on distribution assumptions. In order to gain return rate estimations we used moving averages and exponential smoothing for forecasting based on sales lots and the Holt-Winters-approach for forecasting based on the date of return instead. It was shown that the methods deliver reliable results if return rates do not drastically fluctuate which is why it is crucial to avoid this instance. A former categorization using reason codes in combination with an adjoining determination of low-quality return forecasts could be recommended in that connection. Furthermore, the mentioned actions enable the observation of the consumer's return behavior. Thus, each return can be assigned automatically to the respective customer. Besides the fact that those consumers with a conspicuous fraudulent return behavior are recognized faster (Speights and Hilinski 2005), it is also possible to enhance the forecasting accuracy. Hence, return rates are not only dependent on the sales lot or the date of return anymore and can be determined more reliably.

Beyond, fast deliveries, ideally at the same day of ordering, will surely be a new standard within e-commerce in future. Retailers have to deliver goods, whose purchase was predicted by corresponding software, to distribution centers before they are even ordered, also called predictive purchasing. If that does not happen, the goods have to be taken back. That is why return numbers will increase, too (Gillies 2014). How these expansion options could be integrated into respective models should be a topic for further research. In the age of Big Data forecasts on returns surely will be more and more accurate, but together with that also more complicated. It is an interesting question how complex information can serve as enhancement by still preserving practicability.

References

Agrawal S, Singh RK, Murtaza Q (2014) Forecasting product returns for recycling in Indian electronics industry. J Adv Manage Res 11(1):102–114

Asdecker, B, Weigel A (2013) Der Missbrauch des Widerrufsrecht aus Sicht des deutschen Onlinehandels. *eStrategy Magazin* 13(1):27–30

Brandl N (2014) Fit für Multichannel. Logistik heute 9:48–49
Brown RG (1963) Smoothing, forecasting and prediction of discrete time series. Prentice-Hall, Englewood Cliffs, NJ
Clottey T, Benton WC Jr, Srivastava R (2012) Forecasting product returns for remanufacturing operations. Decis Sci 43(4):589–615
Davis S, Hagerty M, Gerstner E (1998) Return policies and the optimal level of "Hassle". J Econ Bus 50(5):445–460
Dpa-AFX (2012) Verkäufe von Karnevalsartikeln in der Saison 2010/2011 (in Millionen Stück). *Statista*. Retrieved 11 March 2015 from http://de.statista.com/statistik/daten/studie/218513/umfrage/umsatzverteilung-beim-koelner-karneval/
Gillies C (2014) Rundum-Service rund um die Uhr. VerkehrsRundschau 43:26–29
Goh TN, Varaprasad N (1986) A statistical methodology for the analysis of the life-cycle of reusable containers. IIE Trans 18(1):42–47
Goodwin P (2010) The holt-winters approach to exponential smoothing: 50 years old and going strong. Int J Appl Forecast 19:30–33
Hanafi J, Kara S, Kaebernick H (2008) Reverse logistics strategies for end-of-life products. Int J Logistics Manage 19(3):367–388
Harris LC (2010) Fraudulent consumer returns: exploiting retailers' return policies. Eur J Mark 44 (6):730–747
Hess JD, Mayhew GE (1997) Modeling merchandise returns in direct marketing. J Dir Mark 11 (2):20–35
Holt CC (1957) Forecasting seasonals and trends by exponentially weighted averages. ONR Memorandum, 52, Carnegie Institute of Technology, Pittburgh, PA
Hsiao L, Chen Y-J (2012) Returns policy and quality risk in E-business. Prod Oper Manage 21 (3):489–503
Jin X, Hu SJ, Ni J, Xiao G (2013) Assembly strategies for remanufacturing systems with variable quality returns. IEEE Trans Autom Sci Eng 10(1):76–85
Kelle P, Silver EA (1989) Forecasting the returns of reusable containers. J Oper Manage 8 (1):17–35
IW Köln (2010) Verkaufte Karnevalsartikel (in 1.000) in Deutschland 2007/08 und 2008/09. *Statista*. Retrieved 11 March 2015 from http://de.statista.com/statistik/daten/studie/77461/umfrage/verkaufte-karnevalsartikel-in-deutschland-2007-08-und-2008-09/
Krapp M, Nebel J, Sahamie R (2013a) Forecasting product returns in closed-loop supply chains. Int J Phys Distrib Logistics Manage 43(8):614–637
Krapp M, Nebel J, Sahamie R (2013b) Using forecasts and managerial accounting information to enhance closed-loop supply chain management. OR Spectr 35(4):975–1007
Makridakis S, Hibon M (2000) The M3-Competition: results, conclusions and implications. Int J Forecast 16:451–476
Marx-Gómez J, Rautenstrauch C, Nürnberger A, Kruse R (2002) Neuro-fuzzy approach to forecast returns of scrapped products to recycling and remanufacturing. Knowl-Based Syst 15 (1/2):119–128
Mostard J, Teunter R (2006) The newsboy problem with resalable returns: a single period model and case study. Eur J Oper Res 169(1):81–96
Mostard J, de Koster R, Teunter R (2005) The distribution-free newsboy problem with resalable returns. Int J Prod Econ 97(3):329–342
Norek CD (2002) Returns Management: Making Order out of Chaos. Supply Chain Manage Rev 6 (3):34–42
Paletta G (2014) Die Retourenquote ist für Karnevalshändler ein leidliches Thema. *OnlinehändlerNews.de*. Retrieved 11 March 2015 from http://www.onlinehaendler-news.de/handel/allgemein/4122-retourenquote-karnevalshaendler-leidlich.html
Potdar A, Rogers J (2012) Reason-code based model to forecast product returns. Foresight 14 (2):105–120

Pur SS, Weinfurtner S, Wittmann M, Wittmann G, Stahl E (2013) Retourenmanagement im Onlinehandel—Das Beste daraus machen. *ibi research an der Universität Regensburg GmbH*. Retrieved 11 March 2015 from http://www.ibi.de/files/Retourenmanagement-im-Online-Handel_-_Das-Beste-daraus-machen.pdf

Rao S, Rabinovich E, Raju D (2014) The role of physical distribution services as determinants of product returns in internet retailing. J Oper Manage 32(1):295–312

Rosenbaum MS, Kuntze R (2005) Looking good at the retailer's expense: investigating unethical retail disposition behavior among compulsive buyers. J Retail Consum Serv 12(3):217–225

Schmidt RA, Sturrock F, Ward P, Lea-Greenwood G (1999) Deshopping—the art of illicit consumption. Int J Retail Distrib Manage 27(8):290–301

Speights D, Hilinski M (2005) Return fraud and abuse: How to protect profits. Retail Issues Lett 17(1):1–6

Temur GT, Balcilar M, Bolat B (2014) A fuzzy expert system design for forecasting return quantity in reverse logistics network. J Enterp Inf Manage 27(3):316–328

Theil H (1971) Principles of econometrics. Wiley, New York, NY

Toktay LB, Wein LM, Zenios SA (2000) Inventory management of remanufacturable products. Manage Sci 46(11):1412–1426

Van Nunen JAEE, Zuidwijk RA (2004) E-enabled closed-loop supply chains. Calif Manag Rev 46 (2):40–54

Winters PR (1960) Forecasting sales by exponentially weighted moving averages. Manage Sci 6 (3):324–342

A Rollout Algorithm for Vehicle Routing with Stochastic Customer Requests

Marlin W. Ulmer, Dirk C. Mattfeld, Marco Hennig and Justin C. Goodson

Abstract City logistic service providers often have to decide under uncertainty. In some cases, not all customers are known at the time of the decision, but may request service in the course of day. Here, anticipation of possible future requests in current decision making is mandatory to avoid ineffective decisions and to serve a high amount of requests within the service period. In this paper, we present a vehicle routing problem with stochastic customer requests. The objective is to maximize the number of served requests regarding a time limit. Decision points occur dynamically by arriving at a customer. To achieve anticipatory decisions, we define a rollout algorithm (RA) combined with sampling of future requests. Over a limited decision horizon, RA maximizes the immediate reward and the expected rewards-to-go explicitly considering every customer. Later rewards-to-go are estimated by sampling using a cheapest insertion greedy decision policy. We compare the RA with a value function approximation (VFA) benchmark heuristic. VFA allows long-term anticipation on the expense of only implicit customer consideration using key parameters. For instances, where explicit customer consideration is necessary, RA achieves a significantly higher solution quality than VFA.

M.W. Ulmer (✉) · M. Hennig
Technische Universität Braunschweig, Mühlenpfordtstraße 23, 38106 Brunswick, Germany
e-mail: m.ulmer@tu-braunschweig.de

D.C. Mattfeld
Decision Support Group, Business Information Systems, University of Braunschweig, Mühlenpfordtstraße 23, 38106 Braunschweig, Germany
e-mail: d.mattfeld@tu-bs.de

J.C. Goodson
Saint Louis University, 3674 Lindell Blvd, St. Louis, MO 63108, USA
e-mail: goodson@slu.edu

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_16

1 Introduction

In recent years, challenges for city logistic service providers increase. Customers expect fast and reliable service (Ehmke 2012). Courier, express and parcel services dealing with narrow profit margins have to plan and execute their routes under uncertainty. While for parcel delivery static routes are determined by service network design (Crainic 2000), customers may request pick-ups in the whole city and at any point of time (Ulmer and Mattfeld 2013). So, not every pick up may be known in the beginning of the day. Service providers use extra vehicles for pick-ups, adapting their routes dynamically to the new requests during the day. Decisions are made about routing and the confirmation or rejection of new requests aiming for a high service level, i.e., a high number of same-day pick-ups. Here, decisions have a significant impact on future customer services. Routes have to allow a cheap insertion of new requests in the existing tour. So, routing has to consider future requests to be added in the course of the day. Due to working hour regulations, not every pick-up can be served within the day. The dispatcher has to decide, if a request is confirmed and served or has to be rejected. Rejecting an isolated request in one area may allow later serving of many other requests in another. In essence, anticipation of future requests is mandatory to achieve effective and efficient routing.

For anticipation of future events in current decision making, service providers can draw on prognoses of future requests derived from historical data. This data may provide spatial and temporal customer distributions within the service area. To exploit this knowledge, future customer requests can be sampled to evaluate current decisions. Therefore stochastic paths are generated representing all possible future events. For the given problem, a stochastic path contains of a set of new requests with request times distributed over the rest of the service period. For evaluation of the current decisions, a Markov Decision Process (MDP) considering the stochastic paths can be used to calculate the expected future confirmations. Nevertheless, due to computational limitations, this approach is only applicable over a highly limited time horizon. To include expected later rewards-to-go into the evaluation, an efficient estimation is required. The combination of sampling and rewards-to-go estimation can be achieved using a rollout algorithm. RA solves a MDP over a certain number of decision points selecting the best decision for each decision point. Later decisions are made by a base heuristic (Goodson et al. 2015). The outcomes of the base heuristic are used in the MDP to evaluate the current decisions.

In this paper, we present an anticipatory RA for a vehicle routing problem with stochastic customer requests. A vehicle has to serve customers in a service area. A set of customers is known in the beginning and has to be served. During the day, new customers request service. The dispatcher has to decide whether to permanently confirm or reject a request. Considering the time limit, the objective is to maximize the number of customer confirmations.

This paper is organized as follows. In Sect. 2, we give a literature review regarding vehicle routing with stochastic requests. We especially focus on work regarding the sampling of future requests. In Sect. 3, we define the problem. The RA is described in Sect. 4. For a set of instances, RA is applied and compared to benchmark results provided by Ulmer et al. (2014) in Sect. 5. The paper closes with a conclusion and an outlook regarding future research.

2 Literature Review

The literature regarding stochastic and dynamic vehicle routing is vast. Stochastic impacts manifests in stochastic travel times (Ehmke et al. 2015), demands (Goodson et al. 2014), service times (Larsen et al. 2002), and customer requests (Thomas 2007). For a recent overview of stochastic vehicle routing, the interested reader is referred to Pillac et al. (2013). In this review, we focus on problems with stochastic customer requests. An extensive classification of the approaches dealing with stochastic customer requests can be found in Ulmer et al. (2014). Anticipatory heuristics can be classified in policy function approximation, methods of approximate dynamic programming, and sampling approaches.

Policy function approximation aims to identify a set of (straightforward) decision rules and often rely on information aggregation for customer anticipation. So, Thomas (2007) applies a waiting heuristic calculating the center of gravity of possible future customers. The vehicle waits at the customer served right before the center of gravity.

Value function approximation approximates the reward-to-go of future situations by evaluating problem states using simulation. Ulmer et al. (2014) introduced an approach maximizing the current and expected number of future confirmations. For estimation of the rewards-to-go, post decision states are evaluated. Therefore, a post decision state containing of time and a set of customers to serve is described implicit using the key parameters time and slack. Every time-slack tuple is assigned to a value. The values are approximated a priori via simulation.

Sampling approaches simulate a set of stochastic paths online. These paths contain of future requests and allow evaluating current decisions. Decision making is reduced to these sets of stochastic paths. This allows to simplify the outcome space while the detailed level of information within the paths can be maintained. The decision horizon is highly limited in sampling approaches. The more sample paths and decision points are considered, the more the computational effort increases. Bent and Van Hentenryck (2003, 2004) introduce approaches, where sampled customer requests are integrated in a set of routes. The route with the highest similarity to all other routes is selected. This approach is also applied by Flatberg et al. (2007) and Sungur et al. (2010). Ghiani et al. (2009, 2012) use sampling to estimate the level of customer satisfaction over a short-term horizon for a pick-up and delivery problem. Hvattum et al. (2006) sample customer requests using historical data to hedge routing expenses for a real world case study.

VFA and sampling approaches mainly differ in the point of time where states are evaluated regarding rewards-to-go. While VFA evaluates states over all realizations using average state values, sampling calculates the state values online for each realization in each decision point. Therefore, VFA requires aggregation within the states information, whereas sampling is able to consider every state detail in decision making on the expense of a high computational effort. In essence, sampling approaches allow detailed short term anticipation but are generally computational intractable for long term anticipation. Hence, we extend the sampling approach to an anticipatory post decision rollout algorithm. Our RA, defined in Sect. 4, combines short term sampling with a post decision state evaluation using a cheapest insertion greedy decision policy as base heuristic.

3 A Vehicle Routing Problem with Stochastic Customer Requests

In this section, we define the vehicle routing problem using a Markov Decision Process. For a mixed integer formulation of the problem setting, the interested reader is referred to Ulmer et al. (2015).

3.1 Problem Definition

A vehicle serves customers in a service area. It starts the tour in a depot and has to return considering a time limit t_{max}. In the beginning of the day, a set of early request customers already requested service and has to be served. During the day, new requests arrive stochastically in the service area. A decision point occurs by arriving at a customer. The dispatcher has to decide, whether to confirm or reject a request and where to add a confirmed request in the tour. The objective is to maximize the number of confirmations over the day.

The described problem is both stochastic and dynamic and can be defined using a Markov Decision Process (MDP, Bellman 1957) displayed in Fig. 1.

In a MDP, a number of decision points $1, \ldots, K$ occur subsequently. Here, K can be a random variable. In each decision point k, a current system state S_k and a set of decisions $d_1, \ldots d_n$ is given. The application of a decision d_i leads to an immediate reward $R_k(d_i)$ and a known post decision state S_k^d. A stochastic transition $\omega \in \Omega$ leads to the next state S_{k+1}.

For the given problem, a state S_k is defined by the point of time $0 \leq t_k < t_{max}$, the current vehicle location, a set of already confirmed customers and a set of new

Fig. 1 Markov decision process

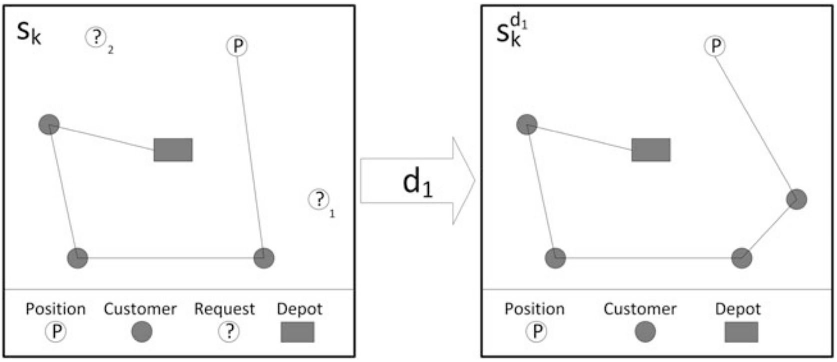


Fig. 2 State, decision, and post decision state for the defined problem

requests. Decisions are made about the subset of requests to confirm and the integration in the existing route. A post decision state S_k^d contains the point of time, the overall set of confirmed customers and the tour. During the travel to the next customer, new requests may arrive stochastically. So, the stochastic transition provides the set of new requests in S_{k+1}.

Figure 2 depicts an exemplary state S_k on the left. Two new requests have arrived. The applied decision d_1 confirms the first request and rejects the second resulting in a reward $R_k(d_1) = 1$ and post decision state $S_k^{d_1}$, shown on the right.

4 A Post Decision Rollout Algorithm

In this section, we develop a post decision rollout algorithm to obtain dynamic routing policies for the vehicle routing problem with stochastic customer requests.

A policy Π is a sequence of decision rules $(X_0^\pi, X_1^\pi, \ldots, X_K^\pi)$, where final decision epoch K may be a random variable and decision rule $X_k^\pi(S_k)$ specifies the decision to select when the process occupies state S_k. Letting Π be the set of Markovian deterministic policies, the objective is to identify a policy $\pi \in \Pi$ that maximizes the expected sum of rewards, conditional on an initial state S_0:

$$\max_{\pi \in \Pi} \mathbb{E}\left[\sum_{k=0}^{K} R_k\left(X_k^\pi(S_k)\right) | S_0\right]. \quad (1)$$

As described in Goodson et al. (2013, 2014), at decision epoch k, a post decision rollout algorithm identifies a decision in $D(S_k)$ that maximizes the sum of the current-period reward plus an estimate of the expected reward-to-go from the corresponding post decision state.

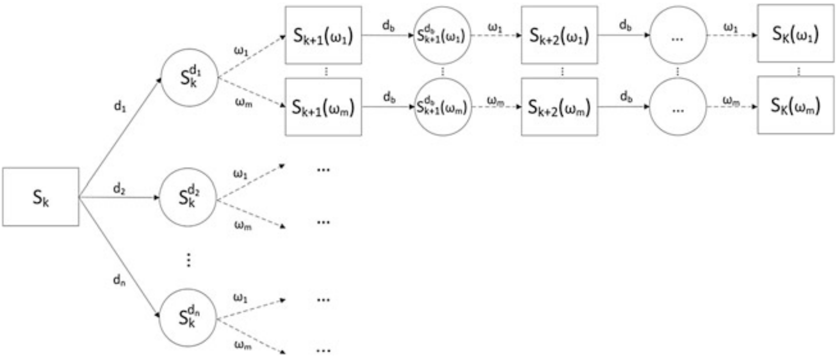


Fig. 3 A post decision rollout algorithm

First, to efficiently apply the algorithm, we reduce the decision space $\hat{D}(S_k) \subset D(S_k)$ drawing on cheapest insertion routing. Hence, decision making is about the subset of requests to confirm. Given r new requests, the number of decisions is $n := \left|\hat{D}(S_k)\right| = 2^r$. The application of a decision leads to a known post decision state S_k as described in Sect. 3.1.

We estimate the expected reward-to-go of each post decision state via a cheapest insertion greedy (CIG) policy. In a given state, the CIG decision rule routes the vehicle via a cheapest insertion heuristic (Rosenkrantz et al. 1974) and accepts the largest subset of customer requests the heuristic can route without violating the time limit—all other requests are rejected. Denoting by π_{cig} the CIG policy, when the process occupies realized state k, our post decision rollout algorithm selects a decision in the set

$$\underset{d \in \hat{D}(S_k)}{\arg\max} \left\{ R_{k(d)} + \mathbb{E}\left[\sum_{j=k+1}^{K} R_j\left(X_j^{\pi_{cig}}(S_j)\right) | S_k \right] \right\}. \tag{2}$$

Because the number and location of potential customer requests may be quite large, as in Bertsekas and Tsitsiklis (1996), we use simulation to estimate the expected rewards-to-go in (2). Figure 3 illustrates the mechanics of Eq. (2), where d_{cig} is the decision returned by policy π_{cig}. From post decision state S_k^d, we randomly generate a set $\{\omega_1, \ldots, \omega_m\} = \hat{\Omega} \subset \Omega$ containing m sample paths of future customer requests. The CIG policy is applied along each sample path and the estimated expected reward-to-go is the average reward accrued across the m simulations. The process is repeated to estimate the reward-to-go for each of the remaining decisions, resulting in a total of nm simulations for the post decision rollout algorithm to identify a decision.

5 Computational Study

In this section, we define the test instances and tune the parameters of the rollout algorithms. We compare the results with the benchmark heuristics provided by Ulmer et al. (2014) and explain for an exemplary case the differences in solution quality of the approaches.

5.1 Instances

Instances are provided by Ulmer et al. (2014). The instance parameters match real world scenarios and are derived from Bent and Van Hentenryck (2004), Hvattum et al. (2006), Thomas (2007), and Meisel (2011). A working day consists of $t_{max} =$ 360 min. We consider two different service area sizes of small $(15\,\text{km} \times 15\,\text{km})$ and large $(20\,\text{km} \times 20\,\text{km})$. The depot is located in the center of the area. The vehicle travels in a constant speed of $25\,\text{km/h}$. Travel times are Euclidean. The expected number of customers is 100. The expected number of early request customers is 25, resulting in a degree of dynamism of 75 % (Larsen et al. 2004).

To confirm the advantages of the RA, we consider two customer distributions in the area. First, customer requests are uniformly distributed (U). Second, customer requests are mainly accumulated in three clusters (3C).

5.2 Parameter Tuning and Benchmark Heuristic

We apply the RA for $m = 1, 2, 4, 8, 16, 32, 64, 128$ sample paths. The according algorithms are described by A_m. As a benchmark, we compare the results with the VFA-approach by Ulmer et al. (2014), and a greedy heuristic. The VFA-approach estimates the value of a post decision state regarding the two key parameters time and slack. So, customer locations are only considered implicitly.

5.3 Solution Quality and Runtime

We run 500 test runs for every instance settings. Experiments are performed on an Intel Core i5-3470 @ 3.2 GHz with 32 GB RAM. The percentage of served customer requests for RA, VFA and Greedy are shown in Table 1. The results are depicted regarding the customer distribution (U, 3C) and the service area size. The solution quality strongly depends on the number of sample paths. For only a few paths, the results are weak and in some cases even inferior to the Greedy approach. Proportional with the number of paths the solution quality increases. So, A_{16} already allows anticipation for all instance settings. A_{128} outperforms the Greedy approach by up to 25.0 %.

Table 1 Served customer requests (in %)

Service area	Large		Small	
Distribution	U	3C	U	3C
A_2	39.5	54.7	53.4	66.2
A_4	41.1	56.0	54.4	68.0
A_8	42.4	57.6	56.9	70.3
A_{16}	43.8	59.3	58.5	72.3
A_{32}	44.5	60.3	59.6	73.9
A_{64}	45.2	61.0	60.4	74.5
A_{128}	45.5	61.4	60.9	74.9
VFA	46.5	59.9	61.3	74.1
Greedy	36.4	55.9	54.5	72.3

In comparison to VFA, RA is superior for 3C, i.e., clustered distributed customers. Here, the explicit customer consideration of RA is advantageous. RA achieves up to 2.5 % better results than VFA. For uniformly distributed customer requests, RA is not able to perform as well as VFA. Here, the implicit customer representation allows a more general anticipation as required by the instance setting. To show the advantages of RA, in the following, we examine an exemplary case with clustered requests in more detail.

An exemplary state for the 3C-distribution and large service area is shown in Fig. 4. The set of (already confirmed) customers is assigned to a route ending in the depot. Two new requests have arrived in opposing regions of the service area. Request 1 lies in an isolated region, request 2 in the third cluster. Due to the time

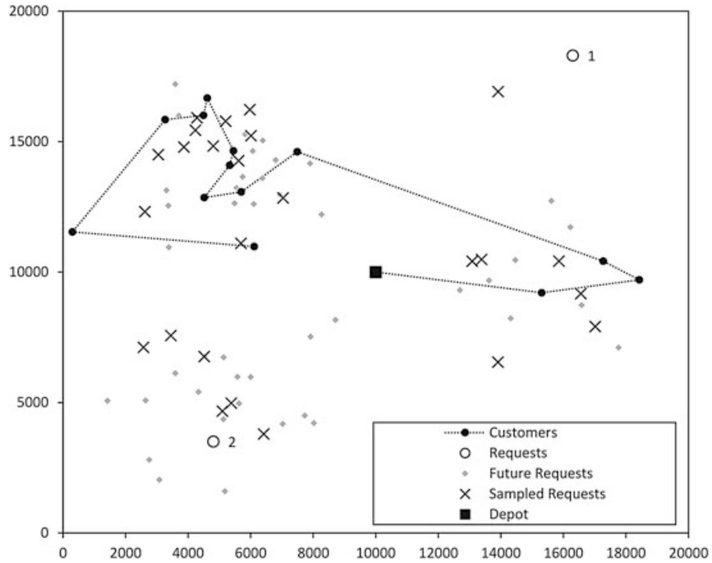


Fig. 4 An exemplary state for 3C-distribution and large service area

Table 2 Runtime per Iteration (in s)

Service area	Large		Small	
Distribution	U	3C	U	3C
A_2	10.30	12.35	5.48	5.29
A_4	24.50	29.33	10.67	9.75
A_8	45.24	59.72	20.31	20.32
A_{16}	90.68	123.04	40.31	40.78
A_{32}	186.51	232.71	80.73	82.21
A_{64}	335.86	424.04	155.46	160.16
A_{128}	701.29	925.79	306.75	320.67

limit, at most one of the requests can be confirmed. Request 1 is cheaper to include in the tour leading to a higher slack. It is expected that VFA chooses the first request. The locations of the (unknown) future requests lay mainly in the three clusters. This is not considered by VFA. Hence, VFA chooses an ineffective solution. RA samples future requests allowing the explicit consideration of request locations in decision making. Confirming request 2, in the course of the six additional sampled requests could be confirmed in this region compared to only one by choosing request 1. Hence, RA confirms request 2, because the sampled customers indicate a future request in the region around request 2. In essence, the explicit consideration of customer locations is advantageous for clustered customer distributions.

The according average runtimes of the RA per iteration are depicted in Table 2. Runtimes for VFA as an a priori algorithm and Greedy are neglectable. Generally for RA, the runtimes for a large service area are higher than for a small service area. This results from longer travel times between the customers and a higher set of customer requests per decision points. Hence, the number of decisions increases exponentially. The same behavior results in the longer runtimes for 3C-distribution. Here, in the long time span while changing the clusters, many new requests accumulate resulting in a large decision space $\hat{D}$.

As expected, the runtime increases nearly linear according to the number of sample paths. For A_{128}, a large service area, and the clustered distribution, the average runtime exceeds 15 min. These high runtimes may impede the online application of the algorithm. To examine the runtime behavior in more detail,

Table 3 Maximal runtime per decision point (in s)

Service area	Large		Small	
Distribution	U	3C	U	3C
A_2	4.51	5.07	0.98	1.10
A_4	12.58	13.90	1.68	1.17
A_8	19.91	27.79	2.04	2.74
A_{16}	38.66	54.06	3.86	5.11
A_{32}	82.32	95.39	8.99	9.99
A_{64}	130.56	142.84	15.19	15.58
A_{128}	290.15	350.78	31.63	27.93

Table 3 shows the average of the maximal runtime per decision point observed in an iteration. The runtimes for the large service area are significantly higher than for the small service area. As mentioned earlier, this results from the high number of possible decisions and the resulting high number of sampled paths. Compared to the overall runtime in Table 2, the results show that in many cases a few decision points cause most of the required runtime. Given 3C-distribution and a large service area, the runtime of A_{128} exceeds 5 min and may be not applicable.

6 Conclusion and Future Research

In many cases, city logistic service providers have to anticipate future customer requests to select effective decisions. In this paper, we presented a post decision rollout algorithm explicitly considering current customer and possible future customer requests. Therefore, the RA uses sampling and a cheapest insertion greedy policy to estimate the value of post decision states. Results show that in cases, where the explicit consideration of customer locations is necessary, RA outperforms the benchmark VFA-approach.

Future work may include modifications of the RA and the extension of the routing problem. The RA can be improved by choosing a more sophisticated base heuristic. Further, for many real-world applications, customers are not rejected, but postponed to a following service period. Generally, service providers dispatch a fleet of vehicles. So, instead of a single vehicle, a set of vehicles has to be considered.

References

Bellman R (1957) Dynamic programming. Princeton University Press

Bent RW, Van Hentenryck P (2003) Dynamic vehicle routing with stochastic requests. In: International joint conference on artificial intelligence, pp 1362–1363

Bent RW, Van Hentenryck P (2004) Scenario-based planning for partially dynamic vehicle routing with stochastic customers. Oper Res 52(6):977–987

Bertsekas DP, Tsitsiklis JN (1996) Neuro-dynamic programming. Athena Scientific, Belmont, MA

Crainic TG (2000) Service network design in freight transportation. Eur J Oper Res 122(2):272–288

Ehmke JF (2012): Integration of information and optimization models for routing in city logistics. In: International series in operations research & management science, vol. 177

Ehmke JF, Campbell AM, Urban T (2015) Ensuring service levels in routing problems with time windows and stochastic travel times. Eur J Oper Res 240(2):539–550

Flatberg T, Hasle G, Kloster O, Nilssen EJ, Riise A (2007) Dynamic and stochastic vehicle routing in practice. In: Operations research/computer science interfaces series, pp 41–63

Ghiani G, Manni E, Quaranta A, Triki C (2009) Anticipatory algorithms for same-day courier dispatching. Transport Res Part E Logistics Transport Rev 45(1):96–106

Ghiani G, Manni E, Thomas BW (2012) A comparison of anticipatory algorithms for the dynamic and stochastic traveling salesman problem. Transport Sci 46(3):374–387

Goodson JC, Ohlmann JW, Thomas BW (2013) Rollout policies for dynamic solutions to the multi-vehicle routing problem with stochastic demands and duration limits. Oper Res 61 (1):138–154

Goodson JC, Thomas BW, Ohlmann JW (2014) A generalized rollout algorithm framework for stochastic dynamic programming, Working Paper

Goodson JC, Thomas BW, Ohlmann JW (2015) Restocking based rollout policies for the vehicle routing problem with stochastic demands and duration limits. Transport Sci

Hvattum LM, Løkketangen A, Laporte G (2006) Solving a dynamic and stochastic vehicle routing problem with a sample scenario hedging heuristic. Transport Sci 40(4):421–438

Larsen A, Madsen O, Solomon M (2002) Partially dynamic vehicle routing-models and algorithms. J Oper Res Soc 53:637–646

Larsen A, Madsen O, Solomon M (2004) The a priori dynamic traveling salesman problem with time windows. Transport Sci 38(4):459–472

Meisel S (2011) Anticipatory optimization for dynamic decision making. Operations research/computer science interfaces series, vol 51

Pillac V, Gendreau M, Guéret C, Medaglia AL (2013) A review of dynamic vehicle routing problems. Eur J Oper Res 225(1):1–11

Rosenkrantz DJ, Stearns RE, Lewis PM (1974) Approximate algorithms for the traveling salesperson problem. Switching and automata theory. In: IEEE conference record of 15th annual symposium on. IEEE, pp 33–42

Sungur I, Ren Y, Ordóñez F, Dessouky M, Zhong H (2010) A model and algorithm for the courier delivery problem with uncertainty. Transport Sci 44(2):193–205

Thomas BW (2007) Waiting strategies for anticipating service requests from known customer locations. Transport Sci 41(3):319–331

Ulmer, MW, Brinkmann J, Mattfeld DC (2015) Anticipatory planning for courier, express and parcel services. In: Logistics Management. Lecture Notes in Logistics, edited by Jan Dethloff, Hans Dietrich Haasis, Herbert Kopfer, Herbert Kotzab and Jörn Schönberger, 313–324

Ulmer MW, Mattfeld DC (2013) Modeling customers in the euclidean plane for routing applications. In: Proceedings of the 14th EU/ME workshop, pp 98–103

Ulmer MW, Mattfeld DC, Köster F (2014) Budgeting time for dynamic vehicle routing with stochastic customer requests. Working paper

A Real-World Cost Function Based on Tariffs for Vehicle Routing in the Food Retailing Industry

Felix Tamke

Abstract In this paper, a real-world cost function for vehicle routing occurring in the distribution planning of a large German food retailing company is considered. In contrast to the widely used standard formulation introduced by Dantzig and Ramser (Manage Sci 6:80–91, 1959), the price structure is based on transport tariffs as the retailer cooperates with several freight carriers. The resulting costs depend on the products transported on a route as well as on the chosen itinerary and therefore, are not completely known in advance. For a deeper understanding, the cost function is formally modeled and afterwards illustrated by different examples. Besides, a detailed description of the underlying real-world issue is given and several problem variants of the Vehicle Routing Problem (VRP), like the VRP with Compartments or the VRP with Time Windows, are derived from the features of the issue. Due to the variety of variants, it can be classified as a member of a recent problem family called Rich VRP.

1 Introduction

The physical distribution of goods is an important part of the supply chain in various industries. On an operative level, the Vehicle Routing Problem (VRP) deals with this task and therefore is one of the most studied optimization problems in Operations Research. Since it has been introduced as "Truck Dispatching Problem" by Dantzig and Ramser (1959), a lot of research has led to a vast amount of different problem variants, each considering another problem emerging from the business world. However, most of them concentrate on just one or a few specific topics, e.g. VRP with Time Windows (VRPTW) (Cordeau et al. 2002), VRP with Pickups and Deliveries (VRPPD) (Parragh et al. 2008) or their combined consideration PDPTW (Dumas et al. 1991). Thus, they are often not sufficient for an

F. Tamke (✉)
Lehrstuhl für Industrielles Management, TU Dresden, 01062 Dresden, Germany
e-mail: felix.tamke@tu-dresden.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_17

application on complex real-world problems (Hartl et al. 2006). Besides this application pull, a push from the scientific community has been noticed as less comprehensive variants of the VRP are viewed as near-complete explored (Hasle and Kloster 2007). Hence, in recent years a new class of problems called real-life VRP's or Rich Vehicle Routing Problems (RVRP) was developed to fill this gap between theory and practice. Here, a combination of several features leads to a better applicability on issues from the business world. The occurring characteristics are very diverse and can be, for instance, vehicle-, driver-, location-, or request-related (Drexl 2012). However, a consistent definition of RVRP's does not yet exist. In a recent approach, Lahyani et al. (2015) formulated one based on a new taxonomy. There, an issue must have a minimum amount of different characteristics to be classified as RVRP. This new approach shows that there is still an open field of research to adapt the academic methods to the practical needs.

An industry, where vehicle routing is an important part for distributing goods, is the food retailing sector (Akkerman et al. 2010). This topic has been studied before by different authors. Some focus on the perishability of the food (Hsu et al. 2007; Osvald and Zadnik Stirn 2008), whereas others consider the combined transport of products with different transport temperatures (Chajakis and Guignard 2003; Ambrosino and Sciomachen 2007). The real-world issue motivating this paper is part of the latter subject and it will be specified in the next section for a better understanding. Afterwards, the occurring problem variants are derived from the characteristics. In Sect. 3, the cost function used by the retail company will be explained in detail as it differs from the usual form and therefore gives insight into the business practice. In the last section, an outlook on future research concludes this paper.

2 Case Study and Relevant VRP Variants

We consider a large German food retailing company which has to distribute its products from a central warehouse to about 120 markets on a daily basis. The company cooperates with several carriers to fulfill this task. Each of them is assigned a proprietary carrier region. This means that the markets are located in just one region and the carriers exclusively are allowed to supply those branches corresponding to their area. Thereby, the scheduling of the routes for the next day is made by the staff of the central warehouse within a tight timeframe of two hours, whereas the carriers provide the capacities, e.g. vehicles or drivers, needed. Hence, capacity shortages regarding the number of vehicles or a violation of the drivers' working hours cannot occur. Besides, there are further particularities. The supply of the markets is only feasible at specific periods, though, there can be more than one —usually one in the early morning and one in the afternoon. Additionally, different handling times exist for unloading the vehicle at the markets. Another important feature is the existence of a heterogeneous set of goods as the company is a full range trader. Especially the transport of food is strictly regulated by law and thus,

Table 1 Commodity classes existing in the case study

CC	Temperature range	Products
1	≤−18	Frozen foods
2	−1.5 to +2	Cheese, cold cuts, meat
3	+13 to +17	Fruits, vegetables, plants
4	No requirements	Non-food items, textiles

the goods are categorized in four commodity classes (CC) based on different transport temperatures as shown in Table 1. To cope with this feature, the load bed of the vehicles can be divided into up to three compartments with flexible sizes. Since cooling units are solely installed at the front and the rear end of the trailer, the middle compartment is not refrigerable. In addition to the supply with products, there are empty transport containers like pallets or stillages which have to be transported back to the depot according to a predefined schedule. From this features, a number of variants of the VRP can be derived and the issue may be classified as RVRP. This will be described below.

Since the transport quantities are large for each market, the capacity of one vehicle is not sufficient and therefore the orders of the markets need to be split among different routes. Hence, this problem is called Split Delivery VRP (SDVRP) (Archetti and Speranza 2012) and the allocation of transport quantities to vehicles becomes part of the optimization problem. Due to the shipment of exclusively whole pallets, those quantities are integer. As explained above, the distinction between commodity classes leads to incompatibilities regarding a spatial combined transport. As a consequence, there have to be either vehicles dedicated to just one commodity class or the load bed of a vehicle has to be separated into compartments. For the latter is economically favorable, the VRP with Compartments (VRPC) (Derigs et al. 2011) is present in the business case. Given the problem characteristics mentioned above, this results in specific constraints like a maximum of one commodity class per compartment or a maximum of three different compartments per vehicle. Another problem arises in the context of compartments as the vehicles are unloaded from the rear end via the last in—first out method. Therefore, goods in a compartment are not reachable until the prior compartments are empty. This issue is part of a wide problem family referred to as VRP with Loading Constraints (Iori and Martello 2010). As mentioned above, there are multiple time windows for the delivery present in the underlying case and thus, it can also be seen as a VRPTW. Resulting from the necessity to carry empties from the markets to the central warehouse and the guideline that the complete delivery has to be done before the pickup, the VRP with Backhauls (VRPB) (Goetschalckx and Jacobs-Blecha 1989) originates. Furthermore, following from the cooperation between the retail company and the carriers, the vehicles do not need to return to the depot. This characteristic is essential for Open VRP's (OVRP) (Li et al. 2007).

Another particular property of the presented case study is the calculation of transportation costs. In most research papers, the cost function is linear and depends just on the distance traveled. However, this does not apply here as transportation costs are based on contracts between the enterprises. These transport tariffs can be complex and are sparsely studied in the academic literature, although they are often used in practice (Drexl 2012). To the best of our knowledge, a more intense examination of tariffs in the context of vehicle routing has not yet taken place. In order to contribute to fill this gap, the real-world cost function emerging from the business case will be considered in greater detail hereinafter.

3 Tariff-Based Cost Function in a Real-World Application

Following from a large number of different objectives in distribution management, there also exist various objectives for VRP's. The most commonly used is the minimization of the total costs. Other objectives could be to minimize the total length of all routes or the number of vehicles employed, a balancing of routes as well as the maximization of profit. Furthermore, a combination of different goals is possible and therefore the implementation of a hierarchical objective system or the usage of multi-criteria optimization methods is necessary (Irnich et al. 2014). Nevertheless, cost minimization is still the most important objective and as it is the single aim of the underlying real-world issue, it will be further examined.

As common, the VRP is modeled as a graph theory problem on a complete graph $G = (V, A)$ with the set of vertices $V = (0, \ldots, n)$ and the set of arcs A. Vertices $i = 1, \ldots, n$ correspond to the markets, while vertex 0 denotes the central warehouse. The arcs represent the infrastructure connecting the markets. Additionally, there are K vehicles which transport the goods. Since the VRP was introduced by Dantzig and Ramser, their objective function has become some kind of standard formulation for many theoretical problems. In a three-index vehicle flow model, which is, unlike the two-index model, capable of a distinction between routes and therefore necessary for the present case, it can be stated as follows (Toth and Vigo 2002)

$$\sum_{i \in V} \sum_{j \in V} c_{ij} \cdot \sum_{k=1}^{K} x_{ijk}. \tag{1}$$

Here, c_{ij} is the cost associated with the usage of arc $(i, j) \in A$ and x_{ijk} is a binary variable with the possible values

$$x_{ijk} = \begin{cases} 1 & \text{, if vehicle } k \text{ travels directly from vertex } i \text{ to vertex } j \\ 0 & \text{, otherwise.} \end{cases} \tag{2}$$

Hence, this formulation requires that the elements of the cost matrix c are known in advance and are independent of anything other than the starting and ending points to calculate the total costs. Unfortunately, this is not given in many real-world applications. Here, costs often are based on transport tariffs and therefore depend on load specifics (e.g. weight, volume), distances, time, or the chosen itinerary. Likewise, the decision variables may be connected in a non-linear way. As a result, real-world objective functions are much more complex than the one shown in (1) (Irnich et al. 2014). This also applies to the cost function present in the business case described above, which will be further studied in the following.

First of all, the price structure among the retail company and the carriers shall be explained as it is highly problem-specific. The retail company only knows prices for the trips from the central warehouse to each supermarket—prices for trips from one branch to another do not exist per se. They are calculated based on a ratio called detour and the particular method will be explained in Sect. 3.2. The price structure is defined by two criteria. On the one hand a distinction is made between routes without (open route) and with (closed route) return of empties. And on the other hand every commodity class $p = 1, \ldots, P$ has its own price. This results in $2P$ different, non-negative prices for open routes po_{ip} as well as closed routes pc_{ip} for each market i. The price structure for three markets is exemplarily shown in Table 2.

Due to the fact that prices for trips between two stores do not exist, the costs for multi-customer routes are not easy to calculate. Therefore, the charging of single-customer routes, which contain solely one branch, is explained first and afterwards an extension to multi-customer routes is made.

3.1 Single-Customer Routes

In the business case, the costs P_k^S of a single-customer route k are determined by the most expensive commodity class transported on the route. As the assumption $pc_{ip} > po_{ip}$ is always valid, the price disparity $(pc_{ip} - po_{ip})$ between a closed and an open route may be seen as a surcharge for backhauls onto the open route price which has to be considered if empty pallets or stillages are carried from the store to the central warehouse. This approach leads to

$$P_k^S = \max_p \left(y_{ikp} \cdot po_{ip} + x_{i0k} \cdot (pc_{ip} - po_{ip}) \right), \tag{3}$$

whereby $y_{ikp} \in \{0; 1\}$ is a binary variable which attains one if the commodity class p is transported on route k to store i and zero otherwise. The procedure shall be illustrated by an example.

Table 2 Price structure for the supply of markets 1, 2 and 3

Market i	po_{ip}				pc_{ip}			
	1	2	3	4	1	2	3	4
1	270	260	240	230	350	330	300	280
2	230	210	200	190	300	275	260	250
3	350	340	325	310	450	430	415	400

Table 3 Examples of single-customer routes

Route k	Market i	Transported quantity of p				Backhaul
		1	2	3	4	
1	1	10	10	0	13	Yes
2	2	0	9	13	5	No
3	3	12	0	0	12	No

The stores from Table 2 are supplied with goods by different single-customer routes as represented in Table 3.

Route 1 is a closed route ($x_{101} = 1$), whereas routes 2 and 3 do not return to the depot ($x_{202} = 0, x_{303} = 0$). Equally, the values of the binary variables y_{ikp} are easy to derive from the shown transport quantities. Having the given data, the price of route 1 can be computed by

$$P_1^{\mathrm{S}} = \max\begin{pmatrix} 1 \cdot 270 + 1 \cdot (350 - 270), 1 \cdot 260 + 1 \cdot (330 - 260), \\ 0 \cdot 240 + 1 \cdot (300 - 240), 1 \cdot 230 + 1 \cdot (280 - 250) \end{pmatrix} = 350 \quad (4)$$

and equate to the maximal price. Likewise, the prices of routes 2 and 3 can be easily calculated as $P_2^{\mathrm{S}} = \max(0, 210, 200, 190) = 210$ respectively

$$P_3^{\mathrm{S}} = \max(350, 0, 0, 310) = 350.$$

3.2 Multi-customer Routes

As an extension to single-customer routes, multi-customer routes contain more than one store and hence the distance between the branches covered by the carrier and the additional stops have to be taken into account. This is realized as follows.

Multi-customer route costs of a route k consist of four parts. The first one is called base price (P_k^{base}) and is the maximum open route price of all stores visited and all commodity classes transported on a route:

$$P_k^{\mathrm{base}} = \max_i\left(\max_p(y_{ikp} \cdot po_{ip})\right). \quad (5)$$

Since the price calculation of the carrier is based on the distance between depot and market, the deviation of the direct distance to the furthermost branch and the length of the route from the warehouse to the last market is decisive for costing. This difference is called detour dt_k and involving the distances d_{ij} plus the binary variable

$$w_{ik} = \begin{cases} 1 & \text{, if branch } i \text{ is approached by vehicle } k \\ 0 & \text{, otherwise} \end{cases} \quad (6)$$

it can be computed in the following way:

$$dt_k = \sum_{i \in V} \sum_{j \in V \setminus \{0\}} x_{ijk} \cdot d_{ij} - \max_i (w_{ik} \cdot d_{0i}). \tag{7}$$

However, a detour up to a permitted distance dt^{perm} is already included in the prices given by the carriers and therefore the detour price P_k^{detour} is omitted if dt_k does not exceed dt^{perm}. Otherwise the base price is multiplied by a coefficient so that

$$\begin{aligned} P_k^{\text{detour}} &= P_k^{\text{base}} \cdot \frac{dt_k - dt^{\text{perm}}}{\max_i (w_{ik} \cdot d_{0i})} \\ &= P_k^{\text{base}} \cdot \left(\frac{\sum_{i \in V} \sum_{j \in V \setminus \{0\}} x_{ijk} \cdot d_{ij} - dt^{\text{perm}}}{\max_i (w_{ik} \cdot d_{0i})} - 1 \right). \end{aligned} \tag{8}$$

Furthermore, a fixed wage of cS for every stop in addition to the first one has to be paid. This leads to a stop price equivalent to

$$P_k^{\text{stop}} = \left(\sum_{i \in V} w_{ik} - 1 \right) \cdot cS. \tag{9}$$

As well as on single-customer routes, it is possible to carry empties on multi-customer routes. Since this is allowed exclusively from the last store in the sequence to the depot, the backhaul price depends on the surcharge of the last market and may be expressed as

$$P_k^{\text{backhaul}} = \max_p \big(x_{i0k} \cdot (pc_{ip} - po_{ip}) \big). \tag{10}$$

Finally, the total price of a route k can be determined by the summation over all components, whereby a distinction has to be made regarding the travelled detour:

$$P_k = \begin{cases} \max_i \left(\max_p \big(y_{ikp} \cdot po_{ip} \big) \right) \\ \cdot \left(\frac{\sum_{i \in V} \sum_{j \in V \setminus \{0\}} x_{ijk} \cdot d_{ij} - dt^{\text{perm}}}{\max_i (w_{ik} \cdot d_{0i})} \right) \\ + \left(\sum_{i \in V} w_{ik} - 1 \right) \cdot cS \\ + \max_p \big(x_{i0k} \cdot \big(pc_{ip} - po_{ip} \big) \big) & \text{, if } dt_k > dt^{\text{perm}} \\ \max_i \left(\max_p \big(y_{ikp} \cdot po_{ip} \big) \right) \\ + \left(\sum_{i \in V} w_{ik} - 1 \right) \cdot cS \\ + \max_p \big(x_{i0k} \cdot \big(pc_{ip} - po_{ip} \big) \big) & \text{, if } dt_k \le dt^{\text{perm}}. \end{cases} \tag{11}$$

This indicates that (3) is a special case of (11) with detour equal to zero and no stop price. As for the single-customer routes, an example shall explain this approach in the following.

The three markets from Table 2 are now supplied by one vehicle. The distances needed to calculate the multi-route costs are given in Fig. 1a and the transport quantities as well as the backhaul characteristics are shown in Table 3.

To show how the detour influences the price, the same quantities are delivered by the two different routes in Fig. 1b respectively 1c. The base price of this market-commodity class combinations is

$$P_1^{\text{base}} = \max(\max(270, 0, 0, 0), \max(0, 0, 0, 190), \max(0, 330, 0, 310)) = 330 = P_2^{\text{base}}. \quad (12)$$

Furthermore, there are $3! = 6$ possibilities to visit all branches in one route. Each of them has a varying length and hence, another detour length. For Route 1 with the sequence $[0, 1, 2, 3]$ in Fig. 1b the detour is calculated by

$$dt_1 = 160 + 125 + 77 - \max(160, 150, 200) = 162, \quad (13)$$

whereas Route 2 (Fig. 1c) with the same configuration but another sequence $[0, 2, 3, 1]$ yields

$$dt_2 = 150 + 77 + 95 - \max(160, 150, 200) = 122. \quad (14)$$

With a given permitted detour of $dt^{\text{perm}} = 50$, the detour prices result in

$$P_1^{\text{detour}} = 330 \cdot \frac{162 - 50}{\max(160, 150, 200)} = 184.8 \quad (15)$$

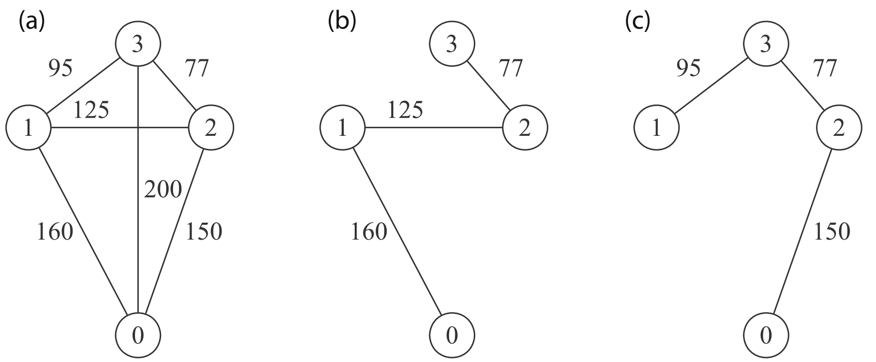


Fig. 1 Distances and routes used in the example. **a** Distances, **b** Route 1, **c** Route 2

Table 4 Transport quantities and backhaul characteristics of a multi-customer route

Market i	Transported quantity of p				Backhaul
	1	2	3	4	
1	9	0	0	0	No
2	0	0	0	9	No
3	0	9	0	6	No

and

$$P_2^{\text{detour}} = 330 \cdot \frac{122 - 50}{\max(160, 150, 200)} = 118.8. \tag{16}$$

It is obvious that they are—as long as the permitted detour is not exceeded—dependent on the route length and therefore on the chosen itinerary. If, for example, a longer detour of 170 is allowed, then both detour prices would be equal to zero and the itinerary would be irrelevant. In contrary, the stop price just corresponds with the number of stops. A stop wage of $cS = 25$ leads to

$$P_1^{\text{stop}} = (3 - 1) \cdot 25 = 50 = P_2^{\text{stop}}. \tag{17}$$

The same holds true for the backhaul price. Since there is no return of empties needed in the example, $P_1^{\text{backhaul}} = P_2^{\text{backhaul}} = 0$ is valid. If a backhaul is necessary, the number of possible itineraries would reduce to $(3 - 1)! = 2$, since the last market in the sequence is fixed. For instance, in the case of a backhaul from market 3, route 2 would be infeasible. The total prices of the routes 1 and 2 are

$$P_1 = 330 + 184.8 + 50 = 564.8 \tag{18}$$

and $P_2 = 498.8$.

4 Conclusion

This paper presented a business case occurring in the vehicle routing of a food retailing company, which can be considered as RVRP. Besides the deduction of several VRP variants, like the VRPC and VRPTW, from given peculiarities, it was especially focused on the exposition of the real-world cost function. This is based on negotiated transport tariffs, which take heterogeneous products as well as the chosen itinerary into account. Therefore, it is more complex and practically relevant than the widely used standard version originating from the first mathematical consideration of Dantzig and Ramser (1959). Yet, it is highly problem-specific and as a consequence, not transferable to other problems. Though, it could get even more complex if the transport across different carrier regions is no longer prohibited. However, an involvement of additional properties is accompanied by the introduction of new variables and constraints—if a consistent mathematical formulation is achievable at all. In this context, it would be possible to model the costs of a route P_k as decision variable and to implement the cost function as a constraint. Thus, nonlinear components like the maximum operators could be eliminated. Again, this is highly problem-specific and can only be discussed further with respect to the complete model.

Furthermore, modeling the cost function is just one part of handling a VRP on which we concentrated here due to the complex calculation of transportation costs. Besides the obviously necessary constraints that insure the feasibility of the solutions in accordance with the given case study, the problem-solving itself, i.e., the finding of feasible and good solutions, constitutes another major task. Since the problem instance in the case study contains ca. 120 markets, which is quite large, and the constraints are diverse, a solution with proven optimality is unlikely to determine. Additionally, the short planning timeframe restricts possible solution methods. Therefore, heuristics, especially metaheuristics, are probably most suitable and thus the subject of future research.

References

Akkerman R, Farahani P, Grunow M (2010) Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges. OR Spectr 32:863–904

Ambrosino D, Sciomachen A (2007) A food distribution network problem: a case study. IMA J Manage Math 18:33–53

Archetti C, Speranza MG (2012) Vehicle routing problems with split deliveries. Int Trans Oper Res 19:3–22

Chajakis ED, Guignard M (2003) Scheduling deliveries in vehicles with multiple compartments. J Global Optim 26:43–78

Cordeau J-F, Desaulniers G, Desrosiers J, Solomon M, Soumis F (2002) VRP with time windows. In: Toth P, Vigo D (eds.) The vehicle routing problem, pp 157–193

Dantzig GB, Ramser JH (1959) The truck dispatching problem. Manage Sci 6:80–91

Derigs U, Gottlieb J, Kalkoff J, Piesche M, Rothlauf F, Vogel U (2011) Vehicle routing with compartments: applications, modelling and heuristics. OR Spectr 33:885–914

Drexl M (2012) Rich vehicle routing in theory and practice. Logistics Res 5:47–63

Dumas Y, Desrosiers J, Soumis F (1991) The Pickup und Delivery Problem with Time Windows. Eur J Oper Res 54:7–22

Goetschalckx M, Jacobs-Blecha C (1989) The vehicle routing problem with backhauls. Eur J Oper Res 42:39–51

Hartl RF, Hasle G, Janssens GK (2006) Special issue on rich vehicle routing problems. CEJOR 14:103–104

Hasle G, Kloster O (2007) Industrial Vehicle Routing. In: Hasle G, Lie K-A, Quak E (eds) Geometric Modelling, Numerical Simulation, and Optimization, pp 397–435

Hsu C-I, Hung S-F, Li H-C (2007) Vehicle routing problem with time-windows for perishable food delivery. J Food Eng 80:465–475

Iori M, Martello S (2010) Routing problems with loading constraints. TOP 18:4–27

Irnich S, Toth P, Vigo D (2014) The family of vehicle routing problems. In: Toth P, Vigo D (eds) Vehicle routing: problems, methods and applications, 2nd edn, pp 1–36

Lahyani R, Khemakhem M, Semet F (2015) Rich vehicle routing problems: from a taxonomy to a definition. Eur J Oper Res 241:1–14

Li F, Golden B, Wasil E (2007) The open vehicle routing problem: algorithms, large-scale test problems, and computational results. Comput Oper Res 34:2918–2930

Osvald A, Zadnik Stirn L (2008) A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food. J Food Eng 85:285–295

Parragh SN, Doerner KF, Hartl RF (2008) A survey on pickup and delivery problems. J für Betriebswirtschaft 58:21–51

Toth P, Vigo D (2002) An overview of vehicle routing problems. In: Toth P, Vigo D (eds) The vehicle routing problem, pp 1–26

Part VII
Transport Management

An Adaptive Large Neighborhood Search for the Reverse Open Vehicle Routing Problem with Time Windows

Kristian Schopka and Herbert Kopfer

Abstract This paper extends the open vehicle routing problem (OVRP) by time windows, truck specific starting positions and a common delivery position. This extension is called reverse OVRP with time windows (ROVRPTW). The ROVRPTW occurs e.g. in dynamic vehicle routing problems, when at planning updates "return trips" to a central depot have to be generated. For generating cost-efficient transportation plans for the ROVRPTW, the well-known adaptive large neighborhood search (ALNS) is modified. A computational benchmark study is carried out, that compares the modified ALNS with other approaches for the vehicle routing problem with time windows and the OVRP, where also new best solutions are included. In addition, the algorithmic performance on the ROVRPTW is evaluated by solving a set of generated instances.

1 Introduction

In modern transportation and shipping, freight forwarders are confronted with increasing customer requirements. In this context, customers expect just-in-time pick-ups and deliveries or transport solutions in nearly real time. Additionally, a high fluctuation of customers' transportation demands can be observed resulting in the situation that many companies do not possess a private truck fleet or the truck fleet of freight forwarders is improper to satisfy the whole customer demand (Tarantilis et al. 2005). To satisfy the demand anyway, freight forwarders employ external carriers that generally invoice based on the traveled distance from the depot to the final customer destination. Consequently, the transport planning problems (TPPs) of freight forwarders relate to the open vehicle routing problems (OVRPs), where adequate and efficient solution procedures have to be developed. In this paper, we focus on an extension of the OVRP with time windows (OVRPTW),

K. Schopka (✉) · H. Kopfer
Chair of Logistics, University of Bremen, Wilhelm-Herbst-Str. 5, 28359 Bremen, Germany
e-mail: schopka@uni-bremen.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_18

where a freight forwarder has to serve a set of delivery locations under the observance of time windows. Therefore, a fleet of trucks, which start their delivery tours at a central depot, is available. In contrast to the vehicle routing problem with time windows (VRPTW), the trucks do not have to return to the depot. This results in a transportation plan (tour plan) that consists of a set of Hamiltonian paths. For a literature review on the OVRPTW, we refer to Repoussis et al. (2007). Several approaches of the literature (e.g. Fu et al. 2005; Repoussis et al. 2007) consider the reversion of tour plans in case of problems with only pick-up nodes. Our approach for tour reversing is different. Here, the trucks are located at truck specific geographical positions that differ from the locations of the pick-up nodes and the central depot. Hence, the travel expenses to the first pick-up node increase the overall costs and have an immediate influence on the generated tour plan. Henceforth, we call this extended problem the reverse open vehicle routing problem with time windows (ROVRPTW) and the tours included in the tour plan are denoted as return trips. Since the OVRPTW is classified as NP-hard (Repoussis et al. 2007), the ROVRPTW is also NP-hard.

Research attention on the ROVRPTW is limited. Some features of the ROVRPTW can be observed by the school bus routing problem (SBRP). A literature review is given by Park and Byung-In (2010). Li and Fu (2002) consider an SBRP, where several school buses start at their overnight parking place, collect students at bus stops and take them to schools within predefined time windows. Whereas the ROVRPTWs assume hard time windows for all locations, the SBRPs generally define only hard time windows at schools. Another area of application arises from the increasing success of car-sharing providers (Ciari et al. 2014). It is imaginable that such providers extend their supply by trucks that are rented on a daily basis for delivering activities. In the morning, the trucks are located (by the former charterer) at random spots in the pick-up area and in the evening, the trucks are parked at the last delivery position (e.g. the carrier's depot). The driver allocation to trucks is consciously not part of this consideration. An additional scenario for the ROVRPTW occurs in the dynamic collaborative vehicle routing problem. In this context, Schopka and Kopfer (2015) introduce a two-step framework (TSF) that organizes a collaborative request exchange among independent freight carriers in dynamic environments by using a periodic re-optimization. Both steps of the TSF are followed by an optimization phase, where several TPPs are solved. Thereby, at every planning update the ROVRPTW occurs and return trips to a central depot of the used trucks (with truck specific capacities/positions) are generated. Because of the identified increasing practical relevance of the ROVRPTW and the lack of adequate planning tools, the contribution of this paper is the development of an effective and robust heuristic approach for the ROVRPTW. Therefore, in Sect. 2 the problem definition with its mathematical formulation is introduced. An extended adaptive large neighborhood search (ALNS) is part of Sect. 3. Computational studies analyze the performance of the ALNS on benchmark instances for the VRPTW and OVRP in Sect. 4, where a computational study on generated ROVRPTW instances is included. Finally, the paper closes with a conclusion in Sect. 5.

2 Problem Definition

The objective of the ROVRPTW is to minimize the overall costs of a set of return trips to a central depot under the restriction of visiting a set of pick-up locations. At each pick-up location a certain demand has to be collected under the observance of hard time windows. Postulating hard time windows can be explained by the customers' expectation of just-in-time pick-up processes. Henceforth, the pick-up process at each customer location is designated as request. To satisfy all requests, a fleet of heterogeneous trucks is available, where each truck owns a specific geographical starting position and a capacity limitation. For each planned return trip, truck specific fixed costs increase the overall costs. Figure 1 illustrates the planning situation, by the requirement of planning return trips for only a subset of the fleet.

The ROVRPTW can be formulated as a mixed integer program (Eqs. 1–11). Let there be n pick-up locations indexed by i. The node set of those is stored as $N_c = \{1, 2, \ldots, n\}$. Furthermore, there exist v trucks indexed by k, where the set of trucks is given by $V = \{1, 2, \ldots, v\}$. The node 0_k represents the truck specific geographical starting position and $n + 1$ represents the central depot, so that the node set of the network is given by $N_k = \{0_k, 1, 2, \ldots, n + 1\}$ for truck k. The matrix d_{ijk} represents the travel time from node i to node j for truck k. Note that one time unit equates to one distance unit. The fixed costs for using truck k are given by e_k. Let α_1 be the multiplier of the fixed costs and α_2 the cost multiplier per used travel time. The pick-up demand at node i is given by l_i and the capacity of each truck k of the heterogeneous fleet is given by L_k. Finally, let $[a_i, b_i]$ denote the pick-up time window for node i and s_i represent its service time. Furthermore, two types of variables are used in the mathematical formulation. The binary decision variable $x_{ijk}, (i, j) \in N_k x N_k, i \neq j, k \in V$ is equal to one, if truck k travels from node i to node j and equal to zero otherwise. Let $t_i \in \mathbb{R}^+$ be the service starting time for all $i \in N_c$. The mathematical model seeks to minimize the sum of the total costs (z). Constraints (2)–(5) and (10) form a multi-commodity flow problem. Constraints (6)–(8) and (11) are the time constraints. The constraint (6) defines the service time for the first pick-up node and makes the sub-tour-elimination constraint (7) applicable. The constraint (8) ensures that the time windows for the pick-up nodes are

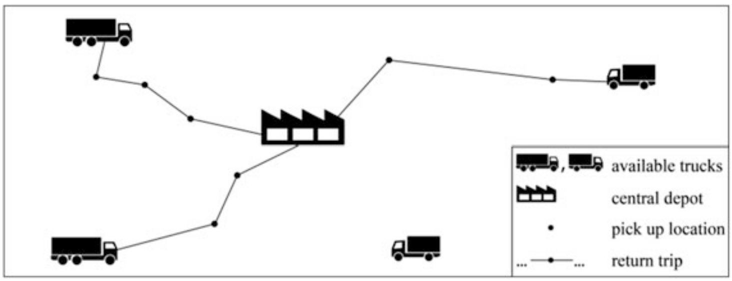


Fig. 1 Planning situation of the ROVRPTW

met. Note that our time constraints allow waiting time before visiting a node. Finally, the constraint (9) excludes an exceeding of the truck capacity.

$$\min z = \alpha_1 \cdot \sum_{j \in N_c} \sum_{k \in V} e_k \cdot x_{0jk} + \alpha_2 \cdot \sum_{i \in N_k} d_{ijk} \cdot x_{ijk} \tag{1}$$

$$\sum_{j \in N_k \setminus \{\mathrm{n}+1\}} \mathrm{x}_{0jk} = 1, \quad \forall k \in V, \tag{2}$$

$$\sum_{i \in N_c} \mathrm{x}_{\mathrm{i,n+1,k}} + x_{00k} = 1, \quad \forall k \in V, \tag{3}$$

$$\sum_{j \in N_k \setminus \{\mathrm{n}+1\}} \sum_{k \in V} x_{ijk} = 1, \quad \forall j \in N_c, \tag{4}$$

$$\sum_{j \in N_k \setminus \{0\}} \mathrm{x}_{\mathrm{ijk}} - \sum_{j \in N_k \setminus \{\mathrm{n}+1\}} \mathrm{x}_{\mathrm{jik}} = 0, \quad \forall i \in N_c, \forall k \in V, \tag{5}$$

$$d_{0jk} - bigM \cdot \left(1 - x_{0jk}\right) \leq t_j, \quad \forall j \in N_c, k \in V, \tag{6}$$

$$t_i + s_i + d_{ijk} - bigM \cdot \left(1 - x_{ijk}\right) \leq t_j, \quad \forall (i,j) \in N_c \times N_c, \forall k \in V, \tag{7}$$

$$a_i \leq t_i \leq b_i, \quad \forall i \in N_c, \tag{8}$$

$$\sum_{i \in N_k \setminus \{\mathrm{n}+1\}} \sum_{j \in N_c} x_{ijk} \cdot l_j \leq L_k, \quad \forall k \in V, \tag{9}$$

$$x_{ijk} \in \{0,1\}, \quad \forall (i,j) \in N_k \times N_k, \quad \forall k \in V \tag{10}$$

$$t_i \in \Re^+, \quad \forall i \in N_c. \tag{11}$$

3 Adaptive Large Neighborhood Search

For solving the ROVRPTW a modified ALNS (Algorithm 1) is used. The benefit of an ALNS, introduced by Ropke and Pisinger (2006), is that a nearly optimal solution can be generated by comparatively less computing time. Pisinger and Ropke (2007) show that the ALNS is able to solve some ROVRPTW-sub-problems. The ALNS is based on the large neighborhood search (LNS) introduced by Shaw (1997). The basic idea is to increase an initial solution (*sInit*) with its objective value ($f_{(sInit)}$) by a ruin and recreate strategy. Therefore, in each iteration a feasible solution (s) is destroyed by a removal operator (line 6) and later repaired by an insertion operator (line 8). This procedure is repeated until a certain stop criterion (e.g. maximal iterations) is met (line 3) and the best solution (*sBest*) is returned. Using large removal and insertion operators which arrange a diversification in the solution space is characteristic for LNS. Particularly, a random number of q requests in the range of $[q_1, q_2]$ is removed. Our ALNS reduces this range to $[q_3, q_4]$ for it_n iterations when a new best solution is found (line 10) and an intensification of the local search is realized. To escape from a local optimum an acceptance criterion is introduced that decides about the acceptance of a worse solution and constitutes the local search framework performed on the master level of the meta-heuristic. An example of such a framework is the simulated annealing (SA) of Kirkpatrick et al. (1983), which is also used by our ALNS (line 13). A feature of the ALNS is the use of different remove and insertion operators. Thereby, the probability of choosing an operator depends on a score list (line 5 and 7). The score list is updated depending on the historic performance of the operators (line 13).

Algorithm 3.1 Modified adaptive large neighborhood search

```
     Input: sInit, scoreList, it_n, q_1, q_2, q_3, q_4,
     Output: sBest
1.   s ← sInit;
2.   sBest ← sInit;
3.   repeat until stop-criterion meet;
4.       s'←s;
5.       choose removal operator (based on scoreList);
6.       remove q = [q_1,q_2] requests from s';
7.       choose insertion operator (based on scoreList);
8.       if f_s'>f_sBest then
9.           sBest←s';
10.          replace q_1 by q_3 and q_2 by q_4 for it_n iterations
11.      if accept(s',s) then
12.          s←s';
13.      update scoreList (dependent on effectivness);
```

Our ALNS uses four different removal operators, where each deletes q requests from the tour plan. The random removal ($r1$) deletes q randomly chosen requests. The idea of the Shaw removal ($r2$) is to remove preferably similar requests, where request specific features of each request $i \in N_c \backslash \{j\}$ are analyzed and rated regarding their similarity (δ_i) to a random chosen request $j \in N_c$. Our $r2$ rates the request features of distance ($|D_i - D_j|$), the difference of the start and end of the time windows ($|a_i - a_j|$, $|b_i - b_j|$) and the similarity of demands ($|l_i - l_j|$) between node j and node i. The scoring scheme of the $r2$ is described in constraint (12). To realize a common dimension for the request features a normalization of each function part is performed and the parts are weighted with β_1, β_2 and β_3. For further information about the original Shaw removal we refer to Shaw (1997). The worst removal ($r3$) deletes the request i from the tour plan that provides the highest traveling costs (c_i). Thereby, c_i is calculated by the savings of skipping node i on the current tour plan

($c_i = d_{i-1,i} + d_{i,i+1} - d_{i-1,i+1}$). Besides those well-known operators, we introduce the sequence removal ($r4$), where the idea is to remove whole request sequences from the tour plan. The main benefit of $r4$ is that by this procedure complete tours can be deleted. Especially, for the ROVRPTW a diversification of considered solution spaces can be realized by replacing whole return trips by trips of other trucks. On ROVRPTW instances we only use a version of $r4$ that deletes whole tours. Contrary, for solving instances of the VRPTW respectively the OVRP, $r4$ randomly chooses one of the following delete strategies: The sequence between two random requests, the sequence between the two longest distances, the complete tour after a random request or the complete tour after the longest distance is removed from the tour plan. This procedure is repeated until at least q requests are deleted.

$$\delta_i := \arg \min_{i \in N_c \setminus \{j\}} \left(\beta_1 \cdot \left|D_i - D_j\right| + \beta_2 \cdot \left(\left|a_i - a_j\right| + \left|b_i - b_j\right|\right) + \beta_3 \cdot \left|l_i - l_j\right|\right) \quad (12)$$

After the removing, the requests are re-inserted into the remaining tour plan by an insertion operator to generate a feasible tour plan. In common those operators identify auspicious insertion positions for the deleted requests with the goal to improve the tour plan. The basic greedy insertion ($i1$) inserts the request with the lowest insertion costs ($c_i = d_{ji} + d_{ih} - d_{jh}$; with j as potential precursor and h as potential successor) first. Contrary, the *regret-n* insertion ($i2$) introduced by Potvin and Rousseau (1993) considers the n auspicious insertion positions for all q requests and the request with the highest diversification among best and n-best insertion position is inserted first. Our *i2* chooses randomly *regret*-2, *regret*-3, *regret*-4 or *regret*-5 with the same probability. To guarantee a diversification in the considered solution space another version of $i1$ (labeled as $i3$) is performed, where a random request is inserted in its most auspicious position, first. To realize a diversification on considered solutions, a noise-factor is introduced that increases or reduces the costs of a potential insertion position factitiously. For more information we refer to Pissinger and Ropke (2007). During the execution of our ALNS each of the presented insertion operators is used twice, once with noise ($i1n$, $i2n$, $i3n$) and once without noise ($i1$, $i2$, $i3$). Additionally, an insertion operator ($i4$) is used that chooses a random request and inserts this request randomly on one of the five most auspicious insertion positions. All insertion operators repeat the described procedures until the tour plan is completed again.

4 Computational Experiments

In this section, the results of extensive computational experiments are presented. The first study analyzes the performance of our ALNS against other algorithms on the VRPTW, where also the solution stability of the ALNS over ten runs is considered. Furthermore, it is analyzed whether the ALNS is able to generate efficient tour plans for problems which consider a Hamiltonian path. To achieve this intention some of the well-known OVRP instances are solved and the results are compared with existing algorithms in literature. This section closes with some computational experiments on new ROVRPTW instances. All test cases are computed on a Windows 7 PC with Intel Core i7-2600 processor (3.4 GHz, 16 GB). Pissinger and Ropke (2007) generate very good results with their parameter settings on related vehicle routing problems. Thereby the most of their parameter settings are adopted. The remaining parameter settings (especially for the new operators) were analyzed in extensive studies. Generally, the test cases include the following parameter settings. The maximum iteration number is limited to 50,000 iterations. The initial solution is generated by a greedy insertion algorithm that uses $i1$. The start temperature of the SA is chosen so that a 10 % worse solution is accepted with a probability of 50 %. Furthermore a standard exponential cooling rate is used with the end temperature of 0.01 for the last iteration. The number of deleted requests lies in the range [q_1 = 10, q_2 = 40], respectively [q_3 = 10, q_4 = 20] for it_n = 50. The removal and insertion operators are chosen randomly by a roulette wheel selection, where the entry probability of selecting the operators is set as follows: $r1$: 25 %, $r2$: 40 %, $r3$: 25 %, $r4$: 5 %, $i1$: 20 %, $i1n$: 15 %, $i2$: 20 %, $i2n$: 15 %, $i3$: 15 %, $i3n$: 10 %, $i4$: 5 %. The probabilities for selecting the operators are updated after 100 iterations, based on their performance on earlier iterations. Here, the weight adjustment function with its weights and scores as described by Ropke and Pisinger (2006) is used. The weights β_1, β_2, and β_3 for $r2$ are chosen randomly in the range [0, 10].

4.1 VRPTW Benchmark Analysis

To evaluate the performance of the ALNS against other algorithms, the instances of Solomon (1987) for the VRPTW are analyzed. This test includes 56 instances [organized in the classes: cluster (C1, C2), random (R1, R2) and random/cluster (RC1, RC2)], where each is solved ten times by our ALNS1 with a single objective function (minimizing distance). The results in Table 1 show that the ALNS1 is able to generate tour plans that minimize the total distance. Thereby, the results of the related approach of Jung and Moon (2002), that also minimize the travelling distance, can be confirmed or improved on five of six instance classes. Furthermore, it can be determined that an adequate stability is realized by the ALNS, where the deviation among the best and the average distance of ten runs lies by 0.67 %. In this context the deviation can be explained by using operators that are dependent on random factors. Table 2 compares the results achieved by the ALNS1 against those of some published VRPTW algorithms which are at most effective on the criteria of used trucks and traveled distance. We restrict to approaches which use a single- (minimizing distance) or a bi-objective function (minimizing the number of trucks and distance). Approaches which consider the hierarchical VRPTW are not included. For a better comparability of our ALNS with approaches that use a bi-objective function, we repeat the previous test study with a version of our ALNS that penalizes each truck usage by fixed costs of 100 distance units (ALNS2), so that a minimum of trucks is aspired. As our results (best of ten runs) show, there exists no

Table 1 ALNS1 results (objective, trucks, time), averaged over classes, on Solomons data sets (best results are marked bold)

Instances	Average of ten runs			Best of ten runs			Jung and Moon (2002)	
	Objective	Trucks	Time(s)	Objective	Trucks	Time(s)	Objective	Trucks
Sol-C1	**828.38**	**10.00**	5.7	**828.38**	**10.00**	4.3	**828.38**	**10.00**
Sol-C2	**589.86**	**3.00**	6.7	**589.86**	**3.00**	5.2	**589.86**	**3.00**
Sol-R1	1186.13	13.24	29.7	**1179.87**	**13.25**	26.6	1179.95	13.25
Sol-R2	884.76	4.85	27.5	879.15	5.09	28.3	**878.41**	**5.36**
Sol-RC1	1356.99	13.05 3	22.4	**1341.55**	**12.75**	20.7	1343.65	13.00
Sol-RC2	1013.83	5.93	24.7	**1000.20**	**6.13**	27.1	1004.21	6.25
Acc.	55106.70	478.10		**54797.35**	**480.00**		54799.02	486.00

Table 2 Number of trucks and travel distance, averaged over classes, for the best solutions found in previous studies and by ALNS1/ALNS2 on Solomon data sets (best results are marked bold)

Authors	Sol-C1	Sol-C2	Sol-R1	Sol-R2	Sol-RC1	Sol-RC-2	Accumulation
Jung and Moon (2002)	**10.00**	**3.00**	13.25	**5.36**	13.00	6.25	486.00
	828.38	**589.86**	1179.95	**878.41**	1343.65	1004.21	54779.02
Ombuki et al. (2006)	10.00	**3.00**	13.17	4.45	12.75	5.63	471.00
	828.48	**589.86**	1203.22	892.89	1370.84	1025.31	55604.83
Tan et al. (2006)	10.00	3.00	**12.92**	3.55	**12.38**	**4.25**	**441.00**
	828.91	590.81	**1187.35**	951.74	**1355.37**	**1068.26**	**56293.06**
Alvarenga et al. (2010)	**10.00**	**3.00**	13.33	4.64	13.00	6.00	**477.00**
	828.37	**589.86**	1196.80	899.90	1341.70	1015.90	**55295.60**
Ghoseiri and Ghan-nadpour (2010)	**10.00**	3.00	12.92	3.27	12.75	3.75	473.00
	828.38	591.49	1228.60	1082.93	1392.09	1162.40	59278.69
Garcia-Najera and Bullinaria (2011)	**10.00**	**3.00**	**13.08**	**4.00**	12.63	5.38	**459.00**
	828.38	**589.86**	**1187.32**	**897.95**	1348.22	1036.65	**55378.61**
Ghannadpour et al. (2014)	**10.00**	3.00	13.42	4.00	13.25	4.00	457.00
	828.38	591.40	1213.14	992.30 13	1381.63	1144.67	57870.00
Dhari et al. (2014)	**10.00**	**3.00**	**12.17**	3.00	**11.75**	**3.25**	**413.00**
	828.38	**589.86**	**1204.19**	982.96	**1366.46**	**1200.62**	**57973.78**
Chiang and Wei-Huai (2014)	–	–	**13.17**	**4.36**	**12.63**	**5.38**	–
	–	–	**1181.36**	**888.36**	**1345.38**	**1014.81**	–
ALNS1	**10.00**	**3.00**	**13.25**	**5.09**	**12.75**	**6.13**	**480.00**
	828.38	**589.86**	**1179.87**	**879.15**	**1341.55**	**1000.20**	**54737.35**
ALNS2	**10.00**	**3.00**	**12.42**	**3.00**	**11.88**	**3.38**	**419.00**
	828.38	**589.86**	**1195.13**	**925.01**	**1360.69**	**1096.43**	**56347.91**

approach that dominates the results generated by ALNS1 or ALNS2. Furthermore, it can be determined that the ALNS1 generates tour plans with the overall minimal distance (54,737 distance units by using 480 trucks), which achieve the superior objective of our ALNS.

4.2 OVRP Benchmark Analysis

This study analyzes the ability of the ALNS to generate tour plans for OVRP instances. Therefore, the OVRP instances (without tour length restrictions) of Christofides et al. (1979) (Chr-1, Chr-5, Chr-11, Chr-12) and Fischer (1994) (Fis-11, Fis-12) are solved ten times by the ALNS1 (minimizing distance). Table 3 compares the objective value (z) and the computing time in seconds (t) of the ALNS1 (best of ten runs) with some of the most effective OVRP approaches (minimizing distance) and the best known solution. It can be determined that the ALNS1 is able to generate tour plans which are as good as comparable approaches. Furthermore, the test case leads to a new best solution on the instance Fis-12. By penalizing each truck usage by 100 distance units in the objective function, the ALNS2 is able to find for each considered instance a tour plan with the minimal number of trucks. Hence, a comparison of ALNS2 with some of the most effective approaches for the hierarchical OVRP is possible. As Table 4 shows, the ALNS2 generates tour plans with a higher gap to the best known solution than ALNS1 does. On average the distances are 0.13 % higher than those of the best known results. The main part of this aberration can be explained by the worse result for the instance Chr-5, where some related approaches also display a weakness. Nevertheless, the ALNS2 also improves the best known solution for the instance Fis-12 by 0.37 %.

4.3 ROVRPTW Analysis

To analyze the ALNS performance on the ROVRPTW, new instances are generated, which are organized in twelve classes, where each consists of four instances (overall 48 instances). Each class is based on a specific instance of Solomon (1987), from which all customer data and the position of the central depot are adapted. The four instances per class differ in the truck specific data, like the truck starting position, the capacity limitation and the fixed costs per truck. All instances and a detailed result analysis can be found on the internet page (http://www.logistik.uni-bremen.de/english/instances/index.htm). Generally, the above mentioned parameter settings are used, with some modifications. The remove operator $r4$ that removes whole return trips from the tour plan is identified as very powerful on the ROVRPTW. Therefore, the probabilities for choosing removal operators are changed to: $r1$: 20 %, $r2$: 15 %, $r3$: 15 %, $r4$: 50 %. Furthermore, the ALNS leads

Table 3 Comparison of effective meta-heuristics with single objective function for the OVRP (best results are marked bold)

Inst.	Best known	Tarantilis et al. (2004)		Tarantilis et al. (2005)		Zachariadis and Kiranoudis (2010)		ALNS1		
	z	z	t	z	t	z	t	z	t	Gap (%)
Chr-1	412.96(6)	**412.96**	98	**412.96**	23	**412.96**	24	**412.96**	13	0.00
Chr-2	564.06(11)	**564.06**	143	**564.06**	53	**564.06**	55	**564.06**	49	0.00
Chr-3	639.26(9)	641.77	330	639.57	128	**639.26**	106	**639.26**	34	0.00
Chr-4	733.13(12)	735.47	613	733.68	279	**733.13**	256	733.64	156	0.07
Chr-5	869.00(17)	877.13	1272	870.26	237	**869.00**	256	869.11	214	0.01
Chr-11	678.54(10)	679.38	318	678.54	141	**678.54**	87	679.41	123	0.13
Chr-12	534.24(10)	**534.24**	363	**534.24**	118	**534.24**	29	535.13	42	0.17
Fis-11	177.00(4)	–	–	–	–	**177.00**	83	**177.00**	47	0.00
Fis-12	761.68(8)	–	–	–	–	761.68	189	**758.72**	101	−0.41
Acc.	5369.87							**5369.29**		−0.01

Table 4 Comparison of effective meta-heuristics for the hierarchical OVRP (best results are marked bold)

Inst.	Best known	Pisinger and Ropke (2007)		Fleszar et al. (2009)		Zachariadis and Kiranoudis (2010)		ALNS2		
	z	z	t	z	t	z	t	z	t	Gap (%)
Chr-1	416.06(5)	**416.06**	23	**416.06**	1	**416.06**	25	**416.06**	19	0.00
Chr-2	567.14(10)	**567.14**	53	**567.14**	1	**567.14**	68	**567.14**	42	0.00
Chr-3	639.74(8)	641.76	128	**639.74**	12	**639.74**	103	639.88	32	0.02
Chr-4	733.13(12)	**733.13**	279	**733.13**	29	**733.13**	190	733.64	156	0.07
Chr-5	893.39(16)	896.08	237	905.96	15	**893.39**	355	901.49	247	0.94
Chr-11	682.12(7)	**682.12**	141	**682.12**	12	**682.12**	85	**682.12**	93	0.00
Chr-12	534.24(10)	**534.24**	118	**534.24**	8	**534.24**	39	535.13	42	0.17
Fis-11	177.00(4)	**177.00**	104	178.09	7	**177.00**	93	**177.00**	47	0.00
Fis-12	769.55(7)	770.17	359	769.66	62	761.68	301	**766.74**	101	−0.37
Acc.	5412.37							5419.2		0.13

Table 5 ALNS results (objective, trucks, time), averaged over classes, on ROVRPTW sets

Instance class	Average of ten runs			Best of ten runs		
	Objective	Trucks	Time(s)	Objective	Trucks	Time(s)
C101	1496.52	11.63	20.8	1482.38	11.25	15.1
C108	1495.70	11.70	25.3	1473.27	11.50	18.0
C201	815.69	3.25	4.6	814.21	3.25	3.2
C208	784.16	3.15	8.4	778.24	3.00	4.8
R101	2569.98	19.00	40.0	2562.32	19.00	37.1
R109	1760.79	12.48	38.4	1721.34	12.00	35.3
R201	1406.72	4.88	19.9	1399.02	5.00	16.7
R209	1048.38	3.60	28.5	1034.10	3.00	29.6
RC101	2278.64	15.00	27.7	2265.11	15.00	29.8
RC108	1671.59	12.00	30.7	1649.65	12.00	29.3
RC201	1542.79	5.00	14.4	1528.86	5.00	14.7
RC208	969.08	3.85	24.6	953.31	4.00	25.1
Accumulation	71360.10	422.10		70655.15	418.00	

to better results, if the range of deleted request is set to [$q_1 = 10$, $q_2 = 60$]. The cost multiplier α_2 is set to 1.0 and the cost multiplier α_1 results from the fix costs determined by the instances for the objective of the ROVRPTW. All other parameters are used as described at the beginning of this section. Table 5 considers the objective values, number of trucks, and computing time, averaged over classes, of the best and the average values of ten runs by our ALNS for the new ROVRPTW instances. The results show that the ALNS is able to generate stable tour plans for the ROVRPTW over ten runs, where a relatively small deviation among the best and average results can be determined. For identifying the best tour plan, the ALNS needs an average computing time of 23.6 s, which is similar to approaches of related TPPs. The manner of the generated tour plans seems plausible, yet it is hard to give conclusions about the solution quality, because of the lack of reference values.

5 Conclusions

In this paper, we introduced the ROVRPTW. Incipiently, we discussed the practical relevance of the ROVRPTW and presented a formulation as a mixed integer program. For generating tour plans, the well-known ALNS was adapted to the ROVRPTW. To increase the solution quality of the ALNS, the sequence removal operator was developed. The remaining paper focused on some extensive computational experiments. First, the ALNS was compared with some of the most effective heuristic approaches for the VRPTW and the OVRP. The studies show that none of the considered approaches dominates the solutions generated by our

ALNS on VRPTW instances. On OVRP instances, the ALNS leads to a comparable solution quality as other approaches. Furthermore, the best-known solutions on the instance Fis-12 are improved. Closing, a new class of ROVRPTW instances was developed and the ALNS was analyzed for its ability to generate tour plans that minimize the transportation costs. Currently, we use the presented ALNS on our main research field of dynamic collaborative TPPs, where we organize the collaboration of independent small- and mid-sized carriers by a stepwise request exchange process. In this context, at each planning update within a periodic re-optimization, the ROVRPTW occurs. The required tour plan depending on return trips to the central depot for each carrier is successfully generated by the ALNS described in this paper. By this fact, we can observe a future research relevance of the ROVRPTW and its approaches.

Acknowledgments The research was supported by the German Research Foundation (DFG) as part of the project “Kooperierende Rundreiseplanung bei rollierender Planung”.

References

Alvarenga GB, Mateus GR, de Tommi G (2010) A genetic and set partitioning two-phase approach for the vehicle routing problem with time windows. Comput Oper Res 34:1561–1584

Chiang T-C, Wei-Huai H (2014) A knowledge-based evolutionary algorithm for solving the multi objective vehicle routing problem with time windows. Comput Oper Res 45:1561–1584

Christofides N, Mingozzi A, Toth P (1979) The vehicle routing problem. In: Christofides N, Mingozzi A, Toth P, Sandi C (eds) Combinatorial optimization pp 315–338

Ciari F, Bock B, Balmer M (2014) Modeling station-based and free-floating carsharing demand. Transp Res Rec: J Transp Res Board 2416:37–47

Dhari A, Zidi K, Ghedira K (2014) Variable neighborhood search based set covering ILP model for the vehicle routing problem with time windows. Procedia Comput Sci 29:844–854

Fischer M (1994) Optimal solution of vehicle routing problems using minimum k-trees. Oper Res 42:626–642

Fleszar K, Osman IH, Hindi KS (2009) A variable neighborhood search algorithm for the open vehicle routing problem. Eur J Oper Res 195:803–809

Fu Z, Englese RW, Li LYO (2005) A new tabu search heuristic for the open vehicle routing problem. J Oper Res Soc 56:552–558

Garcia-Najera A, Bullinaria JA (2011) An improved multi-objective evolutionary algorithm for the vehicle routing problem with time windows. Comput Oper Res 38:287–300

Ghannadpour SF, Noori S, Tavakkoli-Moghaddam R, Ghoseiri K (2014) A multi-objective dynamic routing problem with fuzzy time windows: model, solution and application. Appl Soft Comput Part C 14:504–527

Ghoseiri K, Ghannadpour SF (2010) Multi-objective vehicle routing problem with time windows using goal programming and genetic algorithm. Appl Softw Comput 10:1096–1107

Jung S, Moon BR (2002) A hybrid genetic algorithm for the vehicle routing problem with time windows. In: Proceedings of 2002 genetic and evolutionary computation conference, pp 1309–1316

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Li LYO, Fu Z (2002) The school bus routing problem: a case study. J Oper Res Soc 53:552–558

Ombuki B, Ross BJ, Hanshar F (2006) Multi-objective genetic algorithms for vehicle routing problem with time windows. Appl Intell 24:17–30

Park J, Byung-In K (2010) The school bus routing: A review. Eur J Oper Res 202:311–319

Pisinger D, Ropke S (2007) A general heuristic for vehicle routing problems. Comput Oper Res 34:2403–2435

Potvin JY, Rousseau J-M (1993) A parallel route building algorithm for the vehicle routing and scheduling problem with time windows. Eur J Oper Res 66:331–340

Repoussis PP, Tarantilis CD, Ioannou G (2007) The open vehicle routing problem with time windows. J Oper Res Soc 58:355–367

Ropke S, Pisinger D (2006) An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transp Sci 40:455–472

Schopka K, Kopfer H (2015) Pre-selection strategies for the dynamic collaborative transportation planning problem. Oper Res Pro (2014, to appear)

Shaw P (1997) A new local search algorithm providing high quality solutions to vehicle routing problems. Technical report, Department of Computer Science, University of Strathclyde, Scotland

Solomon MM (1987) Algorithms for the vehicle routing and scheduling problems with time window constraints. Oper Res 35:254–265

Tan KC, Chew YH, Lee LH (2006) A hybrid multiobjective evolutionary algorithm for solving vehicle routing problem with time windows. Comput Optim Appl 34:115–151

Tarantilis CD, Diakoulaki D, Kiranoudis CT (2004) Combination of geographical information system and effective routing algorithms for real life distribution operations. Eur J Oper Res 152:437–453

Tarantilis CD, Kiranoudis CT, Ioannou G, Prastacos GP (2005) Solving the open vehicle routing problem via a single parameter metaheuristic algorithm. J Oper Res Soc 56:588–596

Zachariadis EE, Kiranoudis CT (2010) An open vehicle routing problem metaheuristic for examining wide solution neighborhoods. Comput Oper Res 37:712–723

A Two-Stage Iterative Solution Approach for Solving a Container Transportation Problem

Mengqi Wang, Bingjie Liu, Jiewei Quan and Julia Funke

Abstract The Inland Container Transportation Problem defines the movement of fully loaded and empty containers among terminals, depots, and customers within the same inland area. All kinds of customer requests are organized by one trucking company that owns depots containing a homogeneous fleet of trucks and a sufficiently large set of empty containers. The objective of this study is to minimize the total distance the trucks travel. We present a two-stage iterative solution approach that is capable to optimize around 300 requests. In the first step, the set of requests is divided into subsets, a tabu list prevents returns to recently considered subsets. In the second step, a mathematical problem is solved for each subset. These steps are then repeated and the best known solution is updated so long as certain stopping criteria are not met. The approach is implemented in C++ using IBM ILOG CPLEX. The quality was verified by several computational experiments.

1 Introduction

Intermodal container transportation describes a multi-unit transportation chain in which containers are conveyed with at least two different transportation modes. It is characterized by the use of various means of transportation (i.e. road, rail, water) and the delivery of goods that are in unchanged loading units. A typical transportation chain consists of three segments: pre-haulage, main-haulage and end-haulage. In the main-haulage, containers are shipped from one terminal to another, while in pre- and end-haulage containers are mostly transported by trucks between customers and terminals and vice versa. This paper focuses on the pre- and end-haulage sections of the transportation chain, which are well-known because of their high expenses compared to the short transportation range [truck movements cause between 25 and 40 % of the total transportation costs (Macharis and

M. Wang · B. Liu · J. Quan · J. Funke (✉)
Lehrstuhl für Logistik, Universität Bremen, Wilhelm-Herbst-Str. 5, 28359 Bremen, Germany
e-mail: jfunke@uni-bremen.de

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_19

Bontekoning 2004)]. Because of this issue, we consider a route planning problem for trucking companies arising in the pre- and end-haulage. We present an approach that offers a method to optimize the container transportation problem of a trucking company, where the objective is to minimize the total distance traveled.

The paper is organized as follows: Sect. 2 defines the problem. Section 3 gives an overview of related literature. Our approach is presented in Sect. 4, and its computational results are shown in Sect. 5. Section 6 states the conclusions.

2 Problem Definition

In the Inland Container Transportation Problem (ICT) (Zhang et al. 2010) one trucking company has to move cargo by trucks, while also providing empty containers, which are used as transportation media for cargo. The objective is to minimize costs, which in this case directly correlates to the total travel distance. The fleet of trucks is homogeneous, with each truck having a capacity of one, i.e., a truck can transport at most one container at a time. There are several depots given where the trucks have to start and finish their routes, but the origin and destination depots do not have to be the same. Initially the depots have enough trucks as well as empty containers available to transport all cargo. A transportation request is defined as the transportation of both empty and fully loaded containers, or as a container (un-)packing operation. According to Zhang et al. (2010) each request has to be processed without any interruption, meaning the truck that delivers an empty container stays with it while it is loaded and vice versa. The problem definition distinguishes between four different transportation requests. In inbound transportation requests, the containers are located at terminals and should be moved to depots, terminals, or customers. Inbound requests can be divided into inbound empty (IE) and inbound full (IF) requests. In IE requests, an empty container is picked up by a truck at a terminal, where it is initially located, and has to be carried somewhere. For IF requests, a fully loaded container is picked up by a truck and has to be carried from its initial terminal, to its customer (receiver). After drop off and unpacking, the obtained empty container should be moved to another depot or an alternative location, where it is needed. At outbound transportation requests the containers are located at either terminals, depots, or customers, and should be delivered to terminals. They can be divided into outbound empty (OE) and outbound full (OF) requests. In OE requests, an empty container is located somewhere, has to be picked up by a truck and delivered to a specified terminal. For OF requests, a truck brings an empty container from any location to the specified customer (shipper). After dropping off, the goods can be packed into the container. The fully loaded container should then be delivered to a specified terminal.

Besides a routing problem for trucks, the trucking company is confronted with an allocation problem for empty containers. It has to decide where empty containers should be picked up for OE/OF requests or where they should be delivered to for IE/IF requests. There is a large potential to reduce costs by fulfilling a street-turn, i.e., forwarding empty containers from receivers to shippers, instead of taking the detour via depots, which serve as storage for empty containers. Each reduced distance is called a saving. The potential benefits of savings are enormous, not only reducing costs but also reducing traffic congestion, noise, and emissions (Jula et al. 2006).

3 Literature Overview

Several recently published papers confirm relevance and topicality of tasks taking place at container terminals. Container movement by trucks with time windows at customers' locations is considered by Jula et al. (2005), leading to the proposal of a two-phase exact algorithm based on dynamic programming and two heuristics. If trucking operations involving transportation of loaded and empty equipment (e.g. containers) are considered, two types of requests are distinguished: well-defined requests (the points of origin/destination are given in advance) and flexible requests (either the point of origin or the point of destination has to be chosen from a set of possible locations) (Smilowitz 2006). Imai et al. (2007) presented a heuristic approach based on Lagrangian relaxation to solve a problem containing flexible tasks. Caris and Janssens (2009) extended the problem definition of Imai et al. (2007) by introducing time windows. The authors proposed a two-phase insertion heuristic approach that constructs an initial solution for the extended problem, the initial solution is improved with local search afterwards. Zhang et al. (2010) extended a problem definition of their previous work (2009) by considering multiple terminals, defining the ICT that comprises of four different types of transportation requests with time windows at the points of origin and destination. To solve the ICT, the authors modified a window partitioning approach (Wang and Regan 2002). Sterzik and Kopfer (2013) presented a Tabu Search heuristic approach while Nossack and Pesch (2013) presented a two-stage heuristic method that constructs routes first and improves them afterwards to obtain solutions of good quality for the ICT. Braekers et al. (2013) presented a sequential (assign containers and route requests) and an integrated approach (combine container assignment and route building process) with a single- and a two-phase deterministic annealing algorithm to solve a similar problem that considers a single depot and trucks that can leave containers for unloading. Many variations of container transportation problems have been investigated. Sterzik et al. (2012) considered several cooperating trucking companies and analyzed the benefit that can be achieved by sharing empty containers. Zhang et al. (2011) studied a problem in which the number of empty containers is limited. Whereas all problem definitions mentioned so far are restricted to the transportation of 40-foot containers, recent research investigates the transportation of both 20 and 40-foot containers. Within this formulation, a truck is able to transport up to either one 40-foot container or two 20-foot containers at a time. This problem type is investigated by studies such as Zhang et al. (2013) and Lai et al. (2013).

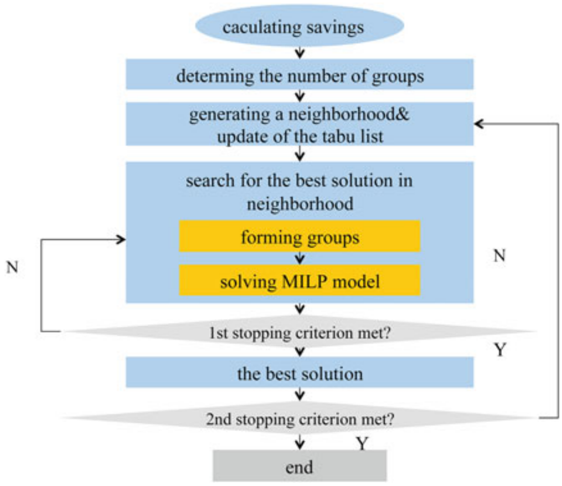
The considered problem definition is a mixture of the ICT and the problem that is considered by Imai et al. (2007) who excluded time windows at the customers' locations. One trucking company has to transport 40-foot containers. A scenario is considered taking place during peak times of shippers and receivers having the same opening hours. In their study, Braekers et al. (2013) concluded that the integrated approach clearly outperforms the sequential one. That is why we decided to investigate an integrated approach. We want to solve large instances with up to 320 requests to give the trucking company a good estimation over an extremely busy horizon. The instances of the ICT consist of 75 (Zhang et al. 2010), the instances of Jula et al. (2005) consist of 100 and the instances of Imai et al. (2007), Caris and Janssens (2009) and Braekers et al. (2013) consist of up to 200 transportation requests. To solve such a large instance, we partition it into groups consisting of all depots and a subset of customers and terminals needed by the requests. An exact mathematical model is implemented to compute the optimal solution to each group afterwards. As we do not consider time windows, we solve an implementation of an exact model of the problem for the groups and do not have to consider relaxation of the vehicles' working time (Imai et al. 2007) or time window discretization (Zhang et al. 2010). To obtain a solution of good quality, it is important to investigate as many grouping possibilities as possible with each grouping possibility having a high potential for improvement. Therefore, we propose a heuristic approach for building groups, while the heuristics of Jula et al. (2005), Caris and Janssens (2009), Sterzik and Kopfer (2013), Nossack and Pesch (2013) and Braekers et al. (2013) assign containers and route trucks in a sequential or an integrated manner. For example, similar to the method of Sterzik and Kopfer (2013), the presented solution approach also uses tabu search and saving techniques but the techniques are used to build groups and thus only indirectly impact on the route building process.

4 Solution Procedure

In case of small test instances, the considered problem can be optimally solved as a mixed-integer linear program (MILP). When it comes to larger test instances, it takes either a lot of time for any commercial solver to solve the resulting MILP, or it runs out of memory, which will be shown in Sect. 5. To solve bigger test instances in an efficient way we propose an algorithm based on a grouping method.

Figure 1 shows the structure of the approach. It is based on a combination of a tabu search algorithm (Glover 1986), a MILP, a saving procedure (Clarke and Wright 1964) and a grouping method. The first problem is a grouping problem, i.e., how to partition the instance into groups which have a great potential for improvement, whereby savings serve as indicator of quality. For each group optimal routes are determined by solving a MILP. The solution to the complete problem is a combination of the solutions to the individual groups. To obtain a solution of high quality, several grouping possibilities need to be investigated. Therefore, these

Fig. 1 Structure of the solution approach



two steps of forming groups and solving a MILP afterwards are iterated and the best known solution is updated until the first stopping criterion is met. A tabu search method is developed to find new grouping possibilities in new iterations. As long as the second stopping criterion is not met, new neighborhoods will be generated. The setting of criteria and other parameters is shown in Sect. 5. In the following section we explain how the two stages work in detail.

4.1 First Step: Forming Groups

The first part of the approach focuses on the potential for improvement. In this part every outbound request should be matched to each inbound one respectively and the saved costs through each connection should be determined. The term *savings* describes the saved distances/costs caused by the connection of an outbound load with an inbound one in comparison to the detour through a depot. The connected outbound and inbound requests are referred to as a *package*. Packages can be operated by the same vehicle. A package can be obtained by combining a container provider (IF/IE request) p with a container demander (OF/OE request) r. The saving s_{pr} of a package is calculated using formula (1); If i and j are locations, then $t(i, j)$ is the distance between these two locations. If i and j are requests, t is computed regarding the respective destination of i/origin of j. d is a savings amount corresponding to the optimal depot location.

$$s_{pr} = t(p, d) + t(d, r) - t(p, r) \quad (1)$$

The next step is to form groups with significant potential for improvement. Therefore, a node set N is introduced and assigned to groups. N represents all locations of the instance, whereby each request indicates new locations. In other words, for an OF request two nodes are inserted into *N*, one referring to the location of its shipper and another to the location of its terminal. Let *A* denote the set of all shippers and *TOF* the set of all terminals of OF requests. An IF request adds two nodes to *N*, one indicating the location of its terminal and another indicating the location of its receiver. Let *E* denote all receivers and *TIF* all terminals of IF requests. Each OE/IE requests introduces one node representing its terminal, with *TOE/TIE* is the set of all terminals of OE/IE requests. Finally, a node is added for each depot with *D* representing the set of all depots.

$$N = A \cup TOF \cup E \cup TIF \cup TOE \cup TIE \cup D \tag{2}$$

The groups are built so that their sizes do not deviate strongly from each other and the number of container providers and suppliers in each instance is as similar as possible. Each location of a request is assigned to exactly one group. The two locations of OF/IF requests are assigned to the same group. Each group contains all depots and at least one of the packages obtained by the savings procedure.

To form groups, first the number of groups is determined. This depends on the size of an instance, i.e., number of nodes *n*. To obtain high quality solutions in short running times, the size of each group should not be too small or too large. A size of 20–40 nodes allows our chosen MILP-solver to compute the optimal solution to a group in an acceptable running time. This is summarized in formula (3). In the first case, there is no need to iterate as there is only one group, thus the complete problem is solved optimally in the next step. That is why the instance in this case is larger: its only iteration is allowed to take more running time.

$$number\ of\ groups = \begin{cases} 1, & 1 \le n \le 60 \\ 3^* \left\lfloor \frac{n-1}{60} \right\rfloor, & n \ge 61 \end{cases} \tag{3}$$

Besides the savings procedure, we need one tabu list that works for generating neighborhoods in our approach. Tabu search is a metaheuristic approach that tries to improve a given solution while searching for a better one in its neighborhood. It uses a memory structure (tabu list) to prevent cycling (repeated visits of the same solution within the solution space). In the first iteration, this tabu list is empty. All packages are investigated and their savings are ordered by their value. The packages with the largest savings are selected and are tried successively to be assigned to the groups. Meaning, the package with the largest saving is selected and has to be checked if it is contained in the tabu list. If it is not contained, it is assigned to a group that has not assigned a package yet. Otherwise, it is forbidden to assign this package. Afterwards, the next package, i.e., the package with the second largest saving, is assigned to a group if it is not contained in the tabu list. This step is repeated until all groups have been assigned one package. A neighborhood is then created. The assigned packages are written into the tabu list, which means this neighborhood is forbidden in the following iterations. In this way, we hope to investigate a large improvement potential.

Afterwards, each group contains a package with the possibly greatest improvement potential, which also means each group contains a container provider and demander. The next step of the approach is to find the best solution in the created neighborhood by randomly assigning the remaining requests. There are many ways to allocate the rest of requests into groups. Every grouping option, i.e. assignment of requests to groups, corresponds to a feasible solution in the neighborhood. As a MILP-solver computes optimal routes for each group, the routes will not change if the elements in a group remain unchanged. Accordingly, many grouping options are examined to find the best solution in the neighborhood. To make each group contain nearly the same amount of container providers and demanders such that each of them gets the same chance for saving, both the container providers and demanders are randomly ordered and allocated to groups regarding this order. In this way, the groups are formed and a solution is found. In the next iteration, the order of both container providers and demanders is changed. Therefore, the grouping process is complicated through the cooperation of the tabu search and random change of the order of requests.

The next step is to solve a MILP. Afterwards, it is possible that the two requests of a package initially allocated to a group are no longer connected in an optimal route. Thus, the assignment of the package offers just an opportunity for saving costs in a group, which may not necessarily be taken.

4.2 Second Step: Constructing Routes

For each group a MILP model is solved to construct routes, assign containers, and obtain the distance minimal (4) solution, where z_{ij} denotes the distance between the locations associated to node i and j.

$$\min z(v) = \sum_{i \in N} \sum_{j \in N} \sum_{f \in F} v_{ijf} * z_{ij} \tag{4}$$

$$\sum_{f \in F} v_{it_i f} = 1, \quad \forall i \in A \tag{5}$$

$$\sum_{f \in F} \sum_{j \in (TIF \cup TIE \cup D)} v_{ijf} = 1, \quad \forall i \in TOF \tag{6}$$

$$\sum_{f \in F} \sum_{j \in (A \cup TOE \cup D)} v_{ijf} = 1, \quad \forall i \in E \tag{7}$$

$$\sum_{f \in F} v_{t_j jf} = 1, \quad \forall j \in E \tag{8}$$

$$\sum_{f \in F} \sum_{j \in (A \cup TOE \cup D)} v_{ijf} = 1, \quad \forall i \in TIE \tag{9}$$

$$\sum_{f \in F} \sum_{i \in (TIE \cup TIF \cup D)} v_{ijf} = 1, \quad \forall i \in TOE \tag{10}$$

$$\sum_{i \in D} \sum_{j \in (TIF \cup TIE \cup TOE \cup A)} v_{ijf} = 1, \quad \forall f \in F \tag{11}$$

$$\sum_{i \in D} \sum_{j \in N} \sum_{f \in F} v_{ijf} = \sum_{i \in N} \sum_{j \in D} \sum_{f \in F} v_{ijf} \tag{12}$$

$$\sum_{i \in N} \sum_{f \in F} v_{ijf} - \sum_{i \in N} \sum_{f \in F} v_{jif} = 0, \quad \forall j \in A \cup E \cup T \tag{13}$$

$$\sum_{i \in N} \sum_{j \in N} v_{ijf} = 0, \quad \forall i,j \in N,\ i = j,\ \forall f \in F \tag{14}$$

$$\sum_{i \in N} \sum_{f \in F} v_{ijf} = \sum_{i \in N} \sum_{f \in F} v_{jif} = 1, \quad \forall j \in A \cup E \cup T \tag{15}$$

$$m_{if} + 1 \leq M * \left(1 - v_{ijf}\right) + m_{if}, \quad \forall i \in A \cup E \cup T, \forall j \in N, \forall f \in F \tag{16}$$

$$v_{ijf} = \begin{cases} 1, & \text{truck } f \text{ drives from } i \text{ to } j \\ 0, & \text{otherwise} \end{cases}, \quad \forall i,j \in N, \forall f \in F \tag{17}$$

$$m_{if} \in \{0, 1, \ldots, |N|\}, \quad \forall i \in N, \forall f \in F \tag{18}$$

The node set N and its subsets are defined as in formula (2), with F denoting the set of trucks. There are two decision variables introduced. The binary variable v_{ijf} (17) indicates whether a truck $f \in F$ drives from node i to node j ($v_{ijf} = 1$) or not ($v_{ijf} = 0$). The integral variable m_{if} (18) ensures the subtour elimination (16), where M is a number greater than the number of nodes (Miller et al. 1960). Constraints (5)–(11) base on the idea that for the different types of nodes only certain types are allowed to be the immediate successor in a truck's route. A similar approach can be found in Sterzik and Kopfer (2013). As shown in formula (2), we distinguish between seven types of nodes and thus constraints. Constraints (11) ensure that after leaving the depot one of the four types of places $A \cup TIF \cup TOE \cup TIE$ is visited by a truck. The same structure applies to constraints (5)–(10) for trucks starting from the different node sets. Constraints (5) restrict a truck, which has already picked up a container from a shipper $i \in A$, to run to its specified terminal $t_i \in TOF$, where t_i denotes the corresponding terminal for a node representing a shipper or receiver $i \in A \cup E$. Similarly, constraints (8) define the drive direction from every terminal $t_i \in TIF$ to its specified receiver $i \in E$. Constraints (12) ensure that all trucks starting from specific depots return to them. Constraints (13) represent the flow conservation. Constraints (14) specify that origin and destination of an edge the truck uses differ. Constraints (15) ensure that customer requests are satisfied exactly once. This model only works as no time windows are given.

5 Computational Results

To test the capability of the solution procedure and the quality of results, 11 test instances are randomly created, whose total number of nodes can be up to 560. Because of the large range of instances and the special grouping method, the parameters must be set variously in different situations. The most important indicators are the number of iterations for the *large loop* (second stopping criterion, see Fig. 1) and *small loop* (first stopping criterion) and the number of groups. A better result requires more iterations in both loops, but due to the transition between C++ and CPLEX, every iteration takes a lot of time. A comprehensive test is carried out. Table 1 shows the parameter settings of the tests. All tests where run on an Intel Core i3, 1.50 GHz machine. The mathematical subproblem was implemented using IBM ILOG CPLEX Studio 12.5.1.

5.1 Analysis in Consideration of Runtime

Table 2 summarizes the structure of the instances. As we wish to analyze runtime, we have to build our test set such that the number of requests rises but the number of OF requests equals the number of IF requests and the number of OE requests equals the number of IE requests. This assumption is under realistic consumptions as OE and IE requests are used to prevent a trade imbalance of empty containers between hub areas (Zhang et al. 2010). Thus the number of IE and IF requests is roughly the same as the number of OE and OF requests. However, our approach can also solve instances differing in the number of inbound and outbound requests, e.g.

Table 1 The parameter settings

	#Nodes	#Groups	#Large loops	#Small loops	#Iterations
Instance group 1	0–60	1	10	#Packages/ (10 * #groups)	Max. 360
	61–120	3			
Instance group 2	121–180	6	5	20	100
	181–240	9	3	20	60
	241–300	12	3	20	60
	301–560	15–27	1	10	10

Table 2 Structure of the test instances

#Instance	#IF	#OF	#IE	#OE	#nodes	runtime appr.	results appr.
I0	10	4	30	6	70	859	298
I1	12	12	8	8	70	1,008	146
I2	14	14	9	9	80	1,050	4,986
I3	16	16	9	9	90	1,119	5,577
I4	17	17	11	11	100	1,403	5,697
I5	19	19	12	12	110	1,640	5,718
I6	21	21	12	12	120	2,567	7,011
I7	24	24	16	16	140	783	8,101
I8	36	36	24	24	210	2,659	11,949
I9	48	48	32	32	280	3,030	14,640
I10	96	96	64	64	560	4,200	23,368

I0 in Table 2. In this case, it is easier for the solver to solve the VRP problem in each group, which is shown by the runtime (859 s compared with 1,008 s for I1). This is because there are less options of combining these two different types of requests as the number of outbound requests is very small. Therefore it is reasonable to investigate the instances with balanced inbound and outbound requests where there are more possibilities for optimization. The parameter setting has a great influence on runtime. For example, after testing instance **I1**, the outcome with three groups (146) shows 13.7 % improvement over the outcome with five groups (166), but also takes more time (1,088 s compared with 801 s). When the nodes are divided into five groups every group contains less nodes, thus the solver needs less time to get an optimal result in each group, but at the same time the potential for improvement is restricted. Moreover, according to the parameter setting, a run, in which the instance is divided into five groups, takes less iterations in small loops and thus needs less runtime. Another key factor for the runtime is the instance's structure. As more requests/nodes demand more runtime, all instances are divided into two categories (indicated by the thick line in Table 2) with different parameter settings. In instance group 1 the computational time grows exponentially. 66 requests/120 nodes already reach a time of about 2,500 s. Thus, the parameters are changed for the second group, resulting in a computational time starting at 783 s for 140 nodes. We conclude that there is a positive correlation between the amount of requests and the time needed to compute the results. This supports the idea that for larger test instances less iterations are allowed.

As mentioned, a linear program solver can already solve the problem, which begs the question if it is essential to develop a grouping method. A test is created to find the limit of such a commercial solver and verify the ability of the presented method. As a result, instances with a range from 140 to 210 nodes make the solver less capable, but our algorithm is functional for at least up to 560 nodes.

5.2 *Quality of Solutions*

The quality of our approach is evaluated by comparing its results with optimal solutions (obtained by using CPLEX) and with solutions without optimization.

5.2.1 Comparison with Optimal Solutions

The quality is tested by comparing it with the optimal solution. Our program performs poorly when solving small instances, as it has longer running times and worse results than the CPLEX solver. But CPLEX has a limit of instances' sizes, which means our program is better at solving larger instances. An example is given for instance **I7** in Table 3. **I7** is divided into six groups, each containing at most 10 trucks. Our program solves the problem with 5 large and 20 small loops which took 171 % more time than CPLEX. Moreover, the result of our program for this instance is 9.5 % worse than the optimal solution obtained by CPLEX.

5.2.2 Compared with Result Without Optimization

The situation *without optimization* refers to a situation where every truck's route contains exactly one request. In this case, empty containers are directly transferred between depots and requests' locations. Since there is no saving through fulfilling a street-turn, the total distance is the maximal distance that is needed to fulfill all requests. Table 4 compares the results of our program with and without optimization for instance **I1**. The total distance without optimization is 673, while our program obtains a solution with a distance of 146. We achieved a great improvement, reducing the distance by 78 %.

Table 5 investigates what percent of containers from IF/IE requests are used for a following outbound request for instance **I1.** An empty container which is needed by an OE request originates at the sets *TIE*, *E* or *D*. Table 5 shows that only 25 % of all empty containers come from depots *D*. By contrast, empty containers from the other

Table 3 Comparison of instance **I7**

	Running time	Results
Our program	783	8,101
CPLEX solver	289	7,396
Comparison	171 % ↑	9.5 % ↓

Table 4 Comparison of instance **I1**

	Results of I1
Our program	146
Without optimization	673
Improvement	78 %

Table 5 Allocation of containers to OE requests for instance **I1**

Route	#Routes	Percentage (%)
$Ti{-}Tj,\ i \in IE,\ j \in OE$	4	50
$E{-}Ti,\ i \in OE$	2	25
$D{-}Ti,\ i \in OE$	2	25
Sum	8	100

two origins combined have a majority percentage of 75 %. In this case, this means that 75 % of empty containers come directly from an inbound request to an outbound request. Using this method, the routes are greatly combined in our program. Thus our goal is achieved: minimization of the total distance of the trucks.

6 Conclusions

In this study we present a solution to the ICT based on a grouping method to optimally solve a mathematical problem within each group as well as the heuristic tabu search procedure. In the evaluation, 11 different instances were tested. Computationally, the solution approach performs well, solving relatively large instances (320 requests) with largely optimized solutions considering both the results before optimization and the percentage of reduced empty container movements.

Algorithmically, there are two improvement suggestions: the first is the introduction of aspiration criteria for tabu search procedure, which allow solutions of the first several large iterations to be released again, as these results are obviously better than those found in the following iterations. The second suggestion is that in the search for the best solutions, parameters could be adjusted such that more solutions in the first several neighborhoods should be investigated and less in the following neighborhoods. Finally, with regard to the instances, time constrains for pick-up and delivery for customers and terminals should be considered.

References

Braekers K, Caris A, Janssens GK (2013) Integrated planning of loaded and empty container movements. OR Spectr 35(2):457–478

Caris A, Janssens GK (2009) A local search heuristic for the pre-and end-haulage of intermodal container terminals. Comput Oper Res 36(10):2763–2772

Clarke G, Wright JW (1964) Scheduling of vehicles from a central depot to a number of delivery points. Oper Res 12(4):568–581

Glover F (1986) Future paths for integer programming and links to artificial intelligence1986. Comput Oper Res 13(5):533–549

Imai A, Nishimura E, Current J (2007) A lagrangian relaxation-based heuristic for the vehicle routing with full container load. Eur J Oper Res 176(1):87–105

Jula H, Dessouky M, Ioannou P, Chassiakos A (2005) Container movement by trucks in metropolitan networks: modeling and optimization. Transp Res Part E Logistics Transp Rev 41 (3):235–259

Jula H, Chassiakos A, Ioannou P (2006) Port dynamic empty container reuse. Transp Res Part E Logistics Transp Rev 42(1):43–60

Lai M, Crainic TG, Di Francesco M, Zuddas P (2013) An heuristic search for the routing of heterogeneous trucks with single and double container loads. Transp Res Part E Logistics Transp Rev 56:108–118

Macharis C, Bontekoning YM (2004) Opportunities for OR in intermodal freight transport research: a review. Eur J Oper Res 153:400–416

Miller CE, Tucker AW, Zemlin RA (1960) Integer programming formulation of traveling sales-man problems. J ACM (JACM) 7(4):326–329

Nossack J, Pesch E (2013) A truck scheduling problem arising in intermodal container transportation. Eur J Oper Res 230:666–680

Smilowitz K (2006) Multi-resource routing with flexible tasks: an application in drayage operations. IIE Trans 38(7):577–590

Sterzik S, Kopfer H (2013) A tabu search heuristic for the inland container transportation problem. Comput Oper Res 40:953–962

Sterzik S, Kopfer H, Yun WY Reducing hinterland transportation costs through container sharing. Flex Serv Manuf J, 1–21 (2012)

Wang X, Regan AC (2002) Local truckload pickup and delivery with hard time window constraints. Transp Res Part B 36:97–112

Zhang R, Yun WY, Moon I (2009) A reactive tabu search algorithm for the multi-depot container truck transportation problem. Transp Res Part E Logistics Transp Rev 45(6):904–914

Zhang R, Yun WY, Kopfer H (2010) Heuristic-based truck scheduling for inland container transportation. OR Spect 32:787–808

Zhang R, Yun WY, Moon IK (2011) Modeling and optimization of a container drayage problem with resource constraints. Int J Prod Econ 133(1):351–359

Zhang R, Yun WY, Kopfer H Multi-size container transportation by truck: modeling and optimization. Flex Serv Manuf J, 1–28 (2013)

Vehicle Routing and Break Scheduling by a Distributed Decision Making Approach

Herbert Kopfer, Christoph Manuel Meyer and Hendrik Braun

Abstract We present a distributed decision making approach for the problem of combined vehicle routing and break scheduling. This problem consists of finding vehicle routes to service a set of customers such that a cost criterion is minimized and legal rules on driving and working hours are observed. In the literature, this problem is mostly analyzed from a central planning perspective. In practice, however, this problem is usually solved coactively by planners and drivers. One possible distribution of tasks is such that the planner does the clustering and routing of the customer requests and instructs the drivers which customers they have to visit in which order. Subsequently, the drivers decide upon their break schedules. We apply a framework for distributed decision making to model this planning scenario and propose various ways for planners to anticipate the process of tour fulfillment performed by the drivers. We analyze whether the way of anticipation has direct impact on the quality and usefulness of the plans representing the planner's instructions for the drivers. Computational experiments performed on the planning situation of a single driver demonstrate that a high degree of anticipation by the planner has a positive impact on the overall planning process.

H. Kopfer (✉)
University of Bremen, Wilhelm-Herbst-Straße 5, 28359 Bremen, Germany
e-mail: kopfer@uni-bremen.de

C.M. Meyer
Bayer Technology Services Gmbh, 51368 Leverkusen, Germany

H. Braun
Business School, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

D. Mattfeld et al. (eds.), *Logistics Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-20863-3_20

1 Introduction

In order to increase safety on the European road network, the European Union implemented Regulation (EC) No 561/2006 on drivers' driving hours in April 2007. This regulation poses restrictions to driving times such that driver fatigue, which is an important cause for road accidents, is prevented. Moreover, Directive 2002/15/EC gives restrictions on drivers' working hours. Together, these legal acts are referred to as EC social legislation. Due to these statutory prescriptions, planners at companies such as logistics service providers and even shippers have to deal with combined vehicle routing and break scheduling problems.

Despite the vast literature on vehicle routing problems (VRPs), combined vehicle routing and break scheduling has drawn only minor attention. Since 2007 there has been some more attention due to the introduction of Regulation (EC) No 561/2006 (Goel 2009; Kok et al. 2010). These papers focus on a central planning perspective, implying that the planner is controlling the complete combined problem. Only Meyer et al. (2011) propose a hierarchical planning approach. In the literature, different strategies have been proposed to deal with the problem of break scheduling within vehicle routing. Some strategies perform explicit break scheduling, in which the rules of Regulation (EC) No 561/2006 are expressly considered in the VRP (Goel 2009; Kok et al. 2010; Zäpfel and Bögl 2008). Others propose implicit break scheduling, in which slack travel time is created by using a lower average speed (e.g. Bartodziej et al. 2009).

The combined problem basically comprises three sub-problems: clustering of the customer requests, routing, and break scheduling. In practice, these sub-problems are distributed over different decision makers: The planner and the drivers. Semi-structured interviews with five medium-sized logistics service providers in Germany (Onken 2009) pointed out that in practice planners are responsible for clustering the customer requests, while the drivers schedule their breaks. In some applications such as parcel services, the routing is carried out by the drivers, but in less-than-truckload transportation, mainly the planners are responsible for the routing. Planners delegate the break scheduling to the drivers in order to reduce the complexity of their planning tasks and the required degree of accuracy.

Using distributed decision structures and delegating decisions to lower planning levels is a common method for addressing complex problems (Laux and Liermann 2005). Moreover, providing lower levels with more autonomy may lead to a better accomplishment of tasks (Windt and Hülsmann 2007). Hence, we investigate strategies for delegating the break scheduling to the drivers. Based on computational experiments for one single driver we analyze the usefulness of the plans constituting the planner's instructions in case of driving time deviations.

We introduce a distributed decision making (DDM) perspective by applying the framework proposed by Schneeweiss (2003). We consider two different types of decision makers: the planner and the drivers. The planner's main task is to determine the customer clusters and routes that will be assigned to drivers. Based on the planner's directives and up-to-date information on actual driving times, the drivers have to decide upon their break schedules.

2 Problem Description

As in traditional VRPs, we assume that there is a fleet of homogeneous vehicles starting from a central depot and delivering goods to a set of customers. Each vehicle makes at most one route and the total customer demand along each route must not exceed the capacity of one vehicle. Each customer may only be visited once and its service must start within a given time window. If a vehicle arrives early at a customer, it has to wait until the opening of the time window. After finishing their route, the vehicles return to the depot. The expected travel times and all customer requests are known in advance. So far, these restrictions are similar to those for the well-known vehicle routing problem with time windows (VRPTW).

Moreover, on their routes drivers have to take breaks and rest periods according to the EC social legislation. We assume that each vehicle is manned by one driver who stays with it for the whole planning period. Therefore, the driving times of the vehicle equal those of its driver. A planning period of one week is considered and we assume that all drivers have just had a weekly rest period. Breaks and daily rest periods have to be included into the vehicle routes such that the legislation is satisfied and the time windows are met. We call this partial planning task 'break scheduling'.

Regulation (EC) No 561/2006 concerns three different time horizons: single driving periods, daily driving times, and weekly driving times. For these time horizons, a complex set of restrictions consisting of basic and optional rules is imposed on the drivers' schedules. The optional rules allow additional possibilities for scheduling breaks and rest periods. For example, twice a week it is allowed to extend the daily driving time by exception from 9 to 10 h.

As driving times are considered to be working times, they are also affected by Directive 2002/15/EC, which contains restrictions on working times and breaks. For example, the directive postulates that after a working time of no more than 6 h workers have to take a break of at least 30 min. If the daily working time exceeds 9 h, the total break time has to amount to at least 45 min. These break times can be divided into parts of at least 15 min each. Consequently, a break that meets the requirements of EC Regulation No 561/2006 also satisfies Directive 2002/15/EC. For a comprehensive description of the legal rules of the EC social legislation, we refer to Kok et al. (2010).

The combined problem comprises the clustering of customer requests, the routing of vehicles, and the planning of breaks and rest periods. These interconnected partial planning problems can be solved either simultaneously or successively. Since in our scenario the tasks are split among different decision makers, we consider the case that these problems are addressed sequentially by planners and drivers. First, the planner carries out the clustering and routing of customer requests. When performing his planning task, the planner tries to minimize the number of vehicles used to serve the customers. His subordinate objective is to minimize the total travel time needed to serve all customers, which is assumed to be a substitute criterion for the total costs.

The routes generated by the planner are then passed on to the drivers. Within their assigned routes, each driver carries out the break scheduling such that the EC social legislation is satisfied on his route.

To avoid infeasibilities in the drivers' subsequent planning, the planner has to take into account the drivers' planning process. This means that when performing his planning task, the planner anticipates the break scheduling that will subsequently be performed by the driver. However, since the relevant legislation is very meticulous and since deviations are to be expected, planners might not anticipate the exact planning process but rather use a simplified approach to anticipate the drivers' planning model.

3 Embedding the Problem into the Framework for DDM

In the framework for DDM presented by Schneeweiss (2003), two decision making units (DMUs) are considered on different levels. The top-level uses a planning model $M^T(C^T, A^T)$ and takes its decision such that it optimizes its criterion C^T over all possible actions $a^T \in A^T$ within its decision field A^T. The top-level's criterion consists of a private criterion C^{TT} and a top-down criterion C^{TB} (which depends on the base-level's behavior), i.e. $C^T = \{C^{TT}, C^{TB}\}$. The top-level derives an optimal instruction $IN^* = IN(a^{T*})$ and communicates it to the base-level. Subsequently, the base-level takes its decision based on the top-level's instruction using its planning model $M^B(C^B, A^B)$ such that its criterion C^B is optimized.

To improve its planning results, the top-level can try to anticipate the base-level's subsequent planning in order to avoid giving an infeasible instruction or to account for the base-level's influence on the top-down criterion. Therefore, the top-level can apply an anticipation function $AF(IN)$, which is a function of the top-level's instruction and gives possible reactions of the expected base-level's behavior. The anticipation function does not need to be a precise representation of the base-level's reaction but can also be an approximation $Exp(M^B(C^B, A^B))$ of the expected base-level's planning model.

Schneeweiss (2003) distinguishes between four different degrees of anticipation: Perfect reactive anticipation, approximately perfect reactive anticipation, implicit reactive anticipation, and non-reactive anticipation. Perfect reactive anticipation means that the base-level's planning model is exactly known and it is anticipated by the top-level without any approximations. In the case of approximately perfect reactive anticipation the base-level's planning model is taken into account only approximately, e.g. by making simplifying assumptions. Implicit reactive anticipation means that only some features of the base model are considered and the anticipation function does no longer explicitly describe the base-level's decision model. These three degrees of anticipation incorporate the base-level's planning behavior as a reaction to the top-level's instruction. In the case of non-reactive anticipation, such an anticipation function does not exist, but only some general features of the base-level may be included in the top-level's criterion.

3.1 General Decision Structure

In the combined problem, the planner constitutes the top-level and the drivers constitute the base-level. In the pure top-down hierarchy where the planner does not account for the drivers' planning, the planner's criterion C^T is to minimize the number of vehicles required to serve all customer requests in a feasible sequence. This criterion is independent of the drivers' behavior and therefore the planner's criterion only comprises his private criterion, i.e. $C^T = C^{TT}$. The planner's decision field A^T comprises all possible transportation plans which ensure that all customers are served while the resulting routes are respecting the time windows and satisfying the capacity restrictions of the vehicles. I.e., in the case that the planner does not anticipate the drivers' planning model, the planner's decision problem results in solving the VRPTW without break scheduling.

After having solved the VRPTW and thus deriving the routes r_k for each vehicle k, the planner passes the routes on to the corresponding drivers who constitute the base-level. The drivers have to solve a break scheduling problem for their own routes while they have to consider the given time windows and EC social legislation rules. We assume that drivers have authentic information on their expected driving times and that they perform their planning according to their level of information which might be more accurate and up-to-date than the information that is available to the planner during planning time. For a single driver k' this results in the decision field A^B which comprises the set of all possible schedules for the route $r_{k'}$ such that the customer time windows are met and the legislation is fulfilled, including all optional legal rules. Clearly, for a given route derived by the planner, A^B could be empty since mandatory breaks will increase the duration of travel times and thus may disable the compliance with time windows.

In order to avoid infeasible routes, the planner must anticipate the drivers' planning behavior which will direct and control the actual execution of the submitted plan. We propose three anticipation functions $AF(IN)$ to be applied by the planner: perfect reactive anticipation, approximately perfect reactive anticipation, and non-reactive anticipation.

3.2 Anticipation Functions

The planner's private criterion C^{TT} is to minimize the number of used vehicles. If there are different routes resulting in the minimum number of vehicles, he will use his top-down criterion C^{TB}, which is to minimize the expected total travel time. However, to estimate the total travel time, he needs to anticipate the scheduling that will be performed by the drivers during the route fulfillment. We propose the following anticipation functions for all relevant degrees of anticipation.

Perfect reactive anticipation means that the planner considers the full planning model used by the drivers, i.e., $Exp(M_B) = M_B$. In this case, the planner expects each driver to solve an entire break scheduling problem, i.e. to generate for the route given to him an optimal solution complying with the EC social legislation and exploiting the full flexibility of the legislation. By anticipating this driver model, the planner is able to instruct routes that allow optimal routes and break schedules for the drivers under the assumption that the driving times assumed by the planner are identical to the driving times realized by the driver.

In case of approximately perfect reactive anticipation, the planner simplifies the driver's planning model. We propose to do this by leaving out the optional rules of the legislation. This means that the planner's anticipated base model is reduced to break scheduling without the optional rules of the EC social legislation. By including only the basic rules of the legislation, the planner provides some slack potentially enabling that the routes he instructs to the drivers allow feasible break schedules even if small deviations from the originally expected driving times occur.

Non-reactive anticipation means that the planner does only consider some general features of the base model. We model this planning approach by assuming that the planner does not explicitly consider the task of break scheduling. Instead, since the planner knows that the drivers have to observe EC regulations, he includes slack time which can be used to insert breaks and rest periods. A reasonable method for including slack time which is proportional to the travel distance is by using a lower speed than the average travel speed (see e.g. Bartodziej et al. 2009).

4 Computational Experiments

The planner's problem is a VRPTW, in which the decision space A^T is restricted by the anticipation function for tour fulfillment; also the top-down criterion C^{TB} is estimated through the anticipation function. For each way of anticipation, we describe the resulting problem and the corresponding mathematical models to be solved. After identifying the different models, we use a commercial solver for generating optimal solutions for the considered routing and scheduling problems.

In case of perfect anticipation, the planner expects each driver to solve his scheduling problem with the objective of minimizing the travel time. The planner's problem can be addressed by solving a VRPTW-EU (VRPTW with EU social regulations), with minimizing the number of vehicles as the primary objective. A mathematical description of this problem can be found in Kopfer and Meyer (2010), who propose a position based ILP-formulation for the VRPTW-EU. All rules of the EC social legislation are considered. The planner's secondary objective is to minimize the total travel time. From the driver's point of view, the instruction passed to him is a route that fulfills all conditions of an optimal solution of a TSPTW-EU (Travelling Salesman Problem with time windows and EC regulations) for the customer requests that have been assigned to him. In case of approximately perfect reactive anticipation, the planner also considers a VRPTW-EU and transfers to each driver a customer cluster together with an optimal solution of the resulting TSPTW-EU. However, now the planner ignores the optional rules. Therefore, the planner's problem is a VRPTW-EU, without considering the optional rules of the EC social legislation (Kopfer and Meyer 2010). With non-reactive anticipation, the planner solves a VRPTW and transfers to each driver a solution of a TSPTW for the driver's customer cluster, but with driving time estimations based on a lower travel speed than the actual average travel speed.

For our experiments on approaches for the distributed solving of combined vehicle routing and break scheduling, we assume that the vehicle clustering and routing has been done by the planner in advance by applying one of the above anticipation functions. The resulting routes are transferred as instructions to the drivers. For simulating the results of the planner's decisions we randomly generate customer clusters and then build routes for these clusters which are optimized according to the planner's anticipation function. For simulating the driver's planning situation we consider random deviations from the originally planned driving times, thus accounting for changing traffic conditions. Then, we perform break scheduling for the deviating driving times according to the driver's optimization criterion. Using this approach for testing, we are able to restrict our computational analysis to experiments performed on single customer clusters, respectively routes. This has the advantage that we only have to solve specific problems of TSP-type (TSPTW-EU, TSPTW-EU with basic rules, TSPTW) instead of VRP-type (VRPTW-EU, VRPTW-EU with basic rules, VRPTW). Problem instances of TSP-type with a problem size of up to seven nodes can easily be solved to optimality with a commercial solver for MIP-problems by using the models for the TSPTW-EU presented in Kopfer and Meyer (2010). We have solved all problem instances with CPLEX 12.1, which has been run on a PC with a 3.40 GHz CPU and 16.0 GB RAM [Core(TM) i7-2600 CPU; 3.40 GHz; 16,0 GB RAM].

Although there are test-sets for different kinds of routing and scheduling problems presented in literature, almost none of them refer to the EC social legislation, except for the benchmark instances proposed by Goel (2009), who adjusted the well-known Solomon benchmark instances (Solomon 1987) for the VRPTW to the VRPTW-EU. Since we consequently stick to problems of the TSP-type, we had to create our own test cases by generating clusters of customers including information for each cluster on the customer distance matrix, customer time windows, required service time, and customer demands.

We consider a set of 20 cities located in Germany and a 20×20-matrix representing the travel times between these cities. The travel times between these nodes were calculated based on their geographical positions and the shortest paths in the road network. In order to obtain a wide geographical diversification, the set of cities consists of the capitals of the German federal states and four additional big cities. In detail, the set includes Berlin, Bremen, Dresden, Düsseldorf, Erfurt, Hamburg, Hanover, Kiel, Magdeburg, Mainz, Munich, Potsdam, Saarbücken, Schwerin, Stuttgart, Wiesbaden, Duisburg, Essen, Cologne and Frankfurt am Main. For generating a single problem instance, a number of cities are randomly selected out of the considered set of cities. The selected cities constitute the cluster of customers for the correspondent problem instance. The customer demand is a randomly generated value between one and nine units. Deliveries are possible within time windows which are randomly positioned within the time span from 6 a.m. to 6 p.m. The service time at each customer node is fixed to 30 min and according to the EC social legislation, service is regarded as working time for the driver. For all problem instances the depot was chosen to be in Erfurt since it is close to the center of Germany. The vehicle capacity is determined to 50 units, and the assumption is made that each vehicle travels 65 km/h on average.

The process of generating test cases consists of randomly creating customer clusters with customer demands and time windows according to the above scheme and solving the TSPTW-EU with basic rules for the created clusters. Clusters which turn out to be unsolvable for the TSPTW-EU (e.g., it might be impossible to meet all time windows with a vehicle of the given average speed) are not accepted since they are considered not to represent reasonable and useful components of potential test cases. All other randomly generated clusters and the corresponding vehicle routes which have been built for them by solving the TSPTW-EU are included in the total test set.

We have generated test cases of different size. There are test cases with six nodes selected from the considered set of cities and test cases with seven nodes. Additionally, we have generated test cases with averagely large and with averagely narrow time windows. For narrow time windows, the length of the delivery time window varies randomly between 4 and 8 h and for large time windows it varies between 6 and 10 h. We have randomly generated test instances and checked whether these generated instances passed the test of being feasibly solvable by the "TSPTW-EU without optional rules". We have continued generating test instances until at least 50 instances were generated for each problem class. Altogether, 222 test cases have been generated, accepted and included in the test set: 52, 54, 57, 59 for the combinations (small time windows, 6 nodes), (small time windows, 7 nodes), (large time windows, 6 nodes), (large time windows, 7 nodes), respectively.

In the first step of our experiments we compare the different anticipation functions with respect to their ability to solve the generated test instances without considering their impact on the actual tour fulfilment. Each of the 222 test cases has been generated by checking its feasibility for the TSPTW-EU with basic rules before accepting it as test case. Consequently, this anticipation function is taken as a kind of filter for evaluating other anticipation functions. Applying other anticipation functions, it cannot be ensured that all 222 test instances can be feasibly solved. Table 1 shows the number of test instances which allow feasible planner instructions and the averagely needed computation time for eight different anticipation functions. The perfect reactive anticipation function (i.e. using TSPTW-EU with all rules) is abbreviated with PER, the approximate reactive anticipation function (i.e. using TSPTW-EU with basic rules) is abbreviated with APP, and the non-reactive

Table 1 Feasible instructions and run-time of anticipation functions

	PER	APP	IMP-58.5	IMP-50	PER-58.5	PER-50	APP-58.5	APP-50
Number of solved instances	222	222	221	198	201	177	169	96
Average CPU time (s)	146	57	11	15	142	126	53	48

anticipation function (i.e. using TSPTW with reduced average speed) is abbreviated with NON. We have tested two further non-reactive anticipation functions NON-58.5 respectively NON-50 by using 58.5 km/h, respectively 50 km/h for the reduced average speed. Reducing the value for speed provides more flexibility to the driver's operations. Obviously, the driver's flexibility does not depend linearly on the amount of speed reduction. That is why we decided to choose the value for the experiments with a medium speed to be closer to 65 km/h than to 50 km/h. In addition to the above anticipation functions we have considered hybrid anticipation functions which are based on a speed reduction for the PER-function and the APP-function: PER-58.5, PER-50, APP-58.5, and APP-50.

Of course, all generated test instances are solvable (i.e. yield feasible planner instructions) in case of using the perfect anticipation. Due to the design of the experiments, the only difference between perfect and approximate perfect anticipation is that perfect anticipation leads to an improvement of the average value of the objective function of the test cases. For our 222 test instances this improvement amounts to 3.9 %. For NON-58.5 the number of solvable instances comes very close to that of PER and APP while for NON-50 already 11 % of the test cases are insolvable. In case of hybrid anticipation functions the difference between perfect and approximate anticipation becomes evident. For a speed of 58.5 km/h, approximate anticipation only solves 84 % of the test instances solved by perfect anticipation, and APP-50 solves only 54 % of the instances solved by PER-50. The anticipation function with the least set of solvable instances is APP-50. Compared to PER and APP, the application of APP-50 only yields a percentage of 43 % solvable instances. The computation times required by the application of the considered functions differ a lot. For all functions which are applying perfect respectively approximate anticipation, the average CPU time is 138 s., respectively 53 s. Since the functions NON-58.5 and NON-50 imply solving the pure TSPTW without EC social regulations, they can be solved relatively fast (averagely in 13 s. CPU time).

In the second step of our experiments we analyze usefulness, i.e. the realizability, of the generated solutions in case of travel time deviations. Now we assume that at the time of tour fulfilment there will be some deviations of the actual driving times from the times which have been presumed by the planner during the planning phase. Further, we assume that the driver is aware of these deviations. This means that, compared to the planner, the driver has more detailed and up-to-date information, since he is attendant at the time when the route is actually executed. Based on refreshed information, the driver will do his break scheduling with his own optimization criterion. The optimization criterion applied by the driver will be called realization function in the following.

The driving time deviations are simulated by applying the Monte Carlo method. For the Monte Carlo method, the travel duration is specified in form of a Weibull distribution with shape-parameter k = 1.35, scale-parameter λ = 0.35 and a mean of 1.0. The resulting values scatter between 0.75 and infinite, while the number of events over 2.0 tends to be zero. These values are used as a factor of the expected driving times which results in a typical Rayleigh distribution for the simulated travel duration. Driving time deviations are simulated for each arc in the network and calculated before the driver's break scheduling is executed. If a simulation experiment reveals that an instruction which has been transferred by the planner to the driver cannot be feasibly scheduled by the driver due to the deviated driving times, then this route is considered to be not accomplishable. Please keep in mind that the planner's instruction is a route which has been successfully solved with the anticipation function of the planner and that the driver may have a realization function which differs from the planner's anticipation function. If it turns out that the transferred route can be feasibly scheduled by the driver in spite of the deviation of the driving times it is considered accomplishable.

In our experiments the sets of solvable problem instances are almost identical for PER, APP and NON-58.5 while for all other anticipation functions only a (more or less) small subset of the test cases can be solved successfully (see Table 1). That is why the experiments on the usefulness of the transferred routes (planner instructions) are restricted to the comparison of those three anticipation functions whose usefulness can be evaluated for (nearly) all 222 test cases. Additionally, APP-50 is considered to show the effects for a function with the smallest quota of solvable routes. Table 2 shows the number of test instances which turned out to be accomplishable for different anticipation functions and realization functions. The values in brackets present the quota of accomplishable routes to transferred routes. This quota can be considered as a measure for the average reliability of solutions which have been generated by a given anticipation function and scheduled by a given realization function. The case that the planner makes use of all rules of social legislation at a normal speed while the driver applies only the basic rules is omitted in Table 2 since this does not make sense in reality.

Since the experiments on the anticipation functions PER and APP are both performed on the same set of planner instructions (set of solvable instances) and since IMP-58.5 could be tested almost on the same set of instructions (only one test case is missing because it was not solvable with IMP-58.5), the quotas for accomplishable routes of these three functions can be directly opposed to each

Table 2 Number of accomplishable test instances

Realization	Anticipation				
	PER	APP	IMP-58.5	…	APP-50
PER	204 (92 %)	204 (92 %)	197 (89 %)	…	94 (98 %)
APP	–	186 (84 %)	179 (81 %)	…	94 (98 %)

other. The entries in Table 2 show that the quota for accomplishable routes (i.e. the average reliability of the planner's instructions) is always above 80 %. If the planner and the driver both apply PER, the average reliability amounts to 92 %; if they both apply APP, the reliability goes down to 84 % but if the driver applies PER for instructions built by means of APP the reliability goes up again to 92 %. The reliability of both, APP and NON-58.5 increases by 8 % in case that the driver uses the realization function PER instead of APP. All eight anticipation functions shown in Table 1 have been tested considering driving time deviations and different realization functions. The highest degree of reliability could be achieved if the planner applies the function APP-50. In that case the high reliability of 98 % is paid by the fact that only less than half (94/222) of the original test cases could have been feasibly solved in the planning phase.

5 Conclusions and Outlook

Explicit break scheduling is more complex than non-reactive break scheduling but allows for finding better results and, provided that there will be no deviations of the travel times, guarantees feasible break schedules for the resulting vehicle routes. Due to the design of our experiments, approximate anticipation is in the role of a reference function for the other anticipation functions. That is why perfect anticipation is slightly disadvantaged since it cannot fully exploit its potential for highly complex instances which are not solvable for approximate anticipation and the other investigated ways of anticipation. Nevertheless, in our experiments on explicit break scheduling, perfect anticipation (PER) achieves the same degree of reliability as approximate anticipation (APP) but is able to achieve a higher solution quality. Experiments with the approximate and non-reactive anticipation function show that it may help a lot if the driver uses a more accurate planning model than the planner. But, of course, the application of such a complex planning model like the route scheduling problem with all mandatory and optional rules of the EC regulation is a complicated decision process which cannot always be expected of the driver. As a next step, our research will concentrate on performing tests where none of the anticipation functions will be favored by being used as a filter. Then, research will focus on dynamic scenarios. In such scenarios the driver will not only perform his individual break scheduling once at the beginning of his tour. Instead, he can adjust his scheduling dynamically whenever new information on expected driving times is available. Already in the near future, it is expected that the dynamic re-scheduling of breaks will be performed by some intelligent features of trucks which are assisting drivers in their break decision making.

Acknowledgments While carrying out the research, the co-author Hendrik Braun was supported by the University of Western Australia and the Australian Government with an International Postgraduate Research Scholarship (IPRS), the Australian Postgraduate Award (APA) and the University of Western Australia Top-Up Scholarship.

References

Bartodziej P, Derigs U, Malcherek D, Vogel U (2009) Models and algorithms for solving combined vehicle and crew scheduling problems with rest constraints: an application to road feeder service planning in air cargo transportation. OR Spect Quan Approaches Manage 31(2):405–429

Goel A (2009) Vehicle scheduling and routing with drivers' working hours. Transp Sci 43(1): 17–26

Kok AL, Meyer CM, Kopfer H, Schutten JMJ (2010) A dynamic programming heuristic for the vehicle routing problem with time windows and european community social legislation. Transp Sci 44(4):442–454

Kopfer H, Meyer CM (2010) Ein Optimierungsmodell für die wöchentliche Tourenplanung unter Einbeziehung der EU-Sozialvorschriften. Zeitschrift für Betriebswirtschaft 80(7–8):755–775

Laux H, Liermann F (2005) Grundlagen der Organisation: die Steuerung von Entscheidungen als Grundproblem der Betriebswirtschaftslehre, 6th edn. Springer, Berlin

Meyer CM, Kopfer H, Kok AL, Schutten M (2011) Distributed decision making in combined vehicle routing and break scheduling. In: Kreowski H-J, Scholz-Reiter B, Thoben K-D (eds) Dynamics in logistics. Springer, Berlin, pp 125–133

Onken M (2009) Umsetzung der Lenk- und Ruhezeiten-Verordnung—eine empirische Studie, Diplomarbeit, University of Bremen, Bremen

Schneeweiss C (2003) Distributed decision making—a unified approach. Eur J Oper Res 150 (2):237–252

Solomon M (1987) Algorithms for the vehicle routing and scheduling problems with time window constraints. Oper Res 35(2):254–265

Windt K, Hülsmann M (2007) Changing paradigms in logistics—understanding the shift from conventional control to autonomous cooperation and control. In: Hülsmann M, Windt K (eds) Understanding autonomous cooperation and control in logistics. Springer, Berlin, pp 1–16

Zäpfel G, Bögl M (2008) Multi-period vehicle routing and crew scheduling with outsourcing options. Int J Prod Econ 113(2):980–996