International Series in
Operations Research & Management Science

Lluis M. Plà-Aragonés *Editor*

Handbook of Operations Research in Agriculture and the Agri-Food Industry

Springer

International Series in Operations Research & Management Science

Volume 224

More information about this series at http://www.springer.com/series/6161

Lluis M. Plà-Aragonés
Editor

Handbook of Operations Research in Agriculture and the Agri-Food Industry

Springer

Editor
Lluis M. Plà-Aragonés
Department of Mathematics Faculty
of Law and Economics
University of Lleida
Lleida, Spain

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-1-4939-2482-0 ISBN 978-1-4939-2483-7 (eBook)
DOI 10.1007/978-1-4939-2483-7

Library of Congress Control Number: 2015935569

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Foreword

The use of OR in Natural Resources has been with us for several decades now. It has been developed importantly in the areas of forestry, agriculture, and mining. The papers published in the literature show results that go from methodological developments to case studies to actual applications used in industry or government. One area where OR has shown particular success has been in forestry. This has been driven in large part by the fact that forest lands and firms are quite large in both area and financial resources. Thus, there is the possibility and incentive to invest in using sophisticated tools for management. We see in this special issue the fruits of similar developments mainly in the area of agriculture. This is so because Agriculture and the Agri-Food Industry are becoming capital intensive and a more complex business than traditionally was. The academic and practical use of quantitative tools to support decision making in agriculture is shown in its many facets. The papers and the problems in them represent the typical challenges managers face:

- Planning of planting and harvesting.
- Production and logistics.
- Optimizing the supply chain.
- Risks of catastrophic events.
- Hierarchical planning.
- Multiobjective decision making.

These are areas where it is well known that Operations Research has proven its high value, with its tools to solve problems, and perhaps more importantly, in the way it helps, defines the issues, and characterizes the problems to be solved. The use of OR tools has become more important given the increased specialization of the sector, higher competition in a globalized world, need to produce and distribute in an efficient way, and the huge improvement in software and hardware possibilities. Additional challenges imposed by environmental constraints, sustainable development, and healthier, safe and secure products add to the need of sophisticated decision making. The field of quantitative decision making in agriculture (and in lesser volume other areas) is significantly enriched by this special Handbook.

Santiago de Chile, Chile Andres Weintraub

Preface

Many real-world decision-making situations arise from agriculture and related agri-food industries such as fisheries, water management, and irrigation. Methods and applications in *Agriculture and the Agri-Food Industry* are of interest at present in research developments related to the globally critical area of food production, animal welfare, and sustainability and it is expected to increase in the future. Many treatments of this subject fail to describe why and how the concerned OR methods work effectively in the context of practice. The scope of this book is to provide an overview of Operations Research (OR) methods in agriculture *and* a thorough discussion of derived applications in the agri-food industry. Of course, this panoramic book does not claim to offer a detailed and exhaustive view of many OR approaches to agriculture and the agri-food industry. We therefore sought high-quality works from leading researchers in the field that fit with this general scope. As Editor, I'm quite pleased with the result, which has brought together a diverse blend of research topics and modelling and solution approaches for different decisions in agriculture or in the agri-food industry.

Structure of the Book

This book represents a set of stand-alone works that introduce several OR methodologies and captures recent research trends in the application of OR methods in agriculture and the agri-food industry. In this sense, the book can be read in different ways depending on the personal interest of the reader, and so, there is not a unique recommended order for reading the different chapters. On the other hand, I'm extremely grateful to the authors for their outstanding contributions and for their patience, which have led to a final product that far exceeded my expectations.

Table 1 Several characteristics of the chapters of the book

Chapter	OR method	Decision problem	Product	Level	Country
1	SP	Planning	Pig	Supply chain	Spain
2	LP	Planning	Horticulture	Supply chain	USA
3	LP	Planning	Seed corn	Supply chain	Brazil
4	LP	Planning	Apple	Orchard	Chile
5	LP	Planning	Sugarcane	Supply chain	Cuba
6	LP	Planning	Soil	Farm	Chile
7	LP	Planning	Fruit	Supply chain	Spain
8	SIM	DA	Fruit and vegetable	Supply chain	Australia
9	SIM	DA	Fish	Fish farming	Israel
10	MHEU	Planning	Fish	Aquaculture	Spain
11	MO	ALL	Water	Regional	Africa
12	RA	DA	Crop	Farm	Netherlands
13	FOR	DA	Grape	Farm	Australia
14	DEA	EA	Pig	Farms	Spain
15	AHP	Sustainability	Olive	Farm	Spain
16	LP	Location	Beef	Supply chain	Australia
17	LP	Planning	Pig	Farms	Spain
18	LP	Diet	Pig	Farm	Canada
19	MDP	Replacement	Pig and cows	Farm	Denmark

All chapters were rigorously reviewed and I would like to thank the anonymous reviewers for their quality reviews and responsiveness.

It has been difficult to be consistent with the use of the same criteria to decide and place one chapter after the other. However, the link or connection chapter by chapter is given sometimes by the method others by the problem or field of application. Hence, the book starts with seven chapters presenting different planning problems for different agricultural products. Afterwards, three chapters making use of simulation and metaheuristics follow, before a set of five chapters dealing with problems solved using multicriteria or multiobjective related methods are presented. The last four chapters of the book are devoted to singular livestock decision problems.

Table 1 shows different characteristics of each chapter in order to help readers organize the reading of the book. Different dimensions are used to classify the content of each chapter:

– *Methodology*: Most of the chapters present and develop mixed integer linear programming models (LP) including several integer variables. The use of commercial software to solve large LP models makes this kind of applications very interesting for practical purpose. However, the adoption of these OR solutions evolves little by little. The rest of the methods employed in the book are simulation (SIM), metaheuristics (MHEU), multiobjective programming (MO), risk analysis (RA), forecasting (FOR), data envelopment analysis

(DEA), multicriteria Analytic hierarchy process (AHP), stochastic programming (SP), and Markov decision process (MDP).
– *Decision problem*: Planning production is the problem with more applications in this book. Most of these problems are solved using LP models. Other examples of problems are decision analysis to assess either risk situations or just management alternatives, efficiency analysis, sustainability, location, the diet problem, and the replacement problem.
– *Agricultural product*: Agriculture produces a variety of products that most of them are presented in one or more chapters of this book. Pig is the most frequent product. There are also chapters dealing with fruits and vegetables, fish, seed corn, olive oil, beef, and horticulture products. It is worth mentioning that even when the described problems seem product-specific (e.g., replacement problem in pigs), the method behind has a wider application to other products/species (e.g., replacement problems in cows, sheep, or other livestock). A couple of chapters are devoted to other topics focused more on the management of natural resources like water and soil impacting on agricultural production. A chapter is devoted to soil management or how to define plots to maximize crop yielding. And another one is devoted to water management in some regions of Africa.
– *Level*: Although several problems are formulated and solved at farm level, applications at supply chain level are becoming more and more common. Furthermore, water management and risk analysis in agriculture are some agricultural problems dealt at regional or national level regularly.
– *Country*: Studies presented have been developed under specific conditions of a country that may be different country to country. There is a wide representation of applications developed in Europe and America, less in Australia and Africa, and unfortunately none from Asia.

The book is primarily a reference for researchers, Ph.D. students, instructors, and advanced practitioners. Depending on the technique, most chapters introduce briefly the method employed before tackling the agricultural problem presented. So, the book is also expected to be useful and appropriate for use as a textbook for certain advanced courses; and due to the interdisciplinary nature of the content, such courses may be taught in a variety of departments including Operations Research, Agriculture, Applied Mathematics, Agricultural or Agronomic Engineering, and Agricultural Economics or Ecosystems.

Lleida, Spain Lluis M. Plà-Aragonés

Acknowledgements

The editor is grateful to the editor at Springer, Camille, for her trust in this book. I'm also grateful to all members of the ORAFM working Group, all contributors to this issue, and the numerous reviewers of the papers in this special volume for carefully reading and commenting on the papers. Their dedicated volunteer work has had a significant impact on the quality of the papers. Finally, my gratitude and thanks to the staff at Springer-Verlag for facilitating administrative matters.

Contents

Chapter 1
Optimal Planning of Pig Transfers Along a Pig Supply Chain

Esteve Nadal-Roig and Lluis M. Plà-Aragonés

1.1 Introduction

A pig supply chain (PSC) is a complex process in which a group of several farms, such as breeding, rearing and fattening, and one or more abattoirs work together to produce pigs. Pigs are slaughtered and converted into pig meat to be distributed among retailers. This is the result of a transformation of the pork sector from the traditional farrow-to-finish farms to a bigger, more industrialized, controlled and efficient pig production systems (Taylor 2006; Nijhoff-Savvaki et al. 2012). Furthermore, concerns about environment, food quality and animal welfare are becoming the new challenges for the pig industry. Modern and intensive production of pigs is becoming more and more specialized. The size of facilities is increasing, and the production process is structured through three phases: the first phase focuses on producing piglets, the second phase focuses on rearing the piglets and the third and last phase focuses on fattening the pigs and delivering them to the abattoir. For each of these phases, a set of specialized farms (i.e. sow farms, rearing farms and fattening farms, respectively) are involved. As a result, private companies and cooperatives tend to integrate farms and abattoirs and coordinate their operations into pork supply chains by using tighter vertical coordination linkages (Rodríguez et al. 2014). Planning simultaneously pig production and transport of animals along the supply chain greatly advances the efficiency of both processes (Mula et al. 2010).

Thus, this chapter presents a general formulation of a stochastic mixed integer linear programming model with the aim to optimize the production planning of a pork supply chain based on a previous seminal proposal (Plà and Romero 2008). The model maximizes the total revenue of the chain. Income depends on animals sold to the abattoir and main cost summarizes animal feeding, doses of insemination,

E. Nadal-Roig (✉) • L.M. Plà-Aragonés
Departament de Matemàtica, Universitat de Lleida, Jaume II, 73, Lleida 25003, Spain
e-mail: enr1@alumnes.udl.cat; lmpla@matematica.udl.es

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_1

labour, transportation and veterinary expenses. A finite time horizon of 3 years is considered on a weekly basis. As a result, the proposed model provides the best solution for production planning, that is, the flow of animals among farms and towards the abattoir, the number of animals to be produced and transferred at each phase and stage, the number of trucks and optimal replacement policy for each sow farm, as well as the optimal delivery of fattened pigs to the abattoir. The formulation presented makes possible to envision new opportunities for operations research methods to be successfully applied to the pork supply chain management optimization (Plà et al. 2014). In this regard, we identify some extensions of the model that we plan to address in the future.

1.2 Modelling Pig Supply Chains

Although the literature in models for supply chain production and transport is vast (Mula et al. 2010), modelling of agricultural supply chains is not so extensive (Ahumada and Villalobos 2009). However, research dealing with pork supply chains agrees on the importance of planning pig production along the entire chain to coordinate productivity and quality improvement strategies (Perez et al. 2010). This is so because most of the literature to support the decision making on the pig sector have only been focused on operations on single farms, while the pork supply chain management involves the coordination of sets of farm units at different phases (Plà 2007; Plà and Romero 2008; Rodríguez et al. 2014). Hence, the modern structure of the pig sector, based on PSCs, requires the new modelling approaches to tackle actual problems. For instance, more than one farm per phase and more than one phase has to be considered. So far, modelling approaches for the pig industry had been developed to mainly improve the productivity of individual farms. Some of these studies made use of Markov decision processes and simulation models (Plà 2007) and focused on a sow farm which is reasonable since it is the most complicated part of the production process. Assumptions of the models imposed by researchers to avoid complexities reduced the interest to practitioners beyond strategic decisions. For instance, the homogeneity of parameters over time or the randomness of parameters like prices were not accompanied with updating methods allowing more precise results for short- or medium-term decisions (Rodríguez et al. 2014). Original strategies to cope with this situation have been presented, like Rodríguez et al. (2009) who considered some constraints aimed at the modelling of a sow farm embedded into a pork supply chain. Other authors (Plà and Romero 2008; Nadal-Roig and Plà 2014) proposed a mixed integer linear programming model to optimize the entire supply chain, taking into account the constraints of companies having the three phases.

The PSC considered in this chapter involves three different farms: sow, rearing and fattening farms (see Fig. 1.1). The PSC model assumes all the farms and the abattoirs are owned by the same company. The transport flow among the different farms including the abattoir, the load to be transported and the structure of the

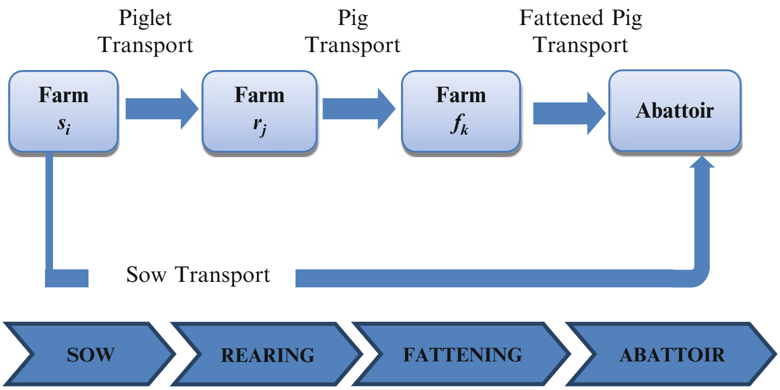


Fig. 1.1 Pig production system. Stages and transport

agents taking part in the PSC model are depicted in Fig. 1.1. Hence, according to that, the first phase produce piglets and takes place on sow farms. Sows are inseminated and are expected to become pregnant. If not, there are limited additional attempts until a successful conception happens, leading to a farrowing and subsequent lactation period. Otherwise, the sow is culled and sent to the abattoir for infertility reasons. The aim of sow farms is to wean the maximum number of piglets to be transferred to rearing farms. The second phase involves piglets transferred to rearing farms to be fed for a specific number of weeks until they reach a weight of around 20 kg. Finally, in the third phase, pigs are transferred to fattening farms. Fattened pigs are delivered to the abattoir once they have reached a marketable weight. Fattening farms are filled and emptied at a time with batches of animals, this is, the so-called all-in-all-out management system. This strategy has been demonstrated useful for disease prevention and control because avoid the contact between animals belonging to different batches and allows the farmer to sanitize the facilities.

1.3 General Formulation of the Model

1.3.1 Mathematical Background on Stochastic Programming

Stochastic linear models provide a suitable framework for modelling decision problems under uncertainty arising in several applications (Wallace and Ziemba 2005). Consider the following general form of a two-stage stochastic programming model that follows the Deterministic Equivalent Model (DEM) proposed in Birge and Louveaux (2011):

$$\text{(SP1)} \quad z_{\text{SP1}} = \min_{x, y_k} \quad c^{\mathrm{T}} x + \sum_{k=1}^{K} p_k q_k^{\mathrm{T}} y_k \tag{1.1}$$

$$\text{s.t.}: \qquad Ax = b \tag{1.2}$$

$$\begin{aligned} T_k x + W_k y_k = h_k \qquad \forall k \in \Omega \\ x, y_k \geq 0 \end{aligned} \tag{1.3}$$

where x is the n_x-vector of the first stage variables, which may include 0–1 variables; y_k is the n_y-vector of the second stage variables for scenario $k \in \Omega$, c is a known vector of the objective function coefficients for the first stage variables, b is the right hand side vector for the first stage constraints, A is the first stage constraint matrix, p_k is the likelihood of the scenario k, h is the right hand side vector for the second stage constraints, q_k^T is the vector of the objective function coefficients for the second stage variables, while T_k is the technology matrix and W_k is the recourse matrix under scenario k, $\forall\ k \in \Omega$.

The structure of the uncertain information in the two-stage stochastic linear model SP1 can be visualized as a tree, where each root-to-leaf path represents one specific scenario, ω, and corresponds to one realization of the whole set of the uncertain parameters linked at the first stage by the non-anticipativity constraints (Rockafellar and Wets 1991). In Fig. 1.2a, the scenarios are shown independently. Solving the problem for each scenario would produce wrong solutions. Thus, non-anticipativity constraints are added to force all the scenarios have the same first stage variables (Fig. 1.2b). The flexibility of these models is related to their multiperiod nature, i.e. besides the first stage variables that represent decisions made in face of uncertainty; the models consider second stage decisions, i.e. recourse actions, which can be taken once a specific realization of the random parameters is observed. Hence, the vector x represents the same decision at the first stage (St_1) for all scenarios while the remaining decision variables y_s are dependent of the corresponding scenario, $s \in S$.

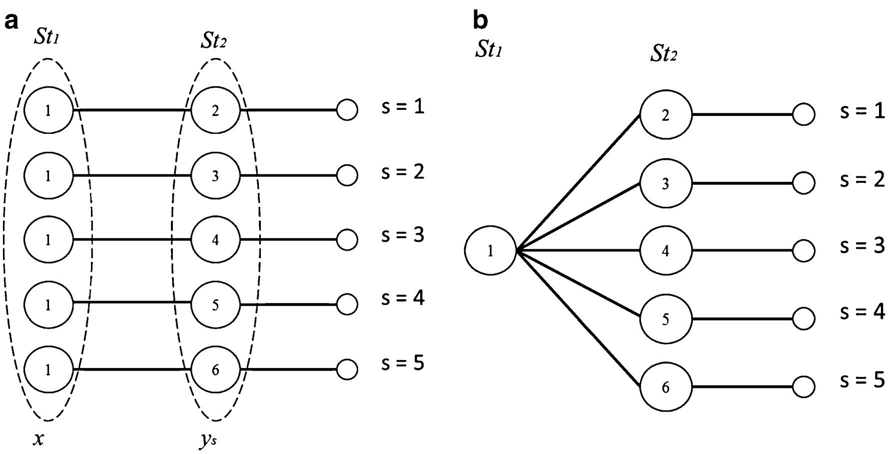


Fig. 1.2 Scenarios of a two-stage stochastic model. (**a**) Individual scenario representation. (**b**) Compact representation

1.3.2 Mathematical Formulation

In our approach, the uncertain parameters are those related with future sale prices. Uncertainty is represented in the model by a set of possible scenarios, S, with corresponding probabilities p_s.

A mixed integer linear programming model was developed to determine optimal planning of pig transfers along a PSC over a finite time period. To present the multiperiod formulation, the following notations were used:

Sets and Indexes

$S = \{s\}$ a finite set of scenarios.
$T = \{t\}$ a finite set of periods in weeks.
$T_1 \subset T$: a subset of T corresponding to the periods of the first stage.
$H = \{h\}$ the set of farms conforming the PSC.
$H = \mathrm{B} \cup \mathrm{R} \cup \mathrm{F}$ disjoint partition of farms, being B the set of sow farms, R the rearing farms and F the fattening farms.
$E = \{e\}$ set of growing stages of pigs expressed in weeks, from birthdate to the delivery to the abattoir.
$E = E_\mathrm{B} \cup E_\mathrm{R} \cup E_\mathrm{F}$ disjoint partition of growing stages of piglets (or pigs) housed in different facilities (B, R or E) being $\mathrm{E_B}$ the lactation period (from the birth to the weaning of piglets), E_R the rearing period (from weaning to the beginning of the fattening) and E_F corresponding to the fattening period.
X: set of physiological states in which sow lifespan is divided.
$X_a \subset X$ set of physiological states in the end of which sows are culled and sent to the abattoir, $\{a\}$.
$X_g \subset X$ set of farrowing states in the end of which farrowing take place and piglets born.
$W = \{w\}$ set of growing stages at the end of the fattening phase when pigs can be sent to the abattoir (marketing time window).

Parameters

IN_{he} initial inventory of pigs of age e and farm h.
K_h farm capacity in number of sows if $h \in \mathrm{B}$ or pigs if $h \in \mathrm{R} \cup \mathrm{F}$.
p_{ij}^{bs} transition probabilities of sows from i to j, with $i, j \in X$ in sow farm $b \in \mathrm{B}$ and scenario s.
LS_{nbts} litter size at parity n, on sow farm b, at week t and scenario s.
$TR_{hh^*ts} = C_{ts} d(h, h^*)$ cost of transport from h to another farm or to the abattoir at week t and scenario s, where C_{ts} is the unitary cost per km of a truck at week t and scenario s, and $d(h,h^*)$ distance in km from farm h to another farm or to the abattoir, $h^* \in H \cup \{a\}$.
Na_{hh^*ts} number of trips from h to another farm or to the abattoir at week t and scenario s.

$CSOW_{hits}$ unitary cost per sow on farm h, physiological state i, week t and scenario s including feeding, doses of insemination, labour and veterinary expenses.
EX_{hets} unitary cost in farm h per piglet/pig, at age e, week t and scenario s, including feeding, labour and veterinary expenses.
p_{hets} sale price per kg of sows ($e=0$; $h \in \mathrm{B}$) or pigs ($e \in \mathrm{W}$; $h \in \mathrm{F}$) sent to the abattoir at week t and scenario s.
ka_h load capacity per truck transporting animals from farm h to the abattoir.
kg_h load capacity per truck transporting animals from farm h to another farm.
π_{bits} steady state inventory of the total number of sows at physiological state $i \in X$ in the sow farm b at week t and scenario s.
D_{ts} minimum demand of the abattoir at week t and scenario s.
AW_{ets} average live weight of pigs at fattening stage e, week t and scenario s.
AW_{its} average live weight of culled sow at state $i \in X_a$, week t and scenario s.

Decision Variables

I_{hets} inventory of piglets on farm h, age e, week t and scenario s.
A_{hts} inventory of pigs on farm h, week t and scenario s to be transferred to the next stage in the chain.
A_{brts} inventory of piglets sent from b to r, at week t and scenario s.
A_{rfts} inventory of piglets sent from r to f at week t and scenario s.
A_{fets} inventory of pigs sent from f to the abattoir at fattening stage $e \in W$, at week t and scenario s.
Nka_{hts} number of trips from $h \in \mathrm{B} \cup \mathrm{F}$ to the abattoir at week t and scenario s.
Nkg_{h1h2ts} number of trips from h_1 to h_2 being either $h_1 \in \mathrm{B}$ and $h_2 \in \mathrm{R}$ or $h_1 \in \mathrm{R}$ and $h_2 \in \mathrm{F}$ at week t and scenario s.

Let us note that farms are of different types, then $H = \{\mathrm{B} \cup \mathrm{R} \cup \mathrm{F}\}$ and this partition of the farms' set is related to the age of pigs growing on them. More formally: $E \times H = \mathrm{E} \times \{\mathrm{B} \cup \mathrm{R} \cup \mathrm{F}\} = E \times \mathrm{B} \cup \mathrm{E} \times \mathrm{R} \cup \mathrm{E} \times \mathrm{F} = E_{\mathrm{B}} \times \mathrm{B} \cup E_{\mathrm{R}} \times \mathrm{R} \cup E_{\mathrm{F}} \times \mathrm{F}$, being $E = E_{\mathrm{B}} \cup E_{\mathrm{R}} \cup E_{\mathrm{F}}$ and $E_{\mathrm{B}} \cap E_{\mathrm{R}} \cap E_{\mathrm{F}} = \emptyset$. Therefore, without loss of generality, in what follows, the use of pairs (e,h) will refer only to $E_{\mathrm{B}} \times \mathrm{B}$ or $E_{\mathrm{R}} \times \mathrm{R}$ or $E_{\mathrm{F}} \times \mathrm{F}$.

1.3.2.1 Structure of the Objective Function

The objective of this model is to get the maximum benefit achieved by optimizing the production planning of the PSC from sow farms to the abattoir. This benefit is represented by the gross margin calculated by the summation of incomes from pigs sold to the abattoir minus the total amount of expenses (such as feeding, doses of insemination, labour and veterinary expenses) and the transportation cost incurred for each farm. The model is formulated in a weekly basis given most of the activities on farm, transportation between phases and to the abattoir occurs regularly at this time frame. Therefore, the objective function is the summation of the total gross margin weighted per scenario of each farm over the time horizon, gm_{hts}.

The gross margin per scenario farm and period is calculated from the income, v_{hts}, minus cost, c_{hts}, and hence:

$$\max\ z = \sum_{s\in S} p_s \sum_{h\in H}\sum_{t\in T} gm_{hts} = \sum_{s\in S} p_s \sum_{h\in H}\sum_{t\in T} (v_{hts} - c_{hts}) \tag{1.4}$$

where the income per scenario is the sale value of culled sows π_{bats} and fattened pigs A_{fets} sent to the abattoir according to the sale price and total pig weight at each marketable stage, that is: $v_{hts} = p_{h0ts} \cdot AW_{ts} \cdot \pi_{bats}$ if $h \in \mathrm{B}$ or $v_{hts} = \sum_{e\in W} p_{hets}\ AW_{ets} A_{hets}$ if $h \in \mathrm{F}$. Notice that $v_{rts} = 0$ because not marketable product is produced. The costs are computed as transport cost and the rest of costs including feeding, doses of insemination, labour and veterinary expenses:

$$c_{hts} = \sum_{h^*\in H\cup\{a\}} TR_{hh^*ts} Na_{hh^*ts} + \sum_{i\in X} CSOW_{hits}\pi_{hits} + \sum_{e\in E} EX_{hets} I_{hets} \tag{1.5}$$

Total transport cost per week and scenario is calculated according to the number of trips needed to transfer pigs from one farm, h_1, to another one, h_2, or to the abattoir, a. This total cost depends mainly on the distance between these farms, $d(h_1,h_2)$, in km, therefore:

$$\begin{aligned}\sum_{h\in H}\sum_{h^*\in H\cup\{a\}} TR_{hh^*ts} Na_{hh^*ts} &= \sum_{h\in H-R} TR_{hats} Nka_{hts} + \sum_{b\in B}\sum_{r\in R} TR_{brts} Nkg_{brts} \\ &\quad + \sum_{r\in R}\sum_{f\in F} TR_{rfts} Nkg_{rfts}\end{aligned} \tag{1.6}$$

1.3.2.2 Constraints of the Model

The different constraints affecting the planning of transfers along the PSC including deliveries to the abattoir can be formulated as:

$$\sum_{i\in X} \pi_{bits} \le K_b \qquad b\in B,\ t\in T,\ s\in S \tag{1.7}$$

$$\sum_{e\in E} I_{hets} \le K_h \qquad h\in H-B,\ t\in T,\ s\in S \tag{1.8}$$

$$\pi_{bits} - \sum_{j\in S} p_{ji}^{bs} \pi_{bjts} = 0 \qquad i\in X,\ b\in B,\ t\in T,\ s\in S \tag{1.9}$$

$$I_{b1ts} \le \sum_{n\in X_g\subset X} \pi_{bnts} \cdot LS_{nbts} \qquad b\in \mathrm{B},\ t\in T,\ s\in S \tag{1.10}$$

$$I_{he0s} = IN_{he} \qquad e\in E,\ h\in H,\ s\in S \tag{1.11}$$

$$I_{be+1ts} = I_{bet-1s} \qquad b\in \mathrm{B};\ e\in \mathrm{E_B}\backslash\{|E_B|\},\ \ t\in \mathrm{T}\backslash\{1\},\ \ s\in S \tag{1.12}$$

$$I_{re+1ts} = I_{ret-1s} \qquad r \in \mathrm{R};\ \ e \in \mathrm{E_R} \backslash \{|\mathrm{E_R}|\},\ \ t \in \mathrm{T} \backslash \{1\},\ \ s \in S \tag{1.13}$$

$$I_{fe+1ts} = I_{fet-1s} \qquad f \in \mathrm{F};\ \ e \in \mathrm{E_F} \backslash \mathrm{W},\ \ t \in \mathrm{T} \backslash \{1\},\ \ s \in S \tag{1.14}$$

$$I_{fe+1ts} = I_{fet-1s} - A_{fet-1s} \qquad f \in \mathrm{F};\ \ e \in \mathrm{W} \backslash \{|\mathrm{W}|\},\ \ t \in \mathrm{T} \backslash \{1\},\ \ s \in S \tag{1.15}$$

$$I_{r|E_B|+1ts} = \sum_{b \in B} A_{brt-1s} \qquad r \in \mathrm{R};\ \ t \in \mathrm{T} \backslash \{1\},\ \ \ s \in S \tag{1.16}$$

$$\sum_{r \in R} A_{brts} = A_{bts} \qquad b \in \mathrm{B};\ \ t \in \mathrm{T},\ \ \ s \in S \tag{1.17}$$

$$I_{f|E_R|+1ts} = \sum_{r \in R} A_{rft-1s} \qquad f \in \mathrm{F};\ \ t \in \mathrm{T} \backslash \{1\},\ \ s \in S \tag{1.18}$$

$$\sum_{f \in F} A_{rfts} = A_{rts} \qquad r \in \mathrm{R};\ \ t \in \mathrm{T},\ \ \ s \in S \tag{1.19}$$

$$A_{f\left|\mathrm{E_F}\right|ts} = I_{f\left|\mathrm{E_F}\right|ts} \qquad f \in \mathrm{F};\ \ t \in \mathrm{T},\ \ \ s \in S \tag{1.20}$$

$$\pi_{bits} \leq Nka_{bts}ka_b \qquad b \in \mathrm{B},\ \ i \in X_a,\ \ t \in T,\ \ s \in S \tag{1.21}$$

$$A_{fets} \leq Nka_{fts}ka_f \qquad f \in \mathrm{F},\ \ e \in W,\ \ t \in T,\ \ s \in S \tag{1.22}$$

$$A_{h_1h_2ts} \leq Nkg_{h_1h_2ts}ka_{h_1} \qquad h_1 \in \mathrm{B} \cup \mathrm{R},\ \ h_2 \in \mathrm{R} \cup \mathrm{F},\ \ t \in T,\ \ s \in S \tag{1.23}$$

$$\sum_{f \in F} A_{fts} \geq D_{ts} \qquad f \in F,\ \ t \in T,\ \ s \in S \tag{1.24}$$

All facilities have a limited capacity. The capacity in sow farms (1.7) depends on the number of sows that can be housed while in rearing and fattening farms (1.8) depends on the maximum number of pigs that can be fed at a time. The abattoir is big enough to accept all pigs produced weekly, so there is no need to limit abattoir capacity, although it would also be possible depending on the case study.

It is assumed that sow farms are operating under a steady state derived from the herd structure at equilibrium (1.9). This is because sow herd dynamics is modelled as a Markov Decision Process (Plà et al. 2009). The number of piglets born alive weekly will depend on the number of sows at farrowing, being $X_g \subset X$ the subset of reproductive states of a sow with a farrowing, and the averaged litter size (1.10). All farms have an initial inventory of piglets or pigs at the beginning of the planning horizon (1.11). This initial inventory affects the flow of animals along the chain in the succeeding weeks and over the time horizon period which is being considered. Pigs which are fed on farms grow from one stage to the next one. We assume that all pigs are fed under the same regime and kept in groups of the same age. Each group grows accordingly to their age and the average live weight, consumption and mean daily gain is known for calculation. Therefore, the inventory must reflect this changing situation week by week over the time horizon. Inventory constraints can be stated for each phase of the supply chain (1.12)–(1.14). Additional constraints are added to represent the time window for marketing fattened pigs representing that not all pigs reach at the same time a marketable weight (1.15). No casualties are considered during the growing process. They could be taken into account when animals are transferred to the following phase in the chain or these constraints could be relaxed by using inequality constraints. The number of piglets to be transferred to the rearing or fattening farms has to be entered the same week. After completing the expected time for the phase, all of them exit also at the same time. For this reason, piglets sent to rearing farms cannot exceed the total number of piglets weaned (i.e. of age $|E_{\mathrm{B}}|$) nor do the pigs starting the fattening phase exceed the number of pigs finishing the rearing phase (1.16)–(1.17). Similarly, this also happens with piglets reared (1.18)–(1.19) and ready to be transferred to fattening farms (i.e. of age $|E_{\mathrm{R}}|$) and pigs fattened (1.20) and ready to be delivered to the abattoir (i.e. of age $|E_{\mathrm{F}}|$). Furthermore, a minimum capacity of the batch (lower bound) could be fixed complementing the upper bound represented by the farm capacity.

Constraints affecting transportation are related to the capacity of each truck. Animals sent to the abattoir are heavier than those transferred between farms and so, different capacities or trucks may apply. Hence, (1.21)–(1.22) represents the number of trucks used to transport culled sows or fattened pigs to the abattoir, respectively, and (1.23) the number of trucks needed to transfer animals among farms. Optionally, a minimum weekly demand to assure some level of operation at the abattoir can be stated by (1.24).

1.3.2.3 Non-anticipativity Constraints

Notice that constraints (1.7)–(1.24) represent s independent scenarios (see Fig. 1.2). We must define the non-anticipativity constraints linking the different scenarios by fixing the same decision variables at the first stage of the model. Hence, the following constraints are added for such purpose:

$$I_{hets} = I_{het1} \qquad h \in H;\ \ e \in E;\ \ t \in T_1,\ \ s \in S \tag{1.25}$$

$$A_{hts} = A_{ht1} \qquad h \in H;\ \ t \in T_1,\ \ s \in S \tag{1.26}$$

$$A_{brts} = A_{brt1} \qquad b \in B;\ \ r \in R;\ \ t \in T_1,\ \ s \in S \tag{1.27}$$

$$A_{rfts} = A_{rft1} \qquad r \in R;\ \ f \in \mathrm{F};\ \ t \in T_1,\ \ s \in S \tag{1.28}$$

$$A_{fets} = A_{fet1} \qquad f \in \mathrm{F};\ \ e \in \mathrm{W};\ \ t \in T_1,\ \ s \in S \tag{1.29}$$

$$Nka_{hts} = Nka_{ht1} \qquad h \in H;\ \ t \in T_1,\ \ s \in S \tag{1.30}$$

$$Nkg_{h1h2ts} = Nkg_{h1h2t1} \qquad h_1 h_2 \in H;\ \ t \in T_1,\ \ s \in S \tag{1.31}$$

1.4 Computational Results

1.4.1 Model Setup and Basic Case

In order to illustrate the suitability of the deterministic model resulting from the consideration of one scenario and the corresponding stochastic extension when considering several scenarios at a time a case study is presented. Basic parameters of the study were taken from standard values under Spanish conditions and recorded in the BD-Porc databank (national record keeping system hosted at http://www.irta.es/bdporc/, Accessed 7 Aug 2014), and do not correspond to a specific farm. The total set of farms per phase owned by a theoretical vertical integrated company are four sow farms, four rearing farms and eight fattening farms plus one abattoir. Since there are three types of origins when collecting animals (sow, rearing and fattening farms) and three types of destinations to deliver them (rearing and fattening farms plus the abattoir), two transfers are feasible between farms of different type (sow to rearing or rearing to fattening farms) and the rest of transports are directed to the abattoir (sow and fattening farms). The transportation cost from sow farms to the abattoir is considered and corresponds to culled sows. The entire set of parameters data is summarized in Appendixes 1, 2 and 3. For all the farms, parameters like farm capacity and initial inventory are required. For simplicity, the herd size of each sow farm will be taken as a parameter combined with the steady state of the herd structure determining accordingly the associated piglet production. Thus, sow farms operate at a constant rate of occupancy of lactation facilities. The road distances between farms and between farms to the abattoir are also required. The inventory of sow, rearing and fattening farms is given in number of piglets per lactation, growing or fattening stage, respectively. The abattoir requires a minimum pig demand for this kind of chains, where the product pushes along the chain instead of being pulled by demand. No risk of overflow capacity of the abattoir is considered because slaughtering capacity is larger enough to slaughter all pigs produced. The sale price is based on the historical series recorded by Mercolleida, the main Spanish auction market for pigs (http://www.mercolleida.com/mercados-ganaderos/porcino/, Accessed 14 Mar 2013). Two different series are considered depending on the meat quality of the animal namely whether they come from the fattening farms or from sow farms, that is, commercial pigs or culled sows. Other considerations regarding the value of pigs include carcass classification depending on lean percent, carcass weight and back fat thickness (SEUROP classification is mandatory in EU abattoirs). Sows and pigs are valued assuming experimental distributions of these traits.

The finite time horizon is set to 3 years. The maximum number of parities cycles for sows is nine. According to usual practices in Spain, lactation period and both rearing and fattening stages for piglets are set to 4, 6 and 18 weeks, respectively, as maximum. Capacity of trucks, weight of animals and unitary transport cost are taken into account according to the age of animals transported and distance conveyed. The number of available trucks is not taken into consideration explicitly, only the number of trips required for transportation.

To develop the model, the modelling language IBM ILOG OPL has been used. The solver CPLEX v12.2 solved the model in a laptop computer (Pentium Dual-Core CPU at 2.1 GHz and 4 Gb RAM). Microsoft Excel has been used for storing data, both input parameters and outputs, for its ease of use and flexibility to manage data. The integration into an Enterprise Resource Planning (ERP) can enable a simple adoption of the system by any company through the maintenance and update of an XLS file with the list of parameter like the inventory of animals for each farm, real prices and unitary costs considered by the model.

1.4.2 Basic Example: Deterministic Model

The deterministic model is built reducing the set of scenarios to one, $|S| = 1$, and then the non-anticipativity constraints (1.25)–(1.31) are not necessary. Strictly speaking, decision variables of the proposed model representing the number of animals are integer and non-negative variables. However, the computational time consumed for calculations in preliminary tests when all decision variables related to animals were considered integer, made the pure integer model inappropriate for practical purposes (Rodríguez et al. 2009). Furthermore, the loss of precision considering these variables as real was neglectable. As a consequence, only decision variables related to number of trucks and trips were considered integer relaxing the integrality condition for the rest. Beyond this relaxation, the first stage represents the roller horizon where decisions must be implemented before new environmental changes could be appreciated or taken into account for an update of the solution.

Specific parameters of the linear programming model are detailed in Appendixes 1, 2 and 3. Figures corresponding to the size of the deterministic model are presented in Table 1.1. The maximum average reward of the represented PSC was 3,231 thousands of euros per year with an overall occupancy of the available facilities of 97 %. Moreover, optimal solution provided the scheduling of the number of piglets and pigs to be transferred week by week and the way (from where to where). The occupancy rate was more than 0.97 over the time horizon of planning and never reached the full occupancy. Inspecting the solution was discovered a rational behaviour related to the preferred usage of the nearest farms to the abattoir and so, reducing cost transport while keeping the rest of the operational cost constant.

Sensitivity analysis. To prepare the extension of the model into a stochastic linear programming model and also to value the impact of the uncertainty of model parameters on the optimal solution, two additional cases were considered. The optimistic case, where the sale price paid by the abattoir was increased at 5 %, and the pessimistic case, where the sale price was reduced by 5 %.

The rest of parameters of the model concerning the productivity of the system are maintained. In all cases, an average of 473 piglets produced every week was considered. The system always produces as much piglets as possible regardless of the sale price. The average occupancy of the system is also the same. The variations over time in the sale price and the marketing window have a limited impact on the technical performance of the system not sufficiently relevant on average. Despite this, the farms' occupation taken individually varies considerably from farm to farm, and also affecting directly the farmers' revenue. Figure 1.3a, b show an example of two fattening farms' occupancy throughout the time horizon. Changes

Table 1.1 Report of the size of the deterministic model

CPLEX	Value
Variables	68.057
Constraints	47.655
Non-zero-coefficients	199.088

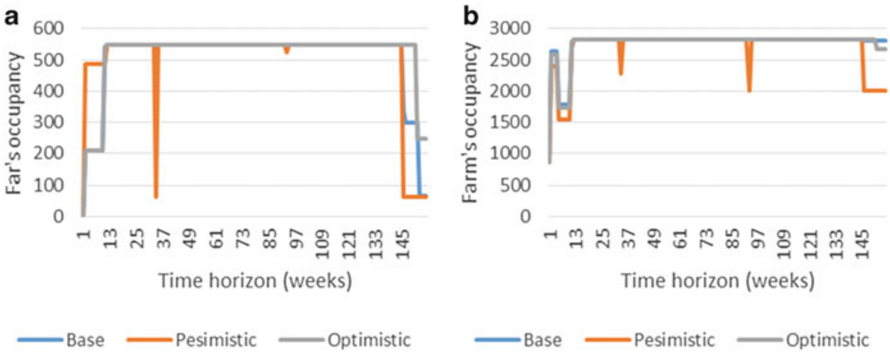


Fig. 1.3 Representation of two fattening farm's occupancy with capacity of 600 (**a**) and 3,000 pigs (**b**). Models: base, optimistic (+5 %) and pessimistic (−5 %) sales prices

Table 1.2 Economic indicators for the three cases in thousands of €

	Base	Pessimistic	Optimistic
Cost of animals	10,755	10,706	10,782
Difference vs. Base		−0.45 %	0.26 %
Cost of transport	38	38	38
Difference vs. Base		−0.33 %	0.29 %
Total cost	10,793	10,745	10,821
Difference vs. Base		−0.45 %	0.26 %
Income	20,488	19,432	21,843
Difference vs. Base		−5.16 %	6.61 %
Benefit	9,695	8,687	11,022
Difference vs. Base		−10.40 %	13.69 %

in the farm occupation can be observed, in particular at the beginning and at the end of the time horizon due to the initial and final inventory. In both cases, the occupancy rate is high.

While technical operation of the PSC is slightly affected by the scenario (either optimistic or pessimistic), it is not the same regarding the economic performance. Main economic outcomes are shown in Table 1.2. Economic indicators such as the benefit and income increase according to the sale price while cost remains almost the same showing a decreasing trend in the pessimistic case and increasing trend in the optimistic case. These variations are also related to the marketing window in which the model tries to achieve the higher benefit by selling the animals at the best price. The sales prices do not affect the production planning committed unless they are lower enough to force the system to not produce piglets. As is shown in Table 1.2 changes of 5 % in the sale price provoke changes of more than the 10 % in the benefit. The overall benefit ranges from 8,687 to 10,721 thousands of euros. Therefore, uncertainty seems to have an important impact on economic results of the whole PSC.

1.4.3 Basic Example: Stochastic Model

Stochastic model formulation requires the generation of a set of scenarios S. In that case, the full model has to include the non-anticipativity constraints (1.25)–(1.31) linking all the scenarios. To illustrate and assess the suitability of the stochastic approach, three scenarios were defined in this example. Again, the sale price as uncertain parameter was considered to be modelled by scenario. Therefore, the optimistic, normal or standard and pessimistic scenarios were defined in correspondence with the values of high, average and low sale prices, respectively. Time horizon was of 152 weeks as with the deterministic example and $T_1 = \{1\}$.

The resolution of this formulation gives an optimal profit (RP) of 3,235 thousands of €/year. The results confirmed also a globally high rate of occupancy. However in that case, a different behaviour for each scenario is observed and reveals the lower occupancy in the pessimistic scenario. The optimistic scenario reached the maximum occupancy of the PSC sooner, and it was maintained more weeks over the time horizon (Fig. 1.4).

Concerning the sales behaviour (see Fig. 1.5) shows how the scenarios tend to take advantage of the marketing window making the sales not steady. Furthermore, the pessimistic scenario shows a singular capability to sell more animals than the rest of scenarios at some weeks to maximize the income.

In addition, just to analyse the importance of the time horizon and final inventory on the outcome of the first 52 weeks different instances for $T = 78$, 104, 130 and 156 were solved. It was observed (data not shown) that the time horizon has a very little influence on the first 52 weeks because in all instances the objective function never reported differences greater than a 0.08 %. Even less is the impact on the

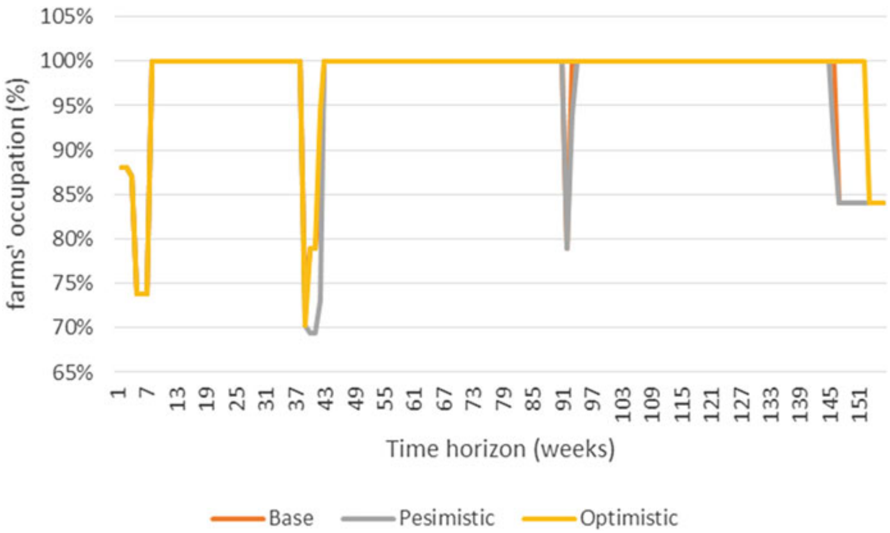


Fig. 1.4 Representation of the behaviour of the occupancy rate of the PSC with three scenarios

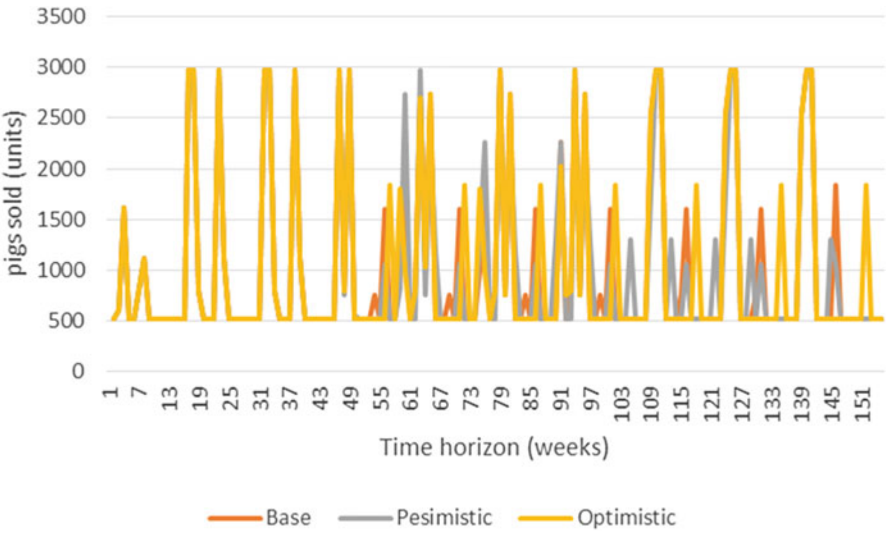


Fig. 1.5 Representation of the sales behaviour regarding three scenarios

expected profit for the first stage period represented by the first week (less than 0.02 % in the worst case).

Another aspect of interest was to see the impact of different number of weeks considered in the first stage. The reason was to consider if the production planning including the transfer of animals could be done biweekly or monthly. Therefore, new instances were solved for a different range of weeks in the first stage. The increment of weeks in the first stage showed a linear reduction in the profit in agreement with the loss of variability represented.

Inspecting the solution of the deterministic models with respect to the stochastic one, with the first 4 weeks as the first stage, we compute the expected value of perfect information (EVPI), defined through the following expression:

$$\mathrm{EVPI} = \sum_{s \in \Omega} p_s \Phi^s - RP$$

being Φ^s the optimal value of the deterministic model when it was solved (separately) for each scenario s in Ω and RP the optimal value of the stochastic model. For our study, the EVPI = 9,801–9,707 = 94 thousands of euros. EVPI measures the value of knowing the future with certainty. This is how much the farmer would be ready to pay this year to obtain perfect information about the dynamic behaviour of future sale price.

Additionally, the Value of the Stochastic Solution (VSS) was computed. Roughly speaking, it measures how good or bad results to use the optimal solution of the stochastic model instead of the deterministic one. Then, the VSS is defined as VSS = RP−EEV, where EEV is the expected value assuming expected yields and expected parameters fixing the optimal values at the first stage. In our case, the $\mathrm{VSS} = 9{,}707 - 9{,}663 = 44$ thousands of euros, this is the cost of ignoring uncertainty in choosing a decision.

1.5 Conclusions and Future Work

Despite the advantages of the stochastic solution shown in the previous section, the preliminary results of the model themselves indicate opportunities for improvements, mainly in two areas. First one, the management of an important amount of farms involved in a PSC by a better coordination among them; second one, the required relaxation of the integrality condition for several variables reducing the computational time and making feasible and possible the use of the model in practical condition for a PSC company.

The practical extension of the model considering more breeds and other sanitary constraints to fit particular PSC companies will make the model more complex. Hence, the resolution of such instances will require more computational power and/or the parallelization of the model. Our contribution then, emphasizes the importance and complexity of new decision-making tasks regarding the modern organization of the pork sector, rationalize the flow of animals over the chain providing a planning tool capable of updating the flow conveniently anticipating changes or reacting face to them.

Finally, the presented model is flexible, allowing a deterministic or stochastic formulation. The stochastic version can deal with the uncertainty of some parameters like the sale price and complemented with a more accurate growth and reproductive performance modelling like litter size, mortality rate or culling rates, but also likely changes in feed cost, labour, medicines, etc.

Appendix 1: General Parameters

Parameter	Value
Farms (in units)	
Sow farms	4
Rearing farms	4
Fattening farms	8
Time horizon (in weeks):	156
Sows' physiological states:	10
Production stages for piglets/pigs (in weeks)	
Sow farms	4
Rearing farms	6
Fattening farms	18
Transportation capacity (in units):	
From sows to rearing farms	700
From rearing to fattening farms	700
From sows/fattening farms to the abattoir	240
Animal cost (in Euro/week)	
Sows	4,85
Piglets in sow farms	1,874
Piglets in rearing farms	2,66
Piglets in fattening farms	4,382
Transportation cost (euro/trip)	1

Appendix 2: Capacity of Farms

Farm #	Type	Capacity (in units)	Initial stock (in units)
1	Sow	1.200	1.125
2	Sow	600	492
3	Sow	1.450	1.309
4	Sow	2.400	2.155
5	Rearing	300	101
6	Rearing	800	759
7	Rearing	2.800	2.633
8	Rearing	3.000	1.537
9	Fattening	1.200	228
10	Fattening	200	82
11	Fattening	548	2
12	Fattening	360	310

(continued)

(continued)

Farm #	Type	Capacity (in units)	Initial stock (in units)
13	Fattening	1.663	18
14	Fattening	200	116
15	Fattening	1.208	582
16	Fattening	2.834	865

Appendix 3: Distances Among Farms and Between Farms and Abattoir

	Distance from sow farm # (km)			
To rearing farm #	1	2	3	4
1	18	5	36	33
2	14	6	42	34
3	48	31	36	5
4	18	5	36	33

	Distance from rearing farm # (km)			
To fattening farm #	1	2	3	4
1	204	199	186	204
2	0	7	36	0
3	52	56	34	52
4	49	55	46	49
5	45	46	10	45
6	45	45	9	45
7	40	44	24	40
8	51	55	36	51

From fattening farm #	To Abattoir (km)
1	205
2	5
3	47
4	44
5	41
6	40
7	35
8	45

References

Ahumada O, Villalobos JR (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 196(1):1–20

Birge JR, Louveaux FV (2011) Introduction to stochastic programming, 2nd edn. Springer, New York

Data tables. Downloadable from the webpage https://cv.udl.cat/x/3YD0x3. Accessed 20 June 2013

Mula J, Peidro D, Díaz-Madroñero M, Vicens E (2010) Mathematical programming models for supply chain production and transport planning. Eur J Oper Res 204:377–390

Nadal-Roig E, Plà LM (2014) Multiperiod planning tool for multisite pig production systems. J Anim Sci 92:4154–4160

Nijhoff-Savvaki R, Trienekens JH, Omta SWF (2012) Drivers for innovation in niche pork netchains: a study of UK, Greece and Spain. British Food J 114:1106–1127

Perez C, de Castro R, Simons D, Gimenez G (2010) Development of lean supply chains: a case study of the Catalan pork sector. Supply Chain Manage 11:271–280

Plà LM (2007) Review of mathematical models for sow herd management. Livest Sci 106:107–119

Plà LM, Romero D (2008) Planning modern intensive livestock production: the case of the spanish pig sector. International Workshop in Operations Research. La Havana, Cuba

Plà LM, Faulìn J, Rodríguez SV (2009) A linear programming formulation of a semi-Markov model to design pig facilities. Journal of the Operational Research Society 60:619–625

Plà LM, Sandars D, Higgins A (2014) A perspective on Operational Research prospects for agriculture. J Oper Res Soc 65:1078–1089

Rockafellar RT, Wets RJ-B (1991) Scenario and policy aggregation in optimisation under uncertainty. Math Oper Res 16:119–147

Rodríguez SV, Albornoz VM, Plà LM (2009) A two-stage stochastic programming model for scheduling replacements in sow farms. TOP 17:171–189

Rodríguez SV, Faulin J, Plà LM (2014) New opportunities of Operations Research to improve pork supply chain efficiency. Ann Oper Res 219:5–23

Taylor DH (2006) Strategic consideration in the development of lean agri-food supply chains: a case study of the UK pork sector. Supply Chain Manage 11(3):271–280

Wallace S, Ziemba WT (eds) (2005) Applications of stochastic programming. Society for Industrial Mathematics, Philadelphia

Chapter 2
Planning the Planting, Harvest, and Distribution of Fresh Horticultural Products

Nicholas Mason, Héctor Flores, J. René Villalobos, and Omar Ahumada

2.1 Introduction

Many significant advances have been made in agriculture over the past century; we now have the ability, in theory, to feed the entire world population. Nonetheless, "there still remain great hunger, health and environmental concerns remaining to be addressed" (Hazell and Wood 2008). These are not problems that can be solved simply by increasing agricultural production (Alexandratos and Bruinsma 2006), especially considering the environmental issues that if left unchecked could adversely affect food supply in the future (M.E.A. 2005). Even though increasing agricultural yields and developing better varieties have great importance, a significant breakthrough can be made through better management of agricultural supply chains. The potential for better resource efficiency should not be overlooked, especially in view that one-third of the food produced for human consumption is estimated to be lost or wasted globally; some of the loss can be attributed to a lack of coordination of the different actors of the supply chain (Gustavsson et al. 2011). Thus, the issue of how to efficiently meet the demand with the production available is of utmost importance when high levels of perishability are present in the underlying product, as it is the case in fresh produce. For this reason the planning of supply chain will play an increasingly key role in the definition of those products that are successfully marketed.

The supply chains of agricultural commodities are already going through a transformation that will affect the welfare of US farmers, the economy, and also the health of the overall population. This transformation will result in the redefinition of roles for current players in the supply chain and the level of control they

N. Mason • H. Flores • J.R. Villalobos (✉) • O. Ahumada
InternationalLogistics and Productivity Improvement Laboratory, Arizona State University, Tempe, AZ, USA
e-mail: rene.villalobos@asu.edu

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_2

maintain in their particular operations (Cook 2011). Nonetheless, in order to address the global challenges ahead of us and to keep up with the changes occurring in agricultural supply chains, greater efficiency must be achieved by all parties involved. It is because of this strong need to increase the efficiency of the supply chain that planning models will become of increasing importance to farmers, intermediaries, and final distributors of agricultural commodities. Planning tools and information technology for each of these key players must become increasingly refined and applied, in order to drive non-value-adding costs out of the value chain.

In this chapter we will give a brief historical overview of planning and optimization models in agricultural supply chains, together with an analysis of the new tendencies observed in the market and consolidation of supply chains, which will be discussed in Sect. 2.1. The second section will give a brief literature review of the most recent models made for planning activities within the supply chain. The third section in this chapter will make use of specific mathematical optimization models and planning methods to illustrate the functionality and importance of planning tools for agri-food supply chains (ASCs); this section will be based in a bottom-up approach, starting from the perspective of the grower, moving up to the marketing and distribution of products to finally give a whole-chain perspective. Finally, we will provide a brief discussion of the planning models, the current gaps in research, and the opportunities for improvement.

2.1.1 What Makes Fresh Horticultural Supply Chains Different?

The term agri-food supply chains (ASCs) has been used to describe all activities from production to distribution that bring agricultural products from the farmland to the table (Aramyan et al. 2006); we will use the term ASC to refer to the conjunction of these activities. The supply chain of agri-foods, as any other supply chain, is a network of organizations and individuals working together in different processes and activities in order to bring products and services to the market (Christopher 2005). However, in addition to the problems common to most supply chains, ASCs must also deal with factors such as food quality, safety, and weather-related variability (Salin 1998). ASCs must also manage issues related to limited shelf life, which restricts the amount of time that most products can spend in storage and therefore the capacity of holding inventory as a buffer for variability (Makeham and Malcolm 1993). What is more, compounding the issues of variability and perishability, we have very long lead times from the moment that planting is made, until harvest (Lowe and Preckel 2004).

Among the problems encountered in ASCs, perishability is particularly critical for horticultural products, whose shelf life is significantly lower than that of traditional crops. Moreover, we must remember the greater economic context of the farm business and the position of farmers who are subject to the forces of the market and have little control over prices and the exact timing and yields of their crops (Ahumada and Villalobos 2009a; Makeham and Malcolm 1993), all of this while working with relatively small profit margins (Lowe and Preckel 2004). Complexity is further compounded when the supply chain has a large amount of stakeholders at various echelons (such as farmers, shippers, and distributors) that must coordinate their actions to avoid losses caused by a mismatch of supply and demand (Kader 2002).

2.1.2 New Tendencies in ASCs

The complexity level of agricultural supply chains has seen a dramatic increase over the past few decades. The increased competition and sophistication of the supply chain has forced newer innovations in the marketing and commercialization strategies of its players. No other fresh produce industry has seen such complexity increase as the European. Over the past few decades, the complexity and competitiveness of its food market industry has slowly consolidated the value chain of the products at both the producer and the retail side. For example, up until the early 1970s, multiple food retailers in Britain each had combined market share of 20 %, trailing both the cooperative and independent sectors (Morelli 1999). However, by 1971, the multiple sectors had 44 % of the food market overtaking the cooperative and independent sectors (Morelli 1999). As of 2009, the four largest food retailers in Britain accounted for 75.6 % of the total grocery sales (Garcia 2007).

External factors have been an indirect catalyst for the changes seen in the European produce industry. Consumers have become more aware and increasingly concerned over the quality and standard of food products, specifically fresh produce. This in turn has translated to higher quality standards for food retailers, their suppliers, and more importantly, farmers. In this case, the latter was left with the responsibility of revising their operational and organizational tools in order to improve the quality and reliability of the products and meet the demands of the end consumer. To face these challenges, investments were made in areas such as infrastructure improvements, quality control programs, and integration of value-added practices; in general, operations were made possible by the implementation of centralized platforms.

The magnitude of these investments was often too large for individual farmers to handle and thus many had to resort to alliances and partnerships. In the case of the European farmers, the formation of cooperatives became the best solution. Individual farmers formed alliances and created their own cooperatives, which they used as platforms to launch more complex operations. Through these platforms, the farmers were able to offer added value to their products, such as repackaging, processing, maturation, etc., in order to differentiate their products from typical commodities. Furthermore, they were able to coordinate logistics operations on a grander scale, and more importantly, the farmer-owned cooperative structure gave them a greater amount of leverage to their negotiating position. Today, private cooperative structures have allowed European farmers to become primary players in the global fresh produce industry, as it provides a way for farmers to innovate and maintain a high quality on products and service standards. They have also become role models for farmer organizations in other regions of the world in their hopes to develop sophisticated supply chains relative to those in Europe.

In the United States, the fresh produce industry has not yet reached the level of maturity as the one observed in Europe. While the general tendency of the market place is toward more complex and dynamic structures, the domestic conditions still allow independent farmers to compete effectively. Nonetheless, the transition toward vertically integrated supply chains has been a model to follow for many farmer organizations. Since the enactment of the North American Free Trade Agreement (NAFTA) in 1994, the commercial boundaries between the United States, Canada, and Mexico liberated many of the commercial hurdles that had been present in earlier decades. This allowed the increase of agricultural exportations into the United States, which led to a dramatic increase of the Mexican presence in the domestic fresh produce industry. All these developments have created a new playing field for all the parties involved in the industry.

More recently, the US market has shown signs of consolidating operations further in the case of the large-scale farmers, who are capable of outcompeting other smaller and midsize firms (Diamond and Berham 2012). As a response, small-scale farmers have capitalized on growing consumer interest on food provenance. Meanwhile, midsize firms which are too big to have direct marketing to the consumer, yet too small to compete in price and variety with the larger producers (Stevenson 2008), have responded through searching for more direct sales from farms to retailers while achieving some degree of product differentiation and supply chain collaboration (Diamond and Berham 2012).

2.1.3 Why Should We Focus in the Optimization of ASCs?

In the United States, the recent consumption growth of fresh agricultural products has been impressive, with a consumption increase of nearly 25 % from the years 1970 to 1997 (Jones Putnam and Allshouse 1999), and although the per capita consumption in recent years has not grown in a significant manner (Stewart 2010), the projected per capita increase in consumption by 2020 may be higher than 7 % (Blisard et al. 2003).

Supplying for the increased demand of fresh fruits and vegetables in the United States and the rest of the world is a challenge since the current supply chains might not be ready to deliver the quality and quantity at the time needed. For instance, in a study by the Food and Agriculture Organization (Gustavsson et al. 2011), it was estimated that about one-third of the food produced for human consumption is lost or wasted globally, which amounts to about 1.3 billion tons per year. The same study estimated that the per capita food waste in Europe and North America was about 95–115 kg/year which was attributed to consumer behavior and the *lack of coordination* of the different players in the supply chain. This issue is further accentuated in the developing world, where the amount of food wasted before the product reaches the consumer is significantly higher.

The technology and tools for increasing the efficiency of ASCs have been researched in the past; however, their implementation has been very limited due to their mathematical formulation, which contrasts with the intuition of traditional decision-makers, their limitations on capturing the whole system dynamics, and the added complexity inherent of integrated models (Ahumada and Villalobos 2009a; Higgins et al. 2009; Lucas and Chhajed 2004). Therefore, it is our intention to review some of the most relevant research that has been done in agricultural supply chains and illustrate some ways in which operations research can be used to aid in the management of the supply chain.

2.2 Literature Review

2.2.1 Historical Overview of Mathematical Models in ASCs

The use of mathematical models and operations research tools for agricultural planning is not a new concept. As observed in the earliest review of mathematical models in farm planning made by Glen (1987), some optimization models and decision support tools have been developed for applications in crop planning as early as 1954. In the earliest publication found of such methods, Heady (1954) advocated the use of a linear programming model as a simple alternative for planning and budgeting agricultural production. Thereafter, we observe that mathematical models directed to agricultural planning started to become more widespread during the 1970s and 1980s with just a few publications during the decade of the 1960s.

By the end of the 1980s, it can be seen that the knowledge of mathematical formulations and optimization solutions was more widespread. The interest in these models grew significantly during the decade of 1990, where we can observe that many additional publications started to appear as it is illustrated by Ahumada and Villalobos (2009a) who provide the most comprehensive compilation of planning models for ASCs. In their review they catalogued the various research topics by mathematical modeling approach, perishable and nonperishable crops, and planning scope of the models (operational, tactical, and strategic). As it can be seen from this compilation, mathematical models for ASCs are becoming more widespread and are starting to cover a much more complex array of problems. In particular, Ahumada and Villalobos (2009a) observe that perishability of fresh products and risk management are two themes which are quickly starting to gain importance and visibility among the academic community.

The importance of specialty crops and mathematical models including perishability features is further accentuated by the latest literature made in the field of mathematical models for agricultural products by Zhang and Wilhelm (2009). In this review an emphasis is made on specialty crops and the models made for the management of these crops. They conclude with a call for further research in supply chain design, an issue they emphasize, which is becoming increasingly important. Furthermore, Zhang and Wilhelm (2009) state that, so far, mathematical models have remained relatively small and within the reach of commercial solvers; however, as the industry grows and problems become larger (with more stakeholders, locations, and further collaboration), basic research will become increasingly important to ensure solvability.

2.2.2 Scope of the Review

Even though optimization models can be used to model separate specific activities and have been used extensively in the past, the modeling of specific problems in isolation is likely to be of little value given the complex and interdependent nature of agricultural operations (Ahumada and Villalobos 2009a). Therefore, the review makes an emphasis on those articles which consider more than one echelon of the supply chain; these include articles that model interactions between different stakeholders, as well as those considering strategic decisions for laying out infrastructure for the ASC. Moreover, since the aim of this research is to identify the state of the art on planning tools, only models created since the year 2000 will be considered for review; for further reference, we will direct the reader to the excellent literature reviews made by Ahumada and Villalobos (2009b), Higgins et al. (2004), Lucas and Chhajed (2004), and Zhang and Wilhelm (2009).

2.2.3 Review of Models on Supply Chain Planning

In this section we present some works dealing with broad-scale supply chain decisions published starting in the year 2000. A list of several articles that fall within this category is found in Table 2.1 in the following page. Table 2.1 presents the leading authors and the year of the publication, followed by a brief description of the research paper; it was also of interest to identify whether perishability is explicitly taken into account, the modeling approach used by the authors, and the assumptions about decision-makers in the models. The *perishability of the crops* (*PER*) indicates whether perishability is explicitly considered or not (X for yes). The *modeling approach* (*MA*) refers to the analytical tools used to analyze and solve the optimization problem at hand; these can be tools such as mixed integer programming, stochastic programming, or other approaches. Finally, the *decision-maker* (*DM*) can be either centralized or decentralized (D for decentralized). The decision-maker is of particular interest because the most realistic case is that of a decentralized decision-maker since most ASCs will actually be composed of independent decision-makers needing to be coordinated (Ahumada and Villalobos 2009a).

Table 2.1 List of articles reviewed

Author(s)	Description	PER	DM	MA
Flores and Villalobos (2013)	Developed an opportunistic shipment policy for extended commercialization of fresh produce toward secondary markets	X		Stochastic
Ahumada et al. (2012)	Models the planting, harvesting, and distribution decisions of a grower/shipper who has control over the allocation of its products and wishes to optimize its profits under uncertain yield and price conditions	X		Stochastic
Xia and Qi (2011)	Making decisions for location of processing facilities, transportation modes, production, and storage of perishable agricultural products	X		MIP
Zhao and Wu (2011)	Models the interaction between one supplier and one retailer when the supplier has stochastic production using a revenue-sharing contract	X	D	News vendor
Rong et al. (2011)	Creates a framework to model the perishability of products throughout a supply chain that can be incorporated into a mixed integer program	X		MIP
Kazaz and Webster (2011)	Determines the amount of land to lease for the production of fruit and how much fruit to buy in the open market when prices and yields are stochastic			Stochastic
Yandra et al. (2010)	Proposes a multi-objective genetic algorithm for the optimization of the supply chain of coconut oil for the production of biodiesel in Indonesia			GA, MIP
Cai et al. (2010)	Modeled the interaction of seller/buyer of perishable items when there is a need to keep the product fresh during transportation	X	D	Nonlinear
Dharma and Arkeman (2010)	Uses a GA to coordinate the output of eight distinct models in horticulture, optimizing overall supply chain costs	X		GA, MIP
Quadra et al. (2009)	Models a supply chain consisting of resource suppliers, growing, and distribution for several specialty crops. Environmental factors such as pesticide use and crop rotation receive an especial focus in this model			MIP
Ahumada and Villalobos (2009b)	Models the planting, harvesting, and distribution of perishable crops to several markets by a grower/shipper with centralized control of production and distribution	X		MIP
Frayret et al. (2008)	Uses agent-based models to coordinate entities within the lumber supply chain		D	Various (agents)
Lodree and Uzochukwu (2008)	Determines the appropriate amounts of product to deliver by a producer in a VMI contract when demand is stochastic and the product deteriorates	X		Heuristics
Burer et al. (2008)	Examined contract dynamics between suppliers and retailers in the agricultural seed industry		D	News vendor

(continued)

Table 2.1 (continued)

Author(s)	Description	PER	DM	MA
Sanchez (2007)	Developed a methodology to identify and establish a logistics platform for the commercialization of fresh produce	X		Various
Thorburn (2006)	Integrated four distinct models using agent-based modeling to assess the implications of policy changes in the supply chain of Australian sugarcane		D	Various (agents)
Widodo et al. (2006)	Models a growing and harvesting to plan for harvesting of perishable crops using the ripening of the crops to create a final value function	X		DP
Ortmann et al. (2006)	Measured the flow of fruit, originating in South Africa, and assessed the total flow capacity of all warehouses, cold-storage facilities, and packing houses			Network
Schepers and Van Kooten (2006)	Models the interaction and collaboration between growers and retailers of tropical fruits to obtain appropriate product ripeness at the selling point	X	D	Nonlinear
Kazaz (2004)	Determines the amount of olive trees to lease or olives to buy in the open market before the season starts to produce olive oil in a future period			Stochastic
Higgins et al. (2004)	Developed an integrated model for harvesting, transportation, and machinery selection in the sugar mill industry to optimize profits		D	Various (agents)
Rantala (2004)	Models strategic decisions for closing, opening, and expanding greenhouses, warehouses, and customer outlets for the Finnish seedling supply chain			MIP
Gigler et al. (2002)	Uses a dynamic programming approach to model a full agricultural supply chain considering the changes in quantity and quality of the products	X		DP

In order to better catalogue the research done in supply chain optimization models in agriculture, we divide them into three main categories. These three categories, to the judgment of the authors, are the ones that best capture the overall focus of the research done in ASCs; these areas are dependent on the magnitude of the scope for the research and are interaction models, location/allocation models, and integrated industry models.

2.2.3.1 Buyer/Seller Interaction Models

The attributes that characterize these models are the explicit modeling of interactions and decisions made by distinct players in the supply chain. These models focus on direct interactions between suppliers and distributors of agricultural commodities, using methods such as the news vendor problem, nonlinear optimization, and game theory principles to achieve a state of equilibrium and greater benefit for all stakeholders. Generally, very specific types of interactions are depicted between the involved parties.

A good illustration for the peculiarities about supply chain interactions when dealing with perishable agricultural commodities was made by Schepers and Van Kooten (2006); they proposed a framework in which cost-sharing contracts between growers and distributors can be used to allow products at the correct ripeness level to be available at the selling point. Other articles which cover similar research are Burer et al. (2008), examining contract dynamics between suppliers and retailers in the agricultural seed industry; Cai et al. (2010), who added the considerations of transit time and transportation technology to the interaction of a seller (producer) and buyer (distributor) of perishable items when there is a need to keep the product fresh throughout its transit time; and Lodree Jr. and Uzochukwu (2008) who created a model consisting of a producer selling its product under a VMI contract when demand is stochastic.

2.2.3.2 Location/Allocation Optimization Models

These research papers explore the possibility of modeling a large part of the supply chain including the planting/harvesting plans and the infrastructure decisions such as usage of warehouses, packing facilities, or transportation modes. The scope of these models is larger than those of the previous section. Although the papers in this section consider a wider set of decisions and much larger-scale problems, they do not focus in specific interactions between key players of the supply chain; instead, these models consider mostly strategic and tactical decisions assuming that centralized control can be achieved for the operations under consideration.

Some articles which fall into this category are those of Gigler et al. (2002) who used a dynamic programming approach to coordinate the handling, processing, and distribution of agricultural goods explicitly considering the changes in appearance and quantity resulting from the perishability of the goods; Widodo et al. (2006) who modeled a plant growing and harvesting process in order to plan for periodical harvesting and inventory consolidation; and Rantala (2004) who modeled strategic decisions and expanded the modeling framework to tactical and operational decisions.

A further expansion in the field, considering stochastic components, was made by Kazaz (2004b) who addressed the problem of a Turkish company producing olive oil; the objective of the model is to maximize the expected profit subject to demand and the sales price of the olive oil. In a later implementation, a similar model was made by Kazaz and Webster (2011b) who determined the amount of land to lease for the production of fruit with stochastic prices and yields. Finally, Ahumada and Villalobos (2009b) created a model that expands further on the topic of stochastic optimization and considers the planting decisions and resource planning for two types of crops. The model builds over a previous research by Ahumada and Villalobos (2009b) which solved the same problem in a deterministic manner.

It can be seen that in this area of research, high-level models of the supply chain have evolved rapidly, quickly generating larger and more robust models. However, their high complexity makes it hard to model larger and more realistic scenarios in which each player in the supply chain makes independent decisions. We address these types of models in the following section.

2.2.3.3 Full-Chain Integrated Models

In this section we include those models with a broad scope and scale, modeling the full supply chain as well as the interactions between key players in the supply chain. These kinds of models are a minority due to their complexity and large size, which makes them hard to solve through the use of traditional OR models. Similar to the other models from Sect. 2.2.3, integrated models attempt to capture some of the characteristics that interaction models and location/allocation optimization models do; however, the analysis of these models is neither as detailed nor as likely to provide a full optimal solution. Instead, these models usually rely on agent-based models and other methods for dealing with decentralized, complex systems.

Higgins et al. (2004) was one of the first to propose a fully integrated model for an agricultural supply chain, which integrated several models for harvesting, transportation, and machinery selection in the sugar mill industry using an agent-based framework. The final objective of the study was to propose opportunities for enhancing the partnerships between the various players in the supply chain.

Other such research papers in this category are those of Thorburn (2006) who linked four separate models (farm planning, harvesting, transporting, and refining sugar), based on MIP and convex optimization, which were integrated in order to understand the implications of specific change in the sugar industry, and Thorburn (2006) and Frayret et al. (2008) who used a similar approach to assess the feasibility of agent-based models for the Canadian lumber supply chain. Another example of a highly complex system representing an agricultural supply chain was made by Dharma and Arkeman (2010) who used a genetic algorithm approach to coordinate the output of eight distinct models in horticulture: product demand forecast, vegetable supply forecast, planting schedule, aggregate planning, material requirement planning I, material requirement planning II, inventory management, and transportation route.

Although the models created using agent-based methods don’t have strong global optimization capabilities, they are capable of representing the interactions of the various players in the supply chain in a manner that can still assume independent decision-making. This is a great strength for these models since it makes them more appropriate for high-level decision-making than other models.

2.3 Sample Case Studies and Mathematical Models in ASCs

In this section we will illustrate several optimization-based approaches for planning and making decisions along the supply chain. This will be illustrated starting from the planting and harvesting decisions and finalizing the section to the marketing and distribution phase for horticultural products.

2.3.1 Planning the Planting and Harvesting (Grower Perspective)

In order to illustrate the form in which planning decisions are made, we will base our analysis in a full-chain optimization model developed by Ahumada and Villalobos (2009b). This research consists in a deterministic optimization model, which solved a problem for the case of a grower/shipper who has control not only over the planting decisions of crops but also the distribution to the final market. In this model, the perishability of the crops is modeled using a decay function, which is dependent on the distance and time of transportation of the products, as well as the rate of decay. A brief description of the optimization problem is provided below; however, for the full model and analysis, we refer the reader to Ahumada and Villalobos (2009b).

The basic structure of the supply chain being modeled is illustrated in Fig. 2.1 below. The modeling of the supply chain is done in two stages: one where the tactical decisions for planting and harvesting are made, together with all the resource constraints and costs at the production level. Thereafter, once the product is harvested, it enters into the production process, starting in the packing facility and ending at the customer warehouse; the product can then flow either through warehouses or the customer's respective distribution center. The decisions for sending the product through the second stage are directly dependent on the outcomes of the first stage (harvesting of the crops).

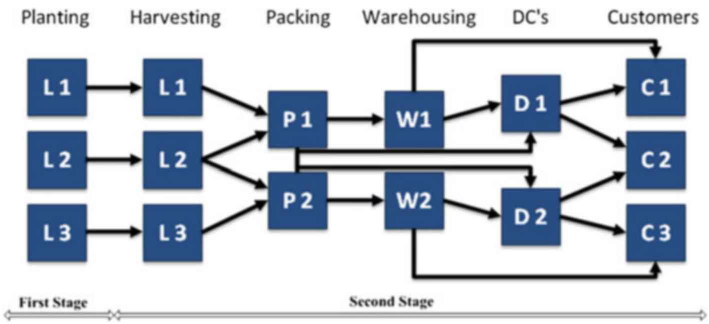


Fig. 2.1 Traditional supply chain for agricultural products

The supply chain illustrated in Fig. 2.1 is modeled as a mixed integer program (MIP), using constraints for the resources used in the first stage as well as constraints for choosing the appropriate routes and transportation modes to the final market. The objective function of the model is to maximize the expected profits for the season.

In addition to the model described above, we will provide examples for two expansions to this model; one expansion for making tactical decisions under uncertainty (Ahumada et al. 2012) and one for making labor allocation decisions for the production of four crops, without distribution decisions (Wishon et al. 2012). The use of these models will allow us to better illustrate the specifics expected from these models when applied to real-world cases.

When making a plan for a full season, a variety of concerns must be taken into account beforehand. For the case of the grower or grower/shipper preparing to make the plantation, three basic constraints are observed, as seen in the previous mathematical formulation. From these, the capital is critical, since farmers may not always have access to the capital, and the financing of the operations season to season is one of the greatest concerns for growers (Makeham and Malcolm 1993). Likewise, land is perhaps the most tangible resource that a grower has as their greatest limitation (at least in the short term). Water is also a concern, although it is only in some cases that this is an important limiting factor.

Of the various resources that a grower must utilize, labor is generally the most problematic; this is particularly true for the case of horticultural products, which are very labor intensive and for which mechanization is rarely a viable option (Emerson 2007). In fact, mechanization is only used when growers are left with no alternative (Schmitz and Seckler 1970). Particularly, for many crops such as trees and other specialty crops, labor can concentrate in a very limited time span, sometimes of a few weeks. It is because of this that the assessment and planning of labor as well as crop mix and the timing of plantations must be closely monitored when making tactical decisions for the season.

To illustrate this, we take the example of a 500-acre farm growing four crops: romaine lettuce, iceberg lettuce, broccoli, and cauliflower. The best planting schedule is determined by using the model obtained in Wishon et al. (2012), for creating a planting and hiring schedule (Table 2.2). However, for this model, we assume that there is a given demand for each product that must be met and that the availability of workers is constant throughout the season. Likewise, we assume that prices are constant for each product and that all production not committed to a contract can still be sold for the same price in the open market. The total profit calculated for a single instance of this problem where the grower commits 50 % of its production to weekly contracts is $120,000.00 and yields the following planning and harvesting schedules.

From this table it becomes apparent that broccoli and cauliflower have a higher lead time than romaine and iceberg lettuce. We can also observe that there is some fluctuation and overlap between various planting and harvesting schedules, which show that even when the product is being harvested, it may still be appropriate to plant another set of acres for a late-season harvest. Moreover, we can see that even though the planting and harvesting schedules may seem to have a lot of variability, the labor requirements can be rather stable, as seen in Table 2.3, which shows the labor requirements and the planned labor for hiring throughout the season.

Table 2.2 Scheduled planting and harvesting activities

Week		*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*
Broccoli	P	0	0	3	3	3	3	3	3	3	0	3	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Cauliflower	P	0	0	3	3	3	3	3	0	3	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Iceberg lettuce	P	0	33	34	25	20	31	26	32	29	30	29	31	31	30	33	0
	H	0	0	0	0	0	0	0	0	0	0	0	17	18	19	17	14
Romaine lettuce	P	0	0	0	3	3	3	7	3	3	3	3	3	3	5	4	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2

Week		*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*	*41*	*42*
Broccoli	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	2	2	2	2	2	2	2	2	2	1	1	1	2	1	0	0
Cauliflower	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	2	2	2	2	2	2	2	2	1	1	1	1	0	0	0	0
Iceberg lettuce	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	13	13	21	20	19	24	23	24	24	24	25	25	25	25	26	0
Romaine lettuce	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	2	2	2	3	5	2	3	3	2	3	2	2	3	4	3	0

Table 2.3 Labor required per week and hiring schedule

Week		*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*	*41*	*42*
Broccoli	P/C	0	0	0	0	0	1	2	2	2	2	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	5	5	4	4	4	4	4	4	5	4	2	3	5	2	0	0
Cauliflower	P/C	0	0	0	0	0	2	4	3	4	4	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	3	3	3	3	3	3	3	2	3	3	1	0	0	0	0
Iceberg lettuce	P/C	0	1	1	1	16	20	15	15	19	18	19	20	19	19	20	19	19	20	6	2	2	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	35	37	39	35	29	27	26	42	40	39	48	47	48	49	48	51	51	50	50	54	0
Romaine lettuce	P/C	0	0	0	0	0	0	2	2	2	4	4	3	3	3	3	2	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	5	5	6	9	13	6	7	7	7	7	7	7	8	10	8	0
Required		0	1	1	1	16	23	23	23	28	28	27	62	62	62	64	62	62	62	62	60	62	62	62	63	62	62	62	62	63	61	62	0
Hired		0	2	2	2	16	23	23	23	28	28	28	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62

Another interesting observation made from the results of the mathematical optimization is the behavior of profits and planting plans as the amount of production committed to contracts is varied. To illustrate this, we take the hypothetical case in which a grower commits a certain production quantity for each week and for each crop (Table 2.4). Under this scenario, each week would have a floor quantity, which must be supplied and is modeled as a constraint in the model (*average required*). We observe that as this quantity is incremented, the profit is severely impacted. This is caused mainly due to one reason: taking a closer look at the models output, we see that for a given amount of available labor, taking flexibility away from the planning process by committing production without proper planning will cause inefficiencies in the labor allocation; this in turn will cause shortages of labor on some weeks (leading to product not being harvested) and overages of labor on other weeks.

Similarly to the impact of different company policies and plans, factors outside of the control of the farmers can impact the decisions made for throughout the season. One clear example is, for instance, that of a labor shortage. Suppose, for example, that growers expect that there will be a labor shortage for the following season and calculate the highest amount of employees they will be able to hire. Depending on this preliminary assessment, a very different crop mix may be observed. To illustrate this, observe Fig. 2.2 below, which shows the projected crop mix for a season as the expected number of workers available for a 500-acre farm is reduced. As we can see, romaine lettuce which is the most labor-intensive crop begins to disappear, while iceberg lettuce quickly gains ground together with broccoli and cauliflower, two less profitable crops but also much less labor intensive.

2.3.2 *Assessment and Management of Risk (Grower/Shipper Perspective)*

Agriculture is a profession and lifestyle that has historically carried a lot of risks. Growers and shippers of agricultural products are exposed to a variety of factors which cause a great amount of variability on their day-to-day operations, the main drivers of variability being weather, market prices, and public policy decisions (Fleisher 1990).

The risks resulting from this variability can be hedged and reduced in many ways, for instance, some traditional forms of reducing risk (Fleisher 1990) are using risk-reducing inputs, product diversification, holding reserves, information, insurance, and forward contracts. Unfortunately, for the case of specialty horticultural crops, many of the alternatives for reducing risk are not available; for instance, reserves can only be held for a very limited time span and forward contracts are not an option for all crops. Because of this, many growers also resort to using insurance as a tool for hedging their production (Backus et al. 1997).

Table 2.4 Impact of contract flexibility on the final profit

	Planned	Average required	Planned	Average required	Planned	Average required	Planned	Average required
% committed	0 %		20 %		40 %		60 %	
Broccoli	0	0	1059	220	1125	450	1234	680
Cauliflower	0	0	1050	130	1050	270	1050	410
Iceberg lettuce	19364	0	17639	2310	17493	4620	17245	6940
Romaine lettuce	2397	0	1649	600	1811	1210	1980	1810
Profit	$1,849,657.60		$1,098,120.93		$447,910.68		$209,723.07	

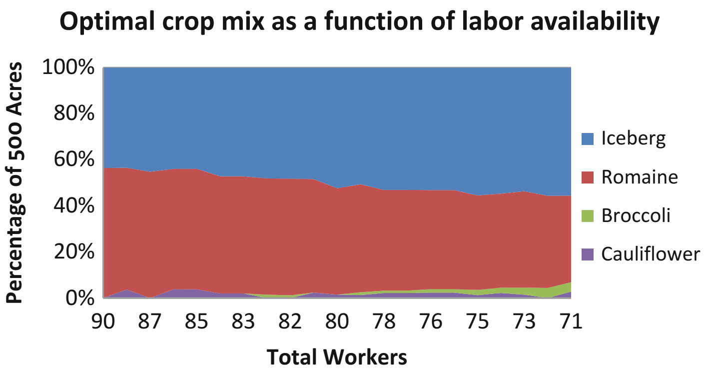


Fig. 2.2 Plan sensitivity to labor shortages

Growers are particularly affected by the fluctuations in market prices and the yields of their crops, which are inherently stochastic which can be modeled through an underlying probabilistic distribution. These features can be analyzed to understand the way in which prices behave at the final markets, the behavior of crop yields at the production points, and finally the interaction between these two variables.

Although statistical analysis of prices and yields can be made from the perspective of the growers, it is of greater interest to analyze how market information at distant locations can also be utilized by a grower/shipper. The availability of information is one of the greatest resources that farmers have to hedge against risk (Makeham and Malcolm 1993), and if the information can be enhanced through statistical analysis and an integrated planning approach, then the benefits to growers can become significant. For instance, a product portfolio is one of the main hedges of producers against risk, but the construction of a suitable portfolio requires the use of information and scenarios.

In order to support this claim, we will compare the output of two models, one that assumes deterministic prices and yields through expected values (Ahumada and Villalobos 2009b) and one which uses the probability distribution of prices and yield to reach an optimal solution through stochastic optimization (Ahumada et al. 2012). A sample planting plan for these two approaches can be seen in Table 2.5, which shows the output for the deterministic model vs. the stochastic formulation of the same problem showing that the planting plan resulting from using the full information can be significantly different. Likewise, the expected profits and worst losses are significantly different for both approaches.

A closer look at these numbers reflects how a plan that is specifically designed for one case only is likely to fail, whereas a plan that is designed to yield a profit under a great variety of scenarios has a greater advantage. In this sense, we say that the stochastic approach is more robust than the deterministic approach (Table 2.6).

Table 2.5 Sample planting plan for deterministic and stochastic approaches

	Deterministic		Stochastic $\lambda = 0$		Stochastic $\lambda = 1$	
Week	Peppers	Tomatoes	Peppers	Tomatoes	Peppers	Tomatoes
1	28	–	22	–	–	–
2	–	–	–	–	20	–
3	–	113	–	117	–	130
4	–	20	20	–	–	–
5	–	–	–	–	–	–
6	–	29	–	–	–	–
7	22	24	–	49	–	37
8	–	–	–	–	20	–
9	–	–	35	–	–	20
10	–	–	81	–	97	–
11	–	85	–	21	–	–
12	23	–	–	29	–	27
13	–	47	–	106	–	123
14	–	–	–	–	–	–
15	–	–	–	–	26	–
16	20	–	–	–	–	–
17	–	–	20	–	–	–
18	–	89	–	–	–	–
Total	93	407	178	322	162	337

Table 2.6 Comparison of results from stochastic program

Model	λ	Profit ($)	Costs ($)	ROI (%)	Worst ($)	CPU/s
Deterministic	0	3,255,643	14,439,900	22.50	37,598,000	197.54
Stochastic	0	5,621,200	14,427,100	38.90	174,526	1,325.4
Stochastic	1	5,619,360	14,434,100	38.90	138,500	1,350.2
Stochastic	10	5,510,680	14,434,500	38.10	153,871	1,485.5

Although the previous example is based on a model that treats the growing, harvesting, and distribution as done by one single decision-maker, there are many cases in which this may not be the case. Under these scenarios, we are interested in modeling the distribution of food products from the moment that they are bought from growers until their delivery to the final customer. In this case, the scope of planning is much smaller since we are no longer concerned about the planting process and the factors that affect it; however, this perspective also allows us take a broader approach for the possible locations where products are purchased, as well as expanding the markets to which we can send our products. The following section will cover two examples of fresh produce distribution and market integration.

2.3.3 Direct Marketing and Distribution Through Logistics Platforms

One of the primary issues facing agricultural producers in today's market environment is the risk associated with their production. The inherent risk present in the farming industry, together with the fact that agricultural commodities are often considered perishable goods, allows the buyer to select from a broad variety of production sources giving the farmers little or no leverage when it comes to negotiating their prices. Consequently, the underlying situation is one in which the farmers assume most of the risks associated with production variability and in turn receive a reduced margin over the final profits.

To counteract their poor bargaining position, many farmers have sought to integrate vertically along the value chain of their products in order to curtail their direct competition, as well as increase their share of the total profit margins. One example of this is the case of European farmers who have asserted themselves as global leaders by coming together through centralized, agricultural cooperatives (Sect. 2.1.2). These cooperatives use "logistics platforms" to engage in value-added practices that allow them to differentiate their products and gain greater ownership of the chain's profit margins. In this manner, the farmers assume a greater role in their product's added value, increase their bargaining leverage, and reduce the risk associated with price variability.

Figure 2.3 exemplifies a common integration often undertaken by farmers within the value chain of the product, from a traditional role to a more complex operational system. The traditional role of the farmer, as it shown in the diagram, is mainly associated with those activities related to production. This leaves the rest of the operations within the value chain, and most importantly its benefits, in the hands of its other members. On the other hand, as one can observe, a vertical integration allows the farmer to assume a greater role of the value-added and distribution activities.

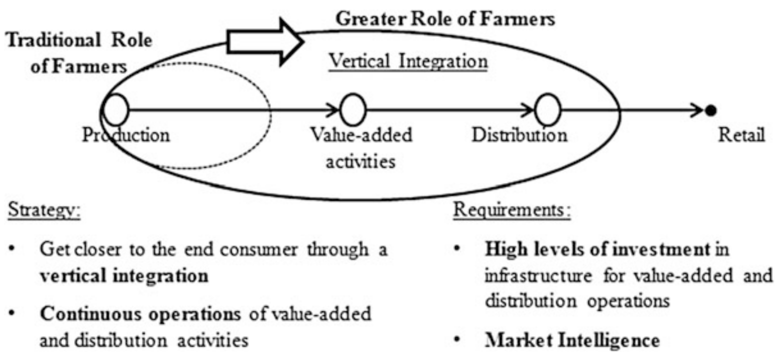


Fig. 2.3 Vertical integration along the product value chain

Vertical integration within the product value chain is a gradual process. In some cases, such as in Europe, the initial market conditions allowed an easier transition into more complex operations. For this case, the traditional structure of the farmers' operations was based on cooperative auctions. Through these cooperatives, the independent farmer had an organized marketplace through which they could sell their products to wholesalers and retailers in simple auctions. The price of the product depended mainly on the conditions of supply and demand factors. Eventually, these common cooperatives merged into centralized entities with established logistics platforms through which they developed value-added activities and centralized their decision-making strategies.

In the literature, a logistics platform can be viewed as homogenous and part of a logistics system controlled by one actor in the supply chain (Aldin and Stahre 2003). This includes concepts for logistics operations, a physical structure, processes and its activities, as well as the information systems needed for design, operations, and reporting (Abrahamsson et al. 2003). The development we see is that platform leaders need to exist for two kinds of coordination purposes. Firstly, there is a need for coordination within the platform where the physical logistics structure does not necessarily need to be centralized as long as the organizational logistics structure is centralized. Secondly, coordination or collaboration is needed between different actors in the supply chain.

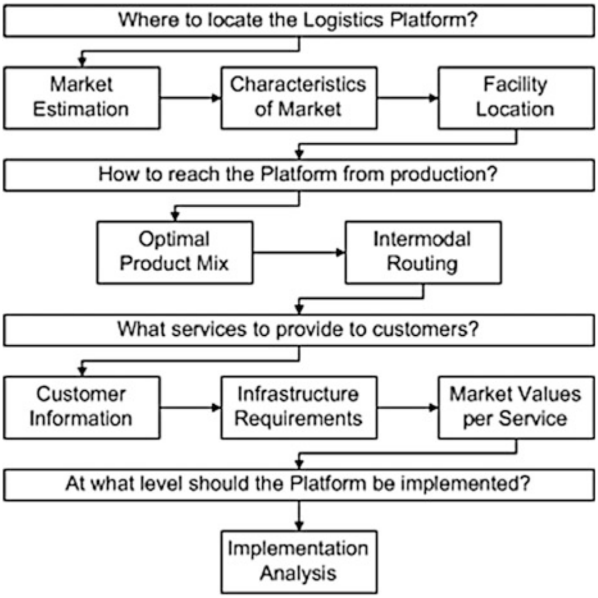
This section develops a case study that illustrates the factors and considerations in the implementation of a logistics platform. The case study involves the development of an enhanced commercialization strategy for Mexican fresh produce growers attempting to increase their marketing reach in the United States. The case study will take the reader through the different decision-making requirements in the design and implementation of a logistics platform. This same case study is applied in Sect. 2.3.4 to develop an alternative marketing strategy.

In this section, our definition of a logistics platform refers to a physical location where several different types of complementary businesses—centralized or decentralized—work together to provide different and better services to consumers. The types of suppliers required in this "logistics campus" are defined by the needs of the customer base inside the influence zone of the platform and can range from redistribution centers to processing facilities.

2.3.3.1 Design of a Logistics Platform

The first step in the design of a logistics platform is to identify the relevant factors involved in the development of an efficient platform, including a methodology for the analysis and strategic design of the platform. The analysis is aimed to find a way to vertically integrate the supply chain to improve its overall efficiency, focusing especially on the case of the fresh produce industry—although the general

Fig. 2.4 Problem decomposition



methodology can be applied to other situations. To achieve this objective the following steps are needed:

- Market estimation
- Facility location problems
- Multimodal transportation route design
- Product mix optimization for shipments
- Identification of customer base requirements for special services
- Strategic design of the logistics platform

The problem can be decomposed in four main questions: where to locate, how to reach the location, what services to provide, and up to what level or the scope of the implementation. The chart in Fig. 2.4 gives a better description of the parts of the study.

The first part of the analysis is related to establishing an objective for the logistics platform. While this might be already defined by the company strategy, the rational initial step is to assess which segments of the market are more convenient. This part includes a market estimation to establish a target region, then research on the characteristics of that region—such as costs, availability of land, subcontractors, distribution center clusters, etc.—and finally a facility location process.

The second part of the analysis is concerned with how to reach the logistics platform once it has been located. This refers to the haul of product from the production site to the logistics platform. Considering there are different types of products, the first step is to define an optimal loading mix based on the characteristics of the product (weight, volume, and compatibility with other products) and finally a multimodal route based in a mixed integer program.

What services to provide to customers is probably one of the most important sections of this study. The main issue here is to identify what types of potential customers are available in the influence zone and which are the requirements or services they have in order to become their supplier or strengthen the level of service. Once these are identified, there are also infrastructure requirements needed in order to provide the different types of services. The way each service is analyzed depends on whether the platform is centralized or decentralized since the objectives of the platform promoter would differ if the rest of the companies are owned by the same entity or not.

This example aims to first find a general location where the platform would operate near optimality. Thereafter, many other factors must be considered, such as existing distribution center clusters, workers' union issues, availability of other services, access to different transportation channels, etc. For practical purposes, the decision will be made in two steps:

1. Find a near optimal location based on a simplified p-median model dictated by demand potential.
2. Manually adjust the location considering other factors.

For this, a model is developed based on the p-median problem. In the p-median problem the objective is to choose p of the m sites to locate a warehouse such that the total demand-weighted (travel distance) cost between the warehouse and the retailers is minimized. For further insight on the plant location problem for this case study, the reader is encouraged to read Sanchez (2007). Finally, once the location is found, a manual selection process should undergo in order to consider more subjective factors mentioned before.

In this particular case study, the p-median model provides five candidate markets for the implementation of the logistics platform. Figure 2.5 presents the results found by Sanchez (2007) for potential locations of a logistics platform for the direct distribution of Mexican fresh produce in the United States. Note that this market identification is not based solely in market size; rather, important competitors such as tomato growers in Florida are taken into account. For instance, Los Angeles seems to be the largest market but it is not; it is just more attractive from the perspective of the Mexican farmers versus those from Florida.

Once the options are available, the decision-maker (or stakeholders) must choose the final region to focus on. It is important that an expert in the business under consideration makes the decision since many other factors (sometimes more subjective) might become important at this point. For instance, even if Los Angeles is the city with the biggest potential market, the decision-maker could opt not to establish a logistics platform there if their current supply chain already serves that zone efficiently or if competition is weak in the zone.

In the case study in question, Mexican farmers agreed on furthering the analysis for the platform to serve the Chicago area. This was in part because of the reasons described before, but also because they believed their current customer base in

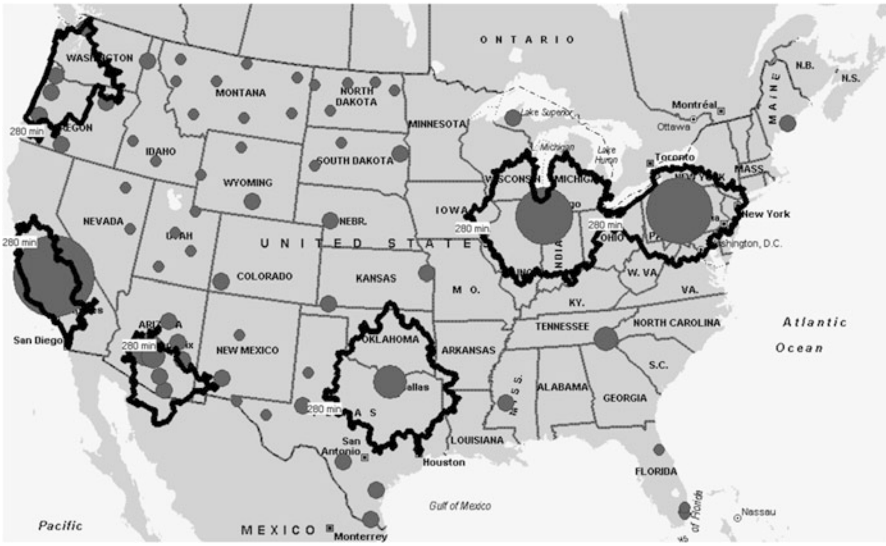


Fig. 2.5 Candidate markets for implementation

Chicago would be more interested in the new services that would be available once the platform is located there.

At this point the target region is identified and for the case study will be the Chicago area with a temporal centroid at Portage, IN, and a market potential of approximately $5,300 million, including two other major cities, i.e., Detroit, MI, and Indianapolis, IN, inside the influence zone.

2.3.3.2 Transportation of Fresh Produce

Once the location of the logistics platform is determined, the next question would be how to reach it from the supply site. For this step, we consider:

- Routes available from the crop site to the platform in all modalities:
  - Land
  - Railroad
  - Water
  - Air
- Location of multimodal terminals in the route
- Cost of transportation for all possible choices
- Handling compatibility of the products being analyzed

While the first three bullets are straightforward, some people might not be aware of the implications of the latter. Some products, including fresh products, have certain characteristics under which they must be handled. In the case of vegetables, handling temperature, humidity, type of packaging, and shelf life must be considered. But they also have certain organic properties which make them compatible with some products and prevent them from being mixed with others. Some examples are products sensitive to chilling injury or freezing injury and ethylene-sensitive/ethylene-producing end products that emit or absorb odors. It is extremely important to consider all factors to find which products can be shipped together and which ones cannot.

The second issue in shipping the product from point to point is the route that the containers will follow. The possibilities are wide and will depend on each situation; in many cases the use of more than one type of transportation—i.e., land, sea, railroad, and air—will be required to minimize cost. The first step is the creation of a network diagram that will help in understanding better each specific situation. The network for the Mexican farmers' case study is presented again in Fig. 2.6.

Using the information obtained of compatibility and product characteristics, an MIP is developed in Sanchez (2007) that aids the decision-maker in finding a cost-efficient mix of products to be shipped. The parameters in the model reflect the cost of transportation of one container from the origin to the destination or market. There are also parameters for the capacity of a container and for the total demand assumed per product.

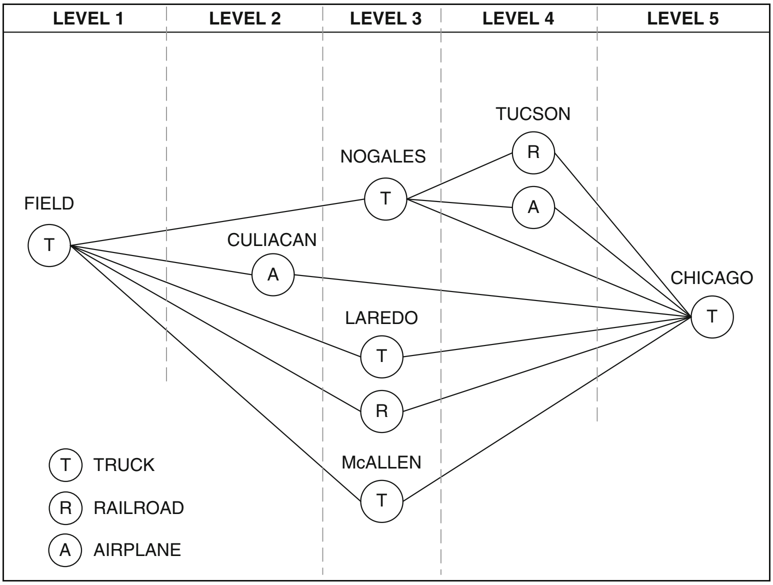


Fig. 2.6 Multimodal transportation network flow

2.3.3.3 Value-Added Services

The types of services and possibilities are wide and depend much on the investors' risk profile. As some companies will not find the risk of establishing some of those services worthwhile, others would. Therefore, the main issue becomes whether the platform would be centralized or not. In the first case, the analysis will have to be done in detail for each type of service by the promoter of the platform, and every business plan developed to start up all new companies required. In the latter, the logistics platform leaders would have to organize and promote the leasing of space to the necessary companies in order to make the services available.

Another issue that concerns the stakeholders of the platform is the investment portion on the redistribution center. The best way to act would be to run a pilot; unfortunately, a platform cannot be built as an experiment. The solution would be to start operations by leasing the redistribution facility and later on start to build the platform as a whole. The order in which farmers should implement the platform is the following:

1. Subcontract redistribution service.
2. Subcontract delivery fleet to serve customers' distribution centers.
3. Lease refrigerated facility and establish redistribution operation.
4. Buy land to establish the logistics platform and the redistribution center.
5. Buy repacking company or establish one for the platform.
6. Subcontract fleet for door-to-door deliveries.
7. Invest in or attract other companies to offer more services.

With this implementation plan, the expansion can be done gradually and still start operations right away. At first, everything could be subcontracted, so the farmers can just ship the product and manage the sales. The delivery fleet can also be subcontracted with the same company if possible or with a third-party logistics provider. A second step would be to subcontract a refrigerated facility and start operations there. This time, it will be required to hire personnel and have an operations staff to run it. Once the market is proven right, it can be chosen to buy land and start the construction of an owned redistribution facility. This land will serve also to build the rest of the platform and start leasing space to the companies that will collaborate in it.

2.3.4 Indirect Marketing and Distribution by Market Speculation

A second approach through which fresh produce owners can increase their commercialization reach is to develop a strategy based on temporal arbitrage opportunities. This would avoid resorting to high-capacity investments, i.e., logistics platforms. The proposed strategy involves reaching secondary markets through intelligence-based operations that require minimum level of investments. The overall objective of the operations is to maintain an attractive balance between the levels of potential returns and associated risks.

The basic premise of an indirect marketing and distribution is the development of a methodology that permits an established farmer with basic local operations to expand his/her commercialization reach into secondary markets, by way of financial engineering tools and statistical analysis. Specifically, this methodology develops decision-making tools that could identify potential opportunities and take advantage of momentary arbitrage opportunities for product placement in secondary markets. Finally, additional aspects of a similar operation, such as risk management policies, allow effective and profitable operations.

2.3.4.1 Design of an Opportunistic Commercialization Strategy

The main argument behind the methodology is that the dynamic characteristics of the fresh produce industry create a prime environment for intermittent commercialization opportunities for those individuals that aim to increase their marketing reach but that do not necessarily want to make an important capital investment. This strategy is based on the market price variability present as a result of the inherent characteristics of the fresh produce industry. These characteristics create market opportunities when one considers inefficient price transmission between markets. In other words, there is a short lag in time for market prices to adjust when the price of a produce item changes. At times, this momentary price differential may be large enough to offset the transaction cost of moving the item from one market to the other. These differentials are in essence momentary arbitrage opportunities.

There are several characteristics of the fresh produce industry that create the economic opportunities within the two-market structure. Among these characteristics are the following:

- Integrated markets
  - Variability at the supply source is indirectly transferred to the consumer market.
  - Constant fluctuation of product prices at wholesale markets.
- Homogenous products
  - Under certain market scenarios, fresh produce can be categorized as undifferentiated, commodity products (without considering those that have received additional value-added activities).
  - Allows for easier ownership of acquired products.

Additionally, given that fresh produce items have very limited shelf life, coupled with high variability at the supply source, fresh produce markets tend to be continuously fluctuating and highly volatile. For this reason, one could make an indirect comparison with the behavior of financial markets.

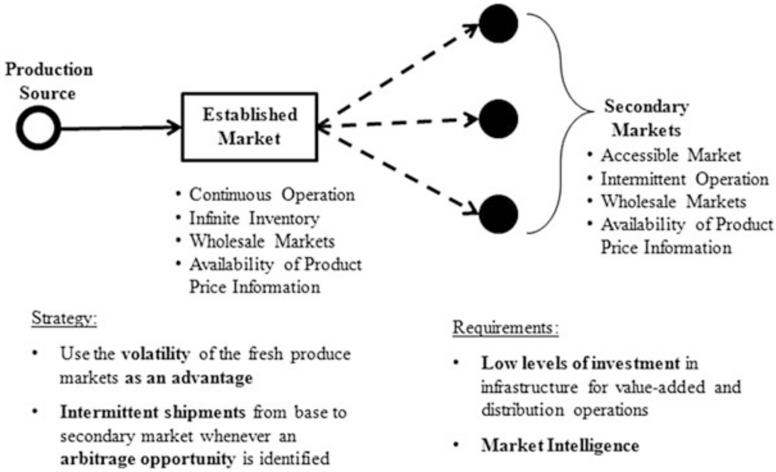


Fig. 2.7 Envisioned operational structure

A potential operational structure that would allow the farmer to take advantage of the market price variability associated with the fresh produce industry is to increase the commercialization reach of the products (Fig. 2.7). In this case, the farmer can use the price variability within two markets in order to identify particular time instances in which their price differential creates an opportunity for a transaction. This transaction is the process of moving a single product from the base to the secondary market under favorable conditions. If the transaction results in a positive profit, then that particular arbitrage opportunity has been correctly identified and captured. If the transaction results in a negative profit, then the opportunity was incorrectly identified.

The main assumption of this operational structure is that the base (primary) market has continuous and established operations, while those in the secondary market are intermittent. In this case, one assumes that the production dedicated to the continuous operations at the established market will be large enough to fulfill sudden opportunistic surges in demand within the secondary market. As a result, it is assumed that that farmer will always have enough inventories at the primary market to take advantage of favorable price differentials as determined by the model. Therefore, for all practical purposes, the inventory at the established market used to supply the secondary market is assumed infinite.

Additional assumptions for the operational structure include (1) the availability of daily wholesale price information at markets (currently reported daily by the US Department of Agriculture), which will be used to calculate the threshold level of the shipment strategy, (2) an accessible secondary market due to the relative ease through which a farmer can access wholesale markets, and (3) a third-party transportation system for which a per item transportation cost is incurred to move the products.

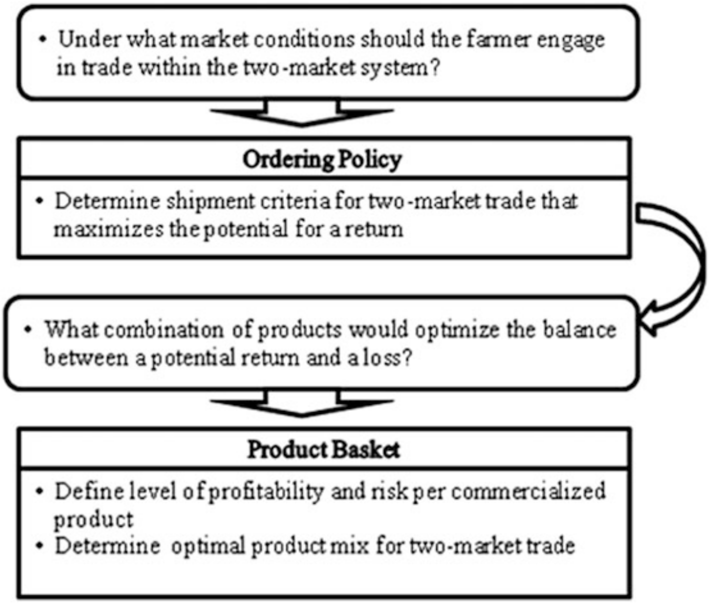


Fig. 2.8 Problem decomposition

Figure 2.8 presents a schematic with the decision methodology underlying the opportunistic commercialization strategy. Each question addresses different aspects of the operation that need to be addressed throughout the planning process of the commercialization expansion.

The question then becomes how one uses the dynamic market characteristics of the fresh produce industry in order to develop a generalized method of operation that allows an entity to identify and take advantage of momentary economic arbitrage opportunities. The purpose is to have positive expectation of return on minimal levels of investment. Furthermore, the farmer would want to limit his/her risk exposure and only speculate on the market whenever the market conditions are favorable.

A case study is presented to demonstrate the development of such commercialization strategy. In this case, six markets were selected. These markets are part of a larger database maintained by the US Department of Agriculture that publishes daily terminal market prices on a variety of fresh produce items. In order to maintain similarity to the case study presented in Sect. 2.3.3, Dallas, TX, was selected as the base market and the rest of the markets in are the potential secondary markets. Ten years of daily terminal market prices were collected from this database and were adjusted by an inflation factor.

The long-run average prices for those markets considered suggest that continuous shipment operations in these structures are not profitable (Table 2.7). As one can observe from the table below, the average long-term prices at the base market in

Table 2.7 Long-term average prices

	Dallas ($)	Boston ($)	Atlanta ($)	Chicago ($)	DC ($)	NYC ($)
Tomato	0.70	0.76	0.70	0.71	0.72	0.66
Squash	0.58	0.46	0.49	0.50	0.53	0.46
Eggplant	0.94	0.86	0.57	0.83	0.55	0.77
Cucumber	0.39	0.37	0.33	0.39	0.31	0.36
Bell pepper	1.07	0.67	0.99	0.97	1.01	0.84

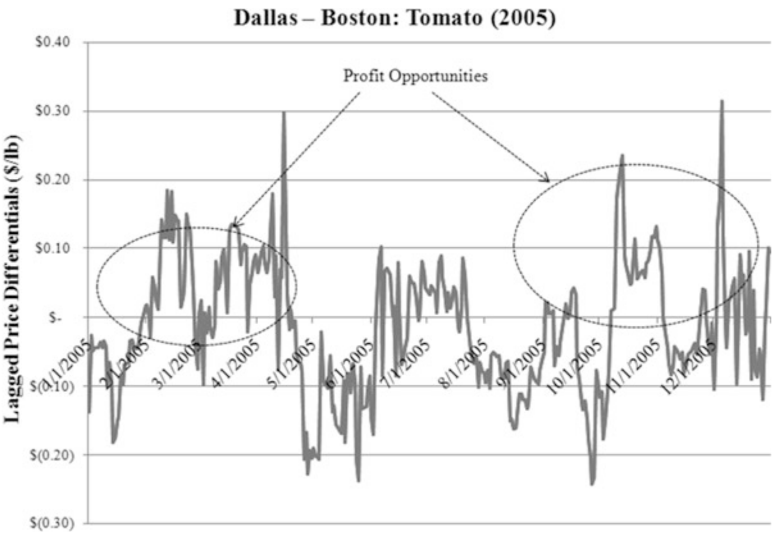


Fig. 2.9 Two-market lagged price differentials

general tend to be close, if not, higher than at the secondary. Consequently, the price differentials between the structures are not large enough to allow continuous profitable transactions. In order to capture the opportunities that are present within the market price differentials, one has to search for specific opportunity windows in which one might be able to observe from gains from engaging in a two-market trade.

For this, one needs to dwell a bit deeper into the market price differentials, in order to identify the specific opportunity windows that indicate potential arbitrage in a two-market transaction. Figure 2.9 presents an in-depth glimpse of the market price differentials for an arbitrary year of 2005 within the Dallas-Boston market structure. The values observed in this graph account for the non-lagged differentials between these markets. This differential is the price of the product (per pound) at the secondary market minus the price at the base market and the cost of transaction, during the same day.

One can gain additional information regarding the behavior of these differentials over the 10-year period by summarizing their statistical characteristics through a fitted theoretical distribution. If one considers a stricter shipment condition

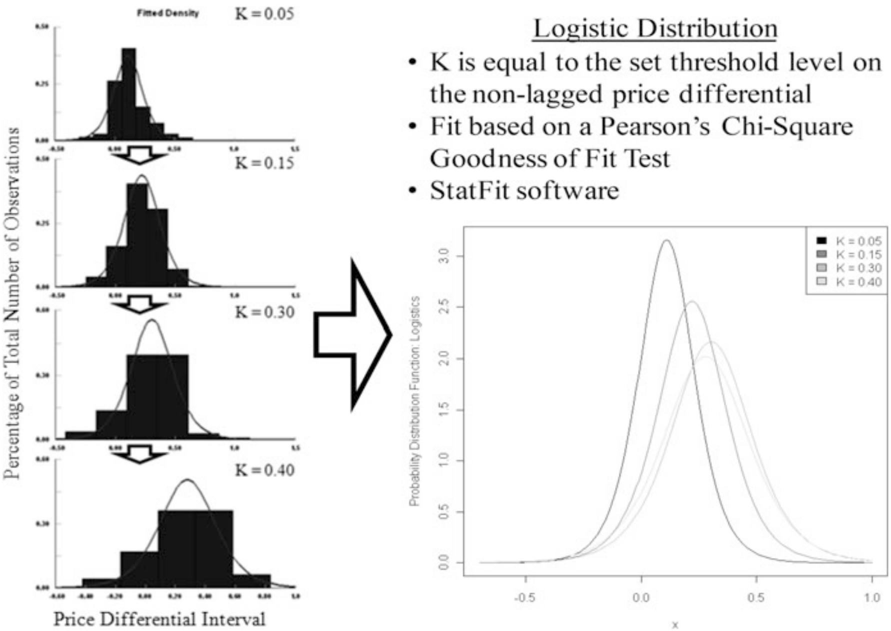


Fig. 2.10 Histogram distribution under various thresholds

(or threshold level) on the non-lagged differentials for the Dallas-Boston market structure, the observed mean and variance of the lagged differentials become increasingly positive, as shown from the histogram distribution in Fig. 2.10. In the right-hand side of this figure summarizes a fitted logistics distribution (based on a Pearson's chi-square goodness of fit test) for each observation histogram on the left, under various threshold levels.

As it can be observed on the figure above, the mean profit of a single, one-time shipment is increased as the threshold condition for the non-lagged differential becomes stricter. However, as the threshold becomes stricter, the number of opportunities also decline greatly. Therefore, a strategic trade-off between the number of opportunities and the expected profits has to be determined in order to maximize revenues. Using the model developed in Flores and Villalobos (2013), one can select the specific market price condition at which an opportunity maximizes the long-term revenues. This model indirectly accounts for this strategic trade-off.

2.3.4.2 Risk Management Strategies

Given that this second approach is dependent on market speculation, it would be necessary to develop a strategy that would reduce the risk associated with market speculation. Thus, the second phase of this case study develops a shipment configuration that would be able to limit the risk exposure of the decision-maker. In this case, the market price characteristics of each individual component (under a defined shipment strategy) are used to hedge the risk. It is assumed that one can apply mean-value portfolio theory to the collection of shipment components in such a way that one can manipulate the overall rate of return and variance. Ultimately, the objective is to determine an optimal shipment configuration that minimizes the variability of the returns for a particular component.

For demonstration purposes, it is assumed that one wants to limit the risk exposure of tomato shipments from Dallas to Boston. It is also assumed that the general objective of the decision-maker is to maximize his/her long-term profits of the shipments during a defined operational period. Thus, the market price information for the rest of the shipment components is limited to the opportunity time windows of tomato within the Dallas-Boston market structure. The main objective would be to reduce the variability of the rates of return for tomato shipments from Dallas to Boston. To reduce the risk, one can use portfolio theory to hedge the risk of only sending products to Boston. In this case, one can reduce the risk of a loss when one sends a shipment to a single market by developing a strategic portfolio of different markets. The goal is to offset the risk in the main target market. This means that whenever an opportunity is identified for a tomato shipment in Boston, a shipment of the same product is also sent to the rest of the secondary markets in a strategic manner (Fig. 2.11).

Again, the objective is to use the market characteristics within each secondary market in order to strategically invest in each for the sake of minimizing the variability observed in the returns. For this approach, the objective is to hedge the risk of sending a tomato container to Boston, by also sending a shipment of the same product to the other secondary markets. The shipment policy used is that for tomato within the Dallas-Boston market structure, which optimizes the long-term profits for this item during the operational period. Thus, the information of rates of return correlation and covariance is limited to the time windows of opportunity of this product over the 3-year period.

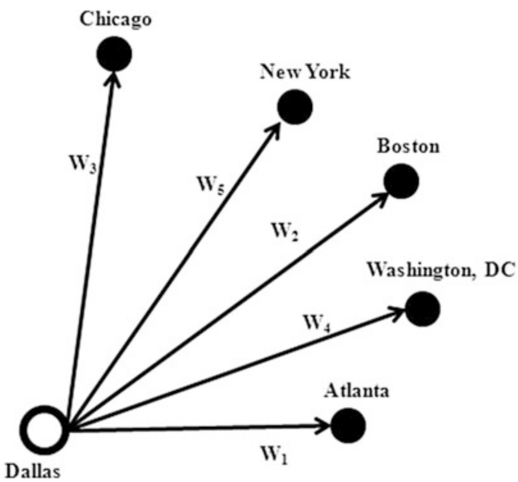


Fig. 2.11 Secondary market configuration

Table 2.8 Secondary market configuration: average rate of return

Two-market structure	Rate of return (Tomato)
Dallas-Atlanta	0.060
Dallas-Boston	0.114
Dallas-Chicago	0.045
Dallas-DC	0.064
Dallas-NY	−0.077

Table 2.9 Secondary market configuration: covariance matrix

	Dallas-Atlanta	Dallas-Boston	Dallas-Chicago	Dallas-DC	Dallas-NY
Dallas-Atlanta	0.024	0.010	0.012	0.013	0.016
Dallas-Boston	0.010	0.035	0.008	0.007	0.014
Dallas-Chicago	0.012	0.008	0.043	0.007	0.014
Dallas-DC	0.013	0.007	0.007	0.020	0.009
Dallas-NY	0.016	0.014	0.014	0.009	0.039

Table 2.10 Variance before/after market configuration

	Variance (only Dallas-Boston)	Variance (market configuration)
Total	0.0346	0.0168

Table 2.8 presents the rates of return of tomato per secondary market under this shipment policy. As one can observe, the highest average rate is observed in Boston, while the lowest details a negative rate in New York. Furthermore, Table 2.9 summarizes the correlation levels observed between each secondary market, which are much higher than those observed in the product mix. In this case, the highest level of correlation with Boston is New York, which approximates 0.378. Overall, Atlanta and DC have the highest level of correlation with 0.571. Since this table is just a representation of the covariance matrix, one can observe that the variance between the Dallas and Boston price markets is 0.0346.

Applying a Markowitz model approach from financial engineering to this case study (Luenberger 1998), it was found feasible to reduce the return variance for tomato shipments to the Boston market by configuring a shipment strategy based on the rest of the secondary markets. The results based on this strategy are found to be satisfactory. Table 2.10 presents the variance of the rates of return before and after the development of a market configuration.

As one can observe from the table above, the overall variance of the shipment in the Dallas-Boston market structure is reduced by combining operations with other secondary markets. It is observed that the variance of the rates of return is reduced by 52 %.

2.4 Conclusions

In this chapter we have provided a brief overview about the evolution of the field of ASC management and optimization, complete with a comment on the recent trends and directions of the field, which is moving toward greater consolidation and toward a vertical integration of operations. Under this perspective, we see that supply chain operations that can be performed in isolation using mathematical models are starting to be performed with greater collaboration between growers and shippers; many of these shared operations can similarly be modeled through mixed integer programming or using other decision tools to optimize the operations of various players in the chain.

Throughout Sect. 2.3 in this chapter, we have described various decisions that are made with the aid of mathematical programming and decision tools to perform specific operations throughout the chain. These operations range from the pure farm planning and planting decisions to the impact of these decisions on the future marketing of the crops. We continue to take a complete approach to the supply chain by analyzing the strategy that growers seeking to integrate vertically could follow in order to penetrate a new market and establish a joint operation. Finally, the perspective of a shipper with access to enough inventory and information to engage into opportunistic marketing of agricultural commodities is illustrated, therefore completing a full overview of operations performed in the chain and their respective technical tools.

2.4.1 Comments on Integrated Farm Planning and Marketing Execution

Upon analyzing the various sections of the supply chain of fresh produce and their modes of operation, we go back to the trends being observed in the recent years for agricultural supply chains. As we have seen before, growers and shippers alike face changing market conditions due to the consolidation of supermarket chains and the new purchasing policies implemented by those chains (Dimitri et al. 2003). The response to this has been a new trend in the US market toward vertically integrated producers, a trend which had already been advanced in the European market.

One clear question that comes into mind (after observing these trends and after reviewing the literature in the subject) is: what is needed to support this emerging business model? Clearly, there will be a need to coordinate the actions of several partners within the cooperative not only in a manner that maximizes the overall profit but also in a way that creates profit and reduces risk for each individual partner. This is a challenging problem which will need to be carefully evaluated, in particular, if we deal with a decentralized system, where the incentives for complete cooperation may not be predominant and individual interests could potentially create unwanted variability at the global level. An example of such problem could be, for example, that several farmers shift a small amount of their production toward a specific time period under the expectation of higher prices; however, if this reasoning is too widespread and uncoordinated, then an oversupply could result, therefore harming all players in the supply chain.

Unfortunately, little research has been done in the topic of decentralized coordination of the supply chain. This emphasizes the need to create the appropriate decision support tools to deal with this emerging complexity, be it with the purpose of profit maximization, risk reduction, or supply chain coordination.

References

Abrahamsson M, Aldin N, Stahre F (2003) Logistics platforms for improved strategic flexibility. Int J Logist Res Appl 6(3):85–106. doi:10.1080/1367556031000123061

Ahumada O, Villalobos JR (2009a) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 196(1):1–20. doi:10.1016/j.ejor.2008.02.014

Ahumada O, Villalobos JR (2009b) A tactical model for planning the production and distribution of fresh produce. Ann Oper Res 190(1):339–358. doi:10.1007/s10479-009-0614-4

Ahumada O, Villalobos RJ, Mason AN (2012) Tactical planning of the production and distribution of fresh agricultural products under uncertainty. Agr Syst 112:17–26. doi:10.1016/j.agsy.2012.06.002

Aldin N, Stahre F (2003) Electronic commerce, marketing channels and logistics platforms—a wholesaler perspective. Eur J Oper Res 144(2):270–279

Alexandratos N, Bruinsma J (2006) World agriculture: towards 2030/2050. Interim report. FAO

Aramyan L, Ondersteijn CJM, Kooten OV, Oude Lansink A (2006) Performance indicators in agri-food production chains. In: Ondersteijn CJ, Wijnands JH, Huirne RB, Kooten OV (eds) Quantifying the agri-food supply chain. Springer, Dordrecht, pp 49–66, http://www.springerlink.com/content/r8874k4373n2441w/

Backus GBC, Eidman VR, Dijkhuizen AA (1997) Farm decision making under risk and uncertainty. NJAS Wageningen J Life Sci 45(2):307–328

Blisard WN, Variyam JN, Cromartie J (2003). Food expenditures by U.S. households: looking ahead to 2020. U.S. Department of Agriculture, Economic Research Service, Washington, DC

Burer S, Jones PC, Lowe TJ (2008) Coordinating the supply chain in the agricultural seed industry. Eur J Oper Res 185(1):354–377

Cai X, Chen J, Xiao Y, Xu X (2010) Optimization and coordination of fresh product supply chains with freshness-keeping effort. Prod Oper Mgmt 19(3):261–278. doi:10.1111/j.1937-5956.2009.01096.x

Christopher M (2005) Logistics and supply chain management: creating value-added networks. Financial Times Prentice Hall, Harlow

Cook RL (2011) Fundamental forces affecting U.S. fresh produce growers and marketers. Choices 26(4):13, http://ideas.repec.org/a/ags/aaeach/120008.html

Dharma RA, Arkeman Y (2010) An intelligent decision support system for horticultue. Jurnal Teknologi Industri Pertanian 19(3). http://medpet.journal.ipb.ac.id/index.php/jurnaltin/article/view/1778

Diamond A, Berham J (2012) Moving food along the value chain: innovations in regional food distribution. USDA. http://www.northcentralsare.org/About-Us/Regional-Initiatives/Scaling-Up-Local-Food/Scaling-Up-Local-Foods-Reports/Moving-Food-Along-the-Value-Chain-Innovations-in-Regional-Food-Distribution

Dimitri C, Tegene A, Kaufman PR (2003) U.S. fresh produce markets: marketing channels, trade practices, and retail pricing behavior (Agricultural Economics Reports No. 33907). United States Department of Agriculture, Economic Research Service. http://ideas.repec.org/p/ags/uerser/33907.html

Emerson RD (2007) Agricultural labor markets and immigration. Choices 22(1):57–66, http://econpapers.repec.org/article/agsaaeach/94471.htm

Fleisher B (1990) Agricultural risk management, illustratedth edn. Lynne Rienner, Boulder

Flores H, Villalobos RJ (2013) Using market intelligence for the opportunistic shipping of fresh produce. Int J Prod Econ 142(1):89–97

Frayret J-M, D'Amours S, Rousseau A, Harvey S, Gaudreault J (2008) Agent-based supply-chain planning in the forest products industry. Int J Flex Manuf Syst 19(4):358–391. doi:10.1007/s10696-008-9034-z

Garcia M (2007) Market trends and retailers' strategies in fresh produce. Cal-Med consortium workshop III presented at the Mediterranean products in a global market place, Barcelona

Gigler JK, Hendrix EMT, Heesen RA, van den Hazelkamp VGW, Meerdink G (2002) On optimisation of agri chains by dynamic programming. Eur J Oper Res 139(3):613–625. doi:10.1016/S0377-2217(01)00191-6

Glen JJ (1987) Mathematical models in farm planning: a survey. Oper Res 35(5):641–666

Gustavsson J, Cederberg C, Sonesson U, van Otterdijk R, Meybeck A (2011) Global food losses and food waste. FAO

Hazell P, Wood S (2008) Drivers of change in global agriculture. Philos Trans R Soc B: Biol Sci 363(1491):495–515. doi:10.1098/rstb.2007.2166

Heady EO (1954) Simplified presentation and logical aspects of linear programming technique. J Farm Econ 36(5):1035–1048. doi:10.2307/1234313

Higgins A, Antony G, Sandell G, Davies I, Prestwidge D, Andrew B (2004) A framework for integrating a complex harvesting and transport system for sugar production. Agr Syst 82 (2):99–115. doi:10.1016/j.agsy.2003.12.004

Higgins AJ, Miller CJ, Archer AA, Ton T, Fletcher CS, McAllister RRJ (2009) Challenges of operations research practice in agricultural value chains. J Oper Res Soc 61(6):964–973. doi:10.1057/jors.2009.57

Jones Putnam J, Allshouse JE (1999). Food consumption, prices, and expenditures, 1970–97 (Statistical Bulletin No. 965). Food and Rural Economics Division, Economic Research Service, U.S. Department of Agriculture. http://www.ers.usda.gov/publications/sb-statistical-bulletin/sb965.aspx

Kader AA (2002) Postharvest technology of horticultural crops. ANR, Oakland

Kazaz B (2004) Production planning under yield and demand uncertainty with yield-dependent cost and price. Manuf Serv Oper Manag 6(3):209–224

Kazaz B, Webster S (2011) The impact of yield-dependent trading costs on pricing and production planning under supply uncertainty. Manuf Serv Oper Manag 13(3):404–417. doi:10.1287/msom.1110.0335

Lodree EJ Jr, Uzochukwu BM (2008) Production planning for a deteriorating item with stochastic demand and consumer choice. Int J Prod Econ 116(2):219–232. doi:10.1016/j.ijpe.2008.09.010

Lowe TJ, Preckel PV (2004) Decision technologies for agribusiness problems: a brief review of selected literature and a call for research. Manuf Serv Oper Manag 6(3):201–208. doi:10.1287/msom.1040.0051

Lucas MT, Chhajed D (2004) Applications of location analysis in agriculture: a survey. J Oper Res Soc 55(6):561–578

Luenberger DG (1998) Investment science. Oxford University Press, New York

Makeham JP, Malcolm LR (1993) The farming game now. Cambridge University Press, Cambridge

Millennium Ecosystem Assessment (2005) Ecosystems and human well-being: our human planet: summary for decision-makers. Island Press, Washington, DC

Morelli C (1999) Information costs and information asymmetry in British food retailing. Serv Ind J 19(3):175–186. doi:10.1080/02642069900000036

Ortmann FG, van Vuuren JH, van Dyk FE (2006) Modelling the South African fruit export infrastructure: a case study. ORiON 22(1):35–57

Quadra PR, Rodriguez JM, Terol MC, Cruz DE (2009) A three level multi-period multi-location and multi-crop sustainable supply chain model. Presented at the IEEE International Conference on Industrial Engineering and Engineering Management, 2009. IEEM 2009, pp. 1414–1418. doi:10.1109/IEEM.2009.5373053

Rantala J (2004) Optimizing the supply chain strategy of a multi-unit Finnish nursery company. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.2847

Rong A, Akkerman R, Grunow M (2011) An optimization approach for managing fresh food quality throughout the supply chain. Int J Prod Econ 131(1):421–429. doi:10.1016/j.ijpe.2009.11.026

Salin V (1998) Information technology in agri-food supply chains. International Food and Agribusiness Management Review 1(3). http://ideas.repec.org/a/ags/ifaamr/34523.html

Sanchez O (2007) Strategic design of a logistics platform for fresh produce. Arizona State University, Tempe

Schepers H, Van Kooten O (2006) Profitability of "ready-to-eat" strategies: towards model-assisted negotiation in a fresh-produce chain. Frontis 15:117–132

Schmitz A, Seckler D (1970) Mechanized agriculture and social welfare: the case of the tomato harvester. Am J Agric Econ 52(4):569–577. doi:10.2307/1237264

Stevenson S (2008) Values-based food supply chains: strategies for agri-food enterprises-of-the-middle. http://citation.allacademic.com/meta/p_mla_apa_research_citation/2/4/4/2/4/p244246_index.html?phpsessid=6fa83ed81b7dc4f98830ebc05a7f22fb

Stewart H (2010) Younger consumers exhibit less demand for fresh vegetables. DIANE Publishing, Darby

Thorburn P (2006) Integrated value chain scenarios for enhanced mill region profitability. Sugar Research and Development Corporation, Brisbane

Widodo KH, Nagasawa H, Morizawa K, Ota M (2006) A periodical flowering–harvesting model for delivering agricultural fresh products. Eur J Oper Res 170(1):24–43. doi:10.1016/j.ejor.2004.05.024

Wishon C, Mason AN, Villalobos R, Flores H, Lujan G, Rodriguez S (2012) Use of MIP for planning temporary immigrant farm labor force (in review)

Xia Z, Qi Lv (2011) Optimal design of agri-food chain network: an improved particle swarm optimization approach. Presented at the 2011 international conference on management and service science (MASS). IEEE, pp. 1–5. doi:10.1109/ICMSS.2011.5998308

Yandra J, Marimin I, Eriyatno HT (2010) An integration of genetic algorithm and fuzzy logic for optimization of agroindustrial supply chain design. http://repository.ipb.ac.id/handle/123456789/28809. Accessed 1 Mar 2012

Zhang W, Wilhelm WE (2009) OR/MS decision support models for the specialty crops industry: a literature review. Ann Oper Res 190(1):131–148. doi:10.1007/s10479-009-0626-0

Zhao X, Wu F (2011) Coordination of agri-food chain with revenue-sharing contract under stochastic output and stochastic demand. Asia-Pac J Oper Res 28(4):487–510

Chapter 3
Production and Logistics Planning in Seed Corn

Rogerio A.R. Junqueira and Reinaldo Morabito

3.1 Introduction

In wide countries like Brazil, the USA, and others, supply chains of seed corn can involve multiple disperse crop fields, processing plants, and demand spots, which require complex and elaborate aggregate production and distribution planning to attend predefined harvest plans and meet forecasted product demands. The limited processing capacities of the plants are also relevant for this tactical planning.

Seed corn enterprises that have experienced rapid growth or operate in a complex and changing environment had a tendency to create simple practical rules, sometimes disconnected, to develop production and logistics plans. For instance, some of them have focused on only a few variables relevant to planning, like the distances between crop fields and processing plants, disregarding other relevant variables, such as the unit production and distribution costs and the goods and services circulation taxes involved.

Optimization approaches have been applied to support production, inventory, and distribution aggregate planning decisions in different agribusiness settings, considering several technical and economical criteria and constraints. As these approaches are incorporated into decision support systems, they become powerful and flexible for the analysis and planning of these systems under different scenarios. Several examples of successful applications in production and logistics planning of agribusiness supply chains can be found in, e.g., Shapiro (1993, 2001), Chen (2004), Higgins and Laredo (2006), Ahumada and Villalobos (2009), INFORMS (2012), and the references therein.

In this chapter, we are concerned with optimization approaches to support tactical planning decisions in production, inventory, and distribution of seed corn.

R.A.R. Junqueira (✉) • R. Morabito
Production Engineering Department, Universidade Federal de Sao Carlos, Sao Carlos, Brazil
e-mail: rarjunqueira@gmail.com; morabito@ufscar.br

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_3

The chapter is organized as follows. In Sect. 3.2, we briefly discuss production and logistics planning of seed corn including the process features, tax planning, and a classification of supply chain systems and some aspects of the production and logistics planning of seed corn. In Sect. 3.3, we review some optimization approaches of the literature based on mathematical programming to support decision making in seed corn supply chains. In Sect. 3.4, we shortly describe some results derived from a case study developed in a Brazilian seed corn company. The emphasis is on the approaches that explicitly consider tax planning. The results show important opportunities for production and logistics cost reduction by using mathematical programming. Finally, in Sect. 3.5, we present some concluding remarks.

3.2 Production and Logistics Planning

Several studies in the literature, such as Chen (2004) and Ahumada and Villalobos (2009), show that the integration of production and distribution functions is very important to achieve high performance in supply chains. Some applications in agri-food industry can be found in the integrated production planning of poultry (Taube 1996), in the decentralizing distribution of ethanol (Yoshizaki et al. 1996), in the aggregate production and distribution of frozen concentrated orange juice (Munhoz and Morabito 2014), in the production planning and trade of lily flowers (Caixeta-Filho et al. 2002), in the tactical planning of breeding farms producing piglets taking into account the sow herd dynamics (Rodriguez-Sanchez et al. 2012), and in the production, inventory, and distribution of sugar and ethanol (Colin et al. 1999; Salassi et al. 2002; Higgins and Muchow 2003; Lopez et al. 2006; Kawamura et al. 2006; Paiva and Morabito 2009).

Besides matching production and distribution, tax planning is another important issue for the optimization of the supply chain performance. For instance, Balaji and Viswanadham (2008) studied a multinational tax integrated model to decide the foreign direct investment or outsource on global network supply chain planning. Basset and Gardner (2010, 2013) proposed models to deal with multinational tax costs to optimize the design of the global supply chain of an agricultural chemical enterprise. Junqueira and Morabito (2012) proposed a production and logistics planning considering circulation taxes in a multiplant seed corn company.

3.2.1 Seed Corn Production Process

Agricultural production is the first phase of the seed corn supply chain as depicted in Fig. 3.1, which illustrates an example of a schematic representation of this supply chain (Junqueira and Morabito 2012). In this phase, the final products to be commercialized by the company are defined, as well as the quality standards

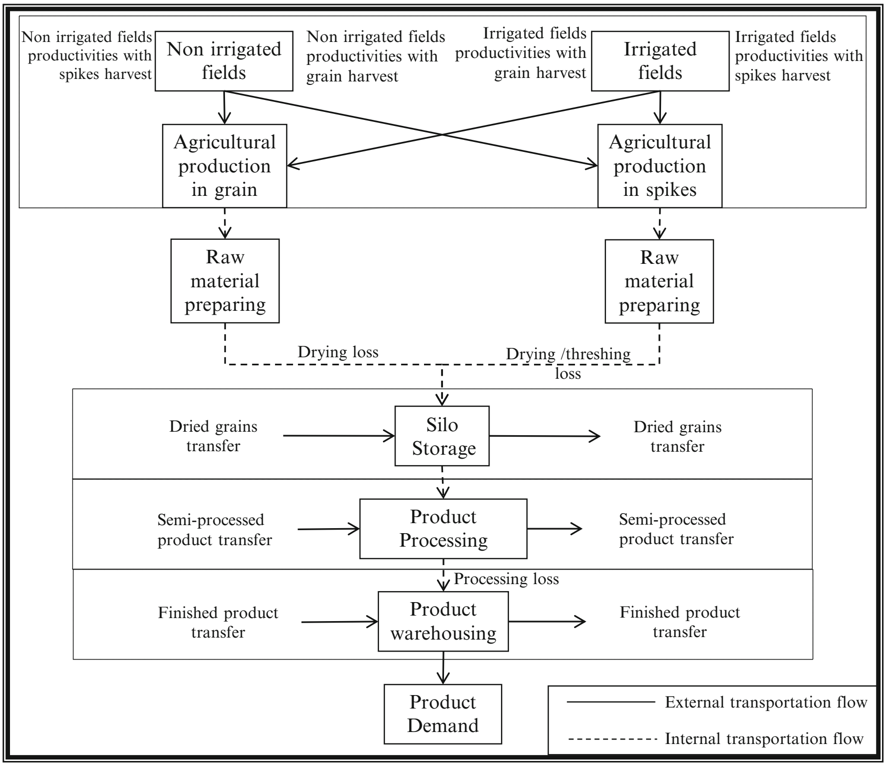


Fig. 3.1 A schematic representation of the seed corn supply chain

desired for them. Seed corns are separated basically into three product types: variety, hybrid, and genetic modified. Variety means that the seed has essentially the same characteristics of its original corn plant, i.e., there are no genetic modifications or improvements in the seed corn. In general, the productivity of a variety seed corn for the customer (i.e., a corn producer or farmer) is relatively low. Hybrid means that there are crossings between different corn plants, generating a seed with different characteristics of its corn plants (matrices), appropriately manipulated as the crossing effects are known. A hybrid seed corn can be simple, double, and triple, and its productivity is higher than the variety. While variety production focuses on low technological producers that remain in the seed market, the focus of hybrid production is mainly on high technological companies with large-scale production levels and which commercialize their products in broader areas. In the case of commercial genetically modified corn seed, after previous laboratorial processes, it can be either a parent of a hybrid or be treated as a variety.

The next phases of seed corn production are crop field harvesting and transportation of the raw material to the plants (Fig. 3.1). For seed production, a spike harvest is recommended as the seed quality is better preserved in terms of germination and vigor (Oliveira et al. 1997). Meanwhile, there are cases in which the enterprises choose to harvest corn seed in grain, especially for more robust products which are less damaged by mechanized grain harvest, such as some varieties and double hybrids. The type of harvest determines some transportation means and equipment used during the initial agro-industrial stages in the processing plants. Transportation must be fast in order to avoid deterioration due to the high water level of the seed. Thus, the raw material (i.e., corn spike or grain) cannot be stored before a drying processing to reduce its humidity.

Like other grain agro-industries, the corn seed process can be divided in two stages: preparation of the raw material and processing of dried grains. The raw material preparation changes humid corn (spikes or grains) into dried grains with a minimum of impurities by means of cleaning, drying, and threshing operations. Depending on the type of raw material, the preparation involves different process operations. The plants can have infrastructure to dry both grains and spikes. This stage enables to storage the corn seed (Fig. 3.1). The dried grain processing transforms the dried seeds into final products, which means dried and clean seed corn, classified by size, chemically treated, bagged, and in germination condition.

The supply of raw material to a processing plant does not come necessarily from a crop field. It can be transferred from another plant or prepared in an independent industrial unit. The demand of seed corn is a critical issue to be considered in production and logistics planning due to its forecasting errors and the complexity of its management. Because of production-scale savings, Brazilian and other wide countries' enterprises seek to commercialize seeds in broader areas to increase their market share. Broader areas mean more diverse climate and production environments, which interfere in the time that the product should be available to the customers. Moreover, farmers of the same area usually demand different product mix.

3.2.2 Tax Planning

Another issue that greatly influences production and logistics plans of seed corn is the taxation of multijurisdictional commerce, which leads managers to consider tax planning (Junqueira and Morabito 2012). In Brazil, for example, tax obligation is created when an incidence hypothesis predicted in laws occurs. However, it is not mandatory to practice acts that result in duty incidence or that make it more costly. A triggering event is a situation predicted in law, which is necessary and sufficient to create this obligation. On the other hand, a tax evasion is characterized when a triggering event occurs and someone tries to hide it or mischaracterize it. Thus, tax planning aims to minimize the total tax payment in operational, administrative, and financial acts or facts, choosing the best alternative among the existing legal ones.

A production and logistics plan can either avoid triggering an event of circulation tax or reduce its due value. In the first case, the plan avoids triggering an event by choosing routes with tax exemption. In the second case, the plan reduces the due value of an event by choosing routes with lower rates. Brazilian laws determine that

Table 3.1 Comparison of transportation costs and circulation taxes

Route	Origin	Destination	Distance (km)	Transportation cost ($/sack)	ICMS cost ($/sack)	Total cost ($/sack)
GO–MT	Rio Verde	Cuiabá	696	1.47	4.80	6.27
MG–MT	Uberlândia	Cuiabá	1,045	2.21	2.80	5.01
SP–MT	Barretos	Cuiabá	1,212	2.57	2.80	5.37
PR–MT	Maringá	Cuiabá	1,338	2.83	2.80	5.63
SC–MT	Chapecó	Cuiabá	1,671	3.54	2.80	6.34
RS–MT	Passo Fundo	Cuiabá	1,856	3.93	2.80	6.73

a circulation tax event occurs when goods leave an organization or establishment (CONFAZ 2006). However, Brazilian states may interfere in the circulation tax adding or adopting different rules. For example, for some states, there are circulation tax exemptions for inner state operations and there are circulation tax reductions of up to 60 % for interstate operations. Moreover, there are circulation tax reductions if the product is made in the south or southeast areas and is sold in the north, northeast, or center-west areas.

A circulation tax exemption for inner state operations can produce a significant impact on the final price of a seed corn. Note that in this way, the legislation aims to promote production and sales inside the state. However, interstate rate differentiation for operations from the south or southeast areas to the north, northeast, or center-west areas adds a trade-off between distance and tax advantages in logistics decisions. The following example of a seed corn supply chain in Brazil illustrates this effect: consider that the transportation unit cost is $0.0988/t.km (SIFRECA 2004), that the product price is $100.00 per 20 kg sacks, and that there are six available routes as presented in Table 3.1, whose origins are production areas in six states (GO, MG, SP, PR, SC, and RS) and whose destination is a demand market in one of the states (MT) (Junqueira and Morabito 2006).

Table 3.1 compares total (transportation + circulation tax) costs per sack for each route with the shortest distance total cost option. For instance, note that for a distance lower than the distance between Barretos (SP) and Cuiabá (MT) (i.e., a distance lower than 1,000 km), the tax is superior to the transportation cost. In the case of a product leaving Maringá (PR), the transportation cost is the same as the tax; however, the total cost is lower than the shortest distance, i.e., departing from Rio Verde (GO). At some point between Maringá (PR) and Chapecó (SC) (i.e., a distance of 1,640 km from Cuiabá (MT)), the point where the total cost is equal to the shortest distance cost is located. The minimum cost occurs if the product is delivered from Uberlândia (MG).

Although this trade-off analysis is focused in the Brazilian tax system, this case shows effects of taxation of multijurisdictional commerce studied, for example, in Shackelford and Shevlin (2001). This taxation introduces additional tax rates and base variations, and it can be separated in two types: multinational and multistate. Some of the models discussed here deal with multistate tax costs.

3.2.3 A Taxonomy of Production and Distribution Models

As pointed out in Chen (2004), several companies manage production and distribution in an independent way and with little or no coordination, which leads to a scenario of increasing holding costs and longer lead times and contradicts a tendency to low inventory and customer responsiveness. Efforts are being made to overcome this contradiction, specially the development of production-distribution optimization models that explicitly integrate production and distribution operations.

A coupled production-distribution (EPD) model can be useful to treat seed corn supply chains as addressed in this chapter. The taxonomy proposed by Chen (2004) includes three main dimensions to classify production-distribution problems in supply chains: (A) decision level, (B) integration structure, and (C) problem parameters.

The decision level can be separated in tactical (A1) and operational (A2) (this taxonomy does not emphasize strategic decisions). A1 EPD models deal with decisions such as how much to produce of each product and to ship to each region in a time period, how much inventory of each product to keep in each region, etc. These decisions are different than strategic ones such as facility location, capacity, and network structure. On the other hand, A2 Operational EPD models deal with detailed scheduling level decisions, such as when and on which machine to process a production lot, when and on which vehicle to deliver products, which route to take for a vehicle for product distribution, etc.

The integration structure can be classified into three types: (B1) integration of production and outbound transportation, (B2) integration of inbound transportation and production, and (B3) integration of inbound transportation, production, and outbound transportation.

The problem parameters are the planning horizon and the nature of demand. This taxonomy considers three variations: (C1) one time period, (C2) infinite horizon with constant demand rate, and (C3) finite horizon but with multiple time periods and dynamic demand.

It also considers five classes of problems based on dimensions A, B, and C:

Class 1: Production-transportation problems—A1, B1, C1
Class 2: Joint lot sizing and finished product delivery problems—A1, B1, C2
Class 3: Joint raw material delivery and lot sizing problems—A1, B2, C2
Class 4: General tactical production-distribution problems—A1, B1 or B3, C1 or C3
Class 5: Joint job processing and finished job delivery problems—A2, B1, C3

The seed corn supply chain models discussed in this chapter can be classified mainly as Class 4. In this class, the number of products of the problem (D1, one product, and D2, multiple products) can be also considered. In the parameter nature of demand, the taxonomy highlights other features like whether the demand is deterministic or stochastic (E1, deterministic, and E2, stochastic) and if backlog has penalty or not (F1, backlog without penalty, and F2, backlog with penalty).

Other characteristics are particularly important to be observed in the models analyzed here:

G: Capacity (G1, limited, and G2, unlimited).
H: Nature of production (H1, deterministic, and H2, stochastic).
I: Number of production stages (I1, one, and I2, more than one). More than one stage allows the transfer between facilities.
J: Penalty to production exceeded (J1, not considered, and J2, considered).
K: Tax costs (K1, not considered, and K2, considered).

3.3 Some Approaches in the Literature

Few studies are found in the literature dealing with the seed corn agribusiness supply chain. In the following, we briefly review two optimization models of the literature based on mathematical programming that do not consider explicitly the tax planning decisions. The remaining of this section is dedicated to a more detailed description of the third optimization approach which considers these decisions.

3.3.1 Production and Transportation in a Single Stage and Time Period

One of the earliest studies in seed corn supply chains found in the literature appears in Zuo et al. (1991). This study addresses an aggregate production-distribution problem for a large seed corn company located in North America. Linear programming and mixed integer programming models were proposed to help the management make production decisions, as the allocation of products (corn hybrids) to available production plants and the transportation of the products to where they are needed by the customers. The models consider production capacity constraints, minimum capacity usage requirements, transportation resource constraints, demand constraints, maximum concentration constraints, among others. The production was not planned by the company as a whole but planned by each company division in parts, usually allocating to the facility that was close to the customers who demand the production. The result was an uneconomical tactical production-distribution system. The models were applied to unify the planning policy of the company, without being restricted by division boundaries or by the proximity of the facility to the customer. They minimize production and transportation costs by allocating the productions of seed corn to the regions that provide maximum yields and then transporting them to the sales regions. They consider production and shipping costs in a single production stage and period.

In the following, we present the baseline model in Zuo et al. (1991) represented as a linear programming formulation. The decision variables of the model are

x_{ij}, the amount of hybrid j produced at facility i, in bags
y_{ijk}, the amount of hybrid j shipped from facility i to sales region k, in bags

The model parameters are

L, the total number of facilities.
M, the total number of hybrids.
N, the total number of sales regions.
A_j, the output-input conversion factor, representing the number of bushels of input needed to produce a bag of hybrid j.
P_{ij}, the unit cost of producing hybrid j at facility i, in \$/bag; this cost is a combination of unit acre grower cost (money paid to the contracted growers), per acre cost (including parent seed cost, fertilizers, chemicals), and per bag cost (including treating material cost, dry and shell wage, harvest utilities).
S_{ijk}, the unit cost of shipping hybrid j from facility i to sales region k, in \$/bag (this cost is a function of unit mileage transportation cost per truck load, number of miles between facility i and sales region k, number of bags that can be loaded on a truck).
C_i, the total capacity available at facility i, in bushels.
CO_i, the minimum capacity usage requirement at facility i, in bushels, $CO_i \leq C_i, \forall i$.
D_{jk}, the demand for hybrid j in sales region k, in bags.
$\max_{ij}$, the maximum amount of hybrid j at facility i, in bags.

This approach searches for an optimal utilization of the available resources at minimum cost, without compromising seed quality, facility capacity, and market demands:

$$\min Z = \sum_{i=1}^{L}\sum_{j=1}^{M} P_{ij}x_{ij} + \sum_{i=1}^{L}\sum_{j=1}^{M}\sum_{k=1}^{N} S_{ijk}y_{ijk}, \qquad (3.1)$$

s.t.

$$\sum_{j=1}^{M} A_j x_{ij} \leq C_i, \qquad i = 1,2,\ldots,L, \qquad (3.2)$$

$$\sum_{j=1}^{M} A_j x_{ij} \geq CO_i, \qquad i = 1,2,\ldots,L, \qquad (3.3)$$

$$x_{ij} - \sum_{k=1}^{N} y_{ijk} \geq 0, \qquad i = 1,2,\ldots,L, j = 1,2,\ldots,M \qquad (3.4)$$

$$\sum_{i=1}^{L} y_{ijk} \geq D_{jk}, \qquad j = 1,2,\ldots,M, k = 1,2,\ldots,N \qquad (3.5)$$

$$x_{ij} \leq \max_{ij}, \qquad i = 1,2,\ldots,L, j = 1,2,\ldots,M \qquad (3.6)$$

$$x_{ij}, y_{ijk} \geq 0, \qquad i = 1,2,\ldots,L, j = 1,2,\ldots,M, k = 1,2,\ldots,N \qquad (3.7)$$

The objective function of the model aims to minimize total production (first term in (3.1)) and distribution (second term) costs. Note that the production costs consider the cost of the land designated to the facility and the agricultural cost of growing the corn hybrid in this area, besides the industrial cost. The transportation cost is a function of mileage per truck to travel from the facility to the sales region. Equation (3.2) represents the capacity constraints, which are defined by the dryer, i.e., the bottleneck of the production processes within a facility in Zuo et al. (1991)'s study. Also the freezing risk was considered in the adjustments of the capacity constraints. Minimum capacity usage requirements are considered in (3.3) in order to ensure that no facility would be closed in accordance with the company's policy. Equation (3.4) corresponds to the transportation resource constraint and balances the flows between production and transportation, i.e., the amount of hybrid produced at a facility has to be greater or equal to the sum of the amounts of this hybrid shipped from that facility to all the sales regions. The demand for various hybrids in each sales region is satisfied due to (3.5). Equation (3.6) relates to maximum concentration constraints and imposes that the concentration of a hybrid production was limited, i.e., no more than a certain amount of a hybrid can be grown at one facility. Finally, (3.7) defines the domain of the variables.

The study also proposed another group of equations that defines if nothing or a large amount of a hybrid can be produced on the facility, "either-or" equations. These equations are defined as

$$x_{ij} = 0 \ \text{ or } \ x_{ij} \geq \min_{ij}, \quad i = 1, 2, \ldots, L, j = 1, 2, \ldots, M$$

where $\min_{ij}$ ($\min_{ij} \leq \max_{ij}$) is the agreeable minimum production amount of hybrid j at facility i, in bags, in order to eliminate solutions involving small amounts which are neither convenient nor economic for handling. By introducing auxiliary binary variables z_{ij} and slightly modifying (3.6), the either-or equations can be easily converted into linear constraints:

$$x_{ij} \geq \min_{ij} z_{ij}, \quad i = 1, 2, \ldots, L, j = 1, 2, \ldots, M$$
$$x_{ij} \leq \max_{ij} z_{ij}, \quad i = 1, 2, \ldots, L, j = 1, 2, \ldots, M$$

but this conversion transforms the linear programming model above into a mixed integer programming model.

The linear and mixed integer models were implemented in IBM MPSX (Mathematical Programming System Extended) package and MIP (Mixed Integer Programming) package used together with heuristics written in FORTRAN language. The real problem had 1,095 constraints and 11,500 variables. When the either-or equations were used, the number of constraints increased to 960, and 480 integer variables were added to the model, which complicates the solution of the model.

Several scenarios were analyzed with different configurations of the models, and sensitivity analyses for them were done varying production and transportation costs, facility capacities, transportation resources, and customer demands. These results were presented to the company staff to show the effects of various constraints to the system operation and total cost. After these analyses, the company identified potential savings of \$5.69 million per year. If some inefficient facilities were closed, as well as good weather forecast and irrigation could be assumed, these savings would increase to \$7.53 million per year. The learning obtained after this work motivated a company-wide reorganization, and \$10 million of cost savings was recognized. The results of this study indicated benefits in applying optimization techniques to decision making in a large-scale seed corn production system.

3.3.2 Production in Two Time Periods (Stochastic)

The fact that growing seed corn is a biological process dependent upon local weather and other conditions during the growing season complicates production planning. In addition, customer's experiences with a particular corn hybrid during a given year influence demand for that hybrid during the next year. Jones et al. (2001) proposed a stochastic dynamic model to integrate the aggregate hybrid seed corn production of an international seed corn enterprise with plants in North and South America, considering production uncertainties and demand uncertainties. The model also considers two periods: a first growing period in North America and a second growing period in South America, which is offset by approximately 6 months. The company took advantage of this second growing season to better manage its production planning process by means of the second-chance production planning, in which decisions of the second period take into account production and demand realizations of the first period.

In the first phase of the planning process, the company determines, for each corn hybrid, how much acreage to plant for the North America production period and makes a contingent production plan for South America. In the second phase of the planning process later in the year, the company updates and finalizes the production plan of the hybrid for South America. At this point, the company knows the average seed corn yields from North America production, and the significant uncertainties remaining are the average seed corn yields from South America production and the customer demand. The objective is to maximize expected gross margin, i.e., expected revenue from seed corn sales less expected costs of production, holding, and shortage.

This problem can be viewed as an extension of the single-period newsvendor (newsboy) problem (Johnson and Montgomery 1974), in which there is a second chance to produce a product to meet a random demand, instead of only one chance as in the newsvendor problem. In this two-period stochastic problem, for each hybrid, the production yields of the two periods are also random variables, besides the hybrid demand. The model treats each hybrid independently of others. Jones et al. (2002) solved this problem by means of a linear programming model discretizing the probability distribution functions of the corn yields and the hybrid demand, i.e., using discrete approximations for each of the yield distributions and the demand distribution. This linear programming model was also reported in Jones et al. (2003) along with the results of the application of the model (with a few changes) to this international hybrid seed corn enterprise.

The parameters of the model are

p, the selling price per unit (80,000 kernels) at the end of period 2
π, the shortage cost per unit for unmet demand at the end of period 2
v, the salvage value per unit for any unsold seed at the end of period 2
c_i, the cost per unit of processing seed at the end of period i (includes holding or shipping as applicable)
K_i, the cost per acre in period i
w_1, the number of units available at the beginning of period 1, which is the quantity of product carried out from the previous year

The discrete approximations for the probability distributions of the random variables yield y_1 and yield y_2 in periods 1 and 2, respectively, are

$$g(y_1) = \{g_{11}, g_{12}, \ldots, g_{1i}, \ldots, g_{1m}\}, \quad \text{where } \operatorname{prob}(y_1 = y_{1i}) = g_{1i}$$
$$g(y_2) = \left\{g_{21}, g_{22}, \ldots, g_{2j}, \ldots, g_{2n}\right\}, \quad \text{where } \operatorname{prob}\left(y_2 = y_{2j}\right) = g_{2j}$$

The discrete approximations for the probability distribution of the random variable hybrid demand D is

$$f(D) = \{f_1, f_2, \ldots, f_k, \ldots, f_p\}, \quad \text{where } \operatorname{prob}(D = D_k) = f_k$$

The decision variables of the model are

Q_1, the number of acres to plant during period one, i.e., the first season acreage choice
X_i, the number of acres to plant during period 2 when $y_1 = y_{1i}$, i = 1,. . .,m, i.e., the second season acreage choice
Z_{ijk}, a dummy variable that reflects sales revenue + salvage − shortage penalty, when $y_1 = y_{1i}$, $y_2 = y_{2j}$, and $D = D_k$.

$$\begin{aligned} \max Z = &-K_1 Q_1 - c_1\left(\sum_{i=1}^{m} g_{1i} y_{1i}\right) Q_1 - K_2 \sum_{i=1}^{m} g_{1i} X_i \\ &- c_2 \sum_{i=1}^{m} g_{1i}\left(\sum_{j=1}^{n} g_{2j} y_{2j}\right) X_i + \sum_{i=1}^{m} g_{1i}\left(\sum_{jk} g_{2j} f_k Z_{ijk}\right) \end{aligned} \tag{3.8}$$

s.t.

$$\begin{aligned} &Z_{ijk} \le p\left(w_1 + y_{1i} Q_1 + y_{2j} X_i\right) - \pi\left(D_k - \left(w_1 + y_{1i} Q_1 + y_{2j} X_i\right)\right) \\ &\quad i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots, p \end{aligned} \tag{3.9}$$

$$Z_{ijk} \leq pD_k + v\left(w_1 + y_{1i}Q_1 + y_{2j}X_i - D_k\right)$$
$$i = 1, \ldots, m, \ j = 1, \ldots, n, \ k = 1, \ldots, p \tag{3.10}$$

$$Q_1 \geq 0 \tag{3.11}$$

$$X_i \geq 0 \quad i = 1, \ldots, m \tag{3.12}$$

$$Z_{ijk} \geq 0 \quad i = 1, \ldots, m, \ j = 1, \ldots, n, \ k = 1, \ldots, p \tag{3.13}$$

The objective function (3.8) maximizes the earnings. The first and third terms represent the first- and second-period planting costs, respectively. The second and forth terms represent the expected cost of processing the first- and second-period harvests, respectively. The fifth term corresponds to the value of revenue plus salvage minus shortage cost, which can be calculated once the seller has realized the demand.

In (3.9), the term $w_1 + y_{1i}Q_1 + y_{2j}X_i$ represents initial inventories plus production of the first and second periods. These equations evaluate total revenue versus shortage costs, and they are tight when the demand is greater than or equal to the supply, i.e., when

$$D_k \geq w_1 + y_{1i}Q_1 + y_{2j}X_i$$

In this case, the optimal solution of the model sells the supply and pays shortage costs for the difference between the demand and the supply. On the other hand, (3.10) is tight when the demand is less than or equal to the supply:

$$D_k \leq w_1 + y_{1i}Q_1 + y_{2j}X_i$$

In this case, the optimal solution of the model sells the demand and pays salvage costs for the difference between the supply and the demand. Equations (3.11)–(3.13) define the domain of the variables.

In order to ensure feasible, finite, and nontrivial solution to the linear programming model (3.8)–(3.13), two assumptions are necessary. The first one guarantees that the expected harvest cost plus the processing cost is greater than or equal to the salvage value of seed in each period:

$$v \leq \frac{K_i}{E(y_i)} + c_i, \ i = 1, 2 \tag{3.14}$$

where $E(y_i)$ denotes the expected value of the random variable y_i. Without this assumption, the expected profit of the producer would be unbounded, i.e., it would not exist as a feasible and finite solution to the problem. The second assumption states that the expected production cost (harvest + processing) must be less than or equal to total gain (product price + shortage cost) obtained by selling seed in each period:

$$\frac{K_i}{E(y_i)} + c_i \leq p + \pi, \quad i = 1, 2 \tag{3.15}$$

If this condition is not satisfied, the optimal solution would be zero production, i.e., a trivial solution would solve the problem. Both assumptions must be applied to both periods (1 and 2).

The linear programming model (3.8)–(3.13) was implemented in the company for each hybrid and generated approximately 1,500 variables and 1,500 constraints. The implementation was stepwise starting by an experiment with four hybrids and data of 2 years. Although the results of the experiments were promising, the senior managers decided to compare the model's performance during a whole season in parallel with the actual method.

After this second validation process, the solutions presented by the model suggested to plant fewer areas, reducing the inventory to carry over. As a final result, the margins increased approximately $5 million on the 18 hybrids studied. Applying the model on the historical data of the company, the old method overestimated demand about 73 % of the time. The model helped to reduce the bias of the forecasting procedures.

3.3.3 Production and Transportation in Multistages and Multi-periods

Junqueira and Morabito (2012) studied the aggregate production and logistics planning of seed corn of a typical Brazilian multiplant company. They presented a linear programming model to support medium-term planning decisions, and they report the results of the model application for a full seed corn season in this case study (more information about this case study is also found in Junqueira and Morabito (2008)). The studied enterprise experienced a rapid growth from one to four facilities in a few years. The planning policy that was once used, sending the seed to the closest facility, could be improved in order to explore opportunities with tax planning.

Basically, the seed corn aggregate production and logistics planning problem can be formulated as a multi-period, multi-product, multiplant two-stage model. Stage 1 corresponds to the harvesting, transportation, and preparation of raw material, resulting in dried grains, while stage 2 corresponds to the transportation (if necessary) and processing of dried grains, resulting in final products. Both stages involve capacity constraints, and there are different suppliers (crop fields), customers (demand markets), and types of harvest (spike and grain). Note that the planning is aggregated in time periods (months), final products (families of variety and hybrid seed corn), and production stages (1 and 2), besides supply (fields) and demand areas (markets) (for more information about the model formulation, see Junqueira and Morabito (2006)).

Figure 3.1 presented before shows the schematic model of the tactical production and logistics planning problem proposed by the authors. The planning time horizon of the problem is typically of 1 year (i.e., the season from April of the current year to March of the following year), divided into monthly periods. The problem involves the irrigated and nonirrigated crop fields with their raw materials and productivities, the different types of harvest of raw material, the transport of raw materials to the processing plants, the preparation of raw material in the plants, the storage of dried grains in the plants and the transfer of these grains between plants, the processing of semi-processed products in the plants, the warehousing of final products in the plants, and the transfer of these products between plants and the transport of final products to the demand areas, among others.

The study assumed that the planting decisions have already been taken by the company, that is, a plan determining which seed corn is planted in which crop field and period and the required type of harvest. This assumption helps to develop the linear programming model of the next section (instead of more complex mixed integer programs) to derive effective tactical plans for the company. As shown in the next section, this planning is useful to determine minimum cost flows from the agricultural fields to the market regions, particularly in cases where the company is increasing the scope of its market, the number of final products, and the number of plants, and in situations where there are changes in the Brazilian state policies of circulation tax exemption.

For instance, it is known that the transfer of dried grains between plants can happen only once, which means that the outputs of the raw material preparation go directly to the silo of the plant and from this silo it is moved directly to the production process. This simplification is reasonable since it disregards only the situations in which the product of the same crop field is transferred in the form of dried grains twice. However, these situations are not desirable in the tactical planning because they involve additional transportation costs. Similarly, in the distribution of the final products, after being processed in the plant, the product is moved to the plant warehouse. Then, the product can either be sent to a market region to meet its demand in this period or be stored in the plant warehouse to meet the demands of the following periods. In this study, the transfer of products between plant warehouses is not considered in the tactical planning. Similarly to the transfer of dried grain between the aforementioned plants, this transfer is not desirable because it involves additional transportation costs.

The level of aggregation of supply and demand areas should be sufficient to characterize the differences between transportation costs and circulation taxes. The level of aggregation of product families should be considered in terms of the customer demands, which depend on specific features, such as differences in product sale prices and adequacy of the product to the customer region. The planting and crop management are also made specifically for the product. The periods should be short enough to avoid processing multiple fields in the same period of time and long enough to not be influenced by the production setup times. The packaging unit of analysis used in the model is a 20 kg sack, as the final products are sold. In the following, index i indicates the crop field, index j indicates the processing plant (UBS), index k indicates the demand market, index h indicates the product, index m indicates the type of harvest, and index t indicates the time period.

The decision variables of the model are

$X_{h,i,j,m,t}$, amount of product h collected in crop field i with harvest type m and transported to plant j in period t for preparing dried grain [20 kg sack]
$Y_{h,j,jp,t}$, amount of product h prepared in plant j and transported to plant jp in period t for processing final product [20 kg sack]
$Z_{h,j,k,t}$, amount of product h processed in plant j and transported to demand area k in period t [20 kg sack]
$Ig_{h,j,t}$, amount of product h stored as dried grain in plant silo j in period t [20 kg sack]
$If_{h,j,t}$, amount of product h stored as final product in plant warehousing j in period t [20 kg sack]

The input parameters of the model are

$Ctg_{i,j,m,t}$, freight cost to transport one ton from crop field i with harvest type m to plant j in period t [\$/t]
$Ctp_{j,jp,t}$, freight cost to transport one ton from plant j to plant jp in period t [\$/t]
$Ctd_{j,k,t}$, freight cost to transport a 20 kg sack from plant j to demand region k in period t [\$/sack]
$\mathrm{ICMS}_{j,k}$, circulation tax rate paid by the sale of a 20 kg sack of plant j in demand region k
$P_{h,k}$, sale price of product h in demand region k [\$/sack]
$Cpg_{j,m}$, cost of raw material preparation of plant j with harvest type m to produce dried grains [\$/sack]
Cpp_j, cost of dried grain processing of plant j to produce final products [\$/sack]
$Ig_{h,j,0}$, initial quantity of product h, stored as dried grains in plant j [20 kg sack]
$Ip_{h,j,0}$, initial quantity of product h, stored as final product in plant j [20 kg sack]
Ipf_h, minimum final quantity of product h, stored as final product in the plants [20 kg sack]
$D_{h,k,t}$, quantity demanded of product h in market region k in period t [20 kg sack]
$S_{h,i,m,t}$, quantity supplied of product h by crop field i with harvest type m in period t defined in the harvest plan [20 kg sack]
Eg_m, efficiency of raw material preparation with harvest type m to produce dried grains [%]
Ep, efficiency of dried grain processing to produce final products [%]
$Kg_{j,m}$, capacity of raw material preparation of plant j for harvest type m to produce dried grain [sacks per day]
Ks_j, capacity of storage of plant silo j [sacks per day]
Kp_j, capacity of dried grain processing of plant j to produce final product [sacks per day]

Kw_j, capacity of storage of plant warehouse j [sacks per day]
N_t, number of workdays in period t [days]

The objective function of the model minimizes total production and logistics costs, that is, raw material transportation costs from the crop fields to the plants (first term in (3.16)), raw material preparation costs in the plants (second term), dried grain transferring costs between plants (third term), dried grain processing costs in the plants (fourth term), and final product distribution costs to the markets (fifth term), which are split into transportation costs and circulation taxes:

$$\begin{aligned} TC = &\sum_{h=1,\dots,H}\sum_{i=1,\dots,H}\sum_{j=1,\dots,J}\sum_{m=1,\dots,M}\sum_{t=1,\dots,T}\frac{Ctg_{i,j,m,t}}{Eg_m Ep}X_{h,i,j,m,t} \\ &+ \sum_{h=1,\dots,H}\sum_{i=1,\dots,I}\sum_{j=1,\dots,J}\sum_{m=1,\dots,M}\sum_{t=1,\dots,T} Cpg_{j,m}X_{h,i,j,m,t} \\ &+ \sum_{h=1,\dots,H}\sum_{j=1,\dots,J}\sum_{jp=1,\dots,J}\sum_{t=1,\dots,T}\frac{Ctp_{j,jp,t}}{Ep}Y_{h,j,jp,t} \\ &+ \sum_{h=1,\dots,H}\sum_{j=1,\dots,J}\sum_{jp=1,\dots,J}\sum_{t=1,\dots,T} Cpp_j Y_{h,j,jp,t} \\ &+ \sum_{h=1,\dots,H}\sum_{j=1,\dots,J}\sum_{k=1,\dots,K}\sum_{t=1,\dots,T} Ctd_{j,k,t}Z_{h,j,k,t} \\ &+ \sum_{h=1,\dots,H}\sum_{j=1,\dots,J}\sum_{k=1,\dots,K}\sum_{t=1,\dots,T} P_{h,k}\mathrm{ICMS}_{j,k}Z_{h,j,k,t} \end{aligned} \tag{3.16}$$

It should be noted that the objective function does not include raw material and harvest costs since the model assumes that the harvest plan has already been made. Moreover, it does not include storage costs because it does not acknowledge the possibility of transferring the product only looking for the most viable storage without necessarily having been processed. The constraints of the model basically correspond to mass balancing and capacity limitation. The mass balancing constraints are

$$\sum_{j=1,\dots,J} X_{h,i,j,m,t} = S_{h,i,m,t} \quad h = 1,\dots,H,\ i = 1,\dots,I,\ m = 1,\dots,M,\ t = 1,\dots,T \tag{3.17}$$

$$\sum_{i=1,\dots,I}\sum_{m=1,\dots,M} X_{h,i,j,m,t} + Ig_{h,j,t-1} = \sum_{jp=1,\dots,J} Y_{h,j,jp,t} + Ig_{h,j,t} \quad h = 1,\dots,H,\ j = 1,\dots,J,\ t = 1,\dots,T \tag{3.18}$$

$$\sum_{jp=1,\dots,J} Y_{h,j,jp,t} + If_{h,j,t-1} = \sum_{k=1,\dots,K} Z_{h,j,k.,t} + If_{h,j,t} \quad h = 1,\dots,H,\ j = 1,\dots,J,\ t = 1,\dots,T \tag{3.19}$$

$$\sum_{j=1,\ldots,J} Z_{h,j,k,t} = D_{h,k,t} \quad h = 1, \ldots, H, k = 1, \ldots, K, t = 1, \ldots, T \tag{3.20}$$

$$\sum_{j=1,\ldots,J} If_{h,j,T} \geq Ipf_h \quad h = 1, \ldots, H \tag{3.21}$$

The capacity constraints (and variable domain restraints) are

$$\sum_{h=1,\ldots,H} \sum_{i=1,\ldots,I} X_{h,i,j,m.,t} \leq N_t Kg_{j,m} \quad j = 1, \ldots, J, \ m = 1, \ldots, M, \ t = 1, \ldots, T \tag{3.22}$$

$$\sum_{h=1,\ldots,H} \sum_{jp=1,\ldots,J} Y_{h,j,jp,t} \leq N_t Kp_j \quad j = 1, \ldots, J, t = 1, \ldots, T \tag{3.23}$$

$$\sum_{h=1,\ldots,H} Ig_{h,j,t} \leq Ks_j \quad j = 1, \ldots, J, t = 1, \ldots, T \tag{3.24}$$

$$\sum_{h=1,\ldots,H} If_{h,j,t} \leq Kw_j \quad j = 1, \ldots, J, t = 1, \ldots, T \tag{3.25}$$

$$\begin{aligned} & X_{h,i,j,m,t}, Y_{h,j,jp,t}, Z_{h,j,k,t}, Ig_{h,j,t}, If_{h,j,t} \geq 0 \\ & \quad h = 1, \ldots, H, i = 1, \ldots, I, j = 1, \ldots, J, jp = 1, \ldots, J, \\ & \quad k = 1, \ldots, K, m = 1, \ldots, M, t = 1, \ldots, T \end{aligned} \tag{3.26}$$

Constraint (3.17) ensures that in the harvest plan, each product h in crop field i with harvest technology m is transported in period t and the preparation of the raw material is made in the plants. Constraint (3.18) ensures that product h is either processed in plant j in period t or stored in the plant silo j as dried grains for the next periods. Constraint (3.19) ensures that product h is either delivered to regions of demand in period t or stored in the plant warehouse j as final products for the next periods. Constraint (3.20) ensures that the demand of product h in region k in period t is met. Constraint (3.21) ensures that the desired final stock of product h is met. Constraints (3.22)–(3.25) consider, respectively, the limitations in the capacities of preparing raw material, processing dried grains, storing dried grains, and storing final products of the plants in each period. Furthermore, constraint (3.26) imposes that the decision variables are nonnegative.

In order to solve the model, the authors used the modeling language GAMS 2.0.10.0 and the optimization software CPLEX 7.0.0 (Brooke et al. 1998; Kiup 1993). A basic microcomputer (Intel Pentium-4, 2.20 GHz, 496 MB of usable RAM, and HD of 40 GB) was used for these experiments, and the original model resulted in 4,758 linear equations and 6,679 variables. After the preprocessing of CPLEX, the numbers of lines and columns of the technological matrix were substantially reduced to 625 and 2,138, respectively. The computer runtime (in seconds) required to solve the model was less than 1 s, which is relatively low and quite acceptable to support the decisions in practice, providing flexibility to generate and evaluate different problem scenarios.

3.3.4 Comparison Among the Models

Table 3.2 shows a comparison among the models previously showed. Zuo et al. (1991) was the first work that analyzed the problem with emphasis on the integration between production and transportation, as well as, with concerns on the capacity of multiplants and multi-products. Jones et al. (2003) approached the stochastic nature of seed production and demand balancing shortages and salvage production. They also analyzed the influence of two production time periods, the first in North America and the second in South America, knowing that the production of the first period can be used on the second period. Junqueira and Morabito (2012) proposed a model that can be seen as a more detailed representation of Zuo et al. (1991)'s model considering multiple time periods, tax costs besides production and transportation costs, as well as the division of the production stages into raw material preparation and processing. This separation enables the transfer of dried grains between facilities.

Table 3.2 Comparison of models characteristics

	Zuo et al. (1991)	Jones et al. (2003)	Junqueira and Morabito (2012)
Decision level	Tactical	Tactical	Tactical
Integration structure	Production and outbound transportation	Production	Inbound transportation, production, and outbound transportation
Number of stages	Two (production and outbound transportation)	One	Four (inbound transportation, raw material preparation, processing, and outbound transportation)
Planning horizon	One time period	Two time periods (North America and South America)	Multiple time periods
Number of products	Multiple products	One product	Multiple products
Nature of demand	Deterministic	Stochastic with penalty for demand shortage	Deterministic
Capacity	Limited	Unlimited	Limited
Nature of production	Deterministic	Stochastic with penalty for salvage production	Deterministic
Tax costs	Not considered	Not considered	Included

These studies indicated that the application of optimization models coupling production and distribution can be useful to support the integration between organizational functions of production and commercialization.

3.4 Optimizing Production and Logistics Planning in a Brazilian Seed Corn Company

The model presented in Sect. 3.3.3 (3.16)–(3.26) was applied by Junqueira and Morabito (2012) in a Brazilian seed corn company. After expanding from one to four processing plants (UBS), the company contracted a specialized consultancy to analyze its production and logistics processes. This partnership took more than 1 year, and among other factors, it focused the dispatch policy from the crop fields to the plants. At that time, the multiplant planning method of the company had the same policy of the single plant method, i.e., always transporting the corn seed to the nearest UBS of the field, independently of the demand.

After a number of interactions with the production and sales departments, the team realized that, besides the demand itself, the consideration of circulation taxes had a significant impact over production costs, affected by this dispatch policy. This information motivated the development of the proposed production and logistics planning optimization approach, which was used by the consultancy to support some recommendations to the company managers. The model was also applied to generate and evaluate several scenarios with the company managerial board to analyze tactical decisions, exploring the fact that one of the plants had an important production cost reduction because of its drying system. The next step would be the implementation of the model in the production planning and control system of the company.

In order to measure the benefits of the application of the proposed model, its solution was compared to the company method, which considered only transportation costs. For more details of this company method, the reader can consult Junqueira and Morabito (2008). This comparison is illustrated in Table 3.3 and evaluates the impact of circulation taxes (scenario 1), preparing costs (scenario 2), and the combination of both (scenario 3). In this table, columns "Company solution" and "Model solution" represent the percentage of each component cost in the total cost of each solution, while column "Cost reduction" represents the cost difference (in percentage) between the model and the company solution values.

In scenario 1, the model solution reduced the cost component of circulation taxes by 27 %. On the other hand, transportation costs of raw material increased by 20 %, and transfer costs among plants were generated (1 % of total costs). The model solution reduced the total cost by 12 % with respect to the company solution, which was equivalent to $420,000, disregarding additional preparation costs. This reduction was possible because, among other factors, there was spare capacity at some of the processing plants (especially UBS that pursued circulation tax advantages).

Table 3.3 Comparison between the company dispatching policy and the solution proposed by the model

	1—ICMS			2—Preparation cost			3—ICMS + preparation cost		
	Company solution (%)	Model solution (%)	Cost reduction (%)	Company solution (%)	Model solution (%)	Cost reduction (%)	Company solution (%)	Model solution (%)	Cost reduction (%)
RM transportation	15	20	20	23	34	20	13	20	23
RM preparing	0	0		26	4	−87	15	3	−83
Transfer	0	1		0	0		0	2	
Delivery	33	36	−3	51	62	−1	28	34	−6
ICMS	53	43	−27	0	0		45	41	−30
Total	100	100	−12	100	100	−19	100	100	−23

In scenario 2, the cost component of raw material preparation was reduced by 87 % with the model solution. On the other hand, transportation costs of the raw material increased by 20 %. Considering the priority to minimize preparing raw material and transportation costs in one of the plants and disregarding the costs associated to circulation taxes, the model solution reduced total costs by 19 %, equivalent to \$450,000. Once more, this reduction was possible because, among other factors, there was spare capacity at some of the processing plants (especially in the UBS that pursued lower preparation costs).

In scenario 3, the model solution was able to reduce both the cost component of circulation taxes by 30 % and the raw material preparation costs by 83 %. On the other hand, transportation costs of the raw material increased by 23 %, and transfer costs among plants were generated (2 % of total costs). The overall cost reduction was 23 %, equivalent to \$974,000. Again, this reduction was possible due to spare capacities at some of the plants with circulation tax and preparation cost advantages. These and other comparisons showed important opportunities to reduce total costs using the model. In cases where some model parameters were uncertain or even unknown, sensitivity analysis were straightforwardly performed with the model to consider a number of possible scenarios.

These and other simulations using this approach were performed by the consulting team to support production and logistics tactical decisions to the company. These results were presented to the company decision makers and guided some of these decisions, such as (1) changing the original policy of always transporting seeds to the closest plant of the crop fields because of the impact of circulation taxes and drying costs, (2) activating working shifts in one of the processing plants and deactivating in other due to the same reasons, (3) changing the drying system of one of the processing plants in order to have similar drying costs as others and reducing delivery transportation costs to some states, and (4) building a new spike dryer in one of the plants (UBS GO1) with similar costs to others to reduce transportation costs of raw material from the crop fields to the plant. All these actions were implanted in the seed corn company based on the results obtained with the model simulations, except to (4) which became part of future investments.

3.5 Concluding Remarks

In this study, we reviewed linear programming models of the literature which were applied to different corn seed production and logistics settings. The application of these models resulted in important practical benefits to the studied companies. Besides the economic gains, the integration of production and distribution operations considered in the approaches also provided relevant organizational improvements in these case studies. In particular, the linear model presented in Sect. 3.3.2, based on discrete approximations of the probability distributions of the corn yields and the product demand, resulted in a potential approach to reduce forecasting bias. And the linear model discussed in Sect. 3.3.3 can be an effective optimization tool in the contexts of relevant multijurisdictional costs, as it explicitly considers tax planning decisions that can highly influence the minimization of production and transportation costs and the determination of optimal corn seed flows in the supply chain.

References

Ahumada O, Villalobos JR (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 195:1–20
Balaji K, Viswanadham N (2008) A tax integrated approach for global supply chain network planning. IEEE Trans Automat Sci Eng 5:587–596
Basset M, Gardner L (2010) Optimizing the design of global supply chains at Dow AgroScience. Comput Chem Eng 34:254–265
Basset M, Gardner L (2013) Designing optimal global supply chains at Dow AgroScience. Ann Oper Res 203:187–216. doi:10.1007/s10479-010-0802-2
Brooke A, Kendrick D, Meeraus A, Raman R, Rosenthal R (1998) GAMS a user's guide. GAMS Development Corporation, Washington, DC
Caixeta-Filho JV, Swaay-Neto JMV, Wagemaker AP (2002) Optimization of the production planning and trade of lily flowers at Jan de Wit Company. Interfaces 32:35–46
Chen ZL (2004) Integrated production and distribution operations: taxonomy, models and review. In: Simchi-Levi D, Wu SD, Shen Z-J (eds) Handbook of quantitative supply chain analysis: modeling in the E-business era. Kluwer Academic, Dordrecht
Colin EC, Cipparrone FAM, Shimizu T (1999) Otimização do custo de transporte na distribuição-armazenagem de açúcar. Produção 9:23–30
CONFAZ Convênios do ICMS (2006) http://www.fazenda.gov.br/confaz. Accessed 10 April 2006
Higgins AJ, Muchow RC (2003) Assessing the potential benefits of alternative cane supply arrangements in the Australian sugar industry. Agric Syst 76:623–638
Higgins A, Laredo LA (2006) Improving harvesting and transport planning within a sugar value chain. J Oper Res Soc 57:367–376
INFORMS (2012) Institute for operations research and the management sciences. http://www.informs.org. Accessed 1 Jan 2012
Johnson LA, Montgomery DC (1974) Operations research in production planning, scheduling, and inventory control. Wiley, New York
Jones PC, Lowe T, Traub RD, Kegler G (2001) Matching supply and demand: the value of a second chance in producing hybrid seed corn. Manufact Serv Oper Manage 3:122–137
Jones PC, Lowe T, Traub RD (2002) Matching supply and demand: the value of a second chance in producing seed corn. Rev Agric Econ 24:222–238
Jones PC, Kegler G, Lowe T, Traub RD (2003) Managing the seed-corn supply chain at Syngenta. Interfaces 33:80–90
Junqueira RAR, Morabito R (2006) Um modelo de otimização linear para o planejamento agregado da produção e logística de sementes de milho. Produção 16:510–525
Junqueira RAR, Morabito R (2008) Planejamento otimizado da produção e logística de empresas produtoras de sementes de milho: um estudo de caso. Gestão & Produção 15:367–380
Junqueira RAR, Morabito R (2012) Production and logistics planning considering circulation taxes in a multi-plant seed corn company. Comput Electron Agric 84:100–110
Kawamura MS, Ronconi DP, Yoshizaki H (2006) Optimizing transportation and storage of final products in the sugar and ethanol industry. Int Trans Oper Res 13:425–439
Kiup CAC (1993) Algebraic languages for mathematical-programming. Eur J Oper Res 67:25–51
Lopez E, Miquel S, Pla LM (2006) Sugar cane transportation in Cuba, a case study. Eur J Oper Res 174:374–386

Munhoz JR, Morabito R (2014) Optimization approaches to support decision making in the production planning of a citrus company: a Brazilian case study. Comput Electron Agric 107:45–57
Oliveira JA, Carvalho MLM, Vieira MGGCV, Von Pinho EVR (1997) Efeito do método de colheita na qualidade física, fisiológica e sanitária de sementes de milho. Revista Brasileira de Sementes 19:200–206
Paiva ROP, Morabito R (2009) An optimization model for the aggregate production planning of a Brazilian sugar and ethanol milling company. Ann Oper Res 169:117–130
Rodriguez-Sanchez S, Pla LM, Albornoz V (2012) Modelling tactical planning decisions through a linear optimization model in sow farms. Livestock Sci 143:162–171
Salassi ME, Breaux JB, Naquin CJ (2002) Modeling within-season sugarcane growth for optimal harvest system selection. Agric Syst 73:261–278
Shackelford DA, Shevlin T (2001) Empirical tax research in accounting. J Account Econ 31:321–387
Shapiro JF (1993) Mathematical programming models and methods for production planning and scheduling (Handbook in Operations Research and Management Science. v4). In: Graves SC, Rinnooy Kan AHG, Zipkin PH (eds) Logistics of production and inventory. Elsevier Science Publishers B.V., Amsterdam, pp 371–439
Shapiro JF (2001) Modeling the supply chain. Duxbury, Pacific Grove
SIFRECA (Sistema de informação de fretes) Custo de momento de transporte de milho (2004) http://sifreca.esalq.usp.br/sifreca/pt/index.php. Accessed 10 Nov 2004
Taube M (1996) Integrated planning for poultry production at Sadia. Interfaces 26:38–53
Yoshizaki HTY, Muscat ARN, Biazzi JL (1996) Decentralizing ethanol distribution in southeastern Brazil. Interfaces 26:24–34
Zuo M, Kuo W, McRoberts KL (1991) Application of mathematical programming to a large-scale agricultural production and distribution system. J Oper Res Soc 42:639–648

Chapter 4
Harvest Planning in Apple Orchards Using an Optimization Model

Marcela C. González-Araya, Wladimir E. Soto-Silva, and Luis G. Acosta Espejo

4.1 Introduction

During the 2010 season, Chile became the first apple exporter in the southern hemisphere and reached second place worldwide (11.4 % of the world market), right after China with 14.2 % (Bravo 2011). On the other hand, in the last few years, agriculture in Chile has lost competitiveness, mainly because of a 30 % decline of the US$, of a 300 % increase in the prices of goods such as fertilizers and pesticides (Contreras 2008), of a 10 % decrease of the availability of agricultural workforce, and finally of a 12 % increase of the workers' wages (Domínguez 2007; Alarcón 2008). These factors progressively led to a decrease in the growth rate of the agricultural sector during the period 2005–2008, being surpassed by the National GDP growth (Contreras 2008). For this reason, producers and exporters in the industry are focusing on improving practices and better allocate resources, in order to reduce the costs involved along the supply chain.

In Chile, the apple (*Malus domestica*) is the second most planted fruit species in the country, covering about 13 % of the national fruit area. The Maule Region

M.C. González-Araya (✉)
Departamento de Ingeniería Industrial, Facultad de Ingeniería,
Universidad de Talca, Camino a Los Niches km. 1, Curicó, Chile
e-mail: mgonzalez@utalca.cl

W.E. Soto-Silva
Programa de Magíster en Gestión de Operaciones, Facultad de Ingeniería,
Universidad de Talca, Camino a Los Niches km. 1, Curicó, Chile
e-mail: wsoto@utalca.cl

L.G.A. Espejo
Departamento de Ingeniería Comercial, Universidad Técnica Federico Santa María,
Av. Santa María, 6400, Vitacura, Santiago, Chile
e-mail: luis.acosta@usm.cl

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_4

concentrates 59.8 % of the planted area and the O'Higgins Region 28.8 % of the surface of apple (Bravo 2011). The country counts with about 324,294 ha of fruit trees, of which about 39,000 are apple trees, which represent 12 % of the national agricultural area. The Maule Region itself has approximately 20,000 ha of apple trees, 84 % of which are the red variety and 26 % the green one (INE 2007). Regarding the production of this fruit, from 2000 to 2008, the modernization of irrigation policies, as well as the incorporation of new planting systems and treatments of diseases, enabled a significant increase in the apple yield from 22,492 kg/ha, in 2000, to 39,142 kg/ha, in 2008 (Maule Competitiveness Center 2010). Plus, the apple represents 31.5 % of the fresh fruit exports, with earnings' returns of about $485,829 Million dollars during the 2008/2009 season.

As such, Maule is the main region producing apples in Chile, and, therefore, an improvement of the productivity and quality of the fruit produced in the region would have a significant impact on its economy. In this sense, this research seeks to increase the amount of fruit that could be exported, which would generate major incomes for the regional agricultural sector.

When harvesting apples, one of the main factors affecting the quality of the fruit (and therefore its export) is the fruit ripeness and this must be taken into account when planning the harvest and the efficient use of resources such as equipment (ladders, containers), machinery (tractors, cars, and collectors carriers), and of course workforce, which is critical in this process. Indeed, over the last 10 years, during the harvest season, it has been every year harder to find the required quantity of seasonal workers, creating a 10 % decrease of them (Contreras 2008), and a 12 % increase in their costs over the last 3 years (Alarcón 2008). The main problem with the lack of workers is that the fruit cannot be harvested at the right ripeness and therefore cannot be selected for export, as it should. This issue has reduced the profitability of the agricultural producer, even though apple plantings remained constant over the last 10 years and the yield per hectare increased by approximately 73 % (Maule Competitiveness Center 2010).

This work proposes a mathematical programming model to support planning decisions in an orchard and minimizing the amount of resources used, ensuring the production of good quality fruit for export. Furthermore, this model provides a harvest schedule that minimizes the loss of fruit for not fulfilling the desired quality parameters for export. The constraints of the model seek to fulfill the demands from the fruit packing plants, respect the processing capacity of the plants, availability of production in orchards, and harvesting time accordingly to the planted apple variety. The model was applied to three orchards in the Maule Region, specifically in the province of Curicó. The data was collected during the 2009/2010 harvest season for each of them.

The model here presented improves the tactical planning of a pome orchard, considering harvesting times for each fruit variety and penalizing the quality if the fruit is not harvested at the right time. Previous models have not so far focused on the specific activities of the apple industry.

This chapter is structured as follows: Section 4.2 presents a review of mathematical programming models developed and applied in agriculture, which are the basis for the proposed model. Section 4.3 describes the global management of an apple orchard and its difficulties, whereas Sect. 4.4 focuses on the main factors of harvesting and post-harvesting. Section 4.5 presents the proposed model and the results of computational experiments. Section 4.6 is a summary of the analysis of the main results obtained by applying the model in the orchards studied. Finally, Sect. 4.7 provides the conclusions of this work.

4.2 Crop Planning in Agricultural Sector

Regarding harvest planning in agriculture, a review covering 40 years of research can be found in Bjorndal et al. (2012). Even though within the agricultural sector, optimization models for harvest planning have been proposed for different types of fruit and vegetables, such as orange (Caixeta-Filho 2006), sugar (Higgins and Laredo 2006), grape (Ferrer et al. 2008), and tomato (Van Berlo 1993), there is no reference to authors who would have studied through mathematical modeling, harvest planning in apple orchards.

The different models developed for harvest scheduling reflect the fact that harvesting operations vary from one crop to another (Glen 1987). But most crops must be harvested during a relatively short period of intense activity. Moreover, the techniques that help planning these activities can bring considerable economic profit. Glen (1987) considers that implementing crop-planning models has greatly contributed to the design and development of harvesting systems for agricultural goods.

Regarding natural resources area, Schuster and Allen (2004) proposed a model of production planning for grape juice. Tadei et al. (1995) developed another model in this area, seeking to minimize stock levels to satisfy a known demand, determining the number of workers required for each month of the year. Thus, the model estimates the number of workforce to be subcontracted in the months of high demand. The model is applied to a company that produces perishable goods, some seasonal, some not.

Regarding optimization models for managing pome orchards, Hester and Cacho (2003) developed a dynamic simulation model, which considers economic and biological interrelationships within an apple orchard, in order to improve the incomes on a given planning horizon. Following this research line, Català et al. (2013) presented a model for strategic planning pome orchards, which considers the varieties and planting densities fields, delivering an optimal investment policy for a given orchard. This policy minimizes the conversion of the orchard (replacing crops) as well as the financial requirements for this conversion.

Ferrer et al. (2008) present a model for planning and scheduling harvest operations of grapes for wine. This model considers the costs related to the harvesting activities and these related to the quality loss of grapes, due to either an early or a delayed harvest. Decisions in this model include the amount of grapes to be harvested in different orchard fields for each period. This aspect depends on the maturity of the fruit, the routes of the workers in the orchard, the scheduled harvest time between the different fields, and the number of workers to hire or let go for each period of the harvest season. The authors conducted computational experiments and an application to a vineyard in central Chile.

Moreover, Ahumada and Villalobos (2011a) also developed a planning and distributing model for fresh goods, seeking to increase the profits of the producers during a harvest season. These authors also included the quality of the product in the objective function of the model by estimating the rejection or discarding cost of the shipment. The model considered whether the product quality reached a limit established by consumers as well as a fine for the deterioration of the product all along the supply chain, from harvest to consumers. The authors performed computational experiments with hypothetical case study producers of tomatoes and green peppers, making possible their resolution with commercial optimization software.

In the work of Ahumada and Villalobos (2011b), the authors propose once more a planning and distributing model for fresh goods, seeking to increase the profits of the producers; but in this model, they let producers have some control over decisions related to harvest distribution.

In a later work, Ahumada et al. (2012) proposed a stochastic tactical model for both harvest planning and distribution, incorporating uncertainty in weather and product demand. The aim of this model was to propose more robust harvest plans, allowing different levels of risk exposure. For the computational experiments, the authors used a stochastic version of Ahumada and Villalobos' case (2011b).

Finally, it should be noted that both Ahumada and Villalobos (2009) and Zhang and Wilhelm (2011) point out that few models or planning tools have been proposed to harvest perishable products, and so that this area requires a greater amount of research. Actually, the first authors who mentioned the lack of models to support decisions in the agricultural area were Lowe and Preckel (2004), describing new problems that had arisen and should be solved in the area.

Since 32 % of fresh fruits exported from Chile are apples, the agricultural sector requires a new optimization model for planning harvest the literature has so far not produced. Our work proposes a model of Mixed Integer Linear Programming developed for this purpose, as it will be exposed in the following sections.

4.3 Global Context of Apples Harvest Planning in Chile

When considering apple harvest planning, many factors have to be taken into account; many of them (including all kinds of costs) at a much earlier stage than harvest itself. If fruit quality remains a key factor, others are involved throughout the process of fruit growth in crop management and post-harvest (Gil 2001).

4.3.1 Managing an Orchard

To understand our model, it is relevant to have an overview of the management of an apple orchard.

Table 4.1 Planted area and characteristics of the main varieties of the Maule Region

Category	Variety	Planted area (Hectares)	Harvest dates	Harvest method
Gala	Royal Gala	4,683.3	Second half of February	Selective pickings (at most three harvests per season)
	Galaxi	1,241.5	Mid-February	
Red	Red Chief	1,770.8	Late March–Early April	Strip picking (once-over pick)
	Scarlet	1,370.4	Late March–Early April	
Fuji	Fuji	2,135.1	Second half of April	Selective pickings (between two and four harvests per season)
Granny Smith	Granny Smith	3,114.9	Second half of March	Strip picking (once-over pick)

Source: Center for Pome (2010)

The Maule Region counts with about 51 varieties of apples that, according to agricultural managers, can be mostly classified into four categories. This classification enables to manage more accurate harvest and/or yield indicators, together with an estimation of harvest periods for each of them. Table 4.1 shows the planted area for the main varieties of the Maule Region, in addition to the times and harvest method for each of them.

Granny Smith variety is the only green apple present among the six varieties planted in the Maule Region. Regarding the Gala apples, they do not ripen at the same time (expected color and size), so they are selectively picked (more than one harvest along the field in a harvest season). Generally speaking, for this category, three "selective pickings" are carried out at the most, each selective picking implying browsing the whole field in search of suitable apples for harvest. On the other hand, Red apples do ripen at the same time. For this reason, they are strip picked (only one harvest along the field in a harvest season). Finally, Fuji apples represent about six apple varieties planted in Chile. However, since all have similar features they can be classified into one single category. Just as Gala apples, Fuji apples are selectively picked and, between two and four harvests take place in a harvest season.

4.3.1.1 The Use of Fertilizer

Orchard management also involves fertilizers and nutrients to be spread on the ground. Indeed, to obtain a good production of apples, nutrition, or nutritional balance is a key factor. In apples, nitrogen, potassium, calcium, and magnesium must be well balanced to prevent physiological disorders, and obtain a good fruit quality. For this reason, Yuri and Moggia (2007) noted "care should be taken not to over-fertilize with nutrients such as nitrogen and potassium, which is a common situation in orchards. In this case, it creates a calcium imbalance, provoking disorders such as greasiness and cracking pedicle, since these three elements are competitive with each other to occupy locations within cells."

4.3.1.2 Tradition, Timing, and Location

In 1984, France and Thornley observed that fruit harvesting in the orchards is generally traditionally performed, only based on the experience of those involved in this process, either administrative staff or field workers. This observation is also true for harvest planning in Chile.

On the other hand, apple harvesting in Chile has a time horizon that depends on the varieties planted in the orchard. However, this horizon usually starts in mid-February and ends in late April. In addition, each variety generally involves a 2-week harvest period, only taking into account those varieties planted in the Maule Region.

In most of the country's orchards, harvest areas are divided into fields, each field corresponding to an area with several rows of trees of a single variety. In these orchards, during the daily harvest planning, collection areas, i.e., the rows of fields, are assigned at the beginning of the workday to the workers, who are transferred to these areas to start harvesting the fruit. This decision is typically made by the manager of the orchard, based on the characteristics of the fruit ripening to be harvested, according to the parameters required by the consumer market.

4.3.1.3 Workforce and Costs

As mentioned earlier, one of the critical points in the apple harvest is to hire workers for the season. Agricultural unemployment rate has remained below 4 % (Contreras 2008). The labor shortage is mainly due to the increasing job opportunities in the cities, due to the significant growth experienced by construction and retail. In addition, cultural reasons have discouraged the rural residents to work in the agricultural sector (Contreras 2008). This worrying factor implies that there is more work to do than workers to do it. As a consequence, these workers ask for wages that are often not profitable for producers, making negotiation necessary. Indeed, if the fruit is not harvested on time, its quality deteriorates, damaging sales incomes. For this reason, knowing in advance the number of workers that will be needed for harvest in each period is necessary.

Besides having a good estimation of the number of workers required, it is necessary to estimate the number of crews (working groups) to assign to each field, considering the size and the variety to be harvested in each field. Finally, the organization of equipment needed for the harvest is also required: containers, ladders, among others.

Another aspect that must be taken into account is the important economic changes the apple harvesting business has suffered over the last few years. Contreras (2008) explains that the costs of fertilizers have increased from 1998 to 2008. As such, urea price has increased over 300 %, while Triple Superphosphate (TSP) and mono-ammonium phosphate (MAP) have increased 200 % but now remain constant. These fertilizers are the most used in the Chilean fruit industry.

Table 4.2 Costs involved in the harvest of one hectare of apple

Operating costs for an apple orchard (full production)	
Item	%
Workforce (man-day)	43.6
Agrochemicals	17.4
Fertilizers	5.9
Pesticides	11.6
Machinery (machine-day)	5.9
Freight	4.8
Management (15 %)	10.8
Total	100

Source: Centre for Pome, Universidad de Talca (2010)

As shown in Table 4.2, the workforce represents the most important cost. This cost is related to the number of man/days needed to harvest one hectare of apples. The costs of chemicals come next and are used for orchard management, pre- and post-harvest.

One last aspect that must be taken into account when considering orchard management is the price of the US$. Between 2003 and 2009, the US$ lost 30 % in Chile, with critical consequences for agricultural producers and exporters who receive their incomes in US$ but spend in Chilean Peso, the local currency. As such, their profit margin is greatly reduced.

4.4 Apple Harvesting

During the apple harvest, timing influences the condition and quality of the fruit (Gil 2001). Thus, harvesting an unripe fruit can develop a number of physiological disorders during post-harvest such as bitter pit and/or scalding. On the other hand, by harvesting an overripe fruit, other disorders could be observed during post-harvest such as lenticel, as a consequence of nutritional imbalances (calcium deficiency), pedicle cracking, excessive grease formation, and increased dehydration and yellowing of the fruit. All these elements could block the potential export of the fruit (Gil 2001). These disorders are described in more detail in Table 4.3.

According to Gil (2001), the parameters to be considered in order to obtain a good quality fruit are:

- *Shape of fruit*: Are only accepted those fruits having the characteristic shape of each variety.
- *Caliber*: Are only accepted those fruits that are within the diameter and weight set by the caliber.
- *Maturity*: Integrates parameters such as coverage color, background color, flesh firmness, starch, and acidity.

Table 4.3 Possible physiological disorders observed during the post-harvest

Situation	Disorder	Description
Unripe fruit	Scalding	Brown spots on the surface skin of the fruit
	Bitter pit	Small depressions on the skin of the fruit, bright green
Overripe fruit	Cracking pedicle	Excessive cracking in the stalk of the fruit, i.e., fruit is in the process of decay
	Yellowing	Yellowing of the skin of the fruit surface

Source: Gil (2001)

For each of the categories of apples (Gala, Red, Fuji, or Granny Smith), required maturation parameters have been established for the beginning of the harvest as well as the limits to do it. These ratios are shown in Table 4.4 (ASOEX 2009).

The most relevant parameters among those shown in Table 4.4 are pressure and soluble solids. The pressure indicates the fruit firmness whereas soluble solids, the sweetness of the fruit. These parameters are the most controlled by orchard managers to begin harvesting apples. Thus, to start harvesting Red apples, the pressure should be between 15 lb and 18 lb, while soluble solids should be higher than 11 °Bx (see Table 4.4).

Figure 4.1 shows the maturation behavior of the Gala category, taking into account the parameters of pressure or firmness (Lb) and soluble solids (°Bx). An agribusiness company in the Maule Region provided data over the last five harvest seasons. The days in Fig. 4.1 correspond to the days after the first flowering (DAFF).

As it can be seen in Fig. 4.1, there is an estimated period of 15 harvest days for the category Gala. This period was estimated from the relationship between firmness (Lb), which maximum limit is 20 Lb and minimum 17 Lb, and fruit soluble solids (°Bx), with a minimum limit of 11 °Bx. As shown in the highlighted period in Fig. 4.1, there are about 15 harvest days for the Gala category (131–146 DAFF). However, there are about 5 days where maturity parameters remain stable and ideal values are fulfilled for harvest (133–138 DAFF). After this period, the parameters decline, even though the fruit may still to be harvested. This implies that the fruit harvested before or after the ideal time may not be exported.

The parameters shown in Fig. 4.1 present a similar behavior for the Red, Fuji, and Granny Smith apples. Firmness and soluble solids remain balanced during the same span of days for all categories (about 5 days). But, the beginning of the harvest of these categories may vary if conditions changed in the zone, altering the evolution of these parameters.

As described above, there are time windows in which the most representative parameters by category remain relatively constant. Harvesting can be done during this time window in order to obtain a higher export potential of the apple. Table 4.5 presents the estimated days to harvest each category in Maule Region.

The dates presented in Table 4.5 can change regarding their days in the calendar, but not their duration.

Table 4.4 Maturation parameters for the beginning and end of harvest by category of apples

	Harvest requirements					Harvest limit				Shipping limit
Category	Background color	Acidity (Gr/Lt)	Starch	Soluble solid (°Bx)	Pressure (Lb)	Background color	Pressure (Lb)	Iodine test	Watery heart	Pressure (Lb/Pulg2)
Gala	F2–F3	3.0–4.0	2.0–3.0	>11	17–20	F4	16	4	–	15
Red	–	2.0–4.0	1.5–2.5	>11	15–18	–	14	3.5	30 % Mod	13
Fuji	F2–F3	3.5–4.5	2.0–3.0	>13.5	16–18	F4	14	4	50 % Mod	13
Granny Smith	Green	5.5–7.5	2.0–3.0	>10.5	15–17	Green	14	4	–	13

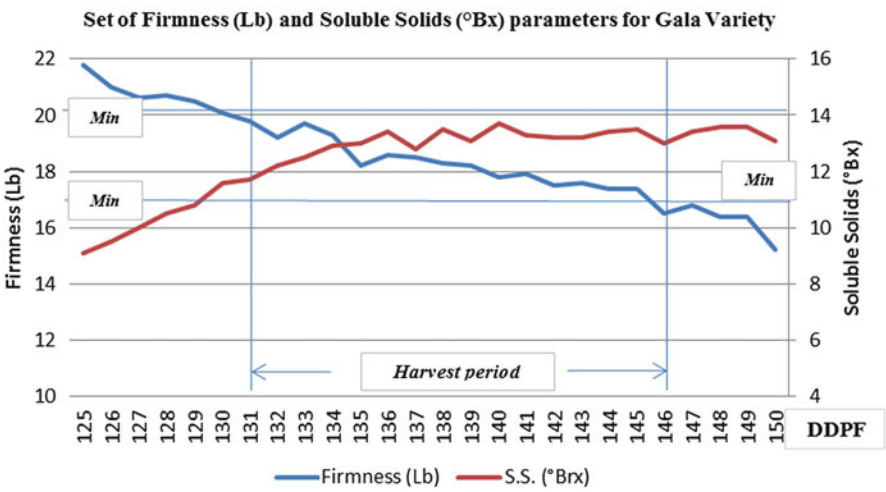


Fig. 4.1 Firmness (Lb) and soluble solids (°Bx) in category Gala (DAFF). *Source*: Agribusiness Maule Region—Orchard A

Table 4.5 Harvest periods by category of apple

	February						***March***						***April***					
Group	5	10	15	20	25	30	5	10	15	20	25	30	5	10	15	20	25	30
Gala																		
Red																		
Granny Smith																		
Fuji																		
Harvest Period																		

Since each orchard counts with different categories of apples, each category reaches the desired harvest parameters at different times.

As mentioned above, for the mathematical model the harvest window for each field of the orchards was estimated. For this estimation, historical data and literature information regarding fruit harvest was considered (Gil 2001; Razeto 1999; Yuri and Moggia 2007; Center for Pome 2009 and 2010).

Besides estimating the duration of the harvest time window for each field, the planning model must include, for each day of the established harvest window, the loss associated with lack of quality. Since the main business of an orchard is to maximize the fruit sale to export, apple harvesting must take place when all the maturation parameters required by the consumer markets meet. Indeed, a regular fruit that does not reach the parameters required for harvest will not be selected for export. Therefore, to include the relationship between quality loss and harvest date, the objective function of the model counts with a cost parameter that reflects the apples lost for not being harvested during the harvest window. In this sense, the model seeks to harvest the greatest possible quantity of kilograms within the time window during which the parameters reach the desired values. For example, for the variety Gala, the model will try to harvest as much fruit within its ideal time window. It here corresponds to the 5 days when the firmness and soluble solids parameters are stable, i.e., days 133–138 DAFF (see Fig. 4.1).

Another important aspect the model considers is that, during the harvest of an apple orchard, the yield of each variety and the harvest productivity of workers vary. At the beginning of each season, orchard managers estimate the amount of kilograms to be harvested for each variety of the orchard and use historical behavior data of the orchard and/or experience of workers.

On the other hand, in the orchards, the productivity of the workers is estimated for each field, according to the variety to be harvested. This estimation, just like the performance of the varieties, is obtained through historical data from previous crops and the agricultural manager's experience. For this estimate, the manager bases on the planting configuration (contours, distances of the trees in the row and between rows), age of the fields, and variety of planted trees.

Regarding the calculation of the productivity of each worker in each of the fields, the average productivity in each field harvested is estimated, distinguishing productivity between permanent and non-permanent workers. It is calculated according to the records of the orchard, by the kilograms yield harvested per day and per field. Permanent workers are known and work for the entire harvest season. There will be different productivities in an orchard, since there are different varieties planted, and since the conditions of each field are different.

4.4.1 Relevant Factors on Post-harvest

During storage, the condition of the fruit can be at most maintained, but never improved, according to the administrator in charge of an agro industry of the Maule Region. However, several actions can be taken from the reception at the processing plants to the shipping of the packed fruit that prevent the deterioration of the condition (Ferreira 2006).

When fruits are needed for rapid consumption, the most wanted apples are the ones with the following characteristics: some progress in terms of softening, increased soluble solids, and reduced titratable acidity. However, when a long shelf life is required, either to be stored in Controlled Atmosphere (CA) or Conventional Cold (CC), a higher content of starch and acid, and a lower concentration of soluble solids is preferred (Gil 2001). These rules are only a guide for decision makers to start harvesting in a particular orchard, considering the next destination of the fruit to be harvested.

4.5 Model for Harvest Planning in Apple Orchards

The formulation of the proposed model for planning a harvest season of apple appears as follows. Indexes, parameters, and decision variables of the model developed are presented in the Annex.

4.5.1 Mathematical Formulation

$$\begin{aligned}&\sum_{t\in T}\sum_{c\in C_1} Q\times HMQ_{ct} + (H_1+F_1)THF_1 + \sum_{t\in T} F_2\times TFV_t + \sum_{t\in T} H_2\times THV_t\\ &+\sum_{t\in T}\sum_{c\in C} J_{c1}\times TFC_{ct} + \sum_{t\in T}\sum_{c\in C} J_{c2}\times TVC_{cjt} + \lambda\sum_{c\in C}\sum_{t\in T}\sum_{k\in K}\sum_{p\in P} A_{ct}X_{ctkp}\end{aligned} \tag{4.1}$$

s.a.

$$\sum_{c\in C} X_{ctkp} \le G_{kpt}, \quad \forall k\in K, p\in P, t\in T. \tag{4.2}$$

$$\sum_{t\in T} X_{ctkp} = D_{ckp}, \quad \forall k\in K, p\in P, c\in C. \tag{4.3}$$

$$\sum X_{ctkp} \le N_{ctk}, \quad \forall c\in C, t\in T, k\in K. \tag{4.4}$$

$$\sum_{t\in T} Y_{ctp} \le MV_c, \quad \forall c\in C, p\in P. \tag{4.5}$$

$$\sum_{t\in T} Y_{ctp} \ge 1, \quad \forall c\in C, p\in P. \tag{4.6}$$

$$X_{ctkp} \ge L_k\times Y_{ctp}, \quad \forall c\in C, t\in T, k\in K, p\in P \wedge D_{ckp}\neq 0. \tag{4.7}$$

$$X_{ctkp} \le M\times Y_{ctp}, \quad \forall c\in C, t\in T, k\in K, p\in P. \tag{4.8}$$

$$\sum_{p\in P} X_{ct1p} \le P_c\times HMQ_{ct}, \quad \forall c\in C_1, t\in T. \tag{4.9}$$

$$\sum_{p\in P} X_{ct2p} \le \mathrm{R}_c\times TFC_{ct} + \mathrm{S}_c\times TVC_{ct}, \quad \forall c\in C_2, t\in T. \tag{4.10}$$

$$X_{ctkp} \le E\times NB_{ctp}, \quad \forall c\in C, t\in T, k\in K, p\in P \tag{4.11}$$

$$\sum_{c\in C_2} HMQ_{ct} \le I_t, \quad t\in T. \tag{4.12}$$

$$\sum_{c \in C_2} TFC_{ct} \leq THF_t, \quad \forall t \in T. \tag{4.13}$$

$$\sum_{c \in C_2} TFC_{ct} \geq N, \quad \forall t \in T. \tag{4.14}$$

$$\sum_{c \in C_2} TFC_{ct} \leq W, \quad \forall t \in T. \tag{4.15}$$

$$THF_t = THF_{t+1}, \quad t = 1, \ldots, |T| - 1. \tag{4.16}$$

$$\sum_{c \in C_2} TVC_{ct} = THV_t, \quad t = 1. \tag{4.17}$$

$$\sum_{c \in C} TVC_{ct} = \sum_{c \in C} TVC_{ct-1} + THV_t - TFV_t, \quad \forall t \in T : t \geq 2. \tag{4.18}$$

$$YV_{ctp} \in \{0, 1\} \quad \forall t, c \tag{4.19}$$

$$X_{ctkp}, Hmaq_{ct} \geq 0 \quad \forall t, p, k, c \tag{4.20}$$

$$THF_t, THV_t, TFV_t, tvc_t, tfc_{ct}, Num\_bins_{ctp} \geq 0 \quad \forall t, p, k, c \tag{4.21}$$

The objective function (4.1) seeks to minimize the costs related to the workforce, goods, and fruit loss due to poor quality. Thus, the objective function can be articulated around the following terms:

Minimizing: machinery costs + permanent workforce turn over costs + the seasonal workforce turn over costs + salary cost + cost for loss of fruit quality.

The cost for loss of fruit quality considers the harvested apple during each harvest season and the percentage of fruit quality loss. Indeed, during the harvest period, in each of the fields, loss of apple quality occurs for not counting with the maturity parameters needed for export. This loss of quality is transformed into monetary units, using the parameter λ (\$/kg). λ corresponds to the amount of money lost by the producer when a kilogram of apple is classified as commercial fruit instead of for export, reducing their income.

Constraint (4.2) presents the amount of harvested fruits in each period and is restricted by the capacity of the processing plant. Constraint (4.3) establishes that the estimated amount of fruit each plant should receive from a certain field and that has been obtained by a specific mode of harvest must be fulfilled. Constraint (4.4) shows that the amount of apples harvested during a season in an orchard, using a harvesting mode, must not exceed the maximum amount that can be harvested for a particular harvest mode at the same time. Constraint (4.5) shows that the harvest of an orchard should be performed during the season, within a time window predetermined by orchard managers. Constraint (4.6) shows that the harvest of a field should be undertaken during at least 1 day within the established harvest window. Constraint (4.7) shows that the amount of apples harvested in each field and for each period, must be greater or equal to the minimum amount of estimated kilograms so the harvest could be profitable in each field. Constraint (4.8) shows that the harvest must be completed at some time within the harvest window of each field, provided that there is fruit to be harvested. In this case, the maximum amount of fruit harvested at a time was used as the value for M (very large number). Constraint (4.9) shows that the amount of apples harvested by machinery in an orchard, in each period, is limited by the productivity of the machine in the field. Constraint (4.10) shows that the amount of apples harvested by hand in an orchard, in each period, is limited by the productivity of each laborer type (permanent or non-permanent) for that field. Constraint (4.11) shows that the amount of kilograms of apples harvested in an orchard, according to the harvest mode and destination facility, depends on the number of bins available for the harvest season. Constraint (4.12) shows that the number of machine-hours available for each period must be respected. Constraint (4.13) shows that the amount of permanent workforce used for a period, considering all the orchards to harvest, must be less or equal to the amount of permanent labor hired for that period. Constraint (4.14) shows that the permanent workforce used during the whole harvest season must be greater or equal to the amount of workers hired by the orchard to do the job. Constraint (4.15) shows that non-permanent workers used in each field and in each period must be less or equal to a maximum value of seasonal workers available in the season period. This value is fixed and predetermined according to historical harvesting data. Constraint (4.16) shows that the amount of permanent workforce estimated for a planning period must be equal in all periods within the harvest season. Constraint (4.17) shows that in the first period, the non-permanent workforce used in all the orchards must be equal to the non-permanent workforce hired in the period. Constraint (4.18) shows the balance of non-permanent workforce. Constraint (4.19) establishes binary decision variables. Constraint (4.20) corresponds to the non-negativity of continuous decision variables. The constraint (4.21) defines integer decision variables.

4.5.2 Computational Experiment

This section shows the impact of the most important parameters in relation to the model behavior. These parameters are: PMC_{ct}, parameter associated with the extension of the harvest period in each field and associated fruit loss mediated by poor quality, and the parameter λ, which is associated with the decrease of the producers' income for switching from an export fruit to a domestic consumption one. Computational times are analyzed, looking for the solution and the type of solution to be provided.

For parameter PMC_{ct}, scenarios presented in Table 4.6 are considered.

Along with the variation of the PMC_{ct} parameter, different values of λ are taken into account, which are $\lambda = 20$, 45, 89, 138, 200, and 500, respectively.

The optimization software used for this computational experiment is ILOG-OPL, version 6.1, with CPLEX-11. The model was solved using a computer with an AMD Turion (tm) 64×2 Mobile Technology TL-60 2.00 GHz, 2 GB RAM, and 512 GB hard drive.

Table 4.6 Scenarios for the parameter of the harvest window (PMC_{ct})

	Extension of the harvest (days)
First scenario	1 day of harvest
Second scenario	5 days of harvest
Third scenario	10 days of harvest
Fourth scenario	15 days of harvest

Table 4.7 Computational time of the different models analyzed

		Computational time (sec)					
		Parameter λ					
Orchard	Harvest window (days)	10	45	89	138	200	500
Orchard A	1	1,986	3,689	2,246	1,876	1,987	2,365
	5	2,435	2,034	1,364	3,543	2,321	2,675
	10	3,546	2,311	1,897	2,432	2,143	1,987
	15	3,429	2,543	1,698	2,675	1,765	2,341
Orchard B	1	2,435	1,956	707	1,987	2,134	239
	5	1,975	2,176	1,230	1,890	2,508	2,134
	10	1,690	1,903	1,898	2,546	1,742	1,865
	15	2,314	2,978	2,235	1,236	1,342	2,124
Orchard C	1	2,480	5	17	3	2	1
	5	2,897	530	2,148	1,890	2,100	3
	10	3,140	2,476	14	13	2,800	28
	15	2,190	480	2,390	13	2,456	2,567

The data considered in this study correspond to three orchard companies: Orchard A, Orchard B, and Orchard C. All problems model a full season, consisting of planning 64 days; the difference being the amount of fields harvested during the 2009/2010 season in each orchard.

Two important aspects to consider in the computational analysis are related to the quality of the solution and the time spent to obtain it. The quality of the solution will be measured with the GAP, which for integer programming problems is defined as the percentage difference between the integer solution found (P) and the relaxed solution of the whole problem (R_P) (Mathioudakis 2007). The equation used for this calculation is:

$$\mathrm{GAP} = \left(1 - \frac{P}{R_P}\right) \times 100$$

The computational time associated with the search for a solution to the various scenarios is measured in seconds.

As shown in Table 4.7, there is an upward trend in computational time associated with finding solutions when a greater harvest window is available in the field. By increasing the value of λ, the computational time associated with finding the solution tends to decrease.

Table 4.8 GAP obtained for the different models analyzed

Orchard	Harvest window (days)	GAP (%)					
		Parameter λ					
		10	45	89	138	200	500
Orchard A	1	0.21	0.00	0.00	0.00	0.00	0.00
	5	0.19	0.18	0.00	0.07	0.10	0.07
	10	0.36	0.28	0.03	0.21	0.25	0.09
	15	0.43	0.47	0.27	0.33	0.24	0.13
Orchard B	1	0.08	0.03	0.00	0.02	0.02	0.00
	5	0.11	0.06	0.00	0.04	0.04	0.04
	10	0.15	0.08	0.09	0.07	0.09	0.04
	15	0.22	0.16	0.21	0.34	0.14	0.08
Orchard C	1	0.09	0.00	0.00	0.00	0.00	0.00
	5	0.05	0.00	0.00	0.00	0.00	0.00
	10	0.04	0.06	0.00	0.00	0.04	0.00
	15	0.04	0.00	0.06	0.00	0.03	0.02

Table 4.8 presents the GAP associated with a search for solutions for each scenario.

Even though the solutions are very close to optimal ones, the model finds solutions with lower GAP when the harvest window decreases and the value of λ increases.

Regarding the value of the objective function, the results are presented in Table 4.9.

As seen in Table 4.9, the value of the objective function decreases as the harvest window increases. This behavior is mainly due to two reasons: first, there will be no apple loss for poor quality; so all apples will be for export. The second reason deals with the workforce. Indeed, by having fewer days to harvest, the workers should harvest more fruit, and therefore the hiring and wages costs will increase. But, with more harvest days, the permanent workforce is enough to do the job.

Table 4.9 also shows that the objective function value increases when the value of λ also increases. Let's not forget that λ represents the loss of income for having a fruit that does not have suitable maturity indices for export, mainly because not harvested at appropriate dates. Therefore, if there is a higher decrease of the incomes, the model seeks to harvest as much fruit in time windows with desirable maturation indices for export. This income decline, or penalty for the fruit quality, is reflected on the objective function, so if this penalty increases, the objective function will too.

4.6 Case Studies

The model presented in Sect. 4.5 was applied to three case studies corresponding to three fruit orchards from the Maule Region. These orchards have different characteristics regarding the number of fields, apple varieties; hectares planted, and

Table 4.9 Value of the Objective Function for different models analyzed

		Objective Function ($ Chilean Pesos)					
		Parameter λ					
Orchard	Harvest window (days)	10	45	89	138	200	500
Orchard A	1	9,550,912	15,808,656	22,907,295	30,632,530	40,336,488	87,060,911
	5	8,111,359	10,092,701	12,175,881	14,459,516	17,343,893	31,237,380
	10	7,785,404	9,506,804	11,545,152	13,813,898	16,690,262	30,575,595
	15	7,304,791	8,928,292	10,958,872	13,237,336	16,103,447	29,994,687
Orchard B	1	34,843,016	52,975,954	74,890,646	99,094,274	129,585,310	276,774,304
	5	31,377,075	36,672,949	42,407,667	48,638,067	56,552,520	94,580,618
	10	30,657,174	35,093,259	40,273,739	46,039,809	53,350,964	88,645,742
	15	30,446,915	33,947,067	39,893,592	46,406,238	54,854,905	87,509,144
Orchard C	1	13,148,806	28,734,185	47,523,273	67,243,355	91,971,172	210,999,986
	5	9,804,846	13,312,194	17,354,998	21,712,374	27,177,428	53,243,900
	10	9,668,894	12,722,521	16,543,208	20,798,064	26,181,760	52,231,900
	15	9,169,338	12,208,521	16,029,208	20,284,064	25,667,760	51,717,900

Table 4.10 Characteristics of orchards taken as case studies

Orchard	Planted hectares	No. of fields	Varieties	Crop estimation season 2009/2010 (kg)
Orchard A	36.4	13	Gala-Red-Granny-Fuji	844,300
Orchard B	39.3	12	Gala-Red-Granny-Fuji	2,186,850
Orchard C	16.9	5	Gala-Red-Fuji	1,489,084

estimated kilograms of apple harvested during the season. Table 4.10 shows the values of the characteristics of the orchard.

The considered planning horizon represents 64 days, time during which all the apple varieties present in an orchard are harvested. This schedule runs approximately from mid-February to mid-April.

The models were run in the same software and hardware as computational experiment. Associated times when applying the model in the case studies for the Orchards A, B, and C are respectively of 1364 s, 1,230 s, and 2,148 s. These times are reasonable considering that labor, supplies, and machinery for the entire harvest season are being planned.

The next subsection describes in detail the results obtained by the model for Orchard C harvest planning. It provides information that helps decision makers.

4.6.1 Analysis Results for Orchard C

Harvest weekly schedule for each of the five fields of Orchard C is presented in Table 4.11 below.

As seen in Table 4.11, weeks 4, 5, and 6 are the ones with the strongest harvesting activity, mainly because this is when the Red variety is harvested, with a total of one million kilograms, which is the greatest number of scheduled kilograms. The amount of kilograms of apple harvested in each field will depend on the variety in each of them, which will affect the type of harvest to be held in the season. For example, Gala apples are selectively picked (about three harvests per season). Whereas, the Red variety is strip picked, i.e., all fruits are harvested on a tree.

The method used for harvesting has implications on the worker productivity in each field. Thus, productivity is given by the apple variety in the orchard, the type of harvest to be performed, and the type of worker currently in the field. Permanent workers have a greater commitment to the job; so will count with a greater productivity than the seasonal or non-permanent worker. The orchard counts with eight permanent workers, who will be there for the 9 weeks harvest. As shown in Table 4.11, week 7 has no scheduled harvest, but these workers must be paid, just for having them for future tasks.

Table 4.11 Weekly harvest schedule

Apples harvested (kg)	Fields × Variety					
	Gala		Red		Fuji	
Weeks	3	5	1	4	2	Total × Week
1	6,490	21,092	0	0	0	27,583
2	64,548	46,477	0	0	0	111,025
3	32,479	67,003	0	0	0	99,482
4	0	0	174,941	386,497	0	561,438
5	0	0	256,052	252,000	0	508,052
6	0	0	21,600	6,794	0	28,394
7	0	0	0	0	0	0
8	0	0	0	0	93,711	93,711
9	0	0	0	0	59,400	59,400
Total × Variety	103,517	134,573	452,593	645,291	153,110	

The productivity of each of the varieties and fields are parameters included in the proposed model. Each orchard manager delivered them. It is important to note that the program presented in Table 4.11 is a summary of a 64-day harvest planning, thus kilograms harvested within 1 week may be obtained in one or more days within it.

Scheduling of non-permanent workers for harvesting tasks is presented in Table 4.12.

As Table 4.12 shows, the largest number of non-permanent workers is needed to harvest the Red variety (fields 1 and 4), which provides about 80 % of the harvested fruit. Most of this harvest should be performed during weeks 4 and 5, which respectively represents 290 workers and 262 days to do the job, including the days of permanent and non-permanent workers.

It is noteworthy that the requirements of the permanent and non-permanent workers in each of the fields, depend on several relevant parameters in the model, such as the amount of kilograms to harvest, harvest type, among others, will depend on their productivity in orchard, particularly in each variety and associated fields. That is why for example, during week 1, when the Gala variety is harvested, the productivity is very low for both permanent and non-permanent workers; thus, hiring non-permanent workers to avoid damaging the quality of the fruit is necessary. By contrast, during week 9, the Fuji variety is harvested. The productivity associated with this variety and associated fields is greater than the productivity of the fields of Gala variety, both permanent and non-permanent staff, mainly due to the characteristics that this field has, such as configuration of the trees, and irrigation type. Therefore, hiring non-permanent workers will not be necessary, and the job will only be performed with permanent workers present in the orchard.

The fruit harvested from different fields can have two destinations: export or domestic market. This is why the export fruit goes to plant 1 and commercial fruit to plant 2. The amount of bins bound to each of the plants during the harvest season is presented below in Table 4.13.

Table 4.12 Planning of permanent and seasonal workers during the harvest season

Man/Days	Data from fields by variety										
	Variety Gala				Variety Red				Variety Fuji		
	Field 3		Field 5		Field 1		Field 4		Field 2		
Weeks	Work days Seas.	Work days Perm	Work days Seas.	Work days Perm	Work days Seas.	Work days Perm	Work days Seas.	Work days Perm	Work days Seas.	Work days Perm	Total working days per week
1	5	35	4	21	0	0	0	0	0	0	65
2	14	39	38	17	0	0	0	0	0	0	108
3	2	25	14	31	0	0	0	0	0	0	72
4	0	8	0	0	87	1	147	47	0	0	290
5	0	0	0	0	122	9	84	47	0	0	262
6	0	40	0	0	0	12	0	4	0	0	56
7	0	16	0	0	0	0	0	0	0	40	56
8	0	0	0	0	0	0	0	0	17	56	73
9	0	8	0	0	0	0	0	0	0	56	64
Total working days by variety	21	171	56	69	209	22	231	98	17	152	1,046

Table 4.13 Bins planning for processing plant 1 (fruit for export) and plant 2 (fruit for the domestic market)

Bins × Plant	Fields × Variety										
	Gala				Red				Fuji		
	3		5		1		4		2		
Weeks	Plant 1	Plant 2	Plant 1	Plant 2	Plant 1	Plant 2	Plant 1	Plant 2	Plant 1	Plant 2	Total × Week
1	19	0	61	0	0	0	0	0	0	0	80
2	189	0	137	0	0	0	0	0	0	0	326
3	5	91	77	118	0	0	0	0	0	0	291
4	0	0	0	0	502	0	1,107	0	0	0	1,609
5	0	0	0	0	601	133	465	258	0	0	1,457
6	0	0	0	0	0	63	0	20	0	0	83
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	273	0	273
9	0	0	0	0	0	0	0	0	140	33	173
Total × variety	213	91	275	118	1,103	196	1,572	278	413	33	4,292

Table 4.14 Summary of the distribution and organization of kilograms to be harvested, workers, and bins during the 2009/2010 season in the Orchard C

		Period planning crop (Days)
Total crop (kilograms)		1,489,048
Permanent labor (worker)		512
Non-permanent labor (worker)		534
Total labor (worker)		1,046
Bins (Unit)	Plant 1	3,578
	Plant 2	716

As seen in Table 4.13, the greatest amount of export fruit is harvested between weeks 4 and 5 (plant 1), where 1,609 and 1,066 bins are, respectively, needed for the harvest. Throughout the harvest season, which lasts 9 weeks, a total of 3,576 bins are required to transport the export fruit.

Table 4.13 also shows the required bins per week for each season for commercial fruit. Most of the bins are needed during week 3 and 5, respectively, with 209 and 391 bins. During week 3, the Gala variety will be harvested, both in fields 3 and 5; and during week 5, the Red variety will be harvested, both in fields 1 and 4. A total of 716 bins are required to transport commercial fruit.

Table 4.14 is a summary of the results that can be generated to help the organization and distribution of labor, bins, and kilograms of fruit to be harvested by season for an orchard, particularly for Orchard C. This information indicates that for the season under study, the Orchard employs on average 20 workers during the harvest days (of which 8 are permanent), with an average extraction of 23,267 kg per workday. Moreover, 1,046 man-days are needed throughout the period of harvest planning, considering both permanent and non-permanent workers. This harvest lasts 42 working days, and to collect orchard apples, a total of 4,294 bins are required, of which 3,578 are for export apples and 716 for commercial ones.

The information provided by the model is presented for each period and/or workday modeled for each orchard. For space reasons, it only shows the totals for each of the relevant items in the analysis and planning of the harvest.

4.6.2 Analysis of the Results Obtained for the Three Orchards

Tables 4.15, 4.16, and 4.17 show the results obtained by applying the proposed model to the three case studies and comparing them with the observed data of the modeled season, 2009/2010 season.

The main cost savings shown in Table 4.15 correspond to a better distribution of workers in the fields, for each apple harvest period, with an estimated average costs saving of 17 %. This is due to the harvest planning model that delivers the optimal

Table 4.15 Summary of the results using labor (Permanent and Non-permanent)

Orchard	Total cost observed for season 2009/2010 (US$) (a) (*)	Optimal total cost obtained through the model (US$) (b) (*)	% Difference (100(a−b)/a)
Orchard A	21,007	16,700	19 %
Orchard B	74,984	64,394	14 %
Orchard C	23,849	19,636	18 %

(*) Observed Dollar: $480.39 pesos (Friday, November 30th, 2012)

Table 4.16 Summary results of man-days

Orchard	Days/man for the season 2009/2010 (a)	Optimal days/man obtained through the model (b)	% Difference (100(a−b)/a)
Orchard A	1,214	976	20 %
Orchard B	2,916	2,651	9 %
Orchard C	1,516	1,049	31 %

Table 4.17 Summary results of quality loss

Orchard	Cost of the loss for the season 2009/2010 (US$) (a) (*)	Cost of the loss obtained through model US$ (b) (*)	% Difference (100(a−b)/a)
Orchard A	9,799	8,646	12 %
Orchard B	28,751	23,883	17 %
Orchard C	18,071	16,491	9 %

(*) Observed Dollar: $480.39 pesos (Friday, November 30th, 2012)

amount of workers for each variety, allowing better organization and distribution of the seasonal workers in the orchard throughout the harvest period analyzed. This planning allows deciding, which needs more workers for harvesting.

Regarding the man-days to be used in each of the orchards (days per workers needed for the entire season), Table 4.16 shows the results of the proposed model, which are compared with the man-days used in the season.

As seen in Table 4.16, there is a reduction of man-days required for harvesting Orchards A, B, and C of 20 %, 9 %, and 31 %, respectively, giving an average reduction of man-days of 20 %. This decrease is due to the fact that planning, distribution, and assignment of workers for the harvest was carried out much more rationally. Since the model solution reduces the number of workers to be used in the harvest season, the number of man-days also decreases.

Regarding the loss of apples for poor quality, Table 4.17 shows that the planning model can help producers improve their incomes since a greater amount of apples is harvested in the time windows, where they reach the desired maturity parameters.

The reduction in costs for loss of apples from the Orchard A, Orchard B, and Orchard C represents, respectively 12 %, 17 %, and 9 %, with an estimated average of 13 % cost reduction for losses to the three orchards.

Considering these improvements in reducing labor forces and increasing quality, reduced overall costs of the Orchard A, B, and C are 18 %, 15 %, and 14 %, respectively, with an estimated average decrease of 16 % in estimated costs for the three Orchards.

The main change in the planning of the orchards, pulling down costs for quality loss, stands in a greater amount of fruit harvested when their maturity parameters are appropriate. This is only possible by effective estimation of the number of required workers.

4.7 Conclusions

The proposed model can work as a support for tactic decisions during the harvest season by organizing and distributing labor and required supplies. As one of the objectives of this model is to provide a plan for the required labor per harvesting tasks within an orchard and in each of the fields, the results of this model work as a proposal when managers need to take decisions regarding staff hiring for the fruit harvest at a given period of time within the season.

Regarding the amount of fruit to be harvested in each of the periods of time, the model provides the quantity of apples to harvest, the day within the time window, and the field. This information helps the decision maker organize the staff, so the estimated amount in each field can be harvested, and prevent damages on the apple quality.

According to the results obtained for the case studies, the solution of the model allows a better distribution of workers in the fields for each apple harvest period, reducing the average labor cost by 17 %. This result is possible because the model allows decreasing the man-days by an average of 20 % in order to perform the harvest. Moreover, the planning provided by the model allows a larger amount of fruit harvested in the time windows, reaching desired maturity parameters. This implies that costs associated to lower incomes are reduced on average to 13 %. All these improvements reduce the average total costs in the orchards of 16 %.

The model can be applied to different orchards that produce pome or stone fruit. Thanks to it, they will be able to choose better tools for decision support on operational planning during the harvest. Even though this is presented as a general model, it provides a basis for a more efficient organization of the resources used in the harvest, since considering planning requirements minimizes costs.

This model allows decision makers (orchard managers in charge of the harvest) plan the whole season, including required goods and workforce. The model includes the experience of orchard managers in charge of the harvest for parameters such as harvest estimation, harvest type, destination of the fruit in each field, and quality requirements for fruit export (quality parameters and ideal dates of harvest).

It is important to note that the model considers two methods of apple harvest: strip picking (all apples harvested in a field) and selective picking (one or more harvests to collect fruit from the fields), which implies different resource planning and labor for their activities. For the harvesting type of selective pickings, the decision maker can decide how many times a field should be harvested in order to collect all the fruit. Based on the model results, the manager can check what type of harvesting is more convenient according to the planning of labor and harvest schedule for the work. This means an improvement of the planning process, since decisions are so far taken based on the experience of those in charge of the orchards to prepare the harvest workday.

In future research, it is recommended to improve the crop estimation method currently used in orchards. This research is important to obtain reliable estimates of harvest, since the modeling results are very sensitive to this parameter. Besides, the optimization model may be extended to modeling the transportation of the bins from the orchard to different process plants, in order to optimize the timing of transfer of the raw material and harvest fields within corresponding time windows.

4.8 Annex

Sets and parameters used in the formulation.

The index sets that are considered in the model are:

K: Set of harvesting modes, $K = \{1$: mechanical harvesting, 2: manual harvesting$\}$.
C_k: Set of fields in an orchard using harvesting mode k, $k \in K$.
C: Corresponds to the total set of fields feasible to be harvested within the orchard, $C = C_1 \cup C_2$.
O: Set of types of existing labor, $O = \{1$: permanent labor, 2: non-permanent labor$\}$.
P: Set of types of fruit processing plants, $P = \{1$: Export, 2: Commercial$\}$.
T: Planning Horizon for the harvest period of the orchard.
A_{ct}: Percentage of apple loss due to poor quality of the fruit in the field c, $c \in C$, in period t, $t \in T$.
H_l: Cost of hiring a labor unit l, $l \in O$.
F_l: Cost of dismissing a labor unit l, $l \in O$.
J_{cl}: Cost of labor l, $l \in O$, for the working day in the field c, $c \in C_2$.
Q: Cost per hour of work for a machine for harvest.
P_c: Productivity for harvesting mechanically the fields c, $c \in C_1$, expressed in kilograms per hour.
R_c: Productivity for harvesting by hand the field c, $c \in C_2$, with permanent labor, expressed in kg/man.
S_c: Productivity for harvesting by hand the field c, $c \in C_2$, with non-permanent labor, expressed in kg/man.
MV_c: Number of days when it is possible to harvest apple in the field c, $c \in C$.
D_{ckp}: Number of apples to be harvested in the field c, $c \in C$, harvested with the mode k, $k \in K$, bound to the processing plant p, $p \in P$, expressed in kilograms.

G_{kpt}: Capacity of the plant p, $p \in P$ for apples harvested with mode k, $k \in K$, in period t, $t \in T$, expressed in kilograms.

N_{ctk}: Maximum amount of apples that can be harvested in the field c, $c \in C$, in period t, $t \in T$, using the harvesting mode k, $k \in K$, expressed in kilograms.

L_k: Minimum amount of apples that can be harvested in a fields using crop mode k, $k \in K$, expressed in kilograms.

I_t: Availability of machines in period t, $t \in T$, expressed in hours.

M: A very large positive scalar.

λ: Parameter to work as a penalty for harvesting fruit without adequate ripening conditions. Furthermore, it is used to transform the fruit loss from kilograms to pesos or monetary units, in order to standardize the units in the objective function.

N: Minimum number of permanent staff required for harvest. This is the number of seasonal workers, which participation should be assured and who are hired prior to the beginning of the job.

W: Maximum number of non-permanent staff allowed in crop planning. This value is constant throughout the season.

E: average load capacity of a bin (in kilograms).

Decision variables.

X_{ctkp}: Apple quantity (in kilograms) harvested in the field c, $c \in C$, in period t, $t \in T$, with harvesting mode k, $k \in K$, and bound to the processing plant p, $p \in P$.

$Y_{ctp} = 1$, if harvested in the field c, $c \in C$, in period t, $t \in T$, bound to the processing plant p, $p \in P$; $Y_{ctp} = 0$, otherwise.

THF_t: Number of workers hired in period t, $t \in T$, that will remain throughout the harvest period.

THV_t: Number of non-permanent workers hired at the beginning of period t, $t \in T$.

TFV_t: Number of non-permanent employees dismissed when the period t, $t \in T$.

TFC_{ct}: Number of permanent employees per field c, $c \in C_2$, in period t, $t \in T$.

TVC_{ct}: Number of non-permanent workers per field c, $c \in C_2$, in period t, $t \in T$.

HMQ_{ct}: Number of machine-hours necessary for harvesting field c, $c \in C_1$, in period t, $t \in T$.

NB_{ctp}: Number of bins required in the field c, $c \in C$, in period t, $t \in T$, which will be bound to the processing plant p, $p \in P$.

References

Ahumada O, Villalobos R (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 196(1):1–20

Ahumada O, Villalobos R (2011a) Operational model for planning the harvest and distribution of perishable agricultural products. Int J Prod Econ 133:677–687

Ahumada O, Villalobos R (2011b) A tactical model for planning the production and distribution of fresh produce. Ann Oper Res 190(1):339–358

Ahumada O, Villalobos R, Mason N (2012) Tactical planning of the production and distribution of fresh agricultural products under uncertainty. Agr Syst 112:17–26

Alarcón R (2008) Evolución del empleo en Chile: Principales resultados de la encuesta CASEN. Chilean Government, Santiago

ASOEX—Asociación de Exportadores de Frutas de Chile (2009) Manual del Exportador. ASOEX, Curicó

Bjorndal T, Herrero I, Newman A, Romero C, Weintraub A (2012) Operations research in the natural resource industry. Int Trans Oper Res 19(1–2):39–62

Bravo MJ (2011) Mercado de la Manzana. ODEPA—Office of Studies and Agricultural Policies, Chile

Caixeta-Filho J (2006) Orange harvesting scheduling management: a case study. J Oper Res Soc 57(6):637–642

Català LP, Durand GA, Blanco AM, Bandoni JA (2013) Mathematical model for strategic planning optimization in the Pome fruit industry. Agr Syst 115:63–71

Center for Pome (2009) Pomáceas. Technical bulletin 9(3) Center for Pome, Universidad de Talca, Talca

Center for Pome (2010) http://pomaceas.utalca.cl/html/index.html. Accessed 3 Oct 2010

Contreras O (2008) Productividad de los Trabajadores Agrícolas. Pontifical Catholic University of Chile, Santiago

Domínguez J (2007) Recursos humanos en la agricultura. Trabajar con la mano de obra. Revista de Extensión de la Facultad de Agronomía e Ingeniería Forestal UC 21:6–10

Ferreira M (2006) Modelos de Localización para optimizar el proceso de cosecha en huerto de pomáceas. Dissertation Escuela de Ingeniería Civil Industrial, Universidad de Talca, Chile

Ferrer J, MacCawley A, Maturana S, Toloza S, Vera J (2008) An optimization approach for scheduling wine grape harvest operations. Int J Prod Econ 112(2):985–999

France J, Thornley J (1984) Mathematical models in agriculture. Butterworths, London

Gil G (2001) Madurez de la fruta y manejo de poscosecha. Editorial Pontificia Universidad Católica de Chile, Santiago

Glen J (1987) Mathematical models in farm planning: a survey. Oper Res 35(5):641–665

Hester SM, Cacho O (2003) Modelling apple orchard systems. Agr Syst 77:137–154

Higgins A, Laredo L (2006) Improving harvesting and transport planning within a sugar value chain. J Oper Res Soc 57:367–376

INE—Instituto Nacional de Estadísticas (2007) Agricultural census. Government of Chile, Santiago

Lowe TJ, Preckel PV (2004) Decision technologies for agribusiness problems: a brief review of selected literature and a call for research. Manuf Serv Oper Manage 6(3):201–208

Mathioudakis M (2007) Linear programming and combinatorial optimization. Lecture 9: semi-definite programming combinatorial optimization, Toronto

Maule Competitiveness Center (2010) Informe sobre Manzanos. Universidad de Talca, Talca

Razeto B (1999) Para entender la fruticultura. Editorial Salesianos S.A., Santiago

Schuster E, Allen S (2004) Controlling the risk for an agricultural harvest. Manuf Serv Oper Manage 6(3):225–236

Tadei R, Trubian M, Avendaño JL, Della Crose F, Menga G (1995) Aggregate planning and scheduling in the food industry: a case study. Eur J Oper Res 87:564–573

Van Berlo MJ (1993) A decision support tool for the vegetable processing industry an integrative approach of market, industry and agriculture. Agr Syst 43:91–109

Yuri J, Moggia C (2007) Avances en Investigación de Poscosecha. Red Agrícola, Ed. No 19

Zhang W, Wilhelm WE (2011) OR/MS decision support models for the specialty crops industry: a literature review. Ann Oper Res 190(1):131–148

Chapter 5
Optimization of the Supply Chain Management of Sugarcane in Cuba

Esteban López-Milán and Lluis M. Plà-Aragonés

5.1 Introduction

The international sugar market is very competitive, and from time to time, it suffers from economic crisis when the price of sugar goes down. This is a serious concern for many large producers (e.g., Brazil and India) and for countries that depend on sugar exports (e.g., Brazil, Australia, and Cuba). After tourism, sugar is the second most important generator of revenue for Cuba, although its importance is decreasing in favor of other activities. Last decade, the Cuban Ministry of Sugar Industries operated 156 sugar mills (http://www.cubagob.cu/des_eco/azucar.htm, accessed 10 Feb 2013). As a result of past sugar market crisis, nowadays there are only around 54 sugar mills operating. The price of sugar in the international market leads to constantly controlling and lowering production costs in order to increase economic efficiency and at the same time increase the chances of some profit. The most important component of production cost in the sugar industry is transportation. This is especially true for developing countries. Cane transport is the largest cost component in the manufacturing of raw sugar as different authors have already shown (Díaz and Pérez 2000; Martin et al. 2001). Australian studies estimated the transportation cost range from 25 to 30 % of the total production cost under their conditions (Higgins 1999, 2002).

The management of the sugar supply chain, and in particular the sugarcane harvest, is a complex logistical operation that involves the cutting and loading of

E. López-Milán (✉)
Mechanical Engineering Department, University of Holguin, Av. XX Aniversario, gaveta 57 CP, Holguín 80100, Cuba
e-mail: elopez@facing.uho.edu.cu

L.M. Plà-Aragonés
Department of Mathematics, University of Lleida, Jaume II, 73, Lleida 25001, Spain
e-mail: lmpla@matematica.udl.es

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_5

cane in the fields, the transportation by truck or train to the sugar mills, and the unloading of the cane in the mill to be processed (Semenzato 1995; Pavia and Morabito 2009; Jena and Poggi 2013). Although the situation may vary from country to country, all these activities may involve different companies or agents that have to operate, collaborating among each other and coordinating with the sugar mill. Hence, the system is organized as a supply chain where the sugarcane supply to the mill has to be assured (Lejars et al. 2008). This supply chain coordination usually starts with the selection of planting dates and cultivars (Scarpari et al. 2008, Scarpari and de Beauclair 2010; Le Gal et al. 2009; Piewthongngam et al. 2009).

Normally, during harvest season, mills operate 24 h a day at a constant level unless there are short programmed shutdowns for maintenance. Each sugar mill has a number of teams that cut cane manually or with several harvester machines to meet a daily quota. The daily quota is imposed naturally by the regular operation at a constant level of the mill to avoid any interruption of sugarcane processing (Jena and Poggi 2013). As people on the field cannot work at night and the processing of cane has to be made as soon as possible to avoid sugar quality deterioration, large storage facilities are cost-ineffective. Depending on the quota and available resources for a particular day, the scheduling is proposed by sugar mill managers based upon their own expertise. In view of day-to-day changes in the amount of cane in the fields, the cane ripeness, the unforeseen failures in machinery, and the performance of harvesters, managers must adapt their schedules daily. Quite often, all these are being done manually, even in developed countries, as Higgins (2006) pointed out. In this sense, recent studies indicated great opportunities to improve the value chain and reduce the cost in the operational and tactical planning to remain competitive (Higgins et al. 2007; Plà et al. 2013).

In order to cope with daily scheduling changes, there has to be a constant and ongoing analysis of the current organization, the available infrastructure, and the future needs. Hence, this chapter shows how computers and mathematical programming can be combined into a practical DSS capable of supporting the operational decisions of sugar mill managers who control the sugarcane supply chain. The DSS has been developed and tested in "Fernando de Dios" mill, a Cuban sugar mill settled in the province of Holguín. The DSS is flexible enough to be adapted to other different conditions of sugarcane harvesting.

5.2 The Sugarcane Supply Chain

The sugar industry is characterized by large sugar mills that are able to take supplies of cane from surrounding farms (Higgins 2002; Grunow et al. 2007). Coordination between producers must be done to avoid problems like the overflow of mill's capacity during the peak of harvest season as reported by Piewthongngam et al. (2009) in Thailand. Stray et al. (2012) among others emphasize the benefits of collaboration between growers and millers. In Cuba, this collaboration is mandatory. The whole sector is controlled by the government through public companies being mill managers who organize and schedule main activities along the sugarcane supply chain. At harvest season, sugarcane must be cut when it is ripe; otherwise, the sugar quality of the cane deteriorates rapidly. As the harvester cuts cane, it fills a truck or a haul-out vehicle next to the harvester. When the truck or the haul-out vehicle is full, they transport the cane to the sugar mill. Generally, sugarcane can be conveyed in two different ways:

1. Direct transportation to the swing bolster for just-in-time processing by road transport only. The truck or the haul-out vehicle completes the distance to the mill. This system is predominant in Brazil (Jena and Poggi 2013).
2. Intermodal transportation (Rizzoli et al. 2002). This means the use in a first step of trucks and tractors (road transport) and, thereafter, in a second step, the train (rail transport). In this case, the sugarcane is transported by road and stored at rail loading stations, where the cane is cut in small pieces, cleaned, and placed into rail wagon bins in a continuous process. Later, the cane is delivered by train to the sugar mill yard to be processed. This system is available in many countries like Australia and Cuba (López et al. 2004, 2006). Depending on the rail network and the train load capacity, different stations may be visited before the mill is reached. Hence, a problem of transport planning may arise (Higgins and Laredo 2006).

Railways connecting loading stations and the mill are almost only devoted to sugarcane transport in most countries. At the mill, sugarcane is unloaded at the same place regardless of the transport mean employed.

The transportation system has to maintain a constant flow of ripe cane to the sugar mill (Higgins 2002). Therefore, the planning of the transport of sugarcane from fields to the sugar mill is a difficult task depending on the carriage means available but at the same time is necessary to avoid the waste of valuable resources (Díaz and Pérez 2000; Higgins and Muchow 2003). The rail system may operate 24 h a day, whereas the harvest period may comprise only a part of the day (in daylight). Then, when the road transport stops working at night, the rail system is the only source of supply. Thus, the rail system acts as storage for cut cane, allowing the creation of a buffer. Mills have minimal storage facilities, so rail transport satisfies the demand of the sugar mill, while road transport either is covering other routes at the same time or is stopped at night. Unless a sugar mill failure or breakdown occurs, railway transportation allows the sugar mill to work 24 h a day without interruption; however, from time to time, the sugar mill is stopped for technical maintenance.

The organization of the operations prior to processing cane affects sugar quality and productivity at the sugar mill. For instance, harvested sugarcane spoils if it is not processed as soon as possible after harvesting. A degradation process begins when sugarcane is cut. This process transforms the sweet juice of sugarcane in acid juice in less than 24 h. This reason makes it necessary to mill as soon as possible the cut cane in the sugar mill, so the cut cane must be milled the same day it is cut. Long time ago, the degradation process was accelerated when sugarcane was set on fire before cutting (this practice is not in use nowadays in most countries).

Therefore, the sugar industry employs the so-called sugarcane freshness as a technical quality indicator (sugarcane freshness is defined as the standard time that sugarcane lasts from being cut in the field until it is processed in the sugar mill). Taking this indicator into consideration, the use of road transport is the best alternative, since it allows the sugarcane to arrive at the sugar mill in optimal conditions to be processed and is therefore preferred. Nevertheless, carriage costs are higher than the ones presented by the railway alternative under agro-industrial conditions.

The second aspect affecting sugar quality is the *pol* of the cane (i.e., an index related to the ripeness that is used for deciding the cutting moment). If necessary, the ripeness of the cane can be assessed by the supply of reapers to the cane plantations. Reapers improve the *pol* of certain cane varieties; however, a disadvantage is that when using these products, the cane must be cut as soon as possible to avoid its earlier deterioration. For this reason, planning the supplies of sugarcane to the mill is related to the agronomical aspects of the crop. In this sense, a balanced arrival of fields to the peak of maturity is required and involves tactical and strategic decisions left aside in this study. Readers interested in these aspects can find interesting the papers of Bezuidenhout and Singels (2007a, b), Grunow et al. (2007), and Le et al. (2008).

In brief, once a set of fields have been selected for harvesting (they are around the maturation peak), the most important operational aspect of cane harvesting is being able to determine the optimal combination of transportation means. The objective is of minimizing the global transportation cost while fulfilling the daily sugar mill supply needs with an acceptable level of quality and avoiding cane losses caused by not harvesting or sugar deterioration.

5.3 General Formulation of the Sugar Supply Chain

The mathematical model formulated is a mixed-integer linear programming (MILP) model dealing with all of the aspects of the operational problem of sugarcane transport for one day. The model is based on a linear programming model developed previously by López et al. (2006) producing schedules of road transports and cutting means. This model is similar to others proposed for production and transport planning reported by Mula et al. (2010). Accordingly, the organization of rail transports is not detailed; only the amount to be served daily to the mill is estimated. It is assumed that cane in storage facilities is available to the mill when needed, i.e., once road transport cannot operate due to inactivity on fields, in particular at night. Quality aspects are considered by means of an opportunity coefficient determined empirically by the decision-maker and by establishing minimum quantities of cane processed just in time to preserve cane freshness.

5.3.1 Decision Variables

The decision variables are represented by X_{ijklm}, where the subscript i represents the origin ($i = 1$ to $i = A$, as storage facilities; $i = A + 1$ to $i = A + B$, as sugarcane fields); j is the destination ($j = 1$, sugar mill; $j = 2$ to $j = A + 1$, as storage facilities); k is the transportation means used ($k = 1$ represents railway transportation and from $k = 2$ to $k = K$ different road transportation means); l is the cutting system ($l = 1$, for $k = 1$ since in railway transportation, the way in which cane is cut is not relevant; $l = 2$ to $l = L + 1$ as a group of harvesting machines; $l = L + 2$ to $l = L + C + 1$ as a group of manual cutters); and m is the time of the day ($m = 1$ to $m = H$, $H \leq 24$).

Then X_{ijklm} is the quantity of sugarcane transported from origin i to destination j by transportation mean k during the hour m and harvested by the group l (X_{ijklm} is expressed in "arrobas," traditional system of weight in Cuba represented by the @ symbol; 1 @ = 25 lb = 11.502 kg).

The decision variables have a combinatorial nature which are not all possible; to define those that will be feasible in the model formulation, the following rules are necessary:

- The variables determining routes (both for road and rail transportation) where an origin is also the destination are not considered.
- In case the origin is a storage facility ($i = 1$ to A), the only destination admitted is the sugar mill ($j = 1$). The storage facilities will not transfer cane between them, and only unloading it in the swing bolster is allowed.
- The sugarcane fields as origins will admit any destination ($j = 1$ to $j = A + 1$).
- The variables presuming the railway transportation ($k = 1$) will only be defined for the combination with the sugar mill ($j = 1$) and the subindex $l = 1$.

In order to represent specific field conditions, other rules affecting decision variables can be considered. In a complementary way, the inclusion of binary variables provides the scheduling of transports and cutting tasks for a day's operation. Namely, binary variables are B_{ilm}, $Y_{il} \in \{0, 1\}$ where the meaning of subscripts are preserved to be consistent with previous notation, but the range is lower to consider only relevant cases:

I: represents the sugarcane fields ($i = A + 1$ to $i = A + B$).
L: is the cutting system ($l = 2$ to $l = L + 1$ as a group of harvesting; $l = L + 2$ to $l = L + C + 1$ as manual cutting).
M: is time of the day ($m = 1$ to $m = H$, $H \leq 24$).

In practice, it is the mill manager who has to maintain and update the abovementioned rules determining the total number of decision variables for feasible solutions.

5.3.2 Main Constraints

Main constraints refer to constraints always present in the different formulations of the problem. The core of the problem can be solved with only these constraints, i.e., taking $H = 1$ (representing one working day and ignoring the schedule hour by hour). These are constraints including continuous variables, and so the model can be solved faster than when binary or integer variables are present. This distinction allows getting a first approach to the level of resources required for a working day, leaving aside the complexities of a detailed scheduling. Thus, the constraints of the mathematical model are classified in the following groups:

- Supply of cane to the sugar mill for a working day
- Capacity of the storage facilities
- Conservation of flow-through storage facilities
- Capacity of transportation by road transportation means
- Production of the sugarcane fields
- Cutting capacity of different teams

5.3.2.1 Supply of Cane to the Sugar Mill

A daily quota of cane has to be cut in the fields and later transported to the mill. It would be desirable to avoid wide variations in this quota. Therefore, limits for upper (Mmax) and lower (Mmin) supply of cane to the sugar mill in a working day are stated.

$$\sum_{i=1}^{A}\sum_{m=1}^{H} X_{i111m} + \sum_{i=A+1}^{A+B}\sum_{k=2}^{K}\sum_{l=2}^{L+C+1}\sum_{m=1}^{H} X_{i1klm} \leq \text{Mmax}$$

$$\sum_{i=1}^{A}\sum_{m=1}^{H} X_{i111m} + \sum_{i=A+1}^{A+B}\sum_{k=2}^{K}\sum_{l=2}^{L+C+1}\sum_{m=1}^{H} X_{i1klm} \geq \text{Mmin}$$

In order to maintain a uniform flow of sugarcane to the mill, the supply per hour is considered. The sugar mill can only process a fixed quantity of cane per hour as maximum (named Smax_m). A problem to avoid is the overflow caused by direct transportation ($j = 1$ and $k \neq 1$).

Maximum processing capacity of the sugar mill by direct transportation:

$$\sum_{i=A+1}^{A+B}\sum_{k=2}^{K}\sum_{l=2}^{L+C+1} X_{i1klm} \leq \text{Smax}_m \quad m = 1, 2, \ldots, H$$

Direct transportation has priority if the aim is to maintain good cane quality. Therefore, a minimum hourly quantity of sugarcane to be transported to the sugar mill is established, in this case Smin_m @ (of course, $\text{Smin}_m < \text{Smax}_m$).

Minimal supply of the sugar mill by direct transport:

$$\sum_{i=A+1}^{A+B} \sum_{k=2}^{K} \sum_{l=2}^{L+C+1} X_{i1klm} \geq \mathrm{Smin}_m \quad m = 1, 2, \ldots, H$$

5.3.2.2 Capacity of the Storage Facilities

Each storage facility has a limited management and storing capacity expressed in @/h. Each one is represented by SP_j, with $j = 2, \ldots, A+1$.

$$\sum_{i=A+1}^{A+B} \sum_{k=2}^{K} \sum_{l=2}^{L+C+1} X_{ijklm} \leq \mathrm{SP}_j \quad m = 1, 2, \ldots, H \text{ and } j = 2, 3, \ldots, A+1$$

5.3.2.3 Conservation of Flow-Through Storage Facilities

The sugarcane delivered to a storage location must come out by train during the day (and even every hour) in order to satisfy the demand of the sugar mill. Hence, for each storage facility:

$$X_{j-1111m} - \sum_{i=A+1}^{A+B} \sum_{k=2}^{K} \sum_{l=2}^{L+C+1} X_{ijklm} = 0 \quad m = 1, 2, \ldots, H \quad j = 2, 3, \ldots, A+1$$

5.3.2.4 Capacity of Transport by Road Transportation Means

Railway transport is able to carry the amount of cane the sugar mill needs in the daily production. A correct operation of the rail system to deliver the cane to the mill is assumed, so that capacity constraints on the transport process will only be associated to road carriage. Thus, we define the coefficient CR_{ijkl} representing transport need (in hours) to carry an @ of cane harvested by harvester group l, from origin i to destination j by the road transport k. This coefficient is irrelevant to the time, m, in which the transportation is carried out, and it is calculated through the following formula:

$$CR_{ijkl} = \frac{D_{ij} \cdot \left(\frac{1}{Vcc_k} + \frac{1}{Vsc_k} \right) + Tc_{kl}}{Cc_k} \tag{5.1}$$

where

D_{ij}: Distances from origin i to destination j
Vcc_k: Speed of the given carriage means k, with load
Vsc_k: Speed of the given carriage means k, without load
Tc_{kl}: Waiting time of carriage means k, with cutting system l
Cc_k: Loading capacity of carriage means k

If the number of carriage means type k available per hour is denoted by TM_k, the related constraint can be written as follows (note that for each working hour, TM_k equals total transport force of transportation means type k expressed in hours of work):

$$\sum_{i=A+1}^{A+B}\sum_{j=1}^{A+1}\sum_{l=2}^{L+C+1} CR_{ijkl}\cdot X_{ijklm} \le \mathrm{TM}_k \quad m=1,2,\ldots,H \text{ and } k=2,3,\ldots,K$$

5.3.2.5 Production of the Sugarcane Fields (in @)

The production of sugarcane fields is estimated in @ and represented by Cap_i. This parameter is set by the mill manager assisted by technicians on the field. However, for tactical or strategic planning, this value can be estimated making use of more refined methods simulating crop production (see, e.g., Bezuidenhout and Singels 2007a, b; Piewthongngam et al. 2009; Stray et al. 2012).

$$\sum_{j=1}^{A+1}\sum_{k=2}^{K}\sum_{l=2}^{L+C+1}\sum_{m=1}^{H} X_{ijklm} \le Cap_i \quad i = A+1, A+2, \ldots, A+B$$

5.3.2.6 Cutting Capacity of Different Teams (in @)

The production of sugarcane fields is cut by different teams having a capacity of cut represented by Prod_l. This parameter is set by the mill manager according to the composition of the different cutting means available. These can be essentially mechanical or manual.

$$\sum_{j=1}^{A+1}\sum_{k=2}^{K} X_{ijklm} \le \mathrm{Prod}_l \quad i=A+1, A+2, \ldots, A+B;$$
$$l=2,3, \ldots, L+C+1 \text{ and } m=1,2,\ldots,H$$

5.3.3 *Constraints for Scheduling Daily Tasks*

As seen, sugarcane harvesting is carried out with groups of harvesting machines (in number: L) and manually with groups of sugarcane cutters (in number: C). Hence, if we consider the workload per hour, the reformulated constraint is as follows:

$$\sum_{j=1}^{A+1}\sum_{k=2}^{K} X_{ijklm} \le \mathrm{Prod}_l B_{ilm} \quad i=A+1, A+2, \ldots, A+B;$$
$$l=2, 3, \ldots, L+C+1 \text{ and } m=1, 2, \ldots, H$$

where $B_{ilm} \in \{0, 1\}$ is the binary variable controlling possible combinations of origin-cutting mean per hour. Aimed at getting a logical and reasonable work plan and scheduling, a group of binary variables are included. The introduction of these binary variables per hour makes it necessary to add new constraints. This set of new constraints refers basically to constraints related to the operation of cutting means used in sugarcane harvesting and transport trips required hour by hour. Therefore, the extra constraints appended have the following meaning:

- Cutting means can work in only one field in 1 h.
- A field can hold up to two groups of harvesters.
- Movements of cutting means between fields are limited to one.
- The work of harvesters cannot overcome the daily hours of work, H.
- A group of harvesters can only work consecutive hours in a field.

5.3.3.1 Cutting Means Can Work in Only One Field in One Hour

Groups of harvesting machines are compounded by several means designed to work together in a field during 1 h; therefore, it is unpractical to divide them to work in more than one field during the same hour.

$$\sum_{i=A+1}^{A+B} B_{ilm} \leq 1 \quad l = 2, 3, \ldots, L + C + 1 \text{ and } m = 1, 2, \ldots, H$$

5.3.3.2 A Field Can Hold up to Two Groups of Harvesters

Each group of harvesting machines is sized to work in a field; therefore, it is unpractical to allow more than two groups placed in the same field. This has to be verified per hour:

$$\sum_{l=2}^{L+C+1} B_{ilm} \leq 2 \quad i = A + 1, A + 2, \ldots, A + B \text{ and } m = 1, 2, \ldots, H$$

But also during the day:

$$\sum_{l=2}^{L+C+1} Y_{il} \leq 2 \quad i = A + 1, A + 2, \ldots, A + B$$

where $Y_{il} \in \{0, 1\}$ is the binary variable representing if the cutting mean l has been working or not on field i during the working day.

5.3.3.3 Movements of Cutting Means Between Fields Are Limited

The task of harvesting groups is to cut cane at their full capacity, and hence, loss of time in movements should be reduced. On the other hand, more than one change between fields during the day would be unrealistic. Then:

$$\sum_{i=A+1}^{A+B} Y_{il} \leq 2 \quad l = 2, 3, \ldots, L + C + 1$$

5.3.3.4 Harvesters Cannot Exceed the Daily Hours of Work, *H*

Working days have a limited number of hours that affect the working hours of harvesting machines:

$$\sum_{m=1}^{H} B_{ilm} \leq HY_{il} \quad i = A + 1, A + 2, \ldots, A + B \quad l = 2, 3, \ldots, L + C + 1$$

5.3.3.5 Harvesting Is Deployed at Consecutive Hours on a Field

If a group of harvesting machines leaves a field to work in another, it makes no sense to come back later to the former. Similarly, idle time or pauses are not allowed in the middle of a working period of time in a day. If a worker stops, he should start working again on the next day.

$$\sum_{m=t+2}^{H} B_{ilm} - (H - 1 - t)B_{il,t+1} + (H - 1 - t)B_{il,t} \leq H - 1$$

$i = A + 1, A + 2, \ldots, A + B \quad l = 2, 3, \ldots, L + C + 1$ and $t = 1, 2, \ldots, H - 2$

5.3.4 *Transport Cost and Quality Aspects*

When planning sugarcane harvest, there are usually two major objectives: quantity of cane harvested and transported and quality of sugar. In this sense, the primary objective is the minimization of daily transportation cost. Hence, the economic coefficients of the objective function (C_{ijklm}) establish the transportation cost of sugarcane, related to the distances and the transportation means used in each case. The economic coefficient of each variable is determined by $C_{ijklm} = c_k \cdot d_{i,j}$, where c_k is the specific economic coefficient related to transport k and d_{ij} is the distance between origin i and destination j. Since C_{ijklm} are just representing transportation costs, the way and time in which the cane is cut do not affect.

However, quality aspects are introduced in the objective function through an opportunity coefficient, Co_i, which represents the preference to cut a sugarcane field i. By default, it is assumed that all fields susceptible of harvest have a similar maturation level (similar *pol* or sugar content) and $Co_i = 1$. If not, this coefficient serves to give priority in the cutting of a field instead of another given the *pol* of the cane. Corresponding priority has to be set accordingly by the manager or the person who knows the *pol* of fields. Note that the preservation of cane freshness can also be considered by establishing minimum quantities of cane processed just in time, i.e., sent directly to the mill by road. To summarize, the objective function representing the transportation cost is:

$$\text{O.F.} : \text{Minimize} \quad \sum_{i=1}^{A+B} \sum_{j=1}^{A+1} \sum_{k=1}^{K} \sum_{l=1}^{L+C+1} \sum_{m=1}^{H} C_{ijklm} \cdot Co_i \cdot X_{ijklm}$$

where

C_{ijklm}: represents the economic coefficients.
Co_i: represents the opportunity coefficients.
X_{ijklm}: are the decision variables.

5.3.5 *Total Constraints and Variables*

The total number of constraints and variables serves to get an idea of the complexity of the model. Nowadays, there exist commercial solvers that can solve very efficiently any linear program. López et al. (2006) detail the calculation of the total number of constraints $H \times (1 + 2A + (1 + 2B)(L + C) + B + K) + 2 + 2B + (L + C)(1 - B)$ and variables $B(L + C) \times [1 + H[(A + 1) \times (K - 1) + 1]] + AH$, emphasizing that $B \times (L + C) \times (H + 1)$ of them are binaries and the rest continuous.

The number of integer variables converts the linear model into a mixed integer linear model (MILP) and raises the difficulty level of reaching an exact solution. Note that the size of the problem is reduced significantly when considering $H = 1$, i.e., one day of work. Because of that, a two-stage approach was introduced to solve the model as explained in the next section.

5.3.6 *Model Implementation*

The package to solve the model was LINDO (Schrage 1997). This was a requirement by the user. With variables representing daily quantities, the system provides a solution in a few milliseconds as reported in López et al. (2004). The whole model as formulated in López et al. (2006) was unsolvable with LINDO due to the complexity of the system and the number of integer variables. Therefore, a two-stage approach was adopted successfully. The first stage consisted in a first round solving the problem for a day without the hourly scheduling. The optimal solution serves to identify the fields to harvest and other resources present in the optimal solution. Not all constraints are saturated nor decision variables needed. This way, the total number of variables and constraints can be pruned smartly before solving successfully the full model in a second stage. Thus, a detailed solution by hour can be obtained from the initial solution for a day. The quantity of cane transported to the sugar mill, the location where cane is collected, the transportation means used, the way cane is cut, and the transportation time are all provided in the solution. In a complementary way, the remaining binary variables provide the scheduling of transports and cutting tasks for a day's operation. In addition, sugar mill managers are able to determine the amount of transportation means required, the exact time of shipment, and the amount of petrol in reserve.

5.4 Embedding the Model into a DSS

The practical use of the model presented in the previous section needs a friendly environment for being used as an operational tool in the sugar industry. For this purpose a DSS was developed and allowed the user to set parameters for the model, solve the model, generate and save reports, and explore different alternatives in a reasonable time. Inherent complexities of the mathematical model are hidden to the intended user who concentrates in the analysis of the problem and proposed solution.

5.4.1 DSS Structure

The structure and system elements of the DSS are presented in Fig. 5.1. The heart of the system is the mathematical model representing the supply chain. The interface was implemented in MS Visual Basic v. 6. It interacts with the user and manages the LINDO library in two ways: to formulate and reformulate the model and to retrieve and show the outcome. The database stores inputs and outputs. Inputs are all of the resources available for daily sugarcane harvesting and processing, like fields, cutting means, road transport means, and storage facilities. The database was implemented in MS Access. Other inputs concern technical characteristics of the mill, as, for example, the daily quota that can be processed as maximum or minimum and the distances to the fields and storages facilities. Outputs are represented by the allocation and distribution of resources along the chain, giving priority to transportation cost minimization and sugar quality. The interface allows the user to interact with the system, modifying inputs and retrieving outputs. Outputs can be displayed summarized by day or hour by hour.

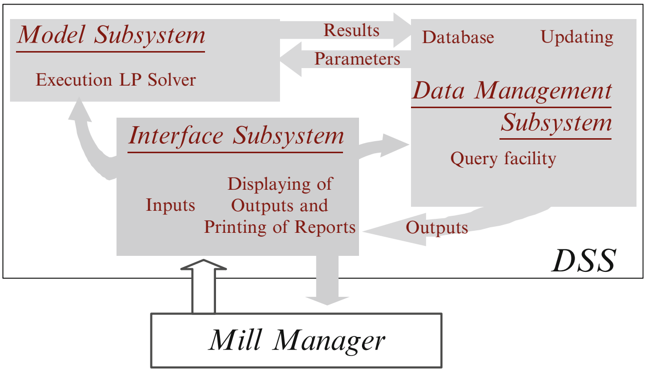


Fig. 5.1 Structure of the DSS

The interface allows also the user to interact with the database to make queries or updates and run the model. This way, inherent complexities of the mathematical model are avoided to the user who can be concentrated in defining all the resources available for harvesting and transport, running the model, and analyzing, approving, and implementing the results. The internal generation of the model is automatically performed from user specifications of model inputs through different menus and tables.

5.4.2 DSS Operation

The DSS was developed accounting for the regular operation of sugar mills in Cuba. In particular, it was tested in a sugarcane supply chain located in the Holguín province (Cuba) that processes cane from 239 fields (Fig. 5.2), representing a total surface of 25,000 ha. All the resources are owned by public companies and involve the mill "Fernando de Dios" and a dedicated railway system serving the mill.

The harvesting labor on the fields is a maximum of 14 h per day, and it corresponds to the maximum daylight duration. The daily train supplies represent about 80 % of cane from a total of 300,000 @ (i.e., 3,400 tm) of sugarcane that can be processed in a working day, and the train can operate also at night. Between 8 and 9 % of the weight of sugarcane is converted in sugar; therefore, the expected production of sugar is 27,300 tm per year during the season (i.e., 100 days per year as maximum). On a regular working day, this sugar mill is supplied with cane cut in different surrounding fields and sent directly by road or transferred by train. The train serves from five storage facilities to the mill. The rail network has a star shape with the mill connected to all the vertices without intermediate stations. This way, congestion is inexistent and organization of rail transports simple (not included in this study). Out of the harvesting season, the use of the train is insignificant and most of the wagons stored for the next season. Regarding road transport means,

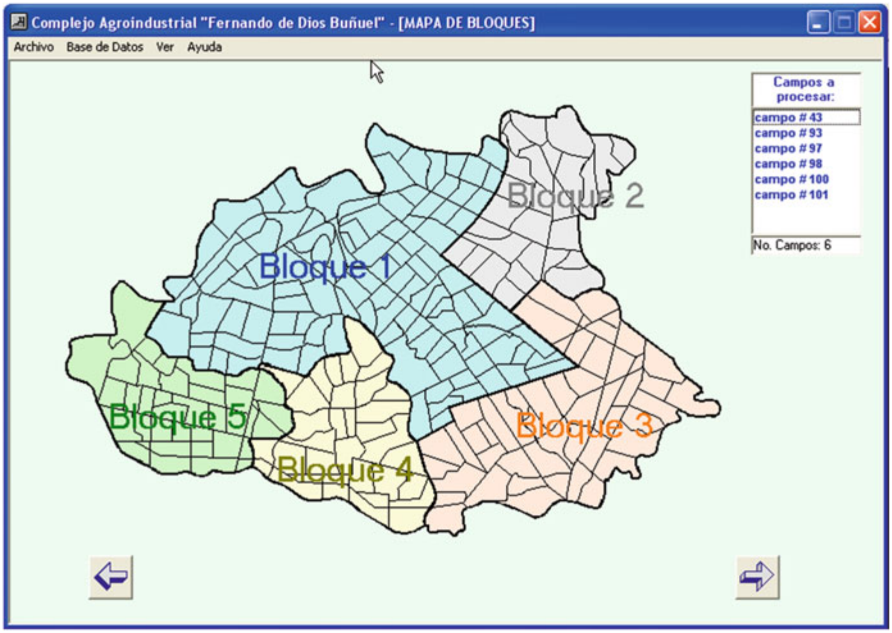


Fig. 5.2 Selection of fields with sugarcane ready to harvest

about 100 of automotive transport means work in the mill. These include 58 ZIL-130 trucks (with and without tows) and 38 MTZ-80 tractors with two tows. The database has to be updated with the units available, new types of vehicles, by unselecting those not present or unavailable at all (e.g., KAMAZ 53212) as shown in Fig. 5.3. The fields ready to be harvested and the rest of available resources have to be selected by the user from the database before launching the model. The estimated sugarcane on fields, opportunity coefficient, available road transportation with the corresponding technical and operational characteristics, number of groups of harvesting means (mechanical or manual), and corresponding work capacity have to be set or revised periodically. For this purpose, the DSS interface has a set of windows and displays where all these parameters can be inspected through different tabs and corresponding figures introduced. Transport means with similar characteristics are grouped, and the user sets the number of them available for the day (e.g., in Fig. 5.3, there are 78 trucks ticked as available). Something similar happens with cutting means where each harvesting machine is represented (15 in total) and corresponding capacities set. In addition, the daily processing capacity of the sugar mill, the minimum quantity of cane to be processed, the maximum and minimum quantities of sugarcane supplied to the sugar mill per hour, and the limited management capacity of each storage facility have also to be set.

The cane is transported to the sugar mill straightaway after being cut or passed through storage facilities to be conveyed by train. Time consumed in load, unload, and road transportation is taken into account through coefficients (1) applied in

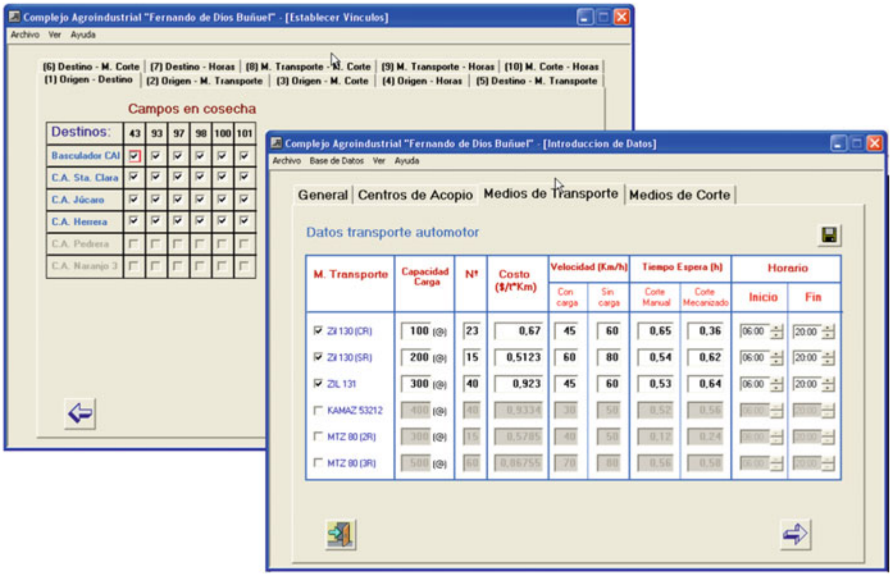


Fig. 5.3 Selection of harvesting means and corresponding technical and operational characteristics

constraints of available transport means. All available arcs connecting road origins and destinations for each feasible combination also have to be set. They can be enabled or disabled through the interface at any time, for example, when a storage place is unavailable. Distances of each feasible path are stored in the database and considered in the calculation of C_{ijkl}.

The solution, as it is shown in Fig. 5.4, represents the daily quantity of sugarcane transported. The solution indicates how direct transportation is not always preferred, because all related constraints are not fully satisfied, and thus, the maximum processing capacity of this kind is not attained. In particular the storage center (i.e., loading station) named *Júcaro* receives 12,499@ of cane to be sent by train, while 14,588@ are sent directly by road to the mill. Generally, direct transportation involves the nearest fields to the sugar mill for which road transportation is cheaper. Also, the nearest storage facility to each field is the preferred one to store cane when needed. On the other hand, fields to be cut (only #87, not yet #88 and #89), road means needed (Zil 130 CR and SR), and cutting teams involved (only #3, although #2 and #5 are available) are displayed with an expected amount of cane.

Daily solution can be used to refine the final outcome and retrieve an hourly scheduling of cutting teams and road transports means. In view of the daily solution, resources not used according to the daily solution can be discarded. Hence, the manager can reduce the number of variables and constraints disabling storage places, harvesting teams, or road transport means. The manager can limit the resources to those appearing in the daily solution and request to the system an hourly solution (Fig. 5.5). In other words, daily solution allows the user to prune

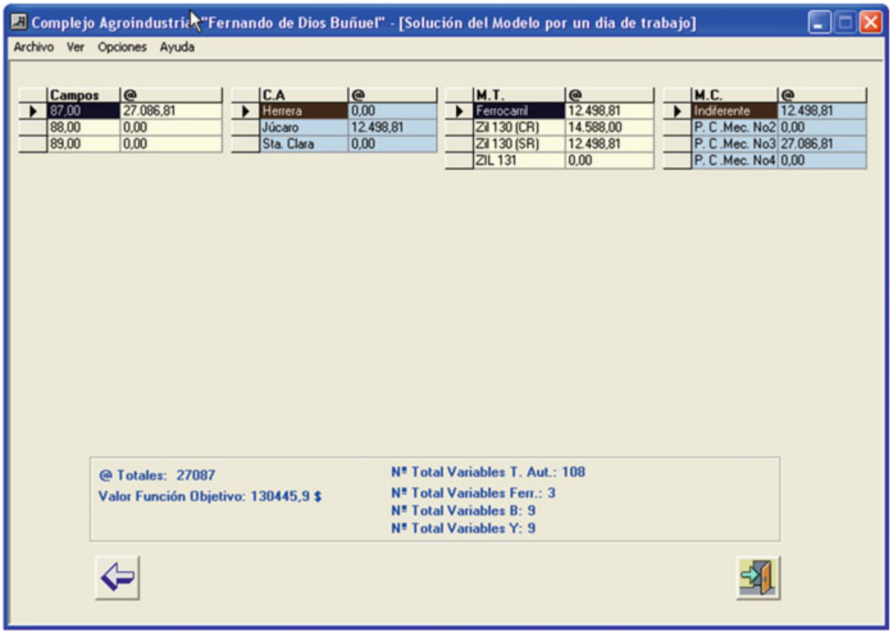


Fig. 5.4 Work plan for a day

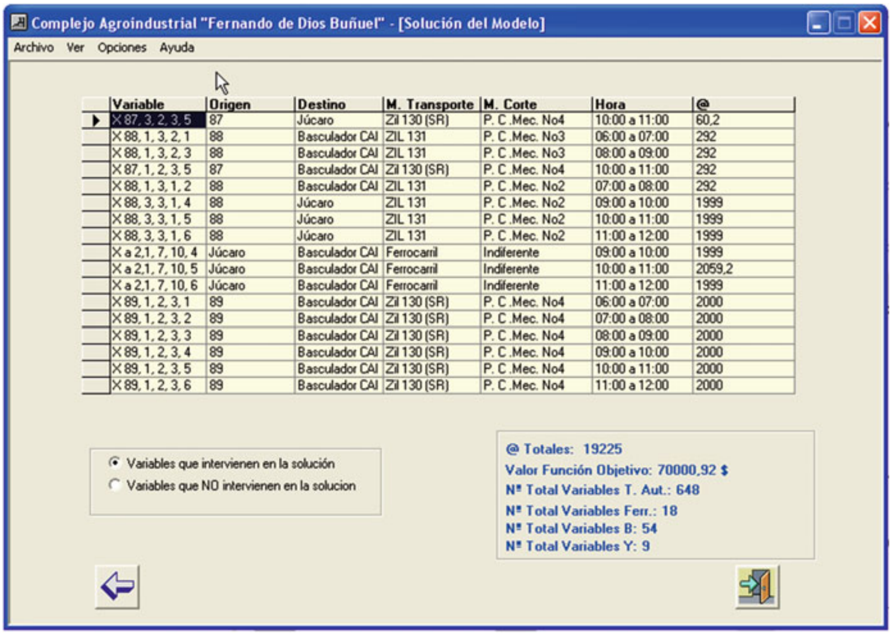


Fig. 5.5 Scheduling of resources per hour

resources that are not necessary for that day, reducing the dimension of the problem and speed up the resolution. Therefore, an hourly solution represents a refinement of this primary daily solution with additional computational savings. As a result, allocations of cutting teams on fields and road transports, means, and paths involved per hour are obtained. This is shown in Fig. 5.5 where the list of nonzero variables is displayed sorted by origin (any other ordering criteria can be selected). The list shows for each origin and destination the transport mean involved, the cutting team, the amount of cane to be cut, and the time (hour of the day). Upon demand by the user, the list is refined and sorted by time showing the schedule of transport means or cutting teams.

5.5 Discussion

The model formulated here represents the particular situation of the Cuban sugar industry. Other sugar supply chain models in literature are focused on other aspects of particular interest in other countries like Thailand, Australia, Venezuela, Brazil, or South Africa. However, the main contribution of this study is the development of a DSS for use daily at mills, covering real needs of mill managers in Cuba. The internal complexity is inherited from its embedded mathematical model. What was more problematic was handling of the large number of variables, although this problem is unappreciated for the user and remains inside the DSS. Other DSSs developed for the sugar industry have been presented by Lejars et al. (2008) and Stray et al. (2012) but with a strategic scope.

The operation with the DSS (and with the embedded model) is as follows. In a preliminary step, a reduced case considering just one day should be solved in order to approximate the optimal solution of the problem with a reasonable computational time (López et al. 2004). The report of solutions is similar to that shown in Fig. 5.3 where resources for a working day are allocated. This first step serves to verify, adjust, and refine the real needs of all kinds of means to perform the harvesting, and thus, superfluous variables can be ignored (unticked) in subsequent runs of the model. This way computational load is lighter instead of having to solve the full instance.

The problem can become more complex depending on the changes in available resources, impacting to the number of constraints from one day to another as López et al. (2006) reported. For instance, the availability of cutting and transportation means, the number of railway stations, the number of fields with matured sugarcane, and the number of working hours may all vary. It is the user who daily has to select the actual parameters from a correctly updated database. If the number of fields and the number of harvesting machines increase and the available transportation means and their working hours increase, then the number of decision variables in the model also increases. Therefore, a regular size of the problem may represent 19,294 constraints and 2,420 variables, of which 1,084 are integers (e.g., when $A = 5$, $B = 9$, $C = 2$, $K = 4$, $L = 6$, and $H = 14$). Solving this kind of complex model daily is seen as very useful in practice by Cuban managers. The DSS presented is capable of solving the problem of cost minimization of sugarcane transport from fields to the sugar mill for one working day. The model determines the capacities of the road and rail transport facilities for transporting cane to ensure an uninterrupted supply to the sugar mill. Moreover, the scheduling of road transportation and harvesting quotas of cutting means are derived from an optimal solution that simplifies the daily task of the mill's managers. Railway transport is not a limiting resource in Cuba, and only daily quotas carried by train are determined. Thus, specific railway scheduling is not considered in detail and kept as a problem apart.

Although the requirement of using LINDO was an initial obstacle because it was impossible to solve the whole model containing all variables and constraints (López et al. 2006), finally it revealed useful in practice when forcing a two-stage solution method. Firstly, a daily solution was obtained summarizing the resources needed for transport and harvesting of the day (subindex m is not considered and many binary variables avoided). This first model represented a reduction of more than 90 % of constraints and variables over original sizes. Secondly, the user refined this solution selecting optimal resources involved in the daily solution to obtain an hourly solution, more practical for operational purposes. This second stage implies that subindex m is enabled and only the resources present at the first solution. Thus, the number of constraints are 10,290 with 1,900 variables, of which 840 are integers (when $A = 3$, $B = 7$, $C = 2$, $K = 4$, $L = 6$, and $H = 14$). Validation tests performed in the mill confirmed this point. In the second stage, the model has become lighter due to the reduction in the number of variables and constraints considered in the formulation.

The mathematical formulation of the model integrates rail and road transport systems emphasizing the reduction of transportation cost. This problem is common to other countries and different solutions are proposed (Higgins 2006). The approach presented here benefits of the use of coefficients C_{ijkl} that account for road transport capacity according to speed, load–unload time, trip time (with and without load), and distance between origin and destination. Hence, a reliable estimation of such coefficients is crucial for the goodness of the optimal solution proposed. The model also controls sugarcane freshness through the minimum supply constraints to the sugar mill with direct transport. Furthermore, the model allows sugar mill managers to schedule daily transport plans automatically, based on either objective criteria or on considerations that have been acquired through professional experience. This was of great value for managers who previously operated only under his/her expertise. They have estimated savings of 8 % in fuel with respect to the traditional scheduling method which represents 23,000 L during the harvesting season. Furthermore, other benefits are the time managers save comparing the computerized system with the old by hand method. Also included are manpower savings valued in the mill of USD 150,000 due to the rational use of road transport means (i.e., less repairs and maintenance activities) and a better welfare of employees. These results confirm the sort of opportunities for value chain research in sugar industries claimed by Higgins et al. (2007).

Professional solvers permit solving huge linear programming models, but managers find them difficult to handle. Because of this, it is helpful to elaborate custom-made software based on a friendly interface that simplifies such complexities, deals with large amounts of variables, and makes the mathematical model transparent to the user. In this way, the reformulation of the problem and its daily update is easier and feasible, saving a lot of time and money for the sugar industry. Combining the capabilities of specific software for solving MILPs with the knowledge and the experience of who are familiar with the “cutting–loading–transportation” system for sugarcane allows the users of such models to make more flexible allocation decisions of harvesting and road transportation means. Moreover, it also provides the scheduling of these resources in place and time according to their daily availability. An open challenge is the integration of different models covering all the decision spectrums for the sugarcane industry as noted by Higgins et al. (2007) and Plà et al. (2013).

5.6 Conclusions

Production per hectare in Cuba is low compared with other figures reported in literature for other countries like Australia or Venezuela (Higgins 2006; Grunow et al. 2007). Reasons for that are the resources and technology involved like irrigation, modern machinery, weather, and seeded varieties. However, the rational use of the scarce resources allows mill managers to achieve a better efficiency productivity and a higher control over production costs remaining competitive. In this sense, the model described in this chapter was developed for planning daily operations in the sugarcane supply chain. The DSS in which the model was embedded was developed and tested in the mill “Fernando de Dios” in the province of Holguín, Cuba. The DSS is of great value for mill managers who previously operated only under his/her expertise by hand. They reported savings of 8 % in fuel with respect to the traditional scheduling method. Furthermore, other benefits are manpower savings valued only in the mill of USD 150,000 due to the rational use of road transport means (i.e., less repairs and maintenance activities) and a better welfare of employees.

Although all inputs are recorded in the database and related to this case, the set of parameters can be easily adapted to represent different situations in the sugar industry where road and rail transports are involved. In general, from mill to mill, the database can change in size according to the number of resources available but not in the kind of these resources. This makes the DSS flexible enough to be applied in other Cuban mills and even foreign mills. New developments involving the enlargement of the model are expected in the future, for instance, including crop planning.

References

Bezuidenhout CN, Singels A (2007a) Operational forecasting of South African sugarcane production: Part 1—System description. Agric Syst 92:23–38

Bezuidenhout CN, Singels A (2007b) Operational forecasting of South African sugarcane production: Part 2—System evaluation. Agric Syst 92:39–51

Díaz JA, Pérez IG (2000) Simulation and optimization of sugar cane transportation in harvest season. In: The proceedings of winter simulation conference, Miami, December 2000, pp 1114–1117

Grunow M, Günther H-O, Westinner R (2007) Supply optimization for the production of raw sugar. Int J Prod Econ 110:224–239

Higgins AJ (1999) Optimizing cane supply decisions within a sugar mill region. J Scheduling 2:229–244

Higgins AJ (2002) Australian sugar mills optimize harvester rosters to improve production. Interfaces 32(3):15–25

Higgins A (2006) Scheduling of road vehicles in sugarcane transport: a case study at an Australian sugar mill. Eur J Oper Res 170:987–1000

Higgins AJ, Laredo LA (2006) Improving harvesting and transport planning within a sugar value chain. J Oper Res Soc 57:367–376

Higgins AJ, Muchow RC (2003) Assessing the potential benefits of alternative cane supply arrangements in the Australian sugar industry. Agric Syst 76:623–638

Higgins AJ, Thorburn P, Archer A, Jakku E (2007) Opportunities for value chain research in sugar industries. Agric Syst 94:611–621

Jena SD, Poggi M (2013) Harvest planning in the Brazilian sugar cane industry via mixed integer programming. Eur J Oper Res 230:374–384

Le Gal P-Y, Lyne PWL, Meyer E, Soler L-G (2008) Impact of sugarcane supply scheduling on mill sugar production: a South African case study. Agric Syst 96:64–74

Le Gal P-Y, Le Masson J, Bezuidenhout CN, Lagrange LF (2009) Coupled modelling of sugarcane supply planning and logistics as a management tool. Comput Electron Agr 68:168–177

Lejars C, Le Gal P-Y, Auzoux S (2008) A decision support approach for cane supply management within a sugar mill area. Comput Electron Agr 60:239–249

López E, Miquel S, Plà LM (2004) El problema del transporte de la caña de azúcar en Cuba. Rev Invest Oper 25:148–157

López E, Miquel S, Plà LM (2006) Sugar cane transportation in Cuba, a case study. Eur J Oper Res 174:374–386

Martin F, Pinkney A, Xinghuo XY (2001) Cane railway scheduling via constraint logic programming: labelling order and constraints in a real-live application. Ann Oper Res 108:193–209

Mula J, Peidro D, Díaz-Madroñero M, Vicens E (2010) Mathematical programming models for supply chain production and transport planning. Eur J Oper Res 204:377–390

Pavia R, Morabito R (2009) An optimisation model for the aggregate production planning of a Brazilian sugar and ethanol milling company. Ann Oper Res 161:117–130

Piewthongngam K, Pathumnakul S, Setthanan K (2009) Application of crop growth simulation and mathematical modeling to supply chain management in the Thai sugar industry. Agric Syst 102:58–66

Plà LM, Sandars D, Higgins A (2014) A perspective on operational research prospects for agriculture. J Oper Res Soc 65:1078–1089. doi:10.1057/jors.2013.45

Rizzoli AE, Fornara N, Gambardella LM (2002) A simulation tool for combined rail/road transport in intermodal terminals. Math Comput Simulat 59:57–71

Scarpari MS, de Beauclair EGF (2010) Optimized agricultural planning of sugarcane using linear programming. Rev Invest Oper 31:126–132

Scarpari MS, Plà LM, de Beauclair EGF (2008) La optimización del cultivo de variedades de caña de azúcar. Rev Invest Oper 29:26–34

Schrage L (1997) Optimization modeling with LINDO, 5th edn. Duxbury Press, ITP, New York
Semenzato R (1995) A simulation study of sugar cane harvesting. Agric Syst 47:427–437
Stray BJ, van Vuuren JH, Bezuidenhout CN (2012) An optimisation-based seasonal sugarcane harvest scheduling decision support system for commercial growers in South Africa. Comput Electron Agr 83:21–31

Chapter 6
A Hierarchical Planning Scheme Based on Precision Agriculture

Víctor M. Albornoz, Néstor M. Cid-García, Rodrigo Ortega, and Yasmín A. Ríos-Solís

Abstract The process for agriculture planning starts by delineating the field into site-specific rectangular management zones to face within-field variability. We propose a bi-objective model that minimizes the number of these zones and maximizes their homogeneity with respect to a soil property. Then we use a method to assign the crops to the different plots to obtain the best profit at the end of the production cycle subject to water forecasts for the period, humidity sensors, and the chemical and physical properties of the zones within the plot. With this crop planning model we can identify the best management zones of the previous bi-objective model. Finally, we show a real-time irrigation method to decide the amount of water for each plot, at each irrigation turn, in order to maximize the total final yield. This is a critical decision in countries where water shortages are frequent. In this study we integrate these stages in a hierarchical process for the agriculture planning and empirically prove its efficiency.

6.1 Introduction

Precision agriculture has modified the decision-making tools used by the farmers in order to plan their production cycle as well as their daily operation. Investing in precision agriculture goods such as humidity sensors or soil samples is interesting when farmers get not only tools for monitoring their fields but also

V.M. Albornoz (✉) • R. Ortega
Departamento de Ingeniería Comercial, Universidad Técnica Federico Santa María, Vitacura, Santiago, Chile
e-mail: victor.albornoz@usm.cl; rodrigo.ortega@usm.cl

N.M. Cid-García • Y.A. Ríos-Solís
Graduate Program in Systems Engineering, Universidad Autónoma de Nuevo León (UANL), San Nicolás de Los Garza, Mexico
e-mail: nestor.cidgrc@uanl.edu.mx; yasmin.riossls@uanl.edu.mx

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_6

a powerful hierarchical system that optimizes the benefits of the yields as the one we present in this study.

One of the main aspects of precision agriculture is to provide farming management methods to respond to within-field variability. Precision agriculture permits the application in a site-specific manner of agronomic practices such as fertilization, weed and pest control, as a function of the information compiled from collected field data.

Physical and chemical soil properties make the soil suitable for agricultural practices. Texture, structure, and porosity influence the movement and retention of water, air, and solutes in the soil, which subsequently affect plant growth and organism activity. Chemical soil properties affect nutrient availability and growing conditions (McCauley et al. 2005). All these properties may be altered by management practices that are usually expensive; so it is imperative to determine which zones of the plots need these practices.

The first problem farmers face (see Fig. 6.1) is how to delineate management zones within the plots before planting the crops to improve the overall yield. More precisely, a management zone is a subregion of a plot that is relatively homogeneous with respect to soil parameters, and for which a specific rate of agricultural inputs is needed (Roudier et al. 2008). For this, soil samples are taken and then analyzed. One of the main contributions of this work is a model that takes as input

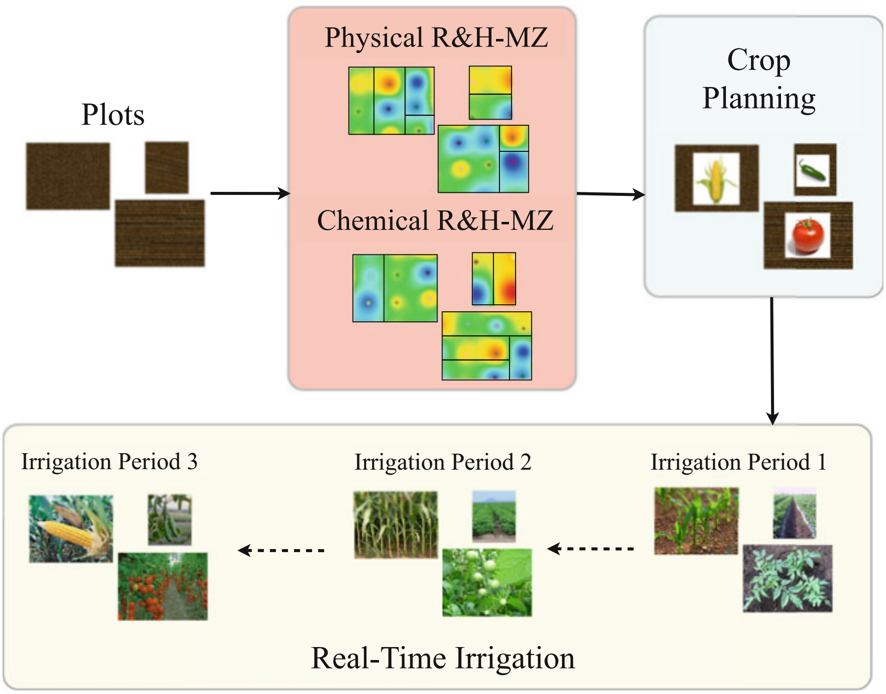


Fig. 6.1 Hierarchical agriculture planning method HAP

the soil samples and delineates the minimum number of rectangular management zones such that the homogeneity within the obtained zones is maximized. Indeed, tiny management zones even if they are rectangular are difficult to operate by agriculture machinery. Therefore, a main issue that is solved in this work is to have the largest management zones that are the most homogeneous possible. We name our methodology as Minimization of *Rectangular and Homogeneous Management Zones (R&H-MZ)*. Two types of management zones are obtained depending on the soil property used during the delineation method: physical and chemical management zones.

The second problem encountered by farmers is to select the crops that they are going to sow into their plots considering the previously delineated management zones. For example, if a plot has several management zones with high amount of phosphorus and nitrogen, then probably it would be better to plant tomatoes or maize because they would save in fertilizers. This problem is known as the *Crop Planning Problem (CPP)*. It becomes quickly a hard problem since there are many parameters to take into account.

Once the CPP problem has been solved and the selected crops are already planted, the decisions the farmers must take are mainly about the optimal amount of water that has to be irrigated to the plots at each irrigation period (see Fig. 6.1). This is another hard problem, named as the *Real-Time Irrigation Problem (RTIP)*, and it considers the phenological stages of the crops, the soil properties of the management zones, the data from the humidity sensors, the evapotranspiration factors, and the previous irrigation decisions. When a drought arises and the total amount of water is not sufficient to irrigate all the crops to optimality, the RTIP decides which crops must be under deficit irrigation or even without irrigation in order to maximize the total final benefits at the end of the production cycle.

In this work we propose an approach named as *Hierarchical Agriculture Planning (HAP)* for helping the decision makers (the farmers) to plan and operate their plots avoiding wastage and maximizing their benefits. The importance of an hierarchical approach resides on the fact that one stage needs as input the results of the previous one.

HAP is composed by a new bi-objective mathematical model that improves the method proposed by Cid-Garcia et al. (2013) to solve R&H-MZ methodology since it considers both the minimum number of management zones and the maximum homogeneity within these zones. One of the contributions of this work is to use the CPP methodology of Cid-Garcia et al. (2014) as a criterion to help the farmer chose between the set of proposals that arise from R&H-MZ. As mentioned, with CPP the farmer obtains an optimal crop pattern. After the crops have been planted on the plots, HAP takes hand of RTIP proposed by Cid-García et al. (2014) to assign them the exact amount of water at each irrigation turn.

As mentioned by Bitran and Hax (1977), to provide effective managerial support for decisions related to production planning, it is useful to partition the set of decisions into a hierarchical framework as we propose in the HAP. Indeed, strategical higher level decisions (management zones) impose constraints on tactic lower level actions (crop planning and real-time irrigation), and lower level decisions provide the necessary feedback to evaluate higher level actions for future production cycles.

The rest of the work is organized as follows. Section 6.1.1 is devoted to related scientific literature. In Sect. 6.2 we present the R&H-MZ methodology. In Sect. 6.3 we present the CPP and how to use it as a discriminator between the set of solutions given by R&H-MZ. In Sect. 6.4 the RTIP is summarized. In Sect. 6.5 we empirically test the R&H-MZ methodology together with the HAP approach on a real instance. Finally, Sect. 6.6 concludes the study.

6.1.1 Literature Review and Terminology

Most of the approaches in literature for determining management zones are based on clustering algorithms. Many of them are based on soil samples information like in our case. For example, Fraisse et al. (2001) and Schepers et al. (2004) use soil and relief information, Carr et al. (1991) base their zoning on topographic maps while methods of Bhatti et al. (1991), Mortensen et al. (2003), or Mulla (1991) need soil sampling. Other approaches are based on yield maps, combining data from several seasons like in Blackmore (2000), Diker et al. (2004), and Pedroso et al. (2010). Some other clustering methods combine soil samples and yield maps: Franzen and Nanna (2003), Hornung et al. (2006), Hornung et al. (2003), and Whelan et al. (2003). In Roudier et al. (2008) they use a watershed segmentation algorithm where the user can introduce morphologies of the desired zones.

Usually, K-means or Fuzzy K-means methods are used for the classification like in Jiang et al. (2011), Li et al. (2005), and Ortega et al. (2002), or principal component analysis with a cluster method (Ortega and Flores 1999). Nevertheless, the choice of the data layers processed by the clustering is an issue as mentioned by Jaynes et al. (2005). Moreover, the resulting fragmentation of the oval-shaped zones due to clustering methods is not an appealing solution for farmers as pointed out by Frogbrook and Oliver (2007), Li et al. (2005), and Simbahan and Dobermann (2006).

Indeed, to the best of our knowledge, Cid-Garcia et al. (2013) are the first to propose a management zone delineation method that directly gives rectangular shape zones. This is important since most of the fertilizing agricultural machinery is towed by tractors. Moreover, most of the irrigation systems are designed in a rectangular pattern. In this work, we improve the results of Cid-Garcia et al. (2013) since instead of minimizing the variance between the fields, we minimize the number of zones. Additionally, we maximize the homogeneity within the management zones. This gives a bi-objective model that offers more practical solutions for the farmers.

In terms of crop planning, Sarker et al. (1997) propose a linear programming model considering land type, alternative crops, crop patterns, input requirement, investment, and output. Nevertheless, they do not take into account that the water is a restriction as we do in this work. Later, Sarker and Ray (2009) formulate the

CPP as a multiobjective optimization model. Mainuddin et al. (1997) propose a crop planning model for an existing groundwater irrigation. Nevertheless, they do not consider the use of humidity sensors as we do in this work. Adeyemo and Otieno (2010) present evolutionary algorithm to solve the multiobjective crop planning model: minimize the total irrigation water, to maximize both the total net income from farming and the total agricultural output. Contrary to our research, water availability is not a restriction.

Ortega Álvarez et al. (2004) propose a nonlinear model solved by genetic algorithms to identify production plans and water irrigation management strategies. They estimate crop yield, production and gross margin as a function of the irrigation depth. In our work, the yields also depend on the irrigation depth, but we manage to have linear restrictions. Moreover, we use the real-time data for the irrigation stage. Sahoo et al. (2006) propose some fuzzy multiobjective linear programming models for land–water–crop system planning. Reddy and Kumar (2008) present a multiobjective approach for the optimal cropping pattern and operation policies for a multi-crop irrigation reservoir system. These authors do not consider water shortages since they try to maximize the yields and to minimize the water.

In Casadesús et al. (2012), Hassanli et al. (2009), Hedley and Yule (2009), and Xu et al. (2011) authors propose schedule irrigation plans according to weather conditions, crop development, and other factors. In this work we propose a mathematical model instead of heuristics. In Alminana et al. (2010) they present models and algorithms to determine water irrigation scheduling by taking into account the irrigation network topology, the water volume, technical conditions, and the logistical operations. Their models do not use the real-time information of humidity sensors like we do in this research.

We use the term field for the whole of land that can be irrigated by a water well or a dam. This field is made up of different plots. In each plot a single crop is to be planted. Each plot is subdivided into physical and chemical management zones (see Fig. 6.2). Notice that all the management zones of a plot must be planted with the same crop.

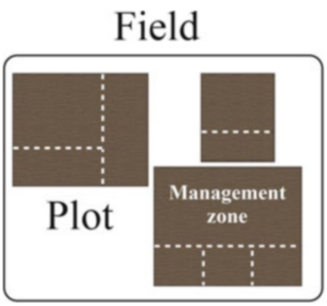


Fig. 6.2 Terms used in this chapter: field, plot, and management zone

6.2 Rectangular and Homogeneous Management Zones

At the begin of the production cycle, two delineations of rectangular and homogeneous management zones are made for each one of the plots. The first one uses chemical soil properties and the second one physical soil properties.

The delineation with chemical soil properties is used to determine the expected amount of nutrients (fertilizers, pesticides, etc.) that the crops require in the whole production cycle, while the delineation with physical soil properties is used to determine the expected amount of water required by the crops in the whole production cycle and the amount of water required by the crops during each irrigation period, respectively.

The R&H-MZ methodology proposed in this work improves the one of Cid-Garcia et al. (2013) since we present a bi-objective problem where the number of rectangular management zones is minimized and the homogeneity within the zones is maximized. R&H-MZ methodology consists of two main stages:

(a) *Instance generation*. In this stage, we process the information from the soil samples that have been taken from the field. These soil samples are approximately equidistant in the field (a GPS detects their position). Then, they are labeled and their positions are translated into the first quadrant of the Cartesian map. Next, the information about each soil property is registered [pH, organic matter rate (OM), amount of phosphorus (P), sand, field capacity, permanent wilting point, etc.]. Soil texture is considered among the most important physical properties and it corresponds to the proportion of three mineral particles (sand, silt, and clay). Then, all the quarters (or potential zones) are computed together with their variances (more details about this stage are given below).

(b) *Mathematical model*. With the input of stage (a), we propose a bi-objective Integer Linear Programming (BILP). The aim of the BILP is to find the minimum set of rectangular management zones such that they cover the whole field and at the same time this set is the one that maximizes the homogeneity within the selected management zones. BILP has a set of optimal solutions that are a trade-off between the two objectives. This set (or Pareto front) is exactly obtained by an ε-constraint algorithm.

We now describe these two stages in more detail. Soil diversity in the field can be observed from a thematic map of a certain property (here we use MapInfo with the default grid and the inverse distance weighting interpolator). In Fig. 6.3 the thematic map of a real plot using phosphorus as chemical soil property and the label of each soil sample (in this case we have 40 soil samples) are presented. In the thematic map we can see the regions at optimal level.

With the thematic map we determine an important parameter which is the smaller management zone allowed, MinWidthQ $\times$ MinLengthQ, where MinWitdthQ is the number of samples in the width of the smaller zone and MinLengthQ is the number of samples in its length. If the diversity of the soil

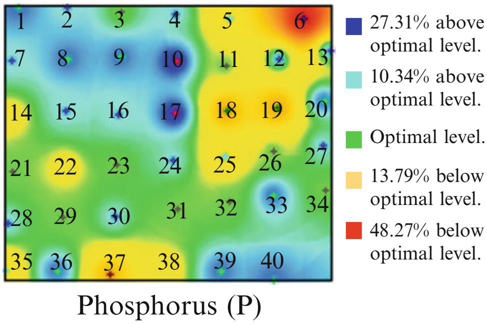


Fig. 6.3 Thematic map for a real plot using phosphorus as chemical soil property

property is high, then the zones should be relatively small (one sample width by one sample length in the worst case).

Once the minimum size of a zone is set, we enumerate all the possible management zones (or quarters) that could be created in this plot. Notice that the soil samples included inside of each potential zone is known. The search of potential zones can be done in Ω(WidthF $\times$LengthF) where WidthF is the number of samples in the width of the field while LengthF is the number of samples in its length. An illustration is given in Fig. 6.4. The left-hand side of this figure shows a plot with nine samples (each one with its label). On the right-hand side, all potential zones are labeled. For this example, we have a total of 36 rectangular quarters (generated by Algorithm 1 presented below). Each quarter shows which samples are included on it, e.g., quarter 1 include only the sample 1 but quarter 30 include the samples 4–9.

The soil samples are almost equidistant, in our example four soil samples (two width for two long) are needed to cover an ha but this number can change according to the farmer's requirements. The total number of potential zones $|Z|$ can be computed by the following formula:

$$|Z| = \left(\sum_{i=1}^{\text{WidthF}-\text{MinWidthQ}+1} i \right) \left(\sum_{j=1}^{\text{LengthF}-\text{MinLengthQ}+1} j \right).$$

The determination of all possible management zones is implemented by Algorithm 1 from Cid-Garcia et al. (2013). The input of this algorithm is the soil samples data, the number of samples in the width of the plot (WidthF), the number of samples in the length of the plot (LengthF), the minimum quantity of samples the width of a quarter must contain (MinWidthQ), and the minimum quantity of samples the length of a quarter must contain (MinLengthQ). The algorithm starts creating the smallest quarters width-wise. Then it checks if there is still some width to cover. After, it checks the length.

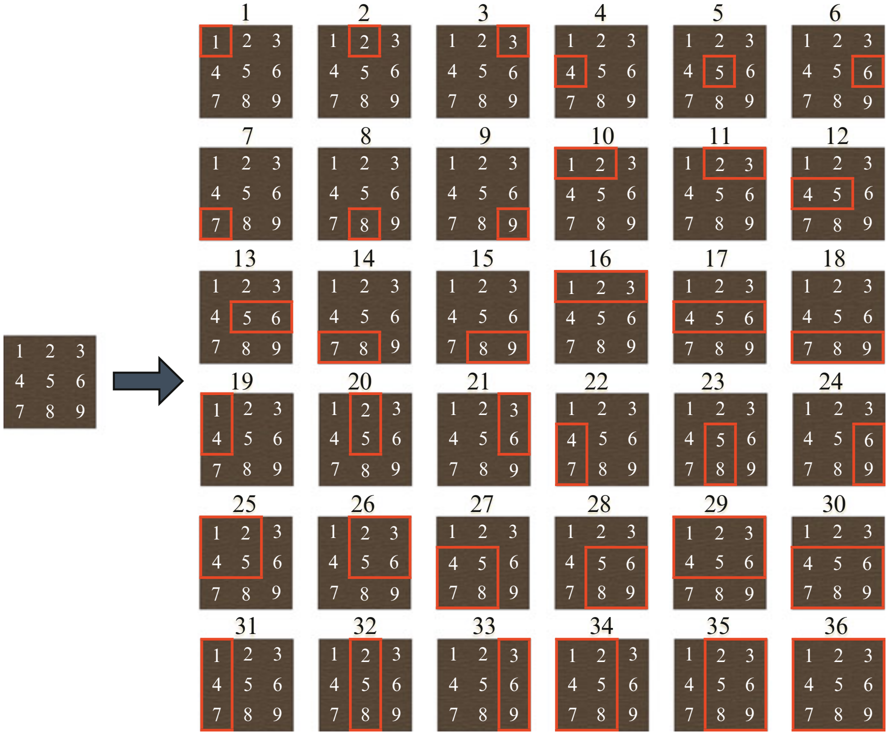


Fig. 6.4 The 36 potential management zones or quarters of a plot

Algorithm 1 Quarters generation of a plot

```
INPUT: WidthF, LengthF, MinWidthQ, MinLengthQ, soil samples
for j = MinWidthQ To WidthF do
    for l = 0 To (WidthF − 1) do
        if (j + l) ≤ WidthF then
            for i = MinLengthQ To LengthF do
                for k = 0 To (LengthF − 1) do
                    if (k + i) ≤ LengthF then
                        creation of a new quarter
                    end if
                end for
            end for
        end if
    end for
end for
```

The result of Algorithm 1 is a correspondence matrix $C = \{c_{zs}\}$, where $c_{zs} = 1$ if quarter with label z covers sample point with label s, $c_{zs} = 0$ otherwise. Once all the potential quarters are determined, we compute the variance of a particular soil property for the set of the samples included in each potential quarter. For example, a quarter that only covers a soil sample would have a variance of 0. A quarter that covers three soil samples would have the variance computed from these three samples. Also, there would be a quarter that covers all soil samples (i.e., there is only one zone that is equal to the plot).

For an example about this instance generation stage, see Sect. 6.5. Next stage is the mathematical model that requires the correspondence matrix of the potential quarters together with their variances. For this purpose, let Z be the set of potential quarters and S the set of soil samples of the field ($|S| = N$). Each quarter z has n_z soil sample points. Farmers do not wish to have tiny management zones because of their machinery; so let LS be the maximum number of zones in the field while LI is the minimum one. The set of decision variables of the BILP model is:

$$q_z = \begin{cases} 1 & \text{if quarter with label} z \text{is chosen,} \\ 0 & \text{otherwise.} \end{cases}$$

The main idea is to cover the plot by a set of non-overlapping quarters. To guarantee a homogeneous zoning delineation we use the relative variance since it has been proved to be a high quality criterion to measure the efficiency of a zoning method (Ortega and Santibáñez 2007; Cid-Garcia et al. 2013). Suppose a set of quarters $Q \subset Z$ is already selected, then the relative variance of Q is $RV(Q) = 1 - \frac{\sum_{z \in Q} \sigma_{w_z}^2}{\sigma_T^2}$, where ${\sigma_T}^2$ is the total variance of all the field and sum of the $\sigma_{w_z}^2$ is the variance within each $z \in Q$ defined as follows:

$$\sum_{z \in Q} \sigma_{w_z}^2 = \frac{\sum_{z \in Q} (n_z - 1)\sigma_z^2}{N - |Q|}. \tag{6.1}$$

Numerator in (6.1) considers the number of samples n_z in quarter z (minus one degree of freedom) as a weight and the denominator takes into account the number of selected quarters (total number N minus the number of quarters $|Q|$). Therefore, we have the following equation where variable $\alpha \in [0, 1]$ must be maximized to have the highest homogeneity within the management zones:

$$\left(1 - \frac{\sum\limits_{z \in Z} (n_z - 1)\sigma_z^2 q_z}{\sigma_T^2 \left[N - \sum\limits_{z \in Z} q_z \right]} \right) \geq \alpha. \tag{6.2}$$

The BILP model to determine the best management zones for a plot is as follows:

$$\left\{\min \sum_{z\in Z} q_z, \max \alpha\right\} \tag{6.3}$$

$$\text{s.t.} \sum_{z\in Z} c_{zs} q_z = 1 \ \forall s \in S \tag{6.4}$$

$$\sum_{z\in Z} q_z \leq LS \tag{6.5}$$

$$\sum_{z\in Z} q_z \geq LI \tag{6.6}$$

$$\left(1 - \frac{\sum_{z\in Z}(n_z - 1)\sigma_z^2 q_z}{\sigma_T^2\left[N - \sum_{z\in Z} q_z\right]}\right) \geq \alpha \tag{6.7}$$

$$\alpha \in [0,1], q_z \in \{0,1\} \ \forall i \in I$$

Bi-objective function (6.3) minimizes the sum of the chosen zones (or potential management zones) and maximizes the homogeneity within each management zones (value of α). Restrictions (6.4) ensure that every point sample s is covered by only one zone, i.e., the whole field is partitioned into non-overlapping zones. Constraints (6.5) and (6.6) limit the number of zones in which the plot will be partitioned. Restriction (6.7), which can be easily linearized, corresponds to the relative variance.

In Sect. 6.5 we prove that the objective functions considered by BILP are conflicting. For the kind of bi-objective problems as BILP, there are many solutions that optimize both objectives. The set of non-dominated solutions (for non-dominated solution there are no other solutions that improve an objective without worsening the other one) represents the trade-off set satisfying both objectives. This trade-off curve is known as the Pareto front and we compute it using the *ε-constraint* method (Marler and Arora 2004; Ehrgott 2005). This *ε-constraint* method optimizes one of the objective functions using the other one as constraint of the model. In our case we have, if we apply the *ε-constraint* method we get the following problem for an α that is fixed and not anymore a decision variable.

$$\begin{aligned} \min \quad & \sum_{z \in Z} q_z \\ \text{s.t.} \quad & (6.4 - 6.6) \\ & \left(1 - \frac{\sum_{z \in Z} (n_z - 1) \sigma_z^2 q_z}{\sigma_T^2 \left[N - \sum_{z \in Z} q_z \right]} \right) \geq \alpha \\ & q_z \in \{0, 1\} \qquad \forall i \in I \end{aligned} \tag{6.8}$$

By using a parametrical variation of the values of α, the efficient solutions of the problem can be obtained. Indeed, in our case α acts as the ε of the method.

Once that we have the Pareto front, the next step is to choose a solution from it that satisfies the farmer's requirements and that guarantees the wished homogeneity within each zone. Experimental results of this bi-objective model in HAP are presented in Sect. 6.5.

6.3 Crop Planning Problem and Selection of the Best Management Zones

We first summarize the CPP problem presented in Cid-Garcia et al. (2014) and then, in Sect. 6.3.2, we show how to use it in order to select the best solution among the Pareto front obtained by R&H-MZ.

6.3.1 Crop Planning Problem

After the chemical and physical management zones have been obtained by R&H-MZ, the second decision in HAP is which crops i to plant in the different plots j by taking into account the soil properties of the physical and chemical management zones that were previously delineated. In this section we present an integer linear programming (ILP) to solve CPP which improves the model presented in Cid-Garcia et al. (2014) since it introduces the chemical and physical management zones of the plots.

Let I be the set of the different crops a farmer could plant, J the set of plots of the farmer's field, $Z^{Ph}(j)$ the set of physical management zones within plot j, while $Z^{Ch}(j)$ the set of chemical management zones of j. Data we use for the ILP mathematical model is described in the following list.

- G_i is the expected benefit of selling a ton (tn) of crop i at the end of the production cycle.
- Cirr_{jz} is the cost of irrigating one cubic meter (m^3) of water in plot j and physical zone $z \in Z^{Ph}(j)$.

- Cseed_i is the cost of buying a kilogram (kg) of seed of crop i.
- Cplant_{ijz} is the cost of planting an hectare (ha) of crop i in plot j, and chemical management zone $z \in Z^{Ch}(j)$.
- ha_j is the number of ha of plot j.
- hac_{jz} corresponds to the number of ha in chemical management zone $z \in Z^{Ch}(j)$ of plot j.
- hap_{jz} corresponds to the number of ha in physical management zone $z \in Z^{Ph}(j)$ of plot j.
- Iseed_i is the quantity of seeds in kg of crop i in the farmer's stock.
- Seed_i is the quantity of seeds in kg needed to plant a ha of crop i.
- Y_i is the expected yield in tn by ha of crop i at the end of the production cycle.
- D_i is the demand in tn of crop $i \in I \subsetneq I$ where is a subset of the crop set I. This is to model situations where a farmer is payed in advance for some yields of a specific crop.
- W is the expected total amount of water in m^3 for all the production cycle.
- W_{ijz} is the expected amount of water in m^3 needed for irrigating a ha planted with crop i in plot j in physical management zone $z \in Z^{Ph}(j)$.

Parameter W_{ijz} can be obtained either by historic data or by deriving it from the Penman–Monteith equation (Allen et al. 2006):

$$ETc_{ij}^{v} = ETo^{v} \cdot Kc_{ij}^{v} \tag{6.9}$$

where ETc_{ij}^{v} is the crop evapotranspiration that represents the amount of water (in mm) required by crop i at phenological stage v for plot j, ETo^{v} is the reference crop evapotranspiration that expresses the evaporating power of the atmosphere (in mm) during phenological stage v. The crop coefficient Kc_{ij}^{v} values change from crop to crop, phenological stage of the crop v, and geographic location j. Then, total expected amount of water consumed by crop i planted in plot j and situated in physical management zone $z \in Z^{Ph}(j)$ throughout the production cycle (W_{ijz}) is the sum of all the expected amount of water required by crop i planted in plot j for each vegetative stage v (ETc_{ij}^{v}) minus the sum of all the stored water in plot j in physical management zone $z \in Z^{Ph}(j)$ at each vegetative stage v (SW_{jz}^{v}):

$$W_{ijz} = \left(10 \sum_{v} ETc_{ij}^{v}\right) - \sum_{v} SW_{jz}^{v}. \tag{6.10}$$

The assignment variables for the CPP integer linear programming model are:

$$\boldsymbol{x_{ij}} = \begin{cases} 1 & \text{if crop } i \text{ is planted in plot } j, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, variables $\mathbf{s}_{\mathrm{i}}$ correspond to the amount of seeds the farmer must buy of crop i in kg, $i \in I$.

$$
\begin{aligned}
\max \sum_{i\in I}\sum_{j\in J} \Bigg[& \boldsymbol{x_{ij}} \cdot G_i \cdot Y_i \cdot \mathrm{ha}_j - \boldsymbol{x_{ij}} \sum_{z\in Z^{Ch}(j)} \mathrm{C\,plant}_{ijz} \cdot \mathrm{hac}_{jz} \\
& - \boldsymbol{x_{ij}} \sum_{z\in Z^{Ph}(j)} \mathrm{Cirr}_{jz} \cdot \mathrm{W}_{ijz} \cdot \mathrm{hap}_{jz} \Bigg] - \sum_{i\in I} s_i \cdot \mathrm{Cseed}_i
\end{aligned} \tag{6.11}
$$

$$
\text{s.t.} \qquad \sum_{i\in I} \boldsymbol{x_{ij}} \le 1 \qquad\qquad j \in J \tag{6.12}
$$

$$
\sum_{j\in J}\sum_{z\in Z^{Ch}(j)} Y_i \cdot \mathrm{hac}_{jz} \cdot \boldsymbol{x_{ij}} \ge D_i \qquad\qquad i \in \bar{I} \tag{6.13}
$$

$$
\sum_{j\in J}\sum_{z\in Z^{Ch}(j)} \mathrm{Seed}_i \cdot \mathrm{hac}_{jz} \cdot \boldsymbol{x_{ij}} \le \mathrm{Iseed}_i + s_i \qquad i \in I \tag{6.14}
$$

$$
\begin{aligned}
& \sum_{i\in I}\sum_{j\in J} \Bigg[\sum_{z\in Z^{Ph}(j)} W_{ijz} \cdot \mathrm{hap}_{jz} \Bigg] \boldsymbol{x_{ij}} \le W \\
& s_i \ge 0, \boldsymbol{x_{ij}} \in \{0,1\} \qquad\qquad i \in I, j \in J
\end{aligned} \tag{6.15}
$$

Objective function (6.11) maximizes the total expected benefits: first term represents the benefits of selling the expected yields of each crop planted in each plot, second one corresponds to the cost of planting the crops in each one of the plots (it includes fertilizers and pesticides that each chemical management zone would require), third term is about the irrigation costs per plot and per physical management zone, finally, we have the cost of buying seeds.

Restrictions (6.12) guarantee unique assignment of a crop i to each plot j. Restrictions (6.13) specify that there are some crops i that must be planted in order to satisfy a certain demand D_i but only for crops in $\bar{I} \subset I$. Restrictions (6.14) determine the amount of seeds needed and also the amount of seeds that must be bought. Restrictions (6.15) establish that the expected amount of water needed to irrigate the crops must be sufficient for the whole production cycle.

Notice that we are supposing that any crop i can be planted on any plot j. In the case where some crops could not be planted in a specific plot (due to soil cycles, or experience), we could easily introduce the notation $J(i)$ corresponding to the set of plots where i can be planted and analogously, $I(j)$ would be the set of crops that can be planted on plot j.

The CPP is a NP-hard problem which has an ILP that is elegant enough to optimally solve real size instances by a branch-and-bound algorithm in less than 1 s as it can be seen in Sect. 6.5.

6.3.2 Selection of the Best Management Zones

After the Pareto front has been obtained by R&H-MZ, the next step is to select the best delineation of chemical and physical management zones in each plot of the

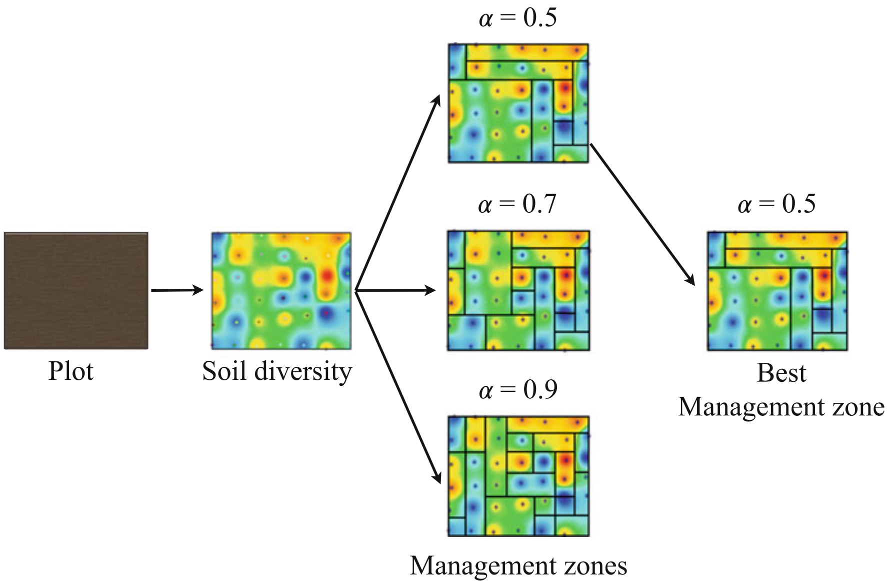


Fig. 6.5 Selection of the best management zone

farmer's field, that is, the delineation that gives the best profit at the end of the production cycle.

It is difficult for the farmer to establish a criterion to chose between a delineation with $\alpha = 0.7$ or $\alpha = 0.5$ (Ehrgott 2005). Therefore, we execute CPP for each solution of the Pareto front of R&H-MZ and selected the management zones that gives the best farmer profit.

Figure 6.5 shows an example about this procedure. We show the delineation resulting of R&H-MZ for a plot using α values of 0.5, 0.7, and 0.9. After executing CPP, we obtain that the best delineation is obtained with alpha value of 0.5. This manner, the farmer does not have trouble to specify which solution is better in the Pareto front because the HAP procedure does it for the farmer.

6.4 Real-Time Irrigation Problem

Suppose that a drought arises once the crops have already been planted in the plots. How to choose the crops that need to be irrigated to optimality, the crops that would be in deficit irrigation, and the crops that would not be irrigated at all (until the point to let a crop die) in order to maximize the benefits at the end of the production cycle?

The production cycle of each crop is divided in different irrigation periods, then, at the beginning of each irrigation period the farmer must decide how much water must be assigned to each plot according to their water requirements in real time, to maintain the maximum crop yield at the end of the production cycle. For this, we use the information of water sensors placed in each one of the physical management zones of the plots. Notice that the aim is to have the maximum possible yields at the end of the production cycle considering that each crop has different vegetative cycles and different needs of water. If there are no water shortages, the crops must be irrigated to optimality. If there are water shortages, then the farmer needs to decide which crops is better to put under deficit irrigation.

The crop water production function (Doorenbos et al. 1986) is given by

$$\frac{Ya_{ij}^{p}}{Ym_{ij}^{p}} = 1 - Ky_{i}^{p}\left(1 - \sum_{z\in Z^{Ph}(J)} \frac{ETa_{ijz}^{p}}{ETc_{ijz}^{p}}\right) \tag{6.16}$$

where the only unknown parameter is Ya_{ij}^{p} that corresponds to the real yields of crop i planted in plot j at period p. Ym_{ij}^{p} is the maximum yield reached by crop i in parcel j at last irrigation period p. When p is the first irrigation period, $Ym_{ij}{}^{p}$ is the harvested yield of crop i under an optimal growing environment, i.e., the yield of the crop is not limited by water, nutrients, pests, nor diseases.

ETc_{ijz}^{p} represents the maximum water requirements of crop i in plot j and physical management zone $z \in Z^{Ph}(j)$ at period p. This is expressed by the sum of all the rates of evapotranspiration in mm per phenological stage v since the last irrigation period minus the sum of all the amount of stored water in plot j in physical management zone $z \in Z^{Ph}(j)$ of each vegetative stage v since the last irrigation period p (Allen et al. 2006). This value can be computed by Eq. (6.9) presented in Sect. 6.3. Based on Eq. (6.10), the amount of water in m^3/ha needed by crop i planted in physical management zone $z \in Z^{Ph}(j)$ of plot j for irrigation period p is

$$ET\,c_{ijz}^{p} = \left(10\sum_{v(p)} ETc_{ij}^{v}\right) - \sum_{v(p)} SW_{jz}^{v}. \tag{6.17}$$

where $v(p)$ represents all the vegetative stages in irrigation period p.

ETa_{ijz}^{p} represents the amount of stored water in the soil of plot j in physical management zone $z \in Z^{Ph}(j)$ at current irrigation period p plus the amount of water supplied at current irrigation period. The water level stored in each one of the physical management zones of the plots, before the current irrigation, is determined by a moisture sensor in real time.

Response factor Ky_{i}^{p} represents the relationship between water and yield for crop i. These values are crop specific and have different values at each irrigation period p.

At the beginning of each irrigation period, the farmer knows the volume of available water. This volume may not be the expected one, therefore some crops are not going to be optimally irrigated. For each irrigation period p a Linear Programming (LP) must be solved to obtain the amount of water to be irrigated in each physical management zone to maximize the expected total benefit. The parameters needed to formulate the LP model for RTIP are listed below:

- G_i^p expected benefit of selling a tn of crop i at the end of the production cycle given that we are at period p. Indeed, this value is known at the beginning of each irrigation period p but it can vary from period to period.
- $\gamma(j) = i$ is a function that indicates that crop i is sown in plot j (this is obtained from the solution of CPP).
- ETc_{ijz}^p corresponds to the real amount of water in m^3/ha that crop i needs in physical management zone $z \in Z^{Ph}(j)$ of plot j at period p. It is easy to compute which vegetative stages v corresponds to crop i at period p, so we omit cumbersome notation.
- hap_{jz} is the number of hectares in physical management zone $z \in Z^{Ph}(j)$ of plot j.
- SW_{jz}^p is the amount of water that already exists in physical management zone $z \in Z^{Ph}(j)$ of plot j at the beginning of irrigation period p. This information is retrieved from the humidity sensors in m^3/ha.
- W^p represents the amount of available water for irrigation period p.
- Ky_i^p is the yield response factor of crop i corresponding at period p.
- $\overline{\mathscr{Y}}_{jz}^{p-1}$ is the maximum crop yield in tn/ha in physical management zone $z \in Z^{Ph}(j)$ of plot j reached at previous irrigation period p.

The variables used in the formulation of the RTIP model are listed below:

- $\boldsymbol{w}_{jz}^{\mathrm{p}}$ is a variable representing the amount in m^3 of irrigated water in physical management zone $z \in Z^{Ph}(j)$ of plot j at period p.
- $\mathscr{Y}_{jz}^p$ represents the current total crop yield in tn/ha reached in physical management zone $z \in Z^{Ph}(j)$ of plot j computed after current irrigation period p.

The LP for RTIP is as follows:

$$
\begin{aligned}
\max \quad & \sum_{i \in I} G_i^p \left(\sum_{\{j \mid \gamma(j)=i\}} \sum_{z \in Z^{Ph}(j)} \mathrm{hap}_{jz} \cdot \mathscr{Y}_{jz}^p \right) \\
\text{s.t.} \quad & \mathscr{Y}_{jz}^p = \overline{\mathscr{Y}}_{jz}^{p-1} \left(1 - Ky_{\gamma(j)}^p \left(1 - \frac{\boldsymbol{w}_{jz}^p + (SW_{jz}^p \cdot \mathrm{hap}_{jz})}{ETc_{ijz}^p \cdot \mathrm{hap}_{jz}} \right) \right) \\
& \qquad i \in I, \{j \mid \gamma(j) = i\}, z \in Z^{Ph}(j)
\end{aligned} \tag{6.18}
$$

$$
\sum_{j \in J} \sum_{z \in Z^{Ph}(j)} \boldsymbol{w}_{jz}^p \leq W^p \tag{6.19}
$$

$$
\begin{aligned}
& \boldsymbol{w}_{jz}^p + (SW_{jz}^p \cdot \mathrm{hap}_{jz}) \leq ETc_{ijz}^p \cdot \mathrm{hap}_{jz} \\
& \qquad i \in I, \{j \mid \gamma(j) = i\}, z \in Z^{Ph}(j)
\end{aligned} \tag{6.20}
$$

$$
\begin{aligned}
& \sum_{\{j \mid \gamma(j)=i\}} \sum_{z \in Z^{Ph}(j)} \mathscr{Y}_{jz}^p \cdot \mathrm{hap}_{jz}^p \geq D_i \qquad \mathrm{i} \in \bar{\mathrm{I}} \\
& \boldsymbol{w}_{jz}^p, \mathscr{Y}_{jz}^p \geq 0 \qquad j \in J, z \in Z^{Ph}(j)
\end{aligned} \tag{6.21}
$$

Objective function maximizes the expected total yield revenues: the price per tn times the total number of ha times the current crop yield (tn/ha), for all crops planted in the different plots of the field.

Restrictions (6.18) correspond to the current yield reached after irrigation period p. It is based on the crop water production function (6.16) where term ${ETa_{ijz}}^{p}$ is equal to $w_{jz}^{p} + (SW_{jz}^{p} \cdot \mathrm{hap}_{jz})$, i.e., the amount of water irrigated at period p plus the already existing water that is indicated by the humidity sensor. By Eq. (6.17), the maximum water requirements are $ETc_{ijz}^{p} \cdot ha_{jz}$. Restriction (6.19) is about the available water the farmer can use for the irrigation of its plots during period p. Restrictions (6.20) indicate that the real amount of water ETa_{ijz}^{p} cannot exceed the maximum (or optimal) amount of water ETc_{ijz}^{p} required by crop i in plot j in physical zone $z \in Z^{Ph}(j)$ at irrigation period p. Restrictions (6.21) determine that current yield reached by the crop $i \in I_0$ at period p must satisfy the demand negotiated beforehand by the farmer.

RTIP is a linear programming that can be solved in an efficient way as we show in Sect. 6.5.

6.5 Experimental Results

In this section we empirically show that the HAP methodology is valid and efficient for a real size instance. For R&H-MZ we use the data from a plot called "Quilaco" presented by Cid-Garcia et al. (2013). CPP and RTIP use crops data from Cid-Garcia et al. (2014) where we use an instance of a field constituted by a set J of 10 plots and a set I of 19 possible crops to be sowed.

The HAP was executed on a Virtual machine with Windows 7 fitted with 1 GB of RAM and a processor Intel Core 2 Duo of 3.06 GHz running on a IMAC equipped with the same processor and 4 GB of RAM. For R&H-MZ and CPP we used the linear integer branch-and-bound algorithm of GAMS/CPLEX 12.2 using default options, except for the optimal criterion fixed at 0. For RTIP we use the linear programming solver of GAMS/CPLEX 12.2 with default parameters. Specific parameters for each stage of the HAP methodology are presented in the following.

6.5.1 Rectangular and Homogeneous Management Zones

In this section two delineations of rectangular and homogeneous management zones are made for each one of the plots. The delineation with chemical soil properties is used in CPP to determine the expected amount of nutrients (fertilizers, pesticides, etc.) that the crops require in the whole production cycle, while the delineation with physical soil properties is used in CPP and RTIP to determine the expected amount of water required by the crops in the whole production cycle and the amount of water required by the crops during each irrigation period, respectively.

The procedure performed in the delineation of management zones is similar for both physical and chemical soil properties. Then, we only present an example for "Quilaco" using as specific chemical soil property the organic matter (OM).

In the instance generation phase of R&H-MZ we create all the possible quarters that could be a management zone in the plot (see Fig. 6.4) using the Algorithm 1 and information about the soil samples of the plot.

Table 6.1 shows the soil samples of "Quilaco." This field has 256 m width and 305.6 m long (around 7.82 ha). There have been taken 40 soil samples that are approximately spaced by 50 m one from each other, so four soil samples are needed to cover an ha. Each soil sample is labeled (first and fourth column of the table) and their positions are translated into a Cartesian map, coordinates (x, y) (second and fifth column of the table). Finally, the information about each chemical soil property is presented: pH, organic matter (OM), phosphorus (P), and sum of bases (SB) determined by the CH3COONH4 method of INIA (2006). A similar table could be presented for the physical soil properties such as field capacity, water holding capacity, and permanent wilting point.

From the thematic map for "Quilaco" of the OM property, we determine that the minimum size of a quarter contains a single sample width (MinWidthQ = 1) per one sample of length (MinLengthQ = 1) since there is a lot of diversity.

Afterwards, the 588 possible quarters are generated and labeled by Algorithm 1. Then, Table 6.2 allows to see the structure of the correspondence matrix of "Quilaco" for organic matter, except by the last column that corresponds to the variance of the different soil samples that are contained in quarter with label z.

Most of the fields are not initially rectangular, so the R&H-MZ method inserts dummy soil samples to fill a rectangle where the field can be contained. This is the reason why Table 6.2 is composed of 42 samples. The dummy samples are also equidistant with respect to the others. Nevertheless, their data about the properties is very high with respect to the real samples. This manner, the mathematical model puts these dummy soil samples alone in a zone or with other dummy samples which facilitates their elimination afterwards.

Table 6.3 presents the experimental results of ε *-constraint* method applied to the R&H-MZ for the "Quilaco" instance. First column is the alpha parameter (ε value) that determines the homogeneity level in each selected quarter. The higher the α, the more homogeneous the management zones. Second column is the number of quarters (zones) used to partitioning the plot (we want to minimize the number of management zones). Last column is the solution time in seconds required by the solver to obtain the optimal solution computed by the branch-and-bound algorithm of GAMS/CPLEX 12.2.

With Table 6.3 we empirically prove that minimizing the number of management zones and maximizing the homogeneity within each zone are conflicting objectives. We can also notice that the computing times are negligible. This implies that we can compute the exact Pareto front in an efficient way which is a remarkable characteristic.

Figure 6.6 shows the exact Pareto front for "Quilaco" using organic matter as chemical soil property. The x-axis represents the value of alpha and the y-axis

Table 6.1 Coordinates and values of the soil properties for each sample of "Quilaco"

Sample	Coordinates	Soil properties				Sample	Coordinates	Soil properties			
	(x, y)	pH	OM	P	SB		(x, y)	pH	OM	P	SB
1	0.00, 9.14	5.2	11.8	8.0	5.89	21	297.68, 166.36	5.6	10.4	4.0	8.26
2	48.97, 8.46	5.5	12.8	4.0	7.97	22	253.87, 160.20	5.4	18.7	11.0	8.88
3	97.52, 5.57	5.2	14.9	10.0	7.63	23	206.99, 157.26	5.6	10.5	11.0	6.03
4	150.52, 9.42	5.4	14.0	7.0	11.44	24	158.29, 155.16	5.5	16.8	3.0	9.48
5	201.07, 8.25	5.5	11.2	4.0	6.36	25	105.27, 153.53	5.4	14.8	5.0	7.85
6	250.24, 0.00	5.4	14.7	4.0	9.31	26	56.47, 156.87	5.5	12.6	5.0	5.38
7	298.57, 84.00	5.6	12.5	6.0	10.03	27	6.15, 151.48	5.4	15.1	7.0	6.50
8	249.94, 78.89	5.6	9.6	4.0	7.99	28	6.33, 204.03	5.4	11.7	5.0	5.88
9	208.71, 73.33	5.5	14.3	6.0	8.20	29	58.83, 205.57	5.5	16.0	4.0	8.09
10	160.73, 66.20	5.5	15.0	6.0	9.23	30	108.59, 207.64	5.4	13.8	4.0	8.18
11	102.69, 59.51	5.4	14.5	5.0	6.64	31	159.65, 203.22	5.6	12.6	3.0	7.95
12	53.66, 58.30	5.4	11.1	6.0	6.00	32	206.04, 199.18	5.4	14.4	6.0	7.50
13	2.81, 52.71	5.3	14.1	5.0	5.67	33	255.23, 205.16	5.4	15.4	5.0	8.23
14	6.93, 101.13	5.3	16.3	6.0	5.51	34	303.14, 212.73	5.7	11.2	5.0	9.51
15	58.25, 105.04	5.4	12.7	7.0	6.36	35	278.06, 242.75	5.2	16.6	22.0	7.30
16	104.05, 107.24	5.4	14.2	6.0	7.80	36	208.60, 243.31	5.5	15.6	8.0	9.21
17	156.53, 111.44	5.5	11.4	5.0	6.72	37	158.68, 247.47	5.5	16.1	5.0	9.51
18	204.49, 114.91	5.5	11.5	8.0	6.11	38	108.00, 249.65	5.4	13.9	6.0	6.90
19	250.37, 119.77	5.4	16.7	6.0	8.75	39	58.16, 253.69	5.5	15.4	5.0	9.69
20	296.17, 124.74	5.5	13.5	5.0	7.81	40	12.72, 254.37	5.4	10.7	4.0	7.71

OM is in %, P in $\mathrm{mg\,kg^{-1}}$, and SB in $\mathrm{Cmol(+)\,kg^{-1}}$

Table 6.2 Correspondence matrix of "Quilaco" for organic matter soil property with the variance of each quarter z

		Sample point s																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…	42	σ_z^2
Potential quarter z	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	…	0	0.00
	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	…	0	0.00
	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	…	0	0.00
	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	…	0	0.00
	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	…	0	0.00
	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	…	0	0.00
	7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	…	0	0.00
	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	…	0	11.04
	9	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	…	0	1.12
	10	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	…	0	2.42
	11	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	…	0	0.12
	12	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	…	0	0.50
	13	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	…	0	0.00
	14	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	…	0	5.76
	15	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	…	0	4.61
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮	⋮
	588	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	…	1	2.07

Table 6.3 Experimental results for R&H-MZ

α	Quarters	Time
0	1	0.296
0.1	3	0.227
0.2	5	0.237
0.3	6	0.291
0.4	7	0.229
0.5	9	0.211
0.6	11	0.241
0.7	14	0.251
0.8	17	0.241
0.9	20	0.231
1	40	0.225

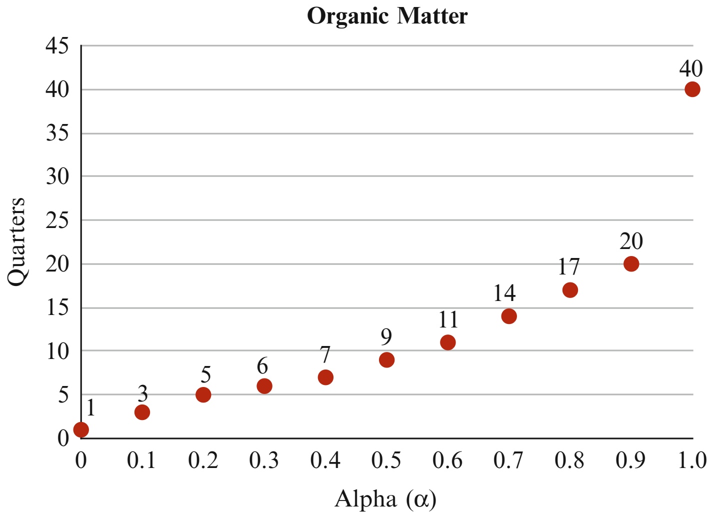


Fig. 6.6 Pareto front for "Quilaco" using organic matter as chemical soil property

shows the optimal number of zones obtained for partitioning the plot. Here we partition the α rank [0, 1] in subintervals of 0.1. We could easily do a more dense partition.

Once that we have the Pareto front, the next step is to choose the solution from this front that satisfies the farmer's requirements and guarantees homogeneity in each selected quarter. In this case, only solutions with α greater or equal than 0.5 guarantee homogeneity in the selected zones (this value is given by an agricultural expert). In Fig. 6.7 is presented the chemical management zones resulting after partitioning the field "Quilaco" using organic matter as soil property and α values of 0.5, 0.7, and 0.9 (left to right maps). We can observe that if α increases, then the number of quarters increases too.

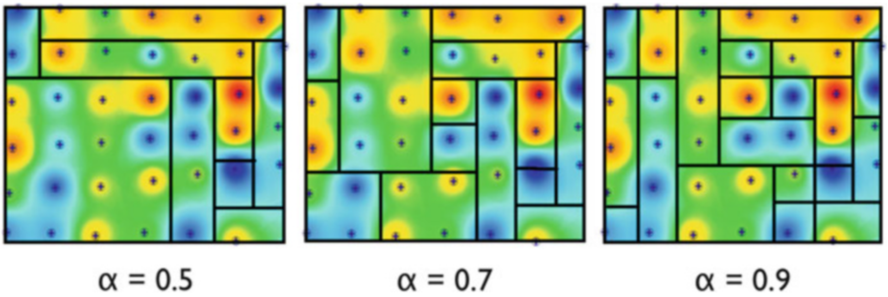


Fig. 6.7 Management zones for "Quilaco" using organic matter as chemical soil property and alpha values of 0.5, 0.7, and 0.9

With R&H-MZ method we have the chemical and physical management zones of all the plots of the field. Then, with this information, HAP decides which crops to plant in each one of the plots.

6.5.2 Crop Planning Problem

In this section we first selected the best delineation of chemical and physical management zones, in each plot of the farmer's field, executing CPP model in each solution of the Pareto front given by R&H-MZ (see Fig. 6.6). Once we have the best chemical and physical management zones, CPP uses this delineation to generate the optimal crop pattern that maximizes the farmer's profit at the end of the production cycle. In this stage, we use the same crops than Cid-Garcia et al. (2014) and 10 plots with their respective best physical and chemical management zones obtained by R&H-MZ.

The total number of ha is 81 and the total expected amount of available water for the whole production cycle is $486,000\,\mathrm{m}^3$ (we assume that we have the maximum limit of water per ha, established by CONAGUA[1]: $6,000\,\mathrm{m}^3$ per ha.) The irrigation costs $Cirr_{jz}$ for each one of the plots j is chosen at random between 1.8 and 2.2 per m^3 (these parameters are based on real data).

General information of plots and the number of their physical and chemical management zones are shown in Table 6.4. First column is the label of the plot, second and third columns are the number of physical and chemical management zones in each plot, and last column is the total number of ha in each plot.

In Table 6.5 the crops used in the instance are presented. They correspond to the crops that can be sown in Michoacán, Mexico, during the production cycle of spring–summer (data of 2008 from SAGARPA[2]). First and second columns are

[1] The Mexican national water commission.

[2] Mexican ministry of agriculture, livestock, rural development, fisheries, and food.

Table 6.4 General information of plots

Plot	$Z^{Ph}(j)$	$Z^{Ch}(j)$	ha_j
1	2	4	10
2	4	3	17
3	2	2	4
4	4	1	7
5	1	2	3
6	2	4	9
7	1	2	6
8	2	3	11
9	2	1	4
10	2	4	10

Table 6.5 Crop data from spring–summer cycle in the state of Michoacán, Mexico

ID	Crop i	Expected yield Y_i	Seed amount Seed_i	Sowing cost Cplant_{ijz}	Seed cost Cseed_i	Expected benefit G_i
1	Sesame TCS	0.60	4	2,318.00	10.00	13,681.80
2	Sesame TMF	0.50	4	8,117.91	10.00	13,681.80
3	Onion BMF	40.60	12,500	69,251.08	0.15	3,381.17
4	Green pepper BMF	24.70	12,500	106,121.91	0.15	4,923.99
5	Strawberry BMF	20.40	85,228	74,543.27	0.11	3,943.97
6	Strawberry GMF	20.40	85,228	48,533.07	0.11	3,493.97
7	Corn grain BCF	4.85	25	10,273.02	17.10	4,373.49
8	Corn grain BMF	5.38	25	10,013.49	17.10	4,373.49
9	Corn grain GCF	4.85	25	9,693.02	17.10	4,373.49
10	Corn grain GMF	5.38	25	10,668.41	17.10	4,373.49
11	Corn grain TCF	2.41	25	10,512.40	17.10	4,373.49
12	Sorghum grain BMF	8.31	324	12,022.65	1.50	3,491.25
13	Sorghum grain GMF	8.31	324	7,891.88	1.50	3,491.25
14	Sorghum grain TMF	4.71	324	6,674.43	1.50	3,491.25
15	Red tomato BMF	38.10	12,500	75,259.74	0.56	2,171.99
16	Red tomato GMF	38.10	12,500	74,440.68	0.56	2,171.99
17	Green tomato BCF	16.20	12,300	50,574.63	0.15	3,416.17
18	Green tomato BMF	17.80	12,300	41,867.90	0.15	3,416.17
19	Green tomato TCF	2.70	12,300	34,056.14	0.15	3,416.17

The 19 crops and their related information were obtained from SAGARPA for the year 2008

the identification number (ID) and name of the crop i. Third column shows the expected yield Y_i of the crop i in tn/ha at the end of the production cycle. Fourth column is the amount of seeds Seed_i in unit/ha needed to sown crop i. The term unit represents kg, plants or packages. Sowing costs Cplant_{ijz} in plot j within chemical management zone $z \in Z^{Ch}(j)$ in $/ha and seed costs Cseed_i in $/unit of crop i are presented in the fifth and sixth columns, respectively. The expected benefit G_i of selling a tn of crop i at the end of the production cycle is shown in the last column.

We assume that the demands that should be satisfied by the farmer of onion BMF, green pepper BMF, corn grain BCF, and red tomato GMF (crops 1, 4, 7, and 16) are all equal to 30. Finally, the stock of seeds Iseed_i for all crops is equal to zero.

When HAP has previous information about the sowing costs Cplant_{ijz} of each crop i, then this cost is specific for each chemical zone $z \in Z^{Ch}(j)$ of each plot j. In this research we take the same sowing costs of crops i for all the chemical zones $z \in Z^{Ch}(j)$ of all the plots j.

To calculate the total expected amount of water supplied W_{ijz} to crop i in each plot j in each physical management zone $z \in Z^{Ph}(j)$ during the whole production cycle, we use Eqs. (6.9) and (6.10). The parameters to compute these equations were obtained from FAO and INIFAP.[3] Crop coefficient values Kc_{ij}^{v} of crop i in plot j for phenological stage v at irrigation period p and the duration of the vegetative cycle of the crops were collected from FAO (Allen et al. 2006). Values for crop reference evapotranspiration ETo^{v} at phenological stage v and for the amount of water stored SW_{jz}^{v} in plot j in physical management zone $z \in Z^{Ph}(j)$ at phenological stage v were obtained from INIFAP located in Zacatecas, Mexico. These values correspond to averages of previous years.

Table 6.6 shows the total expected amount of water W_{ijz} needed by crop i in each physical management zone $z \in Z^{Ph}(j)$ of plot j after computing Eqs. (6.9) and (6.10) (data of ETo^{v} and SW_{jz}^{v} are averages of the last 5 years). In this table we only present the first ten crops of Table 6.5 related to two plots with four physical management zones each one. First column is the plot j and second column is the physical management zone $z \in Z^{Ph}(j)$. The surface hap_{jz} in ha of each physical management zone $z \in Z^{Ph}(j)$ is shown in the third column. The total expected amount of water W_{ijz} in needed by crop i during its production cycle in the plot j in each physical management zone $z \in Z^{Ph}(j)$ is presented in the rest of the columns. The instances are generated such that each crop i consumes the same amount of water in the physical management zones $z \in Z^{Ph}(j)$ with the same ID. For example, crop number 3 (onion) consumes $3,418\,\text{m}^3$ in the physical management zone 1 of plots 1 and 2, while the same crop consumes $4,081\,\text{m}^3$ in the physical management zone 4 of plots 1 and 2.

The result of CPP for this particular instance is an expected income of \$877,690.90 at the end of the production cycle. The crops that should be sown in each plot are presented in Table 6.7. First column indicates the plot while second column the crop sown on it. Notice that all the zones of each plot are planted with the same crop, but the decision about which crop to plant strongly depends on the chemical and physical characteristics of the zones of the plots.

Recall that CPP is used as a method to chose between the different management zones delineations proposed by R&H-MZ. Here we have shown the CPP with the best management zones, that is, the one that gives the best profits in CPP. With this result, we now consider the operational plan of the irrigation decisions.

[3] The Mexican national institute for forestry, agriculture, and livestock. There is a research center INIFAP at every state, and therefore producers can get specific information depending on the geographic location of their fields.

Table 6.6 Total expected amount of water in m^3 supplied by the first 10 crops in each management zone of two plots

Plot j	Zone z	hap_{jz}	Crop (ID) 1	2	3	4	5	6	7	8	9	10
1	1	4	3,845	3,845	3,418	7,372	7,050	7,050	7,418	7,418	7,418	7,418
1	2	5	3,819	3,819	3,871	7,448	6,710	6,710	8,100	8,100	8,100	8,100
1	3	3	2,389	2,389	3,664	5,190	5,232	5,232	5,200	5,200	5,200	5,200
1	4	3	4,220	4,220	4,081	8,136	7,597	7,597	8,442	8,442	8,442	8,442
2	1	1	3,845	3,845	3,418	7,372	7,050	7,050	7,418	7,418	7,418	7,418
2	2	6	3,819	3,819	3,871	7,448	6,710	6,710	8,100	8,100	8,100	8,100
2	3	6	2,389	2,389	3,664	5,190	5,232	5,232	5,200	5,200	5,200	5,200
2	4	5	4,220	4,220	4,081	8,136	7,597	7,597	8,442	8,442	8,442	8,442

Table 6.7 Results of CPP

Plot	Crop
1	Sesame TCS
2	Onion TMF
3	Sesame TCS
4	Corn grain BCF
5	Red tomato GMF
6	Sesame TCS
7	Sesame TCS
8	Sesame TCS
9	Green pepper BMF
10	Sesame TCS

This output is used as input for RTIP

6.5.3 *Real-Time Irrigation Problem*

In this stage it is already known which crops i have been sown in each one of plots j (Table 6.7). The RTIP step of HAP decides the amount of water to be supplied on each plot j during each irrigation period p to maintain the yields as high as possible at the end of the production cycle.

Table 6.8 presents the parameters needed for RTIP at irrigation period $p = 1$. First and second columns correspond to the plot j and the physical management zone $z \in Z^{Ph}(j)$. Third column is the yield response factor ${Ky_{\gamma(j)}}^p$ of crop i sown in plot j at irrigation period p (plots with the same crops and planted at the same time have the same factor), this parameter is obtained from FAO. Fourth column is the amount of stored water ${SW_{jz}}^p$ in m^3/ha in each physical management zone $z \in Z^{Ph}(j)$ of plot j before irrigating at period p. This information is given by the humidity sensors (in this research we used WATERMARK 200SS-V sensors). Fifth column is the amount of water ${ETc_{ijz}}^p$ in m^3/ha needed by crop i in each physical management zone $z \in Z^{Ph}(j)$ of plot j at current irrigation period p. This data is calculated with Eqs. (6.9) and (6.10) using information from INIFAP. Last column is the maximum yield $\overline{\mathcal{Y}}_{jz}^{p-1}$ in tn/ha reached in each physical management zone $z \in Z^{Ph}(j)$ of plot j in the previous irrigation period $p - 1$. When p is equal to zero, $\overline{\mathcal{Y}}_{jz}^{p-1}$ takes the value of the expected harvest yield of crop under an optimal growing environment (this value is given by INIFAP or FAO).

We consider six irrigation periods for our experimental instance. Table 6.9 shows the experimental results for RTIP at period $p = 1$ with available water of $81,000\,\mathrm{m}^3$ which corresponds to a sixth of the total expected amount of water by the production cycle ($486,000\,\mathrm{m}^3$). First and second columns represent the plot and the physical management zone. Third column is the amount of water supplied to the crop in m^3 at current irrigation period. Fourth column indicates whether the crop was irrigated at optimal level or not. Finally, last column is the current expected

Table 6.8 Parameters of RTIP at irrigation period 1

Plot j	Zone z	Yield response factor $Ky^p_{\gamma(j)}$	Stored water SW^p_{jz}	Required water ETc^p_{ijz}	Maximum yield $\overline{\mathcal{Y}}^{p-1}_{jz}$
1	1	0.3	100	393.4	0.6
1	2	0.3	100	424.9	0.6
2	1	0.45	100	1,040.2	40.6
2	2	0.45	100	1,123.5	40.6
2	3	0.45	100	1,010.8	40.6
2	4	0.45	100	1,250.2	40.6
3	1	0.3	100	393.4	0.6
3	2	0.3	100	424.9	0.6
4	1	0.4	100	786.8	4.85
4	2	0.4	100	849.8	4.85
4	3	0.4	100	884.1	4.85
4	4	0.4	100	972.5	4.85
5	1	0.4	100	1,219.2	38.1
6	1	0.3	100	393.4	0.6
6	2	0.3	100	424.9	0.6
7	1	0.3	100	393.4	0.6
8	1	0.3	100	393.4	0.6
8	2	0.3	100	424.9	0.6
9	1	1.1	100	706.4	24.7
9	2	1.1	100	789.9	24.7
10	1	0.3	100	393.4	0.6
10	2	0.3	100	424.9	0.6

yield in tn/ha reached after irrigating the plot. Optimal solutions were computed in less than 1 s. At this period, the total amount of water is enough for irrigating all the crops at optimal level. The total amount of water supplied to irrigate all the crops is only $43,268.6\,\mathrm{m}^3$ which corresponds to 53.4 % of the total available water of the period. Therefore, savings on water are made.

Table 6.10 presents the experimental results of the RTIP throughout the whole production cycle. First column indicates the plot j and the second column the physical management zone $z \in Z^{Ph}(j)$. Third column is the expected harvest yield of crop i planted in plot j under a growing environment, this is the parameter of maximum yield by crop i of plot j only for the first period ($Ym_{\gamma(j)}{}^1$). Fourth column indicates if the irrigation level at period 1 (IL^1) was optimal or not ("–" means that the crop was not irrigated to optimal level). Fifth column shows the current yield of crop i of plot j ($Ya_{\gamma(j)}{}^1$) reached after irrigation at period 1, $Ya_{\gamma(j)}{}^1$ is the parameter $Ym_{\gamma(j)}{}^2$ for the second period. Columns 6 and 7 are the same as above but for period 2, and so on until period 6.

At periods 1 and 2 the crops are in their initial phenological stages, so they do not consume too much water. All the crops are irrigated at optimal level and reach their

Table 6.9 Experimental results of RTIP at irrigation period 1

Plot j	Zone z	Supplied water w_{jz}^{p}	Irrigation level	Current yield $\mathscr{Y}_{jz}^{p}$
1	1	880.20	Optimal	0.60
1	2	2,274.30	Optimal	0.60
2	1	3,760.80	Optimal	40.60
2	2	40,940	Optimal	40.60
2	3	6,375.60	Optimal	40.60
2	4	2,300.40	Optimal	40.60
3	1	586.80	Optimal	0.60
3	2	649.80	Optimal	0.60
4	1	686.80	Optimal	4.85
4	2	749.80	Optimal	4.85
4	3	3,136.40	Optimal	4.85
4	4	872.50	Optimal	4.85
5	1	3,357.60	Optimal	38.10
6	1	1,760.40	Optimal	0.60
6	2	974.70	Optimal	0.60
7	1	1,760.40	Optimal	0.60
8	1	14,670	Optimal	0.60
8	2	1,949.40	Optimal	0.60
9	1	1,819.20	Optimal	24.70
9	2	689.90	Optimal	24.70
10	1	1,173.60	Optimal	0.60
10	2	1,949.00	Optimal	0.60

maximum expected yield at the end of these periods. At period 3 there is no enough water to irrigate all crops to optimum level, and the current expected yield of plot 2 in zones 2, 3, and 4, together with the current yield of plot 4 in zone 4, decrease considerably with respect to the maximum yield.

Since there is no enough water to irrigate the crops to optimal level on their flowering and yield formation stages, the current expected yield of plot 2 in zones 2, 3, and 4 decrease again at period 4 with respect to maximum yield of period 3. At period 5, the current yield of plot 2 in zones 2, 3, and 4, the current yield of plot 5 in zone 1, and the current yield of plot 9 in zones 1 and 2 decrease considerably. At period 6 the crops are in their final phenological stage, so they do not consume too much water (as the initial stages) and all of them are irrigated again to optimal level.

The model must comply with the established demand in the CPP (sesame, green pepper, corn grain, and red tomato), so these crops have priority over the others. The model can let die crops that do not have a fixed demand even if those crops would generate more profit for the farmer. Table 6.11 shows the final yield reached by each crop after each irrigation period. It is verified that the demand established in the CPP is satisfied for each crop at the end of the production cycle (last irrigation period).

Table 6.10 Experimental results of RTIP for the six irrigation periods of the production cycle

			p1		p2		p3		p4		p5		p6	
Plot j	Zone z	$Ym_{\gamma(j)}{}^{1}$	IL^1	$Ya_{\gamma(j)}{}^{1}$	IL^2	$Ya_{\gamma(j)}{}^{2}$	IL^3	$Ya_{\gamma(j)}{}^{3}$	IL^4	$Ya_{\gamma(j)}{}^{4}$	IL^5	$Ya_{\gamma(j)}{}^{5}$	IL^6	$Ya_{\gamma(j)}{}^{6}$
1	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
1	2	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
2	1	40.6	Opt	40.6	Opt	40.6	Opt	40.6	Opt	40.6	Opt	40.6	Opt	40.6
2	2	40.6	Opt	40.6	Opt	40.6	–	25.3	–	14.21	–	9.86	Opt	9.86
2	3	40.6	Opt	40.6	Opt	40.6	–	22.33	–	12.28	–	10.23	Opt	10.23
2	4	40.6	Opt	40.6	Opt	40.6	–	22.33	–	12.28	–	9.08	Opt	9.08
3	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
3	2	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
4	1	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85
4	2	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85
4	3	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85	Opt	4.85
4	4	4.85	Opt	4.85	Opt	4.85	–	0.9	Opt	0.9	Opt	0.9	Opt	0.9
5	1	38.1	Opt	38.1	Opt	38.1	Opt	38.1	Opt	38.1	–	13.88	Opt	13.88
6	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
6	2	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
7	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
8	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
8	2	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
9	1	24.7	Opt	24.7	Opt	24.7	Opt	24.7	Opt	24.7	–	6.512	Opt	6.512
9	2	24.7	Opt	24.7	Opt	24.7	Opt	24.7	Opt	24.7	–	16.854	Opt	16.854
10	1	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6
10	2	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6	Opt	0.6

Table 6.11 Final total yield reached by the crops after each irrigation period

	Final yield (tn)				
Irrigation period	Sesame TCS	Green pepper BMF	Corn grain BCF	Red tomato GMF	Onion TMF
1	30	98.80	33.95	114.30	690.20
2	30	98.80	33.95	114.30	690.20
3	30	98.80	30.00	114.30	464.60
4	30	98.80	30.00	114.30	329.80
5	30	36.39	30.00	41.65	291.52
6	30	36.39	30.00	41.65	291.52

Table 6.12 Final expected profit reached by the farmer at each irrigation period

Period	AW (m^3)	IW (m^3)	IW (%)	Profit ($)
1	81,000	43,269	53.4	3,627,366.19
2	81,000	43,269	53.4	3,627,366.19
3	81,000	81,000	100	2,847,269.31
4	81,000	81,000	100	2,391,531.03
5	81,000	81,000	100	1,796,976.97
6	81,000	49,449	61.1	1,796,976.97

Finally, in Table 6.12 the expected profit achieved by the farmer at each irrigation period after watering the crops is presented (notice that the sowing costs are not considered here). First column indicates the irrigation period. Second column is the amount of available water in m^3 for irrigating crops (AW), and third column is the real amount of irrigated water in m^3 on crops (IW). Fourth column is the percentage (%) of irrigated water (IW) with respect to the total available, and the last column is the expected profit in $ achieved in the period. In the first two periods the crops were irrigated at optimal level; therefore the farmer's expected profit remained at 100 %. However, in periods 3, 4, and 5, there is a greater need of water with respect to the total amount of available water in each irrigation period. Water needed by the crops was not 100 % satisfied causing a decrease of 50.46 % in the farmer's profit that would never be recovered despite that in the period 6 all crops were irrigated at 100 % (see Fig. 6.8). So, at the end of the production cycle the farmer's profit is only 49.54 % with respect to the total expected profit at the beginning of the production cycle.

In Fig. 6.9 the yield reached in each physical management zone $z \in Z^{Ph}(j)$ of plot j after each irrigation period p is shown (periods 1, 3, 4, and 6).

RTIP guarantees to supply only the amount of water needed to satisfy the water requirements of crops and avoid wastage. Thus, water can be stored to future irrigation periods. Moreover, the farmer now has a decision tool that is relevant when water shortages arise.

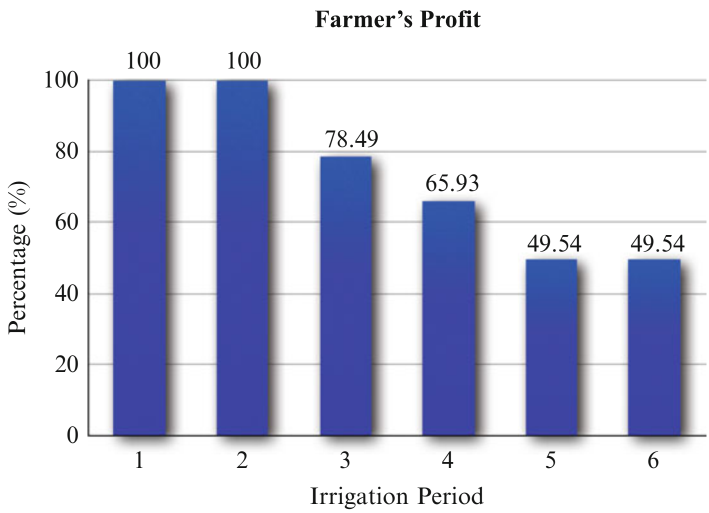


Fig. 6.8 Percentage of farmer's profit after each irrigation period

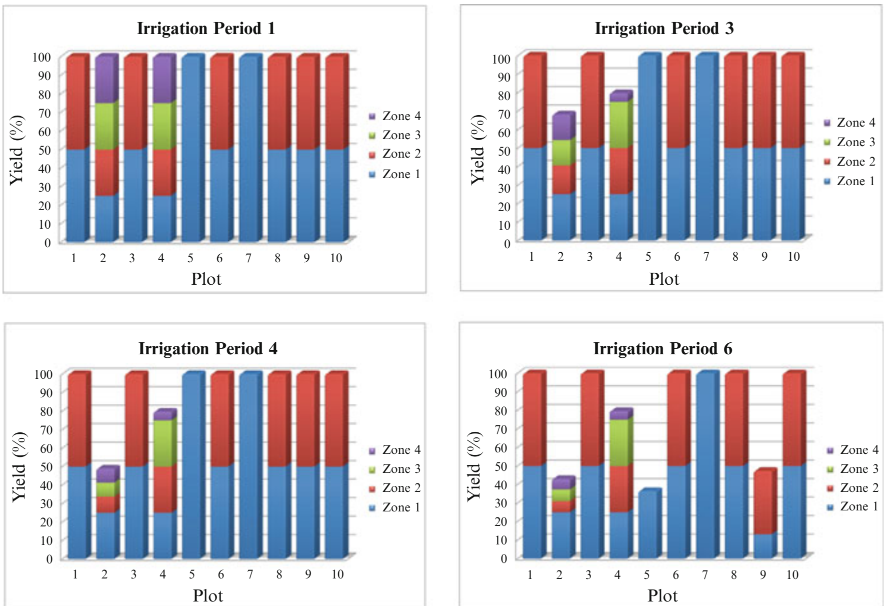


Fig. 6.9 Yield reached in each plot after each irrigation period (periods 1, 3, 4, and 6)

6.6 Conclusions

Physical and chemical soil properties existing in agricultural production plots are an important characteristic that should be considered in the agricultural planning process. Chemical soil properties affect the application of inputs (fertilizers, pesticides, etc.), while physical soil properties are related to water use.

In this work we propose a new approach named as Hierarchical Agriculture Planning (HAP) for helping the decision makers (the farmers) to plan and operate their plots in order to avoid wastage and to maximize their benefits considering the soil diversity. In this hierarchical approach the farmers start by delineating the field into rectangular and homogeneous site-specific chemical and physical management zones to face within-field variability. Then the farmers assign a crop to the different plots to obtain the best profit at the end of the production cycle (CPP). Finally, in each irrigation period the farmer must decide how much and which plots must be watered to maximize the total final yields (RTIP).

Experimental results show that the new hierarchical approach is efficient and practical since optimal solutions are obtained in seconds.

Acknowledgements This study was partially supported by CONACYT (Grant 101857), by DGIP (Grant USM 28.13.69), CATA and CIDIEN of the Universidad Técnica Federico Santa María. Nestor M. Cid-Garcia wishes to acknowledge graduate scholarship from CONACYT. We are grateful to INIFAP for much of the data that we used in this study.

References

Adeyemo J, Otieno F (2010) Differential evolution algorithm for solving multi-objective crop planning model. Agric Water Manag 97(6):848–856

Allen RG, Pereira LS, Raes D, Smith M (2006) Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos, vol 56. Food & Agriculture Organization, Rome

Alminana M, Escudero LF, Landete M, Monge JF, Rabasa A, Sánchez-Soriano J (2010) WISCHE: A DSS for water irrigation scheduling. Omega 38(6):492–500

Bhatti AU, Mulla DJ, Frazier BE (1991) Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. Remote Sens Environ 37(3):181–191

Bitran GR, Hax AC (1977) On the design of hierarchical production planning systems. Decis Sci 8(1):28–55

Blackmore S (2000) The interpretation of trends from multiple yield maps. Comput Electron Agric 26(1):37–51

Carr PM, Carlson GR, Jacobsen JS, Nielsen GA, Skogley EO (1991) Farming soils, not fields: a strategy for increasing fertilizer profitability. J Prod Agric 4(1):57–61

Casadesús J, Mata M, Marsal J, Girona J (2012) A general algorithm for automated scheduling of drip irrigation in tree crops. Comput Electron Agric 83:11–20

Cid-Garcia NM, Albornoz V, Rios-Solis YA, Ortega R (2013) Rectangular shape management zone delineation using integer linear programming. Comput Electron Agric 93:1–9

Cid-Garcia NM, Bravo-Lozano AG, Rios-Solis YA (2014) A crop planning and real-time irrigation method based on site-specific management zones and linear programming. Comput Electron Agri 107:20–28

Diker K, Heermann DF, Brodahl MK (2004) Frequency analysis of yield for delineating yield response zones. Precis Agric 5(5):435–444

Doorenbos J, Kassam AH, Bentvelsen CLM (1986) Efectos del agua sobre el rendimiento de los cultivos, Vol 33. FAO, Roma

Ehrgott M (2005) Multicriteria optimization, vol 2. Springer, Berlin

Fraisse CW, Sudduth KA, Kitchen NR (2001) Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Trans ASAE 44(1):155–166

Franzen DW, Nanna TN (2003) Management zone delineation methods. In: Robert PC (ed) Proceedings of the 6th international conference on precision agriculture and other precision resources management, Minneapolis, 14–17 July, 2002. American Society of Agronomy, Minneapolis, pp 443–457

Frogbrook ZL, Oliver MA (2007) Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. Soil Use Manag 23(1):40–51

Hassanli AM, Ebrahimizadeh MA, Beecham S (2009) The effects of irrigation methods with effluent and irrigation scheduling on water use efficiency and corn yields in an arid region. Agric Water Manag 96(1):93–99

Hedley CB, Yule IJ (2009) A method for spatial prediction of daily soil water status for precise irrigation scheduling. Agric Water Manag 96(12):1737–1745

Hornung A, Khosla R, Reich R, Westfall DG (2003) Evaluation of site-specific management zones: grain yield and nitrogen use efficiency. In: Proceedings of the 4th European conference on precision agriculture. Wageningen Academic Publishers, Wageningen, pp 297–302

Hornung A, Khosla R, Reich R, Inman D, Westfall DG (2006) Comparison of site-specific management zones: soil-color-based and yield-based. Agron J 98(2):407–415

INIA (2006) Métodos de análisis recomendados para los suelos de Chile. Instituto de Investigaciones Agropecuarias, Santiago de Chile

Jaynes DB, Colvin TS, Kaspar TC (2005) Identifying potential soybean management zones from multi-year yield data. Comput Electron Agric 46(1):309–327

Jiang Q, Fu Q, Wang Z (2011) Study on delineation of irrigation management zones based on management zone analyst software. In: Computer and computing technologies in agriculture IV. Springer, New York, pp 419–427

Li X, Pan Y, Zhang C, Liu L, Wang J (2005) A new algorithm on delineation of management zone. In: Proceedings of IEEE international geoscience and remote sensing symposium, IGARSS'05 2005, vol 1, p 4

Mainuddin M, Gupta AD, Onta PR (1997) Optimal crop planning model for an existing groundwater irrigation project in Thailand. Agric Water Manag 33(1):43–62

Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. Struct Multidiscip Optim 26(6):369–395

McCauley A, Jones C, Jacobsen J (2005) soil and water management module 1: Basic soil properties. In: A self-study course from the MSU extension service continuing education series 4481–1. Montana State University, Bozeman

Mortensen DA, Johnson CK, Doran JW, Shanahan JF, Wienhold BJ (2003) Site-specific management zones based on soil electrical conductivity in a semiarid cropping system. Agron J 95(2):303–315

Mulla DJ (1991) Using geostatistics and GIS to manage spatial patterns in soil fertility. In: Proceedings of the automated agricultural for the 21st century. American Society of Agricultural Engineers, St. Joseph, pp 336–345

Ortega JA, Foster W, Ortega R (2002) Definición de sub-rodales para una silvicultura de precisión: una aplicación del método Fuzzy K-means. Ciencia e Investigación Agraria 29(1):35–44

Ortega R, Flores L (1999) Introducción al manejo sitio-específico. In: Ortega R, Flores L (eds) Agricultura de Precisión, Ministerio de Agricultura, Instituto de Investigaciones Agropecuarias. Centro Regional de Investigación Quilamapu, pp 13–46
Ortega RA, Santibáñez OA (2007) Determination of management zones in corn (*Z ea mays* L.) based on soil fertility. Comput Electron Agric 58(1):49–59
Ortega Álvarez JF, Juan Valero JA, Tarjuelo Martín-Benito JM, López Mata E (2004) MOPECO: an economic optimization model for irrigation water management. Irrig Sci 23:61–75
Pedroso M, Taylor J, Tisseyre B, Charnomordic B, Guillaume S (2010) A segmentation algorithm for the delineation of agricultural management zones. Comput Electron Agric 70(1):199–208
Reddy MJ, Kumar DN (2008) Evolving strategies for crop planning and operation of irrigation reservoir system using multi-objective differential evolution. Irrig Sci 26(2):177–190
Roudier P, Tisseyre B, Poilvé H, Roger JM (2008) Management zone delineation using a modified watershed algorithm. Precis Agric 9(5):233–250
Sahoo B, Lohani A, Sahu R (2006) Fuzzy multiobjective and linear programming based management models for optimal land-water-crop system planning. Water Resour Manag 20:931–948
Sarker R, Ray T (2009) An improved evolutionary algorithm for solving multi-objective crop planning models. Comput Electron Agric 68(2):191–199
Sarker RA, Talukdar S, Haque AFM (1997) Determination of optimum crop mix for crop cultivation in Bangladesh. Appl Math Model 21(10):621–632
Schepers JS, Luchiari A, Johnson SH, Liebig MA, Shanahan JF, Schepers AR (2004) Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. Agron J 96(1):195–203
Simbahan GC, Dobermann A (2006) An algorithm for spatially constrained classification of categorical and continuous soil properties. Geoderma 136(3):504–523
Whelan BM, Cupitt J, McBratney AB (2003) Practical definition and interpretation of potential management zones in Australian dryland cropping. In: Robert PC (ed) Proceedings of the 6th international conference on precision agriculture and other precision resources management, Minneapolis, 14–17 July 2002. American Society of Agronomy, Minneapolis, pp 395–409
Xu L, Chen L, Chen T, Gao Y (2011) SOA-based precision irrigation decision support system. Math Comput Model 54(3):944–949

Chapter 7
Optimal Transport Planning for the Supply to a Fruit Logistic Centre

Esteve Nadal-Roig and Lluís M. Plà-Aragonés

7.1 Introduction

Supply chain planning has been studied intensively in recent years (Catalá et al. 2013) in particular for production and transport planning (Mula et al. 2006; 2010), but less in the agri-food industry (Ahumada and Rene Villalobos 2009). Ahumada and Rene Villalobos (2009) distinguish two main types of agricultural supply chains: fresh and non-perishable agri-food chains. They review fresh products paying attention to their logistical complexity, their limited shelf life and the interest of the public on the safety of these products. On the other hand, according to Verdouw et al. (2010) fruit supply chains exhibit some food-specific characteristics such as long lead times, seasonable production, quality variations between producers and plots, fast handling, short delivery time to preserve freshness and special storage conditions and packing demands (Trienekens et al. 2012). Hence, fruit supply chain planning is a complex system involving the interaction of different agents in charge of production, processing, storing and distribution (Fig. 7.1).

The fruit industry is very important in Europe being the EU a major fruit producer. The majority of fruit production in the EU takes place in southern countries like Spain, expecting a significant increase in following years as response to fruit demand (Verdouw et al. 2010). According to the FAOStat in 2011, the rank of Spain in the world for selected fruits was: third one for peaches and nectarines, fifth one for cherries, sixth one for pears and eighth one for plums (FaoStat 2013). Within the EU-27 the role of Spanish fruits is also important being the first producer

E. Nadal-Roig • L.M. Plà-Aragonés (✉)
Departament de Matemàtica, Universitat de Lleida, Jaume II, 73, 25003 Lleida, Spain
e-mail: lmpla@matematica.udl.es

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_7

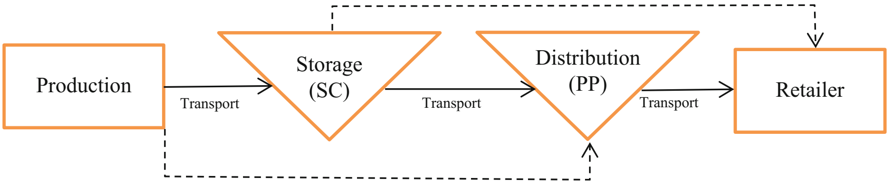


Fig. 7.1 General fruit supply chain structure

of nectarines, the second producer of pears, cherries and peaches, the third producer of plums and the sixth producer of apples (Eurostat 2013). Such a position has stimulated the Spanish fruit industry to evolve, becoming very competitive and looking for a more efficient management of supply chains.

Logistics of fresh fruits is a problem related with the balance between the price achievable in the market and the quality of the product. Quality is related with parameters like sweetness, crunchiness and strengthens, connected in some way with the optimal ripeness of the fruit. The fresh fruit sector is affected by seasonality, understood here as the production of fresh fruits during a limited period of time (Hester and Cacho 2003). This time period is variable depending on the decay of the fruit variety and the admissible means of preservation. There are fresh fruit varieties that have to be consumed rather quickly after harvesting like apricots, cherries and berries in general. Other fruits can complete more slowly the maturation process after harvesting and then enlarge their marketing time window. Even though, there are environmental conditions during storage and transportation that can be used to regulate fruit quality in some extension like cooling, temperature control or controlled atmosphere.

The motivation of this chapter is the PP operating with limited storage capacity, so fruits to be processed have to be transported from intermediate storage centres. This chapter aims to formulate a mixed integer linear programming model to optimise the transport planning of fruit varieties from storage centres (SCs) to a packaging plant (PP) for being processed upon demand to cover daily orders. The main interest of the decision maker is to avoid idle times at the PP and the stock breaking of fruits to be processed. Then, the PP has to maintain a rolling stock to cover committed orders without stopping the processing line. An additional interest concerns the distribution of workload among trucks and drivers available. Depending on the demand, the model may suggest the opening of a controlled atmosphere SC. Then, the model organise the transports from the cooperatives supplying convenient fruit varieties to the PP, maintaining a stock capable of satisfying the daily demand from the customers.

As a case study, the model is applied to a fruit logistic centre (FLC) located in one of the most important production areas of fresh fruit of Spain, in Lleida. However the FLC has special features, the model has been developed in general terms for being applied on most fruit supply chains worldwide.

7.2 Problem Description

Supply chain structure may vary from country to country having different configurations, but sharing characteristics inherent to the fruit industry (Verdouw et al. 2010). A generic fruit supply chain is shown in Fig. 7.1, adapted from the modelling approach presented by Rong et al. (2011). During the harvesting season, the different fruit varieties are usually picked and collected in pallets by farmers who deliver them to the SC or PP, either to be stored or processed (Broekmeulen 1998). Sometimes, SCs are close or part of a PP, depending if the PP is operated by the same company or cooperative or if fruits are distributed quickly or not. Some fruits like apples and pears can be stored for long, others not so, like peaches and some very little like cherries or apricots. However, in all cases, cooling systems is an element to consider for controlling the maturation process and the decay of fruits. This way, apples and pears are available during all year if they are stored in controlled atmosphere while the rest of fruits produced in Europe have a limited marketing time window.

Producers transport harvested fruits to the SC. Regular SC send fruits to the PP in few days or weeks. However, SCs with controlled atmosphere have to be filled with fruits and closed for a longer period. Facilities with controlled atmosphere allow fruits to be stored up to 12 months, but they have to be only opened when all the content is going to be retrieved for processing in the PP. The PP is in charge of washing, sorting and grading of fruits; packaging and labelling in the end of packaging lines. Afterwards, fruits are distributed to retailers to fulfil the day-to-day orders. Operation at PP has to be planned beforehand because ordered fruits have to be processed on time. Transports have to be also planned according to the availability of trucks and drivers even when these activities are outsourced (Hsiao et al. 2010).

A usual fruit supply chain may involve different producers that supplies fruits during the harvesting season to a PP, where they are processed and delivered to the consumer by different retailing channels. The number of PPs may depend on the size of the company and the number of producers, but it is agreed that a PP is the core of the fruit supply chain from a tactical point of view (Blanco et al. 2005). Two main functions are assigned to a typical PP: warehousing and distribution. However, the problem studied here relies on a structure of the supply chain keeping warehousing and distribution apart. Let's consider fruit producers grouped in cooperatives. Individual cooperatives only have storage capacity and thus, the warehouse function is deployed by them. The distribution function, including processing activities, is centralised in the so-called FLC where all orders are concentrated and served. Orders are fulfilled by the fruits stocked in the cooperatives. The FLC manages the logistics of the chain, that is, the planning, implementing and controlling the efficient cost-effective flow and storage of fruits, in-process inventory, distributed fruits and related information from producers and retailers for the purpose of conforming the customer requirements (Van Goor et al. 2003). Thus, according to Manzini and Accorsi (2013) the FLC, as crucial node in the chain, can contain the main source of inefficiency, waste and uncontrollable costs throughout the fruit supply chain.

The long-term storage can be of two types: cooling storage or controlled atmosphere storage. Once a storage is open the preservation chain of fruits is broken and the maturity process progress again making necessary to empty the storage before opening a new one. An issue is the continuous supply of fruits to the centre for a non-stop operation of the packing lines. Fruits have to be sorted out the storages few days before shipping to recover natural properties related to follow-up a good maturity process. Fruits sent by cooperatives to the logistic centre are shipped the same day, but the FLC is who select the suppliers and determines which storage facilities to open. Only a secure inventory is maintained permitting to start up the following day. Then, the logistic centre acts as a PP but without storage capacity which relies on the cooperatives.

The flow of fruits managed from the logistic centre varies along the year. More transport capacity is needed during the harvesting season. There is an increment of transports from fields to cooperatives, among cooperatives and from cooperatives to the FLC (Fig. 7.2). Transports from fields are done by farmers while those to the FLC are planned and controlled by the logistic centre. Out of the harvesting season transports from fields and among cooperatives disappear and only remain the flow from cooperatives towards the FLC. The reason is because not all fruit are available all-round the year. There are perishable fruits with a limited marketing window. No technical means of decay control are feasible for them. However, apples and pears can be preserved in controlled atmosphere and marketed out of the harvesting season.

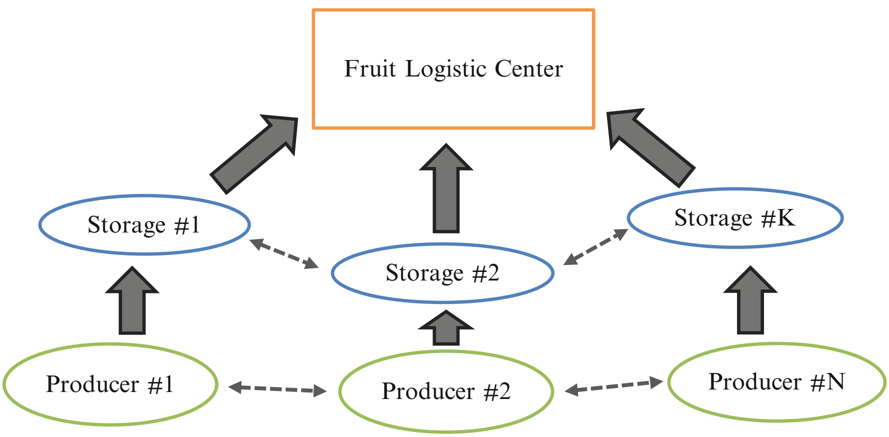


Fig. 7.2 Possible paths followed by fruits from producers till the fruit logistic centre

7.3 Modelling of the Transport Planning Problem of a Fruit Logistic Centre

Transport planning in an FLC is a task with a variable workload depending on the daily demand of fruits, the arrival of new orders and the number of trucks available. The problem modelled here represents the operational planning for a day. The demand of fruits is defined for the next day and the manager makes the planning with which the activity in the FLC will start the following day. However, new orders may arrive or changes in priorities can be introduced. These unforeseen changes may force to refine or redo the original planning again changing the schedules for truck drivers and suppliers. It is in this context that the following model is formulated.

7.3.1 Decision Variables

There are two sets of decision variables according to the quantity of goods to be delivered and the number of trucks needed to perform these operations. The first one, X_{ifcv}, represents the quantity of fruit to be transported in kilograms from the cooperatives of producers to the logistic centre, where the subscripts: i represents the cooperative of producers to procure the fruit ($i = 1, 2, \ldots, |I|$); f represents the variety of the fruit ($f = 1, 2, \ldots, |F|$); c represents the category of the f-fruit ($c = 1, 2, \ldots, |C|$) and v represents the truck to be used ($v = 1, 2, \ldots, |V|$). The second one, the binary variables $Y_{iv} \in [0, 1]$ are defined to represent the expected numbers of trips for the truck v from the cooperative i to the logistic centre.

The total number of decision variables varies from season to season and depending on the number of storages used to preserve and deliver fruits in winter. These fruits are also apples and pears of different varieties and categories.

7.3.2 Objective Function

The primary objective is the minimization of the daily cost of transport. This results in a minimum number of trips that allows the FLC to satisfy the demand. Depending on the unitary cost coefficients, they can be easily adapted to represent distances, load in kilograms or cost in euros.

$$\min C = \min \sum_{i} \sum_{v} c_i Y_{i,v}$$

This is in other words, the minimisation of the sum of the transport cost from the cooperative i to the FLC, given the number of trips to perform, $Y_{i,v}$, are covered by the truck v. A secondary result interesting for practical purpose is the schedule of transports derived from the optimal solution.

7.3.3 Constraints

7.3.3.1 Nature of Daily Demand of Fruits

The daily demand is given by the sum of all customer orders confirmed for the day and that must be delivered. Moreover, a threshold corresponding to a security stock of fruits has to be considered to allow the smooth operation of the FLC during the day and the start-up of the following one. Based on the arriving orders and the experience of the FLC manager, this threshold is stated. If the total demand is represented by D_{fc}, the total quantity for each fruit and category X_{ifcv} transported from all cooperatives must be higher than it:

$$D_{fc} \leq \sum_{i,v} X_{ifcv} \quad f = 1, 2, \ldots, |F|; \quad c = 1, 2, \ldots, |C|$$

7.3.3.2 Number of Loads and Total Load

The various trucks' capacities require constraints on the total load carried by available trucks due to the orders with high volume. Given the capacity of trucks is known, C_v, and the maximum number of trips a truck can do from a specific cooperative of producers, Y_{iv}, a constraint verifying the total amount of fruits transported is taken into consideration:

$$\sum_{fc} X_{ifcv} \leq C_v Y_{iv} \quad i = 1, 2, \ldots, |I|; \quad v = 1, 2, \ldots, |V|$$

This constraint allows the decision maker to detect paths with more demanded trips and then, assigning trucks of more capacity to satisfy the demand given when necessary. Note that different fruits and categories can be transported by the same truck visiting a cooperative.

7.3.3.3 Timetable of Trucks

Trucks are normally used for several trips per day. It is considered that all trips start and finish at the FLC. It is assumed that a truck is driven by the same driver. The number of trucks available may vary. However, the availability of drivers who cannot drive more than a legal number of hours $J(v)$ is more stringent.

On the other hand, depending on the cooperative of producers, the loading and unloading time may vary depending on resources available for such tasks. Regarding the trip time covering the distance from the fruit logistic centre to the cooperative is affected by the type of lorry, the load and the speed to cover the path.

Thus, the total transport time for each truck $\sum_i TT_i Y_{iv}$ must not exceed the available number of working hours of corresponding truck driver:

$$\sum_i TT_{iv} Y_{iv} \leq J(v) \quad v = 1, 2, \ldots, |V|$$

where

$TT_{iv} = \frac{D_i \cdot \left(\frac{1}{Vcc_v} + \frac{1}{Vsc_v}\right) + W_i}{C_v}$ represents the trip time for truck v to cover the path FLC—cooperative i—FLC, being:

D_i: Distances from cooperative i to the FLC.
Vcc_v: Speed of the given carriage means v, with load.
Vsc_v: Speed of the given carriage means v, without load.
W_i: Waiting time at the i-cooperative.
C_v: Loading capacity of truck v.

However, previous constraint can be reinforced taking into account the availability of trucks and the maximum time drivers can be working per day (T_v):

$$\sum_i \sum_v TT_{iv} \cdot Y_{iv} \leq \sum_v T_v$$

7.3.3.4 Multiple Transports Per Truck

Aside the time, the trucks are allowed to make per day a certain number of transportations. These transports are independent of the cooperative to visit. This means, a truck can transport fruit from the same cooperative or not until it reaches its maximum number of daily trips. Therefore,

$$\sum_i Y_{iv} \leq NT_v \quad v = 1, 2, \ldots, |V|$$

This constraint tends to balance the number of trips per truck and hence, the workload of drivers. It can be also specified in terms of total distance covered by day or in total fruit carried per day or simply as stated just in total number of trips per day.

7.3.3.5 Fruit Inventory at the Cooperatives of Producers

The quantity of the fruit to be transported from each cooperative i for fruit f and category c must not exceed the cooperative inventory for this fruit type and category:

$$S_{ifc} \geq \sum_{fcv} X_{ifcv} \quad i = 1, 2, \ldots, |I|; \; f = 1, 2, \ldots, |F|; \; c = 1, 2, \ldots, |C|$$

7.3.4 Size of the Problem

In order to give a view about the problem in terms of size, this section details the total amount of decision variables and restrictions, irrespective of the input data used for the execution of the model:

Total number of constraints:

Nature of daily demand of fruits	$F \times C$
Number of loads and total load	$I \times V$
Timetable of trucks	$V + 1$
Multiple transports per truck	V
Fruit inventory at the cooperatives	$I \times F \times C$

Total #constraints: $FC + IV + V + 1 + V + IFC = FC(1 + I) + V(I + 2) + 1$.

Total number of variables:

Continuous variables procuring fruits to the FLC:

$$X_{ifcv} : I \times F \times C \times V$$

Integer variables representing trips:

$$Y_{iv} : I \times V$$

Total #variables: $IFCV + IV = IV(FC + 1)$.

7.4 Application of the Model: A Case Study

To illustrate the use of the model, a real case is considered from a Spanish company specialised in pome fruit with a similar supply chain structure than that described previously. The main actors of this supply chain are three: the individual farmers, the producer cooperatives where farmers send their production to be stored and a cooperative owning the FLC. Main fruit types are grouped in pome (apples and pears) and stone (nectarines, peaches, cherries and plums).

ACTEL is a Spanish fruit cooperative of second order (i.e. a cooperative of cooperatives, the so-called cooperatives of first order) with one LFC. Different cooperatives of fruit producers (29 in total) are the stakeholders of ACTEL. Individual producers, members of a cooperative, are in charge of the growing, harvesting of fruits. Fruits are sent to the corresponding cooperative for storage while ACTEL, as logistic centre, is in charge of packaging, labelling and distribution to international retailers, exporters and local retailers, wholesalers and food service providers. The FLC is ruled by ACTEL and manage the fruit supply chain. Few of the fruits are sold directly without any processing by producer cooperatives, although most of them are stored only very short.

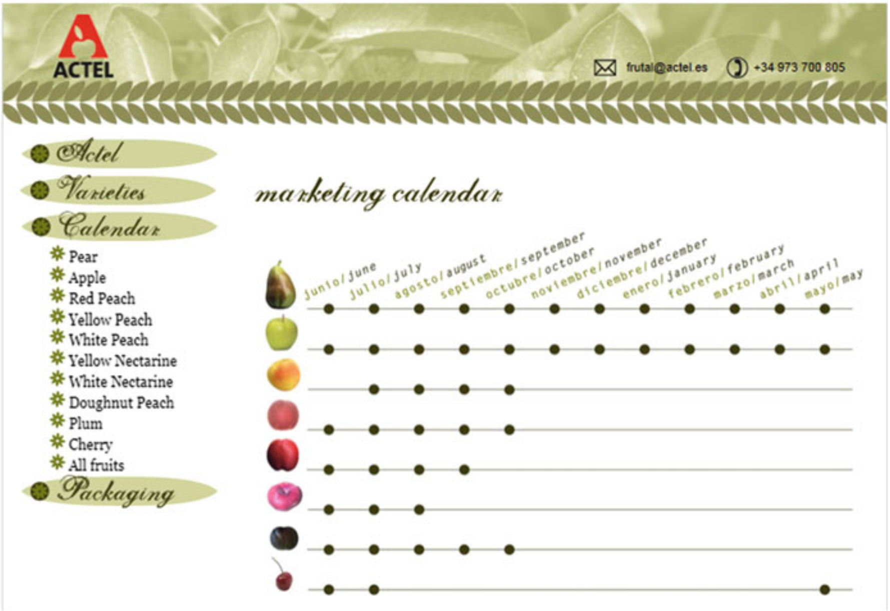


Fig. 7.3 Fruits processed by ACTEL and regular marketing calendar (http://www.actel.org/fruita_cataleg/eng/calendario.html)

In Fig. 7.3, the fruits processed in ACTEL are displayed, as well as the marketing calendar. As shown, July is the most complicated month given all fruits are being harvesting and marketing. On the other side, from November to April only apples and pears are available, thanks to the use of storage facilities under controlled atmosphere. The type of coordination with costumers differs a lot, including spot market, informal long-term relations, formalised contracts and partnerships. Especially, big retailers have specific requirements regarding variety, size, ripeness, certificates, labels and packaging. Fruits can be ready for distribution 24 h after harvesting. However, they are trying new products that include processing like peeled apples for vending machines.

As fruit types have different temperature control protocols and because packaging rates are typically fruit dependent, the different fruit types should be considered as separated commodities. The FLC takes decisions regarding cool storage of fruits and storage under controlled atmosphere for pome fruits. For example, when and how storage facilities has to be filled and closed for fruits being processed later. The FLC organises the transports of fruits to the logistic centre for processing and distribution to fulfil the orders received from customers.

7.4.1 Formulation of the Model

The FLC has an averaged capacity for processing of around 150 ton per day although the maximum stocking capacity is between 4 and 5 ton only. The continuous supply of fruits is necessary during the day to allow the non-stop operation of the FLC. There are 29 cooperatives available to provide fruit to the FLC. Figure 7.4 shows the relative location of cooperatives and their distance regarding the FLC (coordinates 0.0). The exact distance can be found at Appendix 1 as well as the total time per trip (loading, unloading and trip time) from the cooperatives to the FLC.

All the cooperatives have in their stock three varieties of pears (Blanquilla, Conference and Alexandrine) and two of apples (Golden and Red Delicious) representing five different fruits ($f = 5$). Each variety can have until eight different categories ($c = 8$) according to the fruit's size (101, 104, 108, 201, 202, 215, 218 and 220). In Appendix 2, the detailed stock per cooperative, variety and category is shown. Note that this stock corresponds to the winter season.

Daily, the FLC manager sets up the expected demand of fruit per variety and category to deliver to the customers as well as a threshold needed in the FLC to ensure the delivery of fruits. Out of the harvesting season, and in a certain days, it is possible to not have demand or threshold for some categories. Appendix 3 shows the demand and the threshold data used in the model for each variety.

The FLC has outsourced the transport from cooperatives, but the work plan and the schedule of trips are provided by the FLC. The number of trucks available is variable and can be adapted from the needs of FLC from one day to the following day. During the low season (non harvesting months), the FLC uses regularly two different types of trucks with different load capacities and a total of four trucks named T1, T2, T3 and T4. The first type can transport 24 ton each truck (trucks T1 and T3 in the results) and the other one (trucks T2 and T4 as referred in the text) is smaller, that is, of 14 ton. The truck's cost is a daily price, without taking into

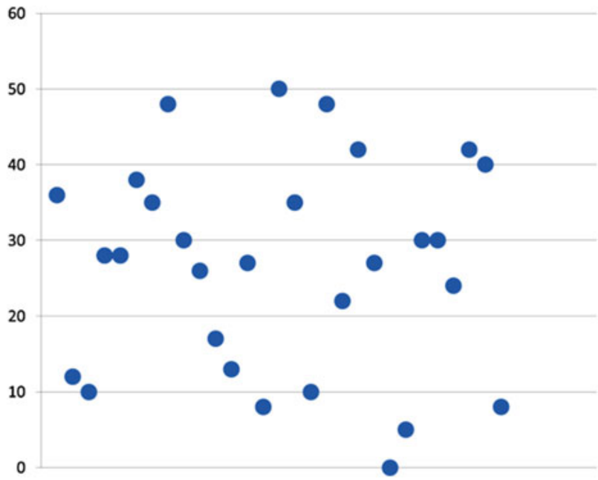


Fig. 7.4 Cooperatives location from the logistics centre

account the kilometres done or the load transported by trucks. At this time, two trucks of each type are used regularly. However, the FLC can have additional trucks available from the transport company if they request them in advance. During peak days at the harvesting season, the FLC can contract more than ten trucks for daily tasks.

7.4.2 *Results and Discussion*

To develop the preliminary version of the model and its execution, the modelling language ILOG OPL and the solver CPLEX v12.2 has been used. The hardware used in the development and test of the model was a laptop computer (Pentium Dual-Core CPU at 2.1 GHz and 4 Gb RAM). Microsoft Excel has been used for storage data, both inputs and outputs of the model due to the easy analytical use.

With the case data provided by the cooperative, the model has 4,756 variables in total (4,650 as continuous and 4,640 as integer). The model finishes in 7:14 s. This allows the FLC manager to get results in a short time and therefore to execute again the model in case the demand changes during the day, to make additional corrections if needed or to explore different alternatives. For instance, the manager can use the model to explore the impact of additional trucks or different number of trips permitted to the same supplier cooperative.

The model shows the optimal transport planning according to the remaining daily demand. As the sum of stock in cooperatives is much higher than the demand in the FLC, all demand is satisfied. Figure 7.5 shows the optimal quantity of fruit to be transported from each cooperative as well as the trucks to be used and the total quantity. Only variables for which the value is different of zero are shown.

On the other hand, the number of trips for each truck and cooperative is shown in Table 7.1. That table shows how only seven cooperatives are visited to load fruits to satisfy the FLC demand.

Furthermore, the smallest trucks T2 and T4 are not used and the biggest ones are preferred reducing in this way the total number of trips required to procure the fruits

Coop	Blanquilla				Conference				Golden			
Code	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
2							24					
3					24							
6							2				22	
14	2								22			
17											24	
23	24											
29	24											

Fig. 7.5 Transport and quantity map for cooperatives procuring to the FLC

Table 7.1 Number of trips per truck and cooperative

Coop code	T1	T2	T3	T4	Trips
2			1		1
3	1				1
6			1		1
14	1				1
17			1		1
23	1				1
29	1				1
Total	4	0	3	0	

to the FLC. Therefore, trucks T1 and T3 are not necessary for that day. This table is useful to explore the increasing or decreasing in the number of trucks available. As the trucks are outsourced, the FLC can act proactively according to the demand expected in future and booking the trucks needed beforehand.

The total number of trips per truck was of four and three. A balanced result as the manager wished. This interaction between the end-user and the system is the main appreciated characteristic of the model because it allows the FLC to save time and money to plan the procurement of fruits for daily operation of the FLC. Furthermore, parameters of the model and results are recorded into an Excel spreadsheet that can be updated automatically by the ERP of the FLC. Even reports and results can be customised according to the intended use by the FLC manager. Although this model was developed to deal with an FLC, the same company owns other plants for which they find also suitable this kind of models like the procurement of a drying forage plant.

7.5 Conclusions

We have presented a mixed integer linear programming developed to support operational decision making in the transport planning for an FLC. We have illustrated the use of the model in a real case satisfying the end-user requirements. FLC manager appreciates the flexibility of the model and saving performed compared to past operation in planning the procurement of the logistic centre.

Although the results from the implementation of the model have been successful, the final adoption of the model is pending of internal adjustments allowing the complete automatisation of the process.

Future work involves the running of the model for the harvesting season, where the number of fruits and categories is bigger, as well as the number of trucks involved in the transportation. Furthermore, from academic point of view the reformulation of the model as a capacitated vehicle routing problem is also in our agenda.

Acknowledgement The authors acknowledge the financial support of the Spanish Research Program (AGL2010-20820 and MTM2009-14087-C04-01).

Appendix 1: Distance Between Cooperatives and the FLC

Coop code	FLC distance	Transportation time (h)
1	36	4.20
2	12	1.40
3	10	1.33
4	28	2.93
5	28	3.93
6	38	3.27
7	35	4.17
8	48	3.60
9	30	3.00
10	26	1.87
11	17	2.57
12	13	2.43
13	27	3.90
14	8	2.27
15	50	2.67
16	35	2.17
17	10	1.33
18	48	4.60
19	22	3.73
20	42	3.40
21	27	3.90
22	0	0.00
23	5	2.17
24	30	4.00
25	30	3.00
26	24	2.80
27	42	4.40
28	40	3.33
29	8	2

Appendix 2: Stock per Cooperative, Variety and Category (in ton)

Coop code	Variety	Category								Total
		101	104	108	201	202	215	218	220	
2	Conference	1,409	0	0	0	0	0	0	0	1,409
2	Alexandrine	66	2	0	0	17	0	0	0	85
3	Blaquilla	209	0	0	0	0	16	0	0	225
3	Conference	1,102	55	0	0	0	146	0	0	1,303
3	Alexandrine	102	0	0	0	0	43	0	0	145
3	Red Delicious	127	40	0	0	0	0	0	0	167
3	Golden	611	0	0	0	0	69	0	0	680
5	Golden	178	47	0	0	0	0	0	0	225
6	Conference	785	0	0	0	0	0	0	0	785
6	Alexandrine	134	0	0	0	0	0	0	0	134
6	Golden	997	0	0	0	0	0	0	0	997
11	Conference	194	0	0	0	0	0	0	0	194
11	Golden	341	0	0	0	0	0	0	0	341
12	Conference	542	0	0	0	0	0	0	0	542
12	Golden	493	0	0	0	0	0	0	0	493
14	Blaquilla	131	8	0	0	0	0	0	0	139
14	Conference	481	30	0	0	0	0	0	0	511
14	Golden	237	0	0	0	0	0	0	0	237
17	Blaquilla	180	0	0	0	0	0	0	0	180
17	Conference	1,750	0	0	0	0	0	0	0	1,750
17	Golden	1,750	0	0	0	0	0	0	0	1,750
21	Blaquilla	149	11	0	0	0	0	0	0	160
21	Conference	261	0	0	0	0	0	0	0	261
21	Golden	489	0	0	0	0	0	0	0	489
23	Conference	684	0	0	0	0	16	0	0	700
23	Golden	653	71	0	2	0	0	0	0	726
24	Conference	343	63	0	0	0	0	0	0	406
24	Golden	520	0	0	0	0	0	0	0	520
25	Conference	220	0	0	0	0	0	0	0	220
25	Golden	353	47	0	0	0	0	0	0	400
26	Golden	261	0	0	0	0	0	0	0	261
28	Golden	291	5	0	0	0	0	0	0	296
29	Blaquilla	353	78	0	0	0	0	0	0	431
29	Conference	225	0	0	0	0	0	0	131	356
29	Golden	290	0	0	0	0	0	0	0	290

Appendix 3: Demand (in ton) and Minimum Stock in the FLC (in kg)

Variety	Minimum stock per category								Demand
	101	104	108	201	202	215	218	220	
Blaquilla	240	60							50
Conference	400	100							50
Alexandrine	80	20							0
Red Delicious									0
Golden	320	80							68

References

Ahumada O, Rene Villalobos J (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 195:1–20

Blanco AM, Masini G, Petracci N, Bandoni JA (2005) Operations management of a packaging plant in the fruit industry. J Food Eng 70:299–307

Broekmeulen R (1998) Operations management of distribution centers for vegetables and fruits. Int Trans Oper Res 5(6):501–508

Catalá LP, Durand GA, Blanco AM, Bandoni JA (2013) Mathematical model for strategic planning optimization in the pome fruit industry. Agric Syst 115:63–71

EuroStat (2013) http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do. Accessed 15 Mar 2013

FAOStat (2013) http://faostat.fao.org/site/339/default.aspx. Accessed 15 Mar 2013

Hester S, Cacho OJ (2003) Modelling apple orchard systems. Agric Syst 77:137–154

Hsiao HI, Kemp RGM, van der Vorst JGAJ, Omta SWF (2010) A classification of logistic outsourcing levels and their impact on service performance: evidence from the food processing industry. Int J Prod Econ 124:75–86

Manzini R, Accorsi R (2013) The new conceptual framework for food supply chain assessment. J Food Eng 115:251–263

Mula J, Poler R, García-Sabater JP, Lario FC (2006) Models for production planning under uncertainty: a review. Int J Prod Econ 103:271–285

Mula J, Peidro D, Díaz-Madroñero M, Vicens E (2010) Mathematical programming models for supply chain production and transport planning. Eur J Oper Res 204:377–390

Rong A, Akkerman R, Grunow M (2011) An optimization approach for managing fresh food quality throughout the supply chain. Int J Prod Econ 131:421–429

Trienekens JH, Wognuma PM, Beulens AJM, van der Vorst JGAJ (2012) Transparency in complex dynamic food supply chains. Adv Eng Inform 26:55–65

Van Goor AR, Amstel MJPV, Amstel WPV (2003) European distribution and supply chain logistics. Wolters-Nordhoff, Groningen

Verdouw CN, Beulens AJM, Trienekens JH, Wolfert J (2010) Process modelling in demand-driven supply chains: a reference model for the fruit industry. Comput Electron Agric 73:174–187

Chapter 8
Simulating Vulnerability in Victoria's Fruit and Vegetable Supply Chain

Leorey Marquez, Andrew Higgins, and Silvia Estrada-Flores

8.1 Introduction

The horticultural industry in Australia is valued at \$3.6 billion per annum (Australian Natural Resource Atlas, www.anra.gov.au). In Victoria, the industry contributed to the state's economy with \$1.3 billion in 2009–2010. Victoria has 8,500 horticulture-related enterprises, employing 50,000 people full-time and up to 100,000 during harvest periods (DPI 2012).

In comparison to other sectors such as livestock and broadacre crops, horticultural production can be regarded as a low emitter of greenhouse gases (GHG). However, post-farm supply chain activities such as freight and storage contribute significantly to the carbon footprint of field-grown fruits and vegetables (F&V). For example, Mithraratne et al. (2008) found that primary production of New Zealand apples exported to the UK represents 20 % of the total carbon footprint, while shipping, retailer and local/consumer distribution chains represent 80 %. In the case of New Zealand, kiwifruit exported to the UK, primary production represents 4 % while 96 % represents GHG emissions from distribution (Hume et al. 2009).

An aspect that has been seldom investigated in published literature dealing with GHG emissions associated to food supply chains is the effect of disruptions, and in

L. Marquez (✉)
CSIRO Mathematics, Informatics and Statistics, Gate 5, Normanby Road, Clayton, VIC 3168, Australia
e-mail: Leorey.Marquez@csiro.au

A. Higgins
CSIRO Ecosystem Sciences, Acton, ACT, Australia
e-mail: Andrew.Higgins@csiro.au

S. Estrada-Flores
Food Chain Intelligence, North Sydney, NSW, Australia
e-mail: silvia@food-chain.com.au

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_8

particular, extreme weather events (EWE) in food freight emissions. It is generally recognised that supply chains are being increasingly exposed to EWE, with varying degrees of damage and consequences. While food distribution systems have been able to cope with infrequent EWE, the IPCC report establishes that global warming will increase the risks of the following phenomena (William et al. 2007; Ugalde et al. 2007):

- Severe storms, flooding, droughts
- More intense, more frequent and longer-lasting heat waves
- Increase in extreme rainfall intensity
- More frequent and more severe cyclones
- A rise in sea level

The Australian Government has pointed out that EWE are likely to have greater impacts today than a century ago. This is because of the greater population density in vulnerable areas, the increase in urbanised areas, and the more extensive networks of sophisticated and costly infrastructure (e.g. transport and energy systems), among other reasons (Parliament of Australia 2008).

Evidence of the effect of EWE on the Australian horticultural sector includes the following:

- In March 2006, Cyclone Larry ruined 200,000 tonnes of bananas, worth an estimated $300 million. In addition to the crop loss, the impact of Cyclone Larry on the Australian banana industry left thousands of Queenslanders out of work and caused banana prices to increase by more than 400 %. It took over 2 years for the banana industry to return to normal harvesting patterns.
- In Victoria, floods in Gippsland in 2007 affected the local supply of some vegetables and herbs (primarily broccoli, cauliflower, carrots, some lettuce and herbs) (Stewart 2007 as cited in Larsen et al. 2008).
- The 2009 floods in Tablelands (QLD) made transport of fruits and vegetables infeasible through the Bruce Highway (Cairns Post 2009). Up to $10 million worth of fresh fruits and vegetables in cold rooms and sheds were stored in cut-off coastal areas.
- Major drought and heatwave conditions increased the severity of the Victorian bushfires in 2009. Severe damage to field and protected crops were registered in several production areas (Food Magazine 2009). For example, in Port Phillip—one of the most important fruit production areas—50–90 % of berry crops and 20–25 % of orchard crops (apples and late season apricots) were lost.

The 2011 Queensland floods provide another example of the significant impact of EWE-related disruptions in the Australian horticultural supply chain. Widespread flooding was also experienced in southwest New South Wales, western Victoria and northern Tasmania. The combination of the January floods and Cyclone Yasi caused almost $2 billion damage to crops, disrupted food supply Australia-wide and increased food prices in South East Queensland. The January 2011 flooding is estimated to have reduced agricultural production by at least A$500–A$600 million in 2010–2011, with significant impacts on the production of fruits and vegetables, cotton, grain sorghum and some winter crops (ABARES 2011). Using 2005–2006 levels, the flood-affected regions in Queensland accounted for 19 % of the total value of Australian vegetable production and 8 % of the total value of fruit and nut production. The key fruits and vegetables affected include beetroot, sweet potatoes, zucchini, mandarins, spring onions, peas, capsicums and chillies. In Victoria, production of vegetables in the flood-affected regions accounts for around 3 % of total Australian production. In the case of fruit, the flood-affected regions of New South Wales and Victoria each accounted for around 13 % of the total value of Australian fruit production.

Evaluating the potential impacts of future EWE events on the F&V is very complex due to the geographically diverse supply chains between farms, processors, distributors and markets. If many of these chains are disrupted and demands are not met through traditional supply chain paths, alternative sources are used, often interstate or overseas. The extent to which these alternatives are used and additional costs incurred depends on disruption to existing chain paths. A supply chain tool that maps out all supply chain paths provides a capability to better understand transport costs, GHG emissions and supply restrictions under future EWE scenarios. Such a tool has not been available in the past primarily due to: lack of complete data sets over a large region (e.g. state of Victoria) for multiple types of crops; confidentiality of available data; and lack of tools to generate transport routes and supply chain paths. This chapter describes how these data limitations are overcome, along with the development of a Supply Chain Database Tool (SCDT). The SCDT is a deterministic model that maps the transport and distribution components of a supply network for F&V and enables the calculation of relative measures of emissions for different scenarios.

The SCDT tool was used to simulate the GHG effect of supply chain disruptions of F&V grown in Victoria. We use the Victorian bushfires scenario where as a consequence of partial loss in some productive areas of Victoria, Victorian buyers are forced to source 25 % of their normal intake of Victorian-grown F&V from interstate suppliers. We demonstrate the capability of simulation technologies in benchmarking the environmental performance of food freight, assessing the environmental costs of supply chain disruptions and identifying new opportunities to decrease the carbon footprint of food distribution systems from farm to fork.

8.2 Literature Review

There have been limited studies in Australia aimed at analysing food freight logistics in a holistic sense. A State of Logistics study was carried out by CSIRO in 2006/2007 (Higgins et al. 2011) which aimed to "Develop and test a methodology that estimates the costs of logistics in Australian food industries, and to apply this methodology to better understand the structure, drivers and challenges of these logistics". Rather than considering all food categories, four different case studies were selected: fresh mango domestic chains, livestock represented by beef and lamb production, field crops including sugar and grain and wine. The project helped to better understand value chains operations such as transport, storage and packaging. The methodology developed can be extended to other food industries in Australia.

A study by Morgan (2009) assessed supply chains of F&V from the perspective of waste and consumption and their impacts on public health in Australia. As with the CSIRO study, case studies were used, primarily due to lack of available large data sets. Morgan considered GHG emissions across the food supply chains through reference to published reports for farming (Rab et al. 2008), distribution and processing and food preparations. The reports cited by Morgan (and Morgan's report itself) provide general statistics rather than a detailed supply chain analysis.

There have been various logistics studies conducted at an industry or sector level. For example, grains logistic costs were extensively addressed in the Royal Commission into the Grains Storage and Transport (1988), though the findings are largely outdated. Internationally, there have been State-of-Logistics (SoL) studies aimed at defining R&D and infrastructure investment priorities, with CSIR (2005) providing a general analysis across the major industry sectors of manufacturing, mining and agriculture of South Africa. Scientists from CSIR also conducted a more detailed analysis on South African fruit logistics (Van Dyk and Maspero 2004) with a focus on providing recommendations for priority investments in infrastructure. In light of the high-level analysis and recommendations from the South African studies, several "more-focused" logistics projects between CSIR and South African industries have been established. To date, there has been no published whole of chain analysis assessing GHG emissions in food systems.

Analysis of F&V GHG emissions at farm scale is far more advanced than post-farm gate. Based on a project by HAL, Rab et al. (2008) and O'Halloran et al. (2008) extensively considered GHG emissions in the Australian vegetable industry by addressing: availability and applicability of emissions factors; limitations on data availability; and features of the production system that have the greatest contribution to GHG emissions. The authors state that their estimation of GHG emissions in the vegetable farming sector (1,047,008 tonne CO_2/year) was about one third of other estimates, highlighting the need to gather more relevant carbon footprint data. At the farm scale, the authors considered farm inputs and their land impact, as well as farm operations (e.g. irrigation, use of machinery). The Australian Farm Institute has released FarmGAS (AFI 2009), a GHG emissions calculator for farmers for use in scenario planning to reduce GHG emissions on their farm. The Victorian DPI website also contains GHG accounting tools for other forms of agriculture (DPI 2012).

Stochastic models have been developed to evaluate the effect of disruptions in multi-echelon supply chains on food quality (Van der Vorst et al. 2009) and levels of inventory (Schmitt and Singh 2009). There are fewer examples of models suited to evaluate the effect of changes in single or multiple parameters of freight systems (e.g. distances between suppliers and buyers, mode of transport, loading efficiency, fuel factor) on transport GHG emissions. For instance, McKinnon and Piecyk (2009) used average fuel efficiency, distances travelled, lading factors and similar data captured in road freight surveys and statistics captured annually by the Department of Transport (UK) to provide a carbon footprint of road freight in the UK. Although this approach allows high-level data on particular food sectors on road freight emissions, it does not allow a complete farm-to-fork assessment, as it neglects the consumer's role on transport.

8.3 Development of an SCDT

8.3.1 Modelling Philosophy and Scope

Figure 8.1 illustrates the principal activities in the Victorian F&V supply chain. As the diagram shows, F&V transport involve complex spatial and dynamic networks in Australia, incorporating many factors, such as: multiple food products and supply chain paths; long supply chains with multiple stages of processing/distribution; specialised transport needs; multiple modes; mixture of domestic and export products; underpinning supply chain relationships; evolving production systems; and climate variability (Higgins et al. 2010). Road transport paths between farms, markets, DCs and supermarkets are also a complex network for food freight (Victorian Department of Transport 2008), which vary substantially with time of year.

The boxes below each supply chain stage in Fig. 8.1 capture the processes that lead to the generation of GHG emissions. Processes highlighted in bold were in the scope of the freight model and encapsulate all of the transport stages from farm to consumer. The F&V freight model was focused on the transport components of the supply chain, including refrigeration within transport where required. However, it did not include energy use of emissions from production, processing, packaging and other processes. Therefore, the outcomes of the model were not representative of the entire product life cycle. The model was developed as a tool to increase understanding of the transport components of F&V supply chains in one specific region (Victoria).

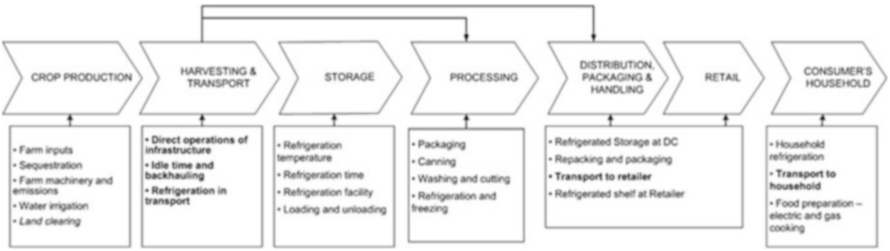


Fig. 8.1 Features of F&V supply chain that contribute to emissions

The transport segments investigated were:

- Movements of F&V produced and consumed in Victoria
- Movements of F&V produced elsewhere in Australia and consumed in Victoria (interstate)
- Movements of F&V produced overseas, transported by ocean freight and consumed in Victoria (imported product)

For the purposes of identifying the major factors affecting CO_2 emissions in Victorian F&V chains, major supermarket chains (MSC) and Melbourne Markets-greengrocer chains (MM) were considered. Cumulatively, these channels account for 97 % of the total fruit and vegetable trade in Australia. Further, exports and volumes leaving Victoria were not considered.

A deterministic methodology was used to identify and assign suitable values (observed or estimated) for each set of variables representing the required components of the supply chain. A full description of this methodology can be found in Marquez et al. (2010). The SCDT model was conceptualised as an MS Access model. Queries are performed on Excel input tables to obtain estimates of GHG emissions produced during the transport of F&V. The tool enabled the combination of parameters and input tables to define different supply chain scenarios.

8.3.2 Data Collection

To construct the SCDT, the following types of information were required:

- Victorian production of F&V
- International imports and exports
- Interstate imports and exports
- Location of processors, supermarkets and distribution centres (DCs)
- Location of consumers

8.3.2.1 Victorian Production of F&V

Primary production data was obtained from the Australian Bureau of Statistics (ABS), which provided production data for 2004–2008. The data, which is partitioned by Natural Resource Management (NRM) region, contained the tonnes of each F&V produced.

Figure 8.2 shows the boundaries of the Victorian NRMs. The ABS data does not specify the exact locations within each NRM region where major production takes place, so assumptions were made as to the coordinates of the origin points.

Two options were available:

1. Assume that production occurs at the geographical centroid of the NRM region
2. Assign a centre on the major production areas

Fig. 8.2 NRM regions for Victoria

For this version of the freight flow model, the geographical NRM centroids were chosen as the origin points. This is because statistics of production per NRM were available, while statistics for production from major growing areas were not. While alternative (2) may be more accurate, there were a number of difficulties with this option in the deterministic approach, namely:

- Data was not available on the actual boundaries of the farm areas
- Classification by growing area could result in different origin points being assigned for different fruit and vegetable items within the same NRM
- The growing region for one item may straddle several NRMs

It was expected that the results of using option (1) would lead to a mixture of over and underestimating distances, which would mostly cancel out one another. Unless restricted by national parks, large bodies of water or extensive urban areas, Victorian farms are expected to be scattered widely within NRMs resulting in growing areas that are non-contiguous and that frequently cross NRM boundaries. With each NRM producing at least 9 of the 45 items of interest, it is highly unlikely that the centroids of all growing areas for the items produced (had these been known) would fall in exactly the same location. Thus, in the absence of any data other than NRM production and boundaries, the NRM centroid remains the best single point estimate of the source of production.

8.3.2.2 International Imports and Exports

Imports and exports were analysed through the following data from the Victorian Department of Primary Industries:

- Volumes (T) of each F&V imported and exported by port for 2004–2008
- Country of origin and destination
- Interstate transfers via port

Emissions from international imports were computed from two sources: (1) sea leg emissions from shipping the volumes from a foreign port to an Australian port, and (2) land leg emissions from transporting the volumes from Australian ports to Victorian DCs.

Representative sea distances between major Australian ports and the nearest port of entry for different countries were obtained using distance calculators available from various websites such as PortWorld (http://www.portworld.com/map/). The distances obtained were based on typical shipping routes used between origin and destination points and do allow for passage through important portals such as the Panama Canal, Suez Canal and Bosporus Strait. Figure 8.3 displays a composite map showing various routes calculated by the PortWorld website for shipping between Melbourne and nine sources of imports. For example, imports from New Zealand only travel 2,700 km to reach Melbourne while those from the UK travel on average 20,200 km.

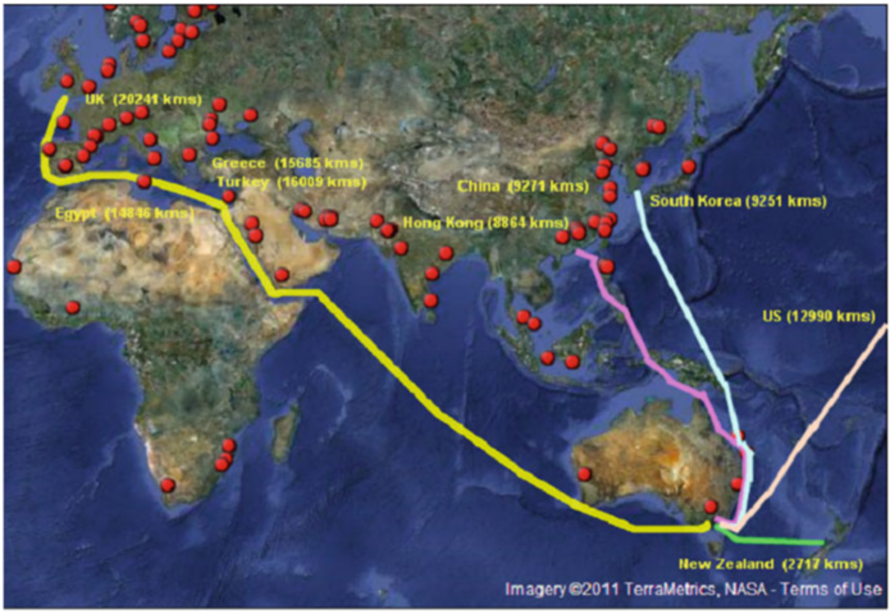


Fig. 8.3 Composite map of shipping routes and distances between Melbourne and import countries from PortWorld website

8.3.2.3 F&V Volumes to and from Interstate

Interstate freight data was obtained from the 2001 ABS Freight Movements survey (Catalogue No. 9220.0), which detailed the tonnes moved by road between states and territories for the year ending 31 March 2001. The volumes transported from Victoria to the other states, and the volumes delivered to Victoria from the other states, were then scaled up to reflect volumes for 2007. The scaling factor used was the ratio between the 2001 total volume of road freight and the total volume for 2007 obtained from Table 17 of the ABS SMVU (2008).

8.3.2.4 Location of Processors, Supermarkets and Distribution Centres

The addresses of the largest seven F&V manufacturers, 800 supermarkets and their corresponding DCs and 540 greengrocers in Victoria were obtained from: (a) the listing of IGA stores (GPS POI 2010c); (b) the listing of Woolworth stores (GPS POI 2010d); (c) the listing of Coles stores (GPS POI 2010b); (d) the listing of ALDI stores (GPS POI 2010a); and (e) the Foodworks store locator (Foodworks 2010). The addresses listed were plotted using Google Maps to obtain their coordinates in latitude and longitude. In addition, the coordinates of the Melbourne Market Authority (MMA) (representing the location of major wholesalers) was also obtained.

8.3.2.5 Location of Consumers

Victoria had 9,298 collection districts in 2006, with an average of about 550 consumers in each. This value was used to represent the locations of households. The concentration of Victoria's population is highest in the Melbourne Metropolitan region.

With the huge volume of data collected and the variety of sources used, it was inevitable for gaps and inconsistencies to occur between the data sets. A number of assumptions were then applied to resolve these issues. Table 8.1 lists the eight principal sources of uncertainties.

8.3.3 General Estimation Procedure

The SCDT enables investigation of relative (rather than absolute) estimates of emissions, indicating the emissions produced from a base scenario based on one set of average values (e.g. average payload, average emissions factors, average distance). The estimates of emissions from the various supply chain legs for this base scenario were then aggregated into a collection of sub-totals (e.g. by F&V, by vehicle type) to enable comparison, i.e. relative contributions of different system attributes.

Table 8.1 Key sources of data uncertainties in the model

Assumption	Likely effect on emissions
Produce takes most "efficient" pathway from producer to consumer in terms of distance, i.e. moves from production to closest processor, DCs, retailers to meet requirement	Underestimate
F&V sourced for processing direct from production (not via MM) and proportional to production volumes for that region	Underestimate
Assuming that produce moves in the shortest road route in all cases	Underestimate
F&V as a proportion of interstate transport/amount of F&V moved interstate	Unknown
Proportion of vehicle types kept constant in different stages of the supply chain	Unknown
Payloads not differentiated by F&V type, i.e. tonne of potatoes requires same transport volume as tonne of lettuce	Unknown
Households would only travel to the nearest supermarket and grocery store to purchase F&V	Underestimate
Attributing all consumer trip emissions to F&V for grocery stores, but only 7.25 % to supermarkets	Likely bias towards supermarkets

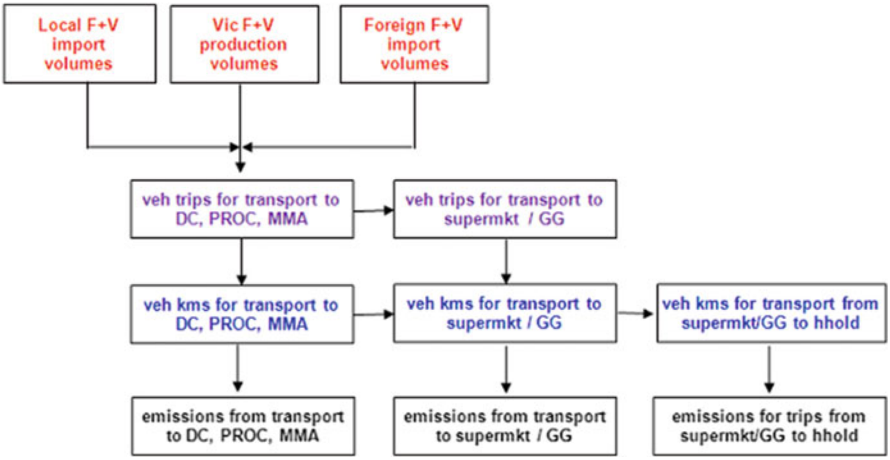


Fig. 8.4 SCDT menu for F&V emissions

Figure 8.4 shows the general procedure used in the SCDT for calculating emissions resulting from the transport of F&V volumes from growing areas to supermarkets to consumer.

The components of the supply chain contributing to emissions start with F&V volumes from Victoria's production regions (NRMs), along with interstate sources for local imports, and foreign ports for international imports. Using the capacities for different modes of transport (LCV, rigid trucks, articulated trucks, ships), the number of trips needed to transport the volumes to the different DC, processing centres (PROC) and the MMA are then estimated. The distances for the various origin–destination pairs (obtained from Goggle Maps) are then incorporated to produce the vehicle-kilometres travelled for these trips. The emission estimates are then obtained by applying the corresponding emission factors for the transport modes used. A similar procedure is applied to obtain the number of trips, vehicle-kilometres travelled and emissions produced when the F&V volumes are transported from the distribution points (DC, PROC and MMA) to the different supermarkets and green grocer locations.

Finally, emissions from household shopping trips are estimated using a number of assumptions on the frequency and distance of grocery trips and the proportion of F&V items purchased based on a representative grocery basket. Full details of the deterministic methodology and assumptions used in the SCDT can be found in the Marquez et al. (2010) report.

8.3.4 Calculating the GHG Emissions

The mathematical formulation of the calculation of GHG emissions for each supply chain leg (e.g. NRM region to DC, DC to supermarket, supermarket to consumer, foreign port to Australian port) for the SCDT was based on individual trips to transport fruit and vegetable items using specific vehicle types (LCVs, rigid trucks, container/ships).

8.3.4.1 Emissions from Road Transport

To estimate emissions from road transport, we define the indices used in the formulation as follows:

- Let $j \in J$ be an individual supply chain link such as between a specific NRM centroid and DC or between a supermarket and a consumer CD (collection district) centroid. It can include interstate and international legs as well.
- Let $v \in V$ be a road transport of a given type, mode and fuel category (e.g. articulated trucks using diesel fuel).
- Let $i \in I$ be an individual fruit or vegetable item (e.g. potatoes, apples, oranges).

If $M_i^{v,j}$ is the total volume (in tonnes) of F&V item $i \in I$ transported over supply chain link $j \in J$ using vehicle $v \in V$ in a given year, then the amount of corresponding emissions produced $E_i^{v,j}$ (in kg) is given by 8.1):

$$E_i^{v,j} = \left(M_i^{v,j}/P^v\right) \times D^j \times (\lambda^v \times F^v \times \alpha^v \times R^v \times \beta^v \times B^v) \tag{8.1}$$

where

- P^v is the average payload (in tonnes) of vehicle type, $v \in V$
- D^j is the distance travelled (in km) on supply chain link, $j \in J$
- λ^v is the multiplier used to account for the combined weight of the vehicle and load
- F^v is the emissions factor for the forward component of the trip (in kg/km) for vehicle type $v \in V$
- α^v is the proportion of trips made by vehicle type $v \in V$ that are refrigerated
- R^v is the emissions factor for the refrigeration component of the trip (in kg/km) for vehicle type $v \in V$
- β^v is the proportion of trips made by vehicle type $v \in V$ that have backhaul
- B^v is the emissions factor for the backhaul component of the trip (in kg/km) for vehicle type $v \in V$. This is usually the same value as F^v

Note that $(M_i^{v,j}/P^v)$ gives the number of trips required to transport the volume while $(M_i^{v,j}/P^v) * D^j$ gives the vehicle-kilometres covered. The emissions formula merely multiplies the vehicle-kilometres covered with the emissions factors from the three components of the trip (forward-delivery, refrigeration and backhaul).

The database model provides the following input tables for the vehicle parameters:

- P^v is given in column *Average* of table *Q_VehPayload*.
- λ^v is given in column *FullLoadMult* of table *P_EmissFactors*. This is currently 1.50 which sets the average weight of the load as 50 % of the weight of the vehicle.
- F^v is given in column *EmissKGPerKm* of table *P_EmissFactors*.
- α^v is given in column *FreqPropn* of table *Q_VehFreqOfRefrigTrips*.
- R^v is given in column *RefrigEmissFctr* of table *P_EmissFactors*.
- β^v is given in column *FreqPropn* of table *Q_VehFreqOfBackhaulTrips*.
- B^v is given in column *BHaulEmissFctr* of table *P_EmissFactors*.

8.3.5 Emissions from Shipping

Similarly, emissions from shipping are estimated by (8.2):

$$S_i^{x,y} = \left[M_i^{x,y} * D^{x,y} * \sigma^C\right] + \left[\left(M_i^{x,y}/P^C\right) * D^{x,y} * \left(\delta^C * \omega^C * \phi^C / \psi^C\right)\right] \tag{8.2}$$

where

- $S_i^{x,y}$ is the total emissions (in kg) from shipping F&V item $i \in I$ between Australian port $x \in X$ and partner country $y \in Y$.
- $M_i^{x,y}$ is the total volume (in tonnes) of F&V item $i \in I$ shipped between Australian port $x \in X$ and partner country $y \in Y$.

- $D^{x,y}$ is the voyage distance (in km) between Australian port $x \in X$ and partner country $y \in Y$.
- P^C is the average payload for 20-ft containers (currently 21 tonnes).
- σ^C is the emissions factor for container (ship) movement (currently 0.014 kg CO_2-e per tonne-km).
- δ^C is the average fuel consumption of the refrigeration unit of the container (currently 300 g-MDO per kWh).
- ω^C is the emissions factor for the refrigeration unit of the container (currently 0.003206 kg CO_2-e per g-MDO. MDO stands for marine diesel oil, the fuel used by generators in refrigerated containers).
- ϕ^C is the average power consumption of the refrigeration unit of the container (currently 4 kW).
- ψ^C is the average speed of the container/ship (currently 38 km per hour).

In (8.2), the first term computes the emissions from the container/ship movement, while the second term calculates the emissions from the refrigeration component.

8.3.6 Aggregations

Using the above notation, we obtain the following basic aggregations:

- Total emissions per year $= \sum_{j \in J} \sum_{i \in I} \sum_{v \in V} E_i^{v,j}$
- Total emissions on supply chain link $j \in J$ for F&V item $i \in I = \sum_{v \in V} E_i^{v,j}$
- Total emissions on supply chain link $j \in J = \sum_{i \in I} \sum_{v \in V} E_i^{v,j}$
- Total emissions from vehicle type $v \in V = \sum_{i \in I} \sum_{j \in J} E_i^{v,j}$
- Total emissions per tonne transported $= \left(\sum_{j \in J} \sum_{i \in I} \sum_{v \in V} E_i^{v,j} \right) \Big/ \left(\sum_{j \in J} \sum_{i \in I} \sum_{v \in V} M_i^{v,j} \right)$

The component nature of the calculations allows more detailed aggregations to be made based on combinations of:

- Supply chain links
- Vehicle types
- Item types
- Specific fruits and vegetables
- Processed and fresh produce
- Victorian production, interstate imports and international imports

- Supermarkets and grocery stores
- Household collection districts
- (Export/import) partner countries

The above formulas can also be expanded to incorporate emissions per month in the case of seasonal effects in state production and interstate trade.

8.4 Modelling Scenarios and Results

8.4.1 Base Scenario

For the case study, a base scenario of fruit and vegetable freight movements was defined using volumes from fiscal year 2007–2008 and involving seven fruit items (apples, grapes, mandarins, oranges, peaches, pears, strawberries) and 28 vegetable items (artichokes, Asian vegetables, asparagus, beetroot, broccoli, Brussels sprouts, butter beans, cabbages, capsicums, carrots, cauliflower, celery, chillies, cucumbers, eggplant, fennel bulb, french and runner beans, garlic, herbs, leeks, lettuce, melons, mushrooms, onions, parsnips, peas, potatoes, pumpkins, radish, silver beet and spinach, snow peas, spring onions, swedes and turnips, sweet corn, tomatoes, watermelons, zucchini and button squash). The base scenario represents the supply chains under the regional crop production levels and transport of 2007/2008.

The scenario covers food transport between the National Resource Management (NRM) regions and export points, DCs for the four MSC (Coles, Woolworth, IGA/Foodworks, Aldi), major food PROC (Simplot, McCain, National Foods, SPC), MMA and listed grocery stores. It also covers imported produce from overseas and the customers' trip to collect food (also known as "the last mile").

Figure 8.5 shows a bar chart comparing the emissions from the top eight countries exporting F&V to Victoria. Notice that most of the high emission sources come from Western and Eastern Europe, where the sea distance covered is highest. As one would expect, GHG emissions are proportional to distance travelled.

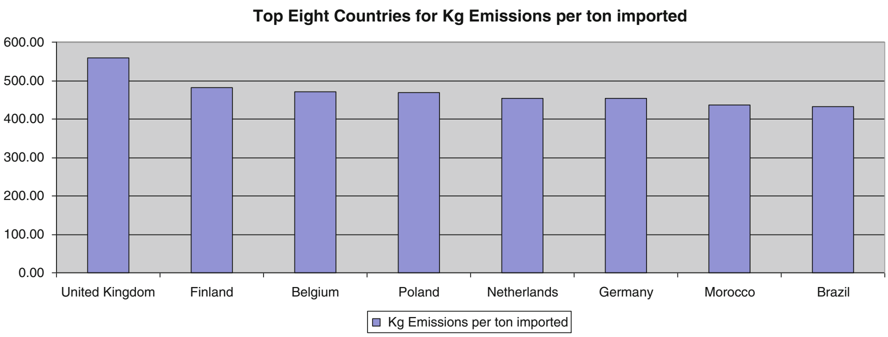


Fig. 8.5 Source countries with highest emissions per tonne of imported F&V

Table 8.2 Comparison of GHG emissions between Victorian and non-Victoria produced F&V

Product type	Destination type	kg of emissions per tonne transported		
		Vic produced	Foreign imported	Interstate supplied
Fruit	Victoria DC	59.57	248.02	215.70
Vegetable	Victoria DC	48.73	216.22	222.45
All	Victoria DC	51.94	222.89	220.62

Table 8.2 compares the average GHG emissions per tonne of Victorian-grown F&V consumed in Victoria compared to F&V sourced from interstate and imported from overseas. As Table 8.2 shows, imported fruits and vegetables showed the highest overall transport emission levels (248 and 216 kg CO_2-e per tonne, respectively) while interstate supplies were not far behind with overall emissions of 221 kg CO_2-e per tonne. Both international and interstate transport presented an emissions profile of almost four times the levels obtained for Victorian-grown fruits and vegetables (60 and 49 kg CO_2-e per tonne, respectively). These results can be attributed to the significantly longer distances required to connect foreign ports and state capitals to Victorian DCs compared to points within Victoria for Victorian-grown fruits and vegetables, particularly the road segments.

Emissions from vegetables were highest for interstate transport while emissions from fruits were highest for international imports. An interesting observation is that the international shipping leg had similar GHG emissions to the interstate road transport leg.

Although international imports produced the highest emission levels of 248 kg per tonne of fruit imported, these emission levels only represent transport (including refrigeration) from the port of departure from the partner country to the port of entry in Australia and ultimately to Victorian DCs. Transport GHG emissions between the overseas growing region and the port of departure in the foreign country are not included. Therefore, these results are likely to underestimate the true values. As noted earlier, the land leg portion of international transport is considerably shorter than interstate transport since most foreign imports use the Port of Melbourne as the port of entry into Victoria, and Victoria DCs are already close to the Port of Melbourne.

Table 8.3 presents the intrastate transport emissions for 28 F&V items delivered to DCs. For the intrastate emissions obtained, the items are presented in descending order of emissions per $1,000 value (at farm gate). The second column of Table 8.3 shows the GHG emissions per tonne of item transported. There is more than a fourfold difference between items with the highest (grapes, 97) and lowest (celery, 18) GHG emissions per tonne. Differences in distances between the growing region and MM/DCs were the main driver of the differences. For example, oranges, mandarins and grapes were amongst the highest as they are primarily grown in Mallee, the furthest NRM region from MM/DCs. By considering product value, the order changes a bit, as F&V with a lower dollar value will have a higher GHG emission per dollar value ratio.

Table 8.3 GHG emissions of 28 F&V items as a function of volume transported within Victoria and farm gate value

Items	Emissions (kg) per tonne transported	Emissions (kg) per 1,000 dollars value
Watermelons	93.64	160.35
Oranges	92.38	159.27
Grapes	97.24	107.68
Carrots	77.86	106.81
Potatoes	43.03	89.64
Melons	93.63	83.75
Tomatoes	41.31	78.69
Sweet corn	67.20	67.88
Mandarins	97.38	59.74
Pears	38.68	56.88
Pumpkins	37.45	56.23
Onions	47.17	49.97
Beetroot	18.16	46.91
Cabbages	42.43	44.20
Peaches	42.71	35.01
Lettuce	32.11	34.97
Capsicums	72.40	32.18
Cauliflower	22.89	32.05
Zucchini and button squash	43.23	30.66
Parsnips	19.71	19.97
Apples	38.49	18.59
Celery	17.56	18.48
Broccoli	31.93	15.89
Cucumbers	21.96	8.38
Asparagus	22.04	4.87
Asian vegetables	19.02	4.20
Mushrooms	26.94	4.15

With the base scenario measures in place, new scenarios can be created in the SCDT by incorporating new data or values for the various parameters representing alternative policy options or transport strategies. The differences between the measures obtained in the new scenario and those from the base scenario are then used to evaluate the relative impact of the new options or strategies on GHG emissions. For investigating the impact of Victoria's vulnerability to EWE increases to GHG emissions from the F&V transport, a EWE scenario was defined and tested in the next subsection.

8.4.2 Victorian EWE Scenario

In this scenario, we investigate the GHG impacts of lost production in Victoria's NRM regions caused by EWE. The analysis simulates losses in production due to bushfires, storm damage and droughts and estimates the increase in GHG levels resulting from the transport of replacement produce. Under normal circumstances, when production is lost in one or more NRM regions, the shortfall in production to meet Victoria's demand (consumption) requirements would be met through other NRM regions in Victoria (as distinct from international imports). Furthermore, when there is insufficient supply from other NRM regions, the shortfall would be compensated by interstate imports, where the GHG emissions are expected to increase due to lengthier freight movements and increased fuel use. In this subsection, we focus on the use of interstate imports as the primary source of replacement for production losses. We believe a basic scenario where lost production from an NRM is replaced by interstate imports can provide important insight into the efficiency of alternative sources of supply during EWE. We also identify the NRM regions most vulnerable to increased GHG emissions resulting from EWE. Included in Victoria's "consumption" were the F&V demand by PROC, as the inability to meet demand locally (or reliably) will impact on their costs of sourcing produce (at some oil / carbon price point, there could be a significant impact).

A representative scenario was set up to investigate the impact on emissions of switching the source F&V volumes from an NRM to interstate. The scenario simulates the condition where 25 % of the annual production of an NRM is lost and the lost volume is replaced by corresponding additional supply from interstate, with each state given equal allocations. This condition may be produced by an extreme weather event causing damage to production areas in an NRM with recovery taking three months or more. The increased distance in transporting the replacement volumes between the interstate capitals and Victorian DCs would result in increased emissions, with the total amount dependent on the actual volumes replaced.

Figure 8.6 presents a bar chart of the estimated overall percentage increase in total emissions from the base scenario when the NRMs lose 25 % of their production and the corresponding replacement volumes are sourced from interstate to the DCs and MM. The line chart displays the volume of lost production re-sourced from interstate. The four NRMs with the highest volumes of production also have the highest replacement requirements if 25 % of production is lost, i.e. Goulburn/Broken (76 M-kg), Port Philip/Westernport (70 M-kg), Mallee (58 M-kg) and North Central (54 M-kg). Consequently, these four NRMs produced the biggest impact on emissions from lost production. The figure shows that the most significant increases on fruit emissions is produced by production loses from Goulburn/Broken (6.24 %), followed by Mallee (3.29 %) and Port Philip/Westernport (1.10 %). For vegetable emissions, the highest increases came from Port Philip/Westernport (2.99 %), North Central (2.20 %) and East Gippsland (1.56 %).

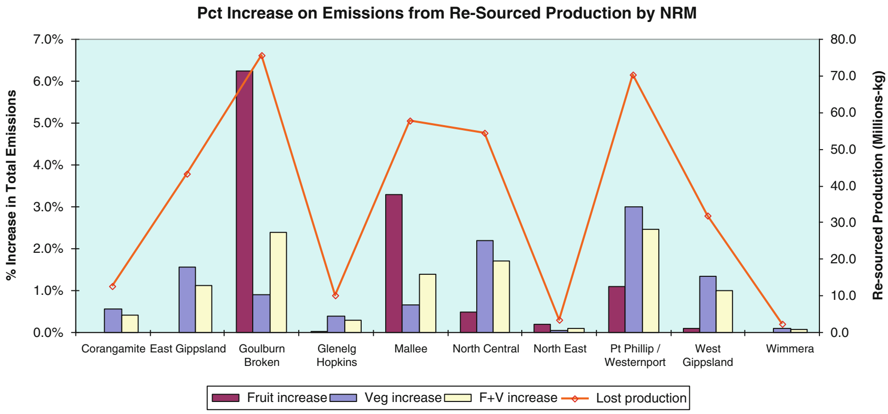


Fig. 8.6 Impact of lost production from NRM on total emissions from transport of F&V

In terms of overall emissions, Port Philip/Westernport (2.46 %), Goulburn/Broken (2.38 %) and North Central (1.72 %) represent the NRMs with the biggest impact on Victoria's fruit and vegetable supply chain. Overall, these do not represent significant gains in GHG emissions (mostly less than 4 %) given a significant loss of production of 25 % in a region. However, this finding is an average based on the simulated volume replacements from interstate sources for lost production from one affected NRM at a time, assuming the remaining NRMs are unaffected. Replacement volumes (equalling the volumes lost) will come from the other states (through their import/export point) in equal proportion. Thus, the gain in GHG is proportional to the lost volume and takes into account the (average) interstate distances travelled by replacement produce.

8.5 Opportunities for Supply Chain Optimisation

There is an extensive literature on optimising food supply chains for efficiencies or costs. See Ahumada and Villalobos (2009), Lucas and Chhajed (2004) and Higgins et al. (2010) for extensive reviews. Limited studies (e.g. Tan and Comden 2012; Romero 2000) consider biophysical and climatic uncertainty, which may affect yield or best time to harvest. However, chain optimisation to increase resilience to EWE is an important issue that has not been addressed in the literature. There is also a need to consider a trade-off between the severe impacts of infrequent EWE versus small impacts of regular yield uncertainty. With such high uncertainty, Pannell (2006) recognises the flat earth economic problem where the benefits of mathematical optimisation have low sensitivity to changes to decision variables. This is particularly an issue with smaller scale operational decisions such as a transport route of an individual vehicle, and tactical decisions such as hectares of different crops that a farmer plants in a given year. For strategic decisions, such as where to place DCs or new growing regions for F&V, there are opportunities for mathematical optimisation to increase resilience to EWE.

Archer et al. (2009) formulates a complexity matrix for supply chains, where opportunities to increase resilience (typically strategic) face both high biophysical and managerial complexity. Archer et al. (2009) goes on to describe the use of agent-based methods to capture the dynamic relationships under high uncertainty (such as EWE and market instability) and described applications in forestry, grains, sugar and canola. In the case of F&V, agents would potentially represent chain actors (farmers, distributors, processors, retailers) along with the relationship between these actors and those currently outside the network (e.g. interstate or overseas suppliers). By linking such an agent-based model with our SCDT, one could optimise contingency F&V supplied between chain actors to minimise costs and GHG emissions or meeting similar demands under a range of EWE scenarios.

Probably, the best optimisation application to maximise the resilience of the F&V industry to EWE is in land use allocation of different crops. The main problem of the Australian banana industry with respect to Cyclone Larry and Yasi was that 90 % of the industry was within 50 km of the Tully region of Queensland, where the cyclones hit hardest. A land use profit map (Marinoni et al. 2012) would show that these are the best regions to grow bananas under normal climatic and soil conditions. When considering the severe impacts of EWE and GHG emissions, other less profitable regions (e.g. Coffs Harbour) could be part of an optimal land use allocation zone for bananas. Models for land use optimisation of different agriculture/horticulture crops are available in the literature (e.g. Rounsevell et al. 2003). These typically optimise the mixture of crops and rotation strategies at each location to maximise combined profitability, subject to resource constraints and climate and land suitability. A future opportunity would be in combining a land use planning model with our SCDT. It would provide a more holistic approach, over time and space, to optimising land allocation of different crops, accounting for alternative supply chain paths in the event EWE and consequential transport costs and GHG emissions.

8.6 Conclusion

With Australia's recent catastrophic flooding and bushfires resulting in considerable damage to life and property, inflicting massive losses to regional economies and causing serious disruptions and price increases in the food supply chain, attention has been focused on tools for analysing and mitigating the impacts of these EWE. This chapter described the development of the SCDT, a simulation and scenario evaluation tool aimed at estimating the impact of transport efficiency measures on the level of GHG emissions produced by various legs in the supply chain. The SCDT was employed in a case study, where a hypothetical EWE scenario is applied to the Victorian F&V supply chain. A base scenario featuring the distribution network of 7 fruits and 28 vegetables consumed in Victoria was defined using 2007–2008 levels of F&V production and imports. A representative EWE scenario was then set up to simulate the condition where 25 % of the annual production of an NRM is lost, and the lost volume is replaced by corresponding additional supply from interstate, with each state given equal allocations.

The results of the case study demonstrate that simulation in general, and the SCDT in particular, can be an effective tool for evaluating emission reduction measures in the food supply chain. The SCDT also helps in understanding the complex interactions between the different components of the supply chain. While the emission estimates produced by the SCDT provide a useful indicator of the relative impacts of transport measures, several data availability and consistency issues restrict the range of analyses that can be performed. Due to the severe lack of suitable data on interstate transport, many questions around the seasonal implications of GHG emissions remain unanswered. This includes better understanding of the inefficiencies of interstate (and intrastate) F&V GHG emissions at a more granular scale such as individual trip movements and companies. Such a more detailed analysis would help identify more tangible strategies to reduce GHG emissions. To do this, the key data requirements are: tonnes of individual F&V transported between each state in each week, as well as movements at company scale. If the latter will be impossible to obtain as complete data sets, then we recommend it be collected in part through surveys.

Although GHG emissions of transport to and from the processor were considered, a lot more information is needed on the freight movements to provide an accurate analysis. In particular, data of F&V from different states or overseas for processing is much needed. Also, GHG emissions of activities within the processor need to be considered to provide a balanced comparison with fresh F&V supply chains.

Acknowledgements The authors acknowledge the support of Kirsten Larsen and the Victorian Eco-Innovation Laboratory (VEIL) for funding the study "Understanding Victoria's Fruit and Vegetable Freight Movements", where part of the research presented in this article was drawn.

References

ABARES (2011) The impact of recent flood events on commodities. ABARES special report, 21 January 2011, ABARES. http://adl.brs.gov.au/data/warehouse/pe_abares99001773/Floods_on_commodities_2011_REPORT.pdf

ABS SMVU (2008) 9208.0 survey of motor vehicle use. Australian Bureau of Statistics, pp 1–48

AFI (2009) FarmGAS calculator to help farmers understand greenhouse emissions. Australian Farm Institute, 4 August 2009. http://farmgas.farminstitute.org.au/publicpages/AFIPublic.aspx?ReturnUrl=/default.aspx. Accessed Feb 2012

Ahumada O, Villalobos J (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 195:1–20

Archer A, Higgins A, Thorburn P (2009) A method for comprehending and adapting complex supply chains in agriculture. J Chain Netw Sci 9(1):9–15

Cairns Post (2009) Fruit left to rot. The Cairns Post, 10 February 2009. http://www.cairns.com.au/article/2009/02/10/27835_local-business-news.html. Accessed Feb 2012
CSIR (2005) Second annual state of logistics survey for South Africa: defining research priorities for developmental logistics. CSIRO Build Environment. ISBN 0-7988-5559-2
DPI (2012) On-farm greenhouse gas accounting tools. Victorian Department of Primary Industries. http://www.dpi.vic.gov.au/agriculture/farming-management/weather-climate/under standing-carbon-and-emissions/frequently-asked-questions/on-farm-greenhouse-gas-accounting-tools. Accessed Feb 2012
Food Magazine (2009) Natural disasters destroy food crops. Food Magazine, 9 February 2009. http://www.foodmag.com.au/Article/Natural-disasters-destroy-food-crops/437366.aspx. Accessed Feb 2012
Foodworks (2010) Foodworks Store Locator, FoodWorks website. http://www.foodworks.com.au/store. Accessed Feb 2010
GPS POI (2010a) ALDI stores in Australia. GPS POI website. http://www.gps-data-team.info/poi/australia/shopping/ALDI_au.html. Accessed Feb 2010
GPS POI (2010b) Coles stores in Australia. GPS POI website. http://www.gps-data-team.info/poi/australia/shopping/Coles.html. Accessed Feb 2010
GPS POI (2010c) IGA stores in Australia. GPS POI website. http://www.gps-data-team.info/poi/australia/shopping/IGA_Store-AU.html. Accessed Feb 2010
GPS POI (2010d) Woolworths stores in Australia. GPS POI website. http://www.gps-data-team.info/poi/australia/shopping/Woolworths.html. Accessed Feb 2010
Higgins A, Miller C, Archer A, Ton T, Fletcher CS, McAllister R (2010) Challenges of operations research practice in agricultural value chains. J Oper Res Soc 61:964–973
Higgins A, Estrada-Flores S, Singh G, Ton T (2011) The state of logistics in the Australian food industries. ASOR Bull 30(1):24–41
Hume A, Barber A, East A, McLaren S, Deurer M, Clothier B, Palmer J (2009) Carbon footprinting for the apple supply chain—report on methodology and scoping study. Landcare research contract report LC0809/157, for MAF and Pipfruit New Zealand
Larsen K, Ryan C, Abraham A (2008) Sustainable and secure food systems for Victoria: what do we know? What do we need to know? Victorian Eco-Innovation Lab (VEIL), pp 1–152
Lucas M, Chhajed D (2004) Applications of location analysis in agriculture: a survey. J Oper Res Soc 55:561–578
Marinoni O, Navarro Garcia J, Marvanek S, Prestwidge D, Clifford D, Laredo L (2012) Development of a system to produce maps of agricultural profit on a continental scale: an example for Australia. Agric Syst 105:33–45
Marquez L, Higgins A, Estrada-Flores S (2010) Understanding Victoria's fruit and vegetable freight movements. Final report submitted to the Victorian Eco-Innovation Lab
McKinnon AC, Piecyk MI (2009) Measurement of CO_2 emissions from road freight transport: a review of UK experience. Energy Policy 37:3733–3742
Mithraratne N, McLaren SJ, Barber A (2008) Carbon footprinting for the kiwifruit supply chain. Report on methodology and scoping study, Landcare Research. Landcare Research contract report LC0708/156
Morgan E (2009) Fruit and vegetable consumption and waste in Australia. Victorian Health Promotion Foundation, Carlton South
O'Halloran N, Fisher P, Rab A (2008) Options for mitigating greenhouse gas emissions for the Australian vegetable Industry. Discussion paper 6. Horticulture Australia Limited project VGO8107
Pannell D (2006) Flat-earth economics: the far-reaching consequences of flat payoff functions in economic decision making. Rev Agric Econ 28(4):553–566
Parliament of Australia (2008) Climate change. More extreme weather. http://www.aph.gov.au/library/pubs/climatechange/theClimate/moreExtreme.htm. Accessed Apr 2011
Rab MA, Fisher P, O'Halloran N (2008) Preliminary estimation of the carbon footprint of the Australian vegetable industry. Horticulture Australia Limited project VGO8107

Romero C (2000) Risk programming for agricultural resource allocation: a multidimensional risk approach. Ann Oper Res 94(1–4):57–68

Rounsevell M, Annetts J, Audsley E, Mayr T, Reginster I (2003) Modelling the spatial distribution of agricultural land use at the regional scale. Agric Ecosyst Environ 95:465–479

Royal Commission into the Grains Storage, Handling and Transport (1988) Australia Government Publishing Service, Canberra. Vols 1–3. ISBN 0 644 0723 2 6

Schmitt AJ, Singh M (2009) Quantifying supply chain disruption risk using Monte Carlo and discrete event simulation. In: Proceedings of the 2009 winter simulation conference, pp 1237–1248

Tan B, Comden N (2012) Agricultural planning of annual plants under demand, maturation, harvest and yield risk. Eur J Oper Res 220(2):539–549

Ugalde D, Brungs A, Kaebernick M, McGregor A, Slattery B (2007) Implications of climate change for tillage practice in Australia. Soil Tillage Res 97(2):318–330

Van der Vorst J, Tromp S, van der Zee DJ (2009) Simulation modelling for food supply chain redesign: integrated decision making on product quality, sustainability and logistics. Int J Prod Res 47(23):6611–6631

Van Dyk FE, Maspero E (2004) An analysis of the South African fruit logistics infrastructure. ORiON 20(1):55–72

Victorian Department of Transport (2008) The Melbourne metropolitan freight movement task. Policy and Communications Division

William D, Friedlingstein P, Amadou TG, Gregory J (2007) Global climate projections. In: Meehl G, Stocker TF (eds) IPCC 4th assessment report, pp 747–845

Chapter 9
Simulation Optimization: Applications in Fish Farming—Theory vs. Practices

Ilan Halachmi

Nomenclature

Simulation optimization problem general form

Minimize	$F(x)$	(Objective function)
Subject to:	$Ax \leq b$	(Constraints on input variables)
	$g_l \leq G(x) \leq g_u$	(Constraints on output measures)
	$l \leq x \leq u$	(Bounds)

where $F(x)$ and $G(x)$ represent output performance measures *obtained from the simulation*. The constraints represented by inequality $Ax \leq b$, and both the coefficient matrix A and the right-hand-side values corresponding to vector b are known. The constraints represented by inequalities of the form $g_l \leq G(x) \leq g_u$ impose simple upper and/or lower bound requirements on an simulation output function $G(x)$ that can be linear or nonlinear. The values of the bounds g_l and g_u are known constants. The vector x is the decision variable that includes continuous and discrete values. All decision variables x are bounded. Each evaluation of $F(x)$ and $G(x)$ requires an execution of a simulation of the system.

B_f	Final body weight of a fish (kg)
B_i	Body weight of a fish (kg) in growing phase i
c	Number of netcages

I. Halachmi, Ph.D. (✉)
The Institute of Agricultural Engineering, Agricultural Research Organization ARO,
The Volcani Center, P.O. Box 6, Bet Dagan 50250, Israel
e-mail: halachmi@volcani.agri.gov.il;
http://www.agri.gov.il/en/people/716.aspx;
http://scholar.google.com/citations?user=PMfuwcIAAAAJ;
http://il.linkedin.com/pub/ilan-halachmi/58/10/221

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_9

c_i	Number of netcages in growing phase i e.g., $c_1, c_2, c_3 = 4, 8, 16$ stands for 4 netcages in the first growing phase 8 netcages in the second growing phase, and 16 netcages in the third phase
D	Fish biomass stocking density (kg/m^3)
D_i	Biomass in growing phase i
DOE	Design of experiment (Kleijnen 2008)
MR	Mortality rate. The number of fish at each phase: $N_{i+1} = N_i \times (1 - \text{MR})$. Survival rate = 1 − MR
$N, c_1, c_2, c_3, S_1, S_2, S_3, P_1, P_2, P_3$	The decision parameters
N_f	Final number of fish in a batch
N_i	Number of fish in a batch in growing phase i
P	Number of sub-batches formed from a batch
RSM	Response surface methodology (Kleijnen 2008)
S	Growth period in a netcage (years)
S.T.	"Subject to the constraints." This term refers to extrema with constraints in mathematical optimization
Si	Growth period in growing phase i
T	Fish age (days in sea)
V	Culture volume (m^3)
V_i	Culture volume in growing phase i a culture volume is a manmade water tank pond or netcage, made of plastic, concrete, soil, etc. (inland aquaculture) or made of a net and located in the sea (marine fish farming), lakes, rivers, seaports, or offshores, e.g., in Ashdod harbor 18 netcages of 2,900 m^3 each and 11 netcages of 2,000 m^3 each can fit along the breakwater. Total number of netcages is 29
$y(t)$	Fish body weight on any given day
λ and μ	Fish arrival and departure rates, respectively (batches/year)
ρ	Expected utilization of a netcage

9.1 Introduction

Aquaculture continues to be the fastest growing animal food-producing sector and to outpace population growth. Per capita supply from aquaculture increased from 0.7 kg in 1970 to 7.8 kg in 2006, an average annual growth rate of 6.9 %. From a production of less than one million tons per year in the early 1950s, production in 2006 was reported to be 51.7 million tons with a value of US$78.8 billion, representing an annual growth rate of nearly 7 %. Increased yields were obtained as a result of intensification, advanced feed formulations, water chemistry, disease prevention, fish treatment, and genetic selection for desirable traits. However, the aquaculture industry has realized that economic viability of aquaculture systems cannot be ensured solely through increased yields. To seek an economically viable solution, the complexity of the system should be conceptualized to consider the interactions among many decision variables and biological factors.

The immense leaps in computational power have greatly benefited both optimization and simulation. Now, large-scale simulation "optimization" routines can be performed on PCs in a fraction of time, comparing with only few years ago. Therefore, nearly every commercial simulation software packages have now included a sort of "optimization." [AutoStat (AutoMod; www.autosim.com)—genetic algorithms; OptQuest (Arena, Crystal Ball, et al.; www.opttek.com)—scatter search and tabu search, neural networks; OPTIMIZ (SIMUL8; www.simul8.com)—neural networks; SimRunner (ProModel; www.promodel.com)—genetic algorithms; Optimizer (WITNESS; www.lanner.com/corporate)—simulated annealing, tabu search—the interested reader can refer to Fu (2002) and Fu et al. (2002)].

However, optimization procedures such as linear programming, nonlinear programming, and (mixed) integer programming—require an explicit mathematical formulation. Such a formulation is generally impossible for problems, where simulation is relevant.

Contrary to the use of mathematical programming software packages, the simulation user has no way of knowing if a global optimum has actually been reached (hence, the quotations around optimization at the beginning of this paragraph). Optimizers designed for simulation embody the principle of separating the optimization method from the simulation model. In such a context, the optimization problem is defined outside the complex system.

Stochastic discrete-event simulation (the so-called simulation) mimics the random spirit of a system. The simulation's (a) stochastic nature, (b) ad hoc heuristic tools, and (c) random numbers generators—do not adequately addressed by the currently implemented optimization algorithms (Fu 2002; Fu et al. 2000; Marco XX).

Therefore, in the aquaculture context, a method is proposed as follow steps:

(a) Build a valid simulation model of the aquaculture system. (For a complete validation process, refer to Law (1990) and
(b) Generate initial guess, "optimal" solution based on classic optimization methods, the known bio-physiology of the specific fish growth function, aquatic conditions, the environment, and the local management practices of the farm (9.1)–(9.9) below)
(c) Feed the "optimal" solution from step (b) and (a) wide mesh of nodes around the solution (based on response surface methodology (RSM) Sect. 9.3.6 DOE and RSM below and (Kleijnen 2008))
(d) Run the simulation, get simulation outputs, fit an objective function ($f(x)$), and if needed fit constraints function ($g(x)$) (RSM equations—refer to Sect. 9.3.6 DOE and RSM below and (Kleijnen 2008)). In our two cases presented here below, 100 scenarios were simulated at this phase
(e) Solve the optimization problem ($f(x)$, $g(x)$ from step (d), check the feasibility of the solution from practical point of view and only if "make sense"—feed the new optimal solution to the simulation model
(f) Compare the previous solution and the new one, stop the iterations when no considerable improvement has been reached

(g) Enrich the simulation with animation and user friendly reports, and validate the simulated solution with data from similar systems (if exist?) and with panel of experts and with the farm workers who are familiar with the system. For complete validation process, refer to Halachmi et al. (2001)
(h) Discuss the proposed solution with the farm management while running simulated scenarios during the meeting till an agreed design has being selected

The above (c)–(f) scheme is presented by Fig. 9.2.

The differences between the proposed method (Fig. 9.2) and the classical simulation–optimisation (Fig. 9.1) are: (a) human inspected selection of a fitted equation that guarantees that the global optimization point is at reached (up to order

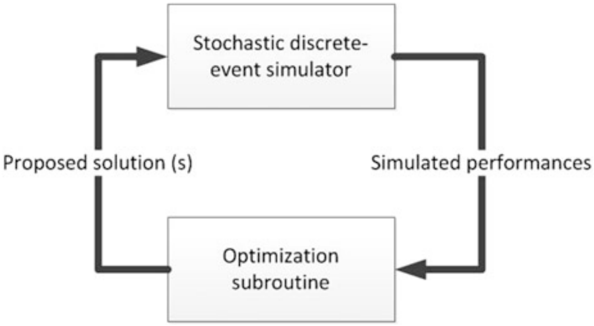


Fig. 9.1 "Classic" optimization for simulation, in nearly every commercial software the simulation kernel engine is separated from the optimization engine

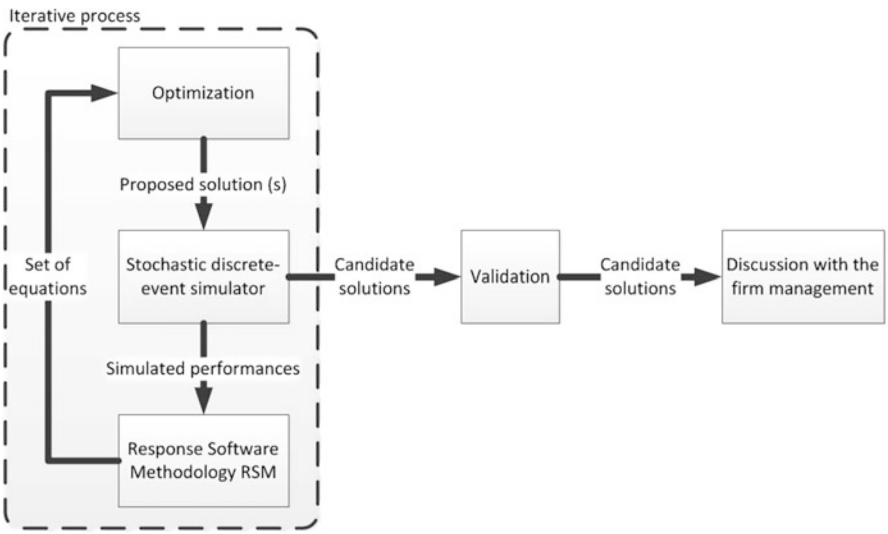


Fig. 9.2 The proposed simulation–optimization procedure. Combining bio-physiology knowledge about the aquaculture system that brings us closer to the global optimal point, then mesh around the global optimum, RSM-based objective function, fine tuning until satisfactory solution has being reached

three polynomial, KKT conditions, etc. see Sect. 9.3.6 RSM). (b) Human inspection of the practical feasibility of a solution generated by the optimization engine before fed into the simulation kernel, and (d) the other way around—human inspection of the practical feasibility of the simulation output before fed into the RSM equations and the optimization engine.

The drawback of the proposed method (Fig. 9.2) is that a designer must be well acquainted with (9.1) the existing and the planned aquaculture system, ins-and-outs (fish biology, biomass density, potential aquatic conditions the will allow a proposed growth function, and physical limitation of the filters and the space) in order to set the correct constraints and boundaries and to reject most of nonpractical solutions at early phase before they enter into the simulation kernel. (2) Such a designer should be able to perform all the tasks, by himself, in-house: (a) constructing the simulation model, (b) validate and alter the simulation model, (c) to formulate the set of optimization functions and to solve them, (d) design and execute the RSM methodology including its all design of experiment (DOE) aspects. (3) Frequently, to perform 2b–2d on-the-spot, at the fish farm facility while discussing selected simulated alternatives with the farm workers and farm managers as they talk.

This book chapter describes the development of Fig. 9.2's equations and procedures and reports their application in two aquaculture farms. While the two farms' numerical values, presented below, reflect local aquatic conditions, the concept (the equations and the methodology) may be applicable elsewhere.

9.2 The State of the Art

9.2.1 Aquaculture

This study address to types of production systems: (a) recirculating aquaculture system (RAS) and marine netcages.

Fish production in an RAS requires less than 10 % of the water (Hamlin et al. 2008) needed by extensive pond systems to produce a given quantity of fish. RAS reduces the effluent waste stream by a factor of 500–1,000 (Timmons et al. 2001). Furthermore, an RAS enables production to proceed at the consumer's "doorstep" (Timmons et al. 2001), and with a high degree of product traceability (Smith 1996; Jahncke et al. 2000). This system also ensures higher levels of biosecurity, and enables year-round production. Unfortunately, (a) an intensive RAS production is strongly dependent on high water quality and minimal fish stressors. (b) The RAS is rather capital intensive (O'Rourke 1996); its profitability relies on maximizing economic productivity per unit volume of rearing water (Brune et al. 2003; Summerfelt et al. 2009). Thus, operating an RAS demands well-coordinated management of the many unit processes and/or operations involved

(Summerfelt 1996; Libey and Timmons 1996). Timmons et al. (2001) lists several RAS ventures that failed, citing poor management as the primary reason.

Fish farming in netcages is a traditional practice for raising fishes. Fish netcages of various shapes and dimensions have been used all over the world. In general, square or rectangular cages are widely used for farming of yellowtail (Harada 1970; Fujiya 1979), salmonids (Møller 1979; Kennedy 1975), and groupers (Chua and Teng 1978). Cylindrical cages are used for marine or brackish water species such as milkfish (Yu and Vizcarra 1979) and rainbow trouts (Tatum 1973). Cylindrical cages can be designed to rotate so as to delay change of nets due to biofouling (Caillouet 1972). Other forms of cages such as orthogonal (Milne 1979; Anonymous 1976) and octagonal (Møller 1979) in shape have been used for salmonid culture in United Kingdom, Norway, and France. The design of the physical structure of a cage is determined by the oceanographic conditions of the culture site and the target species. Each design is site specific and knowledge of the topography, wind force and direction, prevalence of storms or monsoons, wave loads, current velocity, and water depths are important parameters for consideration.

Refer to Weintraub et al. (2007) for operational research (OR) models in both types of fisheries.

9.2.2 Simulation Optimization

Simulation optimization is providing solutions to important practical problems previously beyond reach—optimization via simulation. Refer to Fu (1994, 2002), Fu et al. (2000), Better et al. (2008) for review.

9.2.3 Simulation, Optimization, and Statistical Models in Aquaculture

Modeling aquaculture a fish farm operation and harvesting time requires both equations describing fish growth, and algorithms determining the optimal harvest size at any future time. Three model approaches have been described. Summerfelt et al. (1993) calculated the number of fish available for harvesting by assuming that fish length follows a normal distribution; thus, the number of harvested individuals was determined from the capacity limit of the farm. In the second approach, the size distribution of individuals was considered a discrete, time-varying Markov process in which the number of individuals in each size class could be calculated. The optimal harvesting of individuals from various size classes was determined by dynamic programming (Leung et al. 1990; Leung and Shang 1989; Sparre 1976). The third approach is a variation of linear programming (LP) and dynamic programming (DP). LP has been frequently used in production planning in aquaculture production, such as for cost minimization in oyster farms (Lipschultz and Krantz 1980) and profit maximization in prawn farms (Shaftel and Wilson 1990; Wilson 1991), multicycle and multiponds operation of shrimps farms (Yu and Leung 2005; Yu et al. 2006), salmonid hatcheries (Johnson 1974), and salmonid grow-out farms (Gates et al. 1980; Varvarigos and Home 1986). We refer to Forsberg (1996) who discussed LP vs. DP in the aquaculture context and indicated that neither LP nor DP is adaptable to all aquaculture systems. Other approaches include modeling sinusoidal marketing conditions (Seginer and Halachmi 2008) and operational research (Weintraub et al. 2007). Halachmi (2007) reported on a single application of queuing theory, but it did not combine simulation optimization or address marine netcages. It ended with an analysis of a "what-if?" scenario, rather than a complete optimization methodology. A more recent study (Part I of this study: (Halachmi 2012a) integrated optimization, but (1) dealt with the design of a smaller facility that would handle 250 ton/year, (2) applied predefined parameters, set by the farmers, such as fish arrival frequency—once per month that reduced the space of feasible solutions. Part 2 (Halachmi 2012a), introduced reliability analysis (6σ robust design) into the optimization solver, and addressed the location issue. But both parts, Halachmi (2012a, b) developed queuing-based models to inland recirculating aquaculture system (RAS). Inland RAS operation mode is considerably different since the RAS limitation is the biofilter capacity and RAS's water temperature and water quality can be controlled, parameters that are uncontrollable in marine netcages. Halachmi (2012d) reported on a single application of queuing theory, but it did not combine simulation optimization to address RAS or marine netcages.

We could not find any scientific reported simulation addressing the same problem of designing 1,000 ton/year RAS and 2,500 ton/year netcages fish farms. A queuing network model that had been built in a previous study (Halachmi 2007) but could only handle steady-state situations; moreover, earlier simulation models (for reviews, see Halachmi et al. 2005; Halachmi 2006; Seginer and Halachmi 2008) did not address the same problem. They address smaller farms or ornamental fish farms.

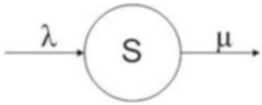
In conclusion, an integrated model (simulation optimization) in aquaculture has not been reported yet. The current study might serve to bridge the gap.

9.3 Model Formalisms

The model is presented in three parts: Sect. 9.3.1 presents the general case, Sect. 9.3.2 develops the equations for three growing phases. Section 9.3.3 presents a detailed examination of two case studies: (a) 2,500 ton/year marine netcages, (b) 1,000 ton/year recirculating aquaculture system. While the numerical values reflect local aquatic conditions, the concept (9.2)–(9.8) may be applicable elsewhere.

9.3.1 Developing the Analytical Concept: General Case

The model developed herein applies queuing-theory terms. A fish culture volume (tank, pond, marine netcage, etc.) can be treated as a queuing system:



(9.1)

where λ and μ are the arrival and departure rates, i.e., the number of fish batches that arrive and depart per year. The expected time spent by a fish in a culture volume is S. When a fish has completed its growth in a culture volume, it leaves that culture volume, and therefore

$$S = \frac{1}{\mu} \tag{9.2}$$

The expected utilization of a culture volume (in queuing terms “service facility”) is

$$\rho = \frac{\lambda}{\mu} \tag{9.3}$$

Over-holding ($\rho > 1$) or idle ($\rho < 1$) culture volumes are not economic.

The number of parts into which a fish batch is divided (P), and the number of culture volumes (c) in a growing phase gives:

$$\rho = \frac{\lambda}{\mu} \times \frac{P}{c} \tag{9.4}$$

Substitute $\rho = 1$, which means we want 100 % culture-volume utilization:

$$\frac{1}{\mu} = \frac{1}{\lambda} \times \frac{c}{P} \tag{9.5}$$

Substituting $S = \frac{1}{\mu}$, the expected time S that a fish spends in growth phase i, gives:

$$Si = \frac{1}{\lambda} \times \frac{c_i}{\prod_1^i Pi} \tag{9.6}$$

where Si is the period in growing phase i, c_i is the expected number of culture volumes in growing phase i, P_i is the number of parts into which a batch is divided as it enters growth phase i (this division is known in fishery practice as grading and sorting events). A batch is divided into smaller parts, P sub-batches, i.e., a batch with $N_1 = 200{,}000$ fish is divided into two batches of $N_2 = 100{,}000$ fish each: $P_2 = 2$.

The entire growing period, $\sum Si$, is usually known, although it depends on local conditions such as water quality, temperature, oxygen, feed, genetic merit, handling, and more:

$$\sum Si = \sum \left(\frac{1}{\lambda} \times \frac{c_i}{\prod Pi}\right) \rightarrow \sum Si = \frac{1}{\lambda}\sum \frac{c_i}{\prod Pi} \rightarrow \frac{1}{\lambda} = \frac{\sum Si}{\sum \frac{c_i}{\prod Pi}} \tag{9.7}$$

The yearly turnover, T (ton per year) is:

$$T = \lambda \times N_{\mathrm{f}} \times B_{\mathrm{f}} \tag{9.8}$$

where λ (batches per year) is arrival and departure rates, i.e., the number of fish batches that depart per year, N_{f} (fish) is the number of fish in a batch at marketing time, and B_{f} (kg) is the fish's final body weight.

Thus, the optimization problem appears to be:

$$\max T = \max \frac{\sum \frac{c_i}{\prod Pi}}{\sum Si} \times N_{\mathrm{f}} \times B_{\mathrm{f}} \tag{9.9}$$

S.T.

$\sum c_i \leq$ upper limit of the number of culture volumes restricted by available space, length of the breakwater or water-carrying capacity.

$D_i = \dfrac{B_i \times N_i}{V_i} \times \dfrac{1}{p_1 \times p_2 \times p_3 \times \cdots \times p_i} \leq$ fish biomass-density limitation, at any growing phase i.

where V_i is the culture volume in growing phase i. D_i is the biomass in growing phase i. Pi is the number of sub-batches formed from a batch; N_i is the number of fish in a batch in growing phase i. B_i is the body weight of a fish (kg) in growing phase i.

The decision parameters were: N, c_1, c_2, c_3, S_1, S_2, S_3, P_1, P_2, P_3.

9.3.2 The Equations for a Special Case: Three Growing Phases

Three growing phases, such as the "4,8,16 layout" (Fig. 9.3), derive $i = 1$ to 3, therefore (from 9.7) above),

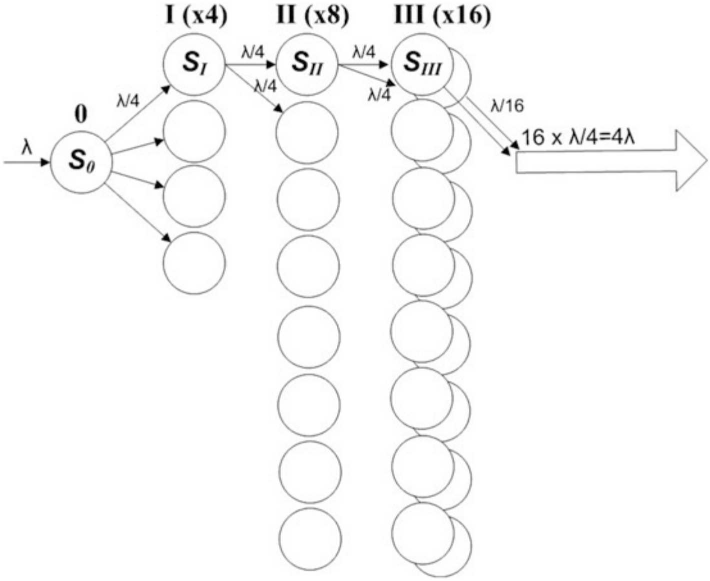


Fig. 9.3 The 4,8,16 layout. Fish culture netcages, 0, I, II, and III growing phases. The zero ("0") growing phase is the fingerling source. Our farm grows phases I to III. Phase I contains four round netcages, 2,900 m^3 each, summing to 11,600 m^3 in total. Phase II contains eight round netcages, 2,900 m^3 each, and one 2,900 m^3 summing to 23,200 m^3 in total. Phase III contains 16 round netcages: six 2,900 m^3 each, and ten 2,000 m^3 each, summing to 37,400 m^3 in total. "Fish traffic": once a month, a batch arrives with 3–6 g fish to phase I and occupies a netcage for 4 months (122 days, till around 60 g). Then, the batch is split into two while size sorting and is moved forward to occupy two netcages in phase II (also 4 months, till around 200 g). Then, each batch in phase II is split again into two netcages in phase III while size sorting. Phase III is also 122 days, reaching around 400 g. Batch splits are $P_1 = 1$, $P_2 = 2$, $P_3 = 2$. In total, each original batch is split into four sub-batches during its lifetime in the sea, while size sorting

$$\frac{1}{\lambda} = \frac{\sum Si}{\left(\frac{c_1}{p_1} + \frac{c_2}{p_1 p_2} + \frac{c_3}{p_1 p_2 p_3}\right)} \tag{9.10}$$

Note that (I) (9.10) and (9.11) apply only for three growing phases, i.e., $i = 1$ to 3; (II) the zero phase is the hatchery, not on our farm, and therefore it is not a 1, 4, 8, 16 system, i.e., not $i = 1$ to 4.

9.3.3 A Special Case: 2,500 ton/year, Three Growing Phases, a Space for Only 29 Netcages Along the Breakwater

The three phases, substituting the given biomass-density limitations and the available space, give the following set of equations:

$$\max\ T = \max \frac{\left(\dfrac{c_1}{p_1} + \dfrac{c_2}{p_1 p_2} + \dfrac{c_3}{p_1 p_2 p_3}\right)}{\sum Si} \times N_{\mathrm{f}} \times B_{\mathrm{f}} \tag{9.11}$$

$$\text{S.T.} \quad \sum c_i = c_1 + c_2 + c_3 \leq 29 \text{ netcages}$$

$D_1 = \frac{B_1 \times N_1}{V_1} \times \frac{1}{p_1} \leq 15\,\mathrm{kg/m^3}$ is fish biomass density at the end of the first growing phase.

$D_2 = \frac{B_2 \times N_2}{V_2} \times \frac{1}{p_1 \times p_2} \leq 20\,\mathrm{kg/m^3}$ is fish biomass density at the end of the second growing phase.

$D_3 = \frac{B_3 \times N_3}{V_3} \times \frac{1}{p_1 \times p_2 \times p_3} \leq 25\,\mathrm{kg/m^3}$ is fish biomass at the end of the final, third growing phase.

In our case, market demands $B_3 = B_{\mathrm{f}} = 0.350$–$0.402$ kg. Experience in our marine conditions suggests $\sum Si = 365 - 420\,\text{days}$, which is given by the fish growth functions (9.1), the harbor's depth and its available space: $V_{\text{small netcage}} = 2{,}000\ \mathrm{m^3}$, $V_{\text{large netcage}} = 2{,}900\ \mathrm{m^3}$, total $\sum c_i = 29$ netcages: $18 \times 2{,}900\ \mathrm{m^3}$ and $11 \times 2{,}000\ \mathrm{m^3}$ netcages can fit along the breakwater. In our case, given a constant fish mortality, the initial amount of fish $N_1 = N_{\mathrm{f}}/\mathrm{MR}$, where MR is the fish mortality rate.

The decision parameters were: N_{f}, c_1, c_2, c_3, S_1, S_2, S_3, P_1, P_2, P_3, and the model was fed into the Matlab optimization toolbox (Coleman et al. 1999). The solution was a 4,8,16 layout.

9.3.3.1 Calculating Fish Growing Periods

The sections below (Sects. 3.3.1–3.3.4) demonstrate the calculations for "4,8,16" potential layout (Fig. 9.3 describes the "4,8,16 layout" details). "4,8,16" stands for 4, 8, and 16 netcages for each of the three successive growing phases; $4 + 8 + 16 = 28$ netcages, while the length of the breakwater allows 29 netcages. One netcage is therefore reserved for daily operational tasks such as temporary storage of small numbers of fish waiting to be transported to market, disease analysis, vaccination activities, etc.

An optimal value of arrival rate, λ, can be calculated as follows:

$$\left(Si = \frac{1}{\lambda} \times \frac{c_i}{\prod Pi}\right) \rightarrow S_{\mathrm{I}} = \frac{1}{\lambda} \times \frac{c_{\mathrm{I}}}{p_{\mathrm{I}}} = \frac{1}{\lambda} \times \frac{4}{1} = \frac{4}{\lambda},\ S_{\mathrm{II}} = \frac{8}{2\lambda};\ S_{\mathrm{III}} = \frac{16}{2 \times 2\lambda} \tag{9.12}$$

$$\sum Si = S_{\mathrm{I}} + S_{\mathrm{II}} + S_{\mathrm{III}} = \frac{4}{\lambda} + \frac{4}{\lambda} + \frac{4}{\lambda} = \frac{12}{\lambda} \rightarrow \frac{1}{\lambda} = \frac{\sum Si}{12} \tag{9.13}$$

where, for a given set of aquaculture conditions, $\sum Si = 365$ days is the total time needed to raise a gilthead seabream from fingerling size (2–3 g) to a final product size of 350 g. Substituting $\sum Si = 365$ into (9.13) suggests that $1/\lambda = 30.4$ days, i.e., the farmer should purchase a fresh fingerling batch every month (30.4 days). Substituting $1/\lambda = 30.4$ into (9.12) suggests growing periods in phases S_{I}, S_{II}, and S_{III} of $4 \times 30.4 = 122$ days each.

9.3.3.2 Fine Tuning. Calculating Facility Utilization and Number of Sub-batch Splits

Full capacity is assured from the model construction bricks (9.4 and 9.5) assume $\rho = 1$ which means 100 % netcage utilization).

Phase 1. Arrival rate $\lambda = 12$ batches/year, a batch grows for 4 months (122 days), $\mu = \frac{365}{122} = 3$ batches/year. From (9.3) $\left(\rho = \frac{\lambda}{\mu} \times \frac{\prod Pi}{c}\right) \rightarrow \rho_1 = \frac{12}{3} \times \frac{1}{4} = 1$, meeting the $\rho = 1$ design criteria.

Phase 2. Arrival rate $\lambda = 12$ batches/year, a batch grows for 4 months (122 days), $\mu = \frac{365}{122} = 3$ batches/year and $P_2 = 2$. From (9.3) $\left(\rho = \frac{\lambda}{\mu} \times \frac{\prod Pi}{c}\right) \rightarrow \rho_1 = \frac{12}{3} \times \frac{2}{8} = 1$.

Phase 3. Arrival rate $\lambda = 12$ batches/year, a batch grows for 4 months (122 days), $\mu = \frac{365}{122} = 3$ batches/year, and $P_3 = 2$. From (9.3) $\left(\rho = \frac{\lambda}{\mu} \times \frac{\prod Pi}{c}\right) \rightarrow \rho_1 = \frac{12}{3} \times \frac{2 \times 2}{16} = 1$.

9.3.3.3 Fine Tuning Calculating Number of Fish in Each Batch

D is the biomass stocking density criterion; say $D = 20$ kg/m^3. V is the volume of a netcage; say $V = 2{,}000$ m^3. N_{f} is the number of fish in a batch at the final stage, and B_{f} is the final body weight of a fish. In the present case, the market demands fish of 350 g. Therefore: $D = \frac{N \times B_{\mathrm{f}}}{V} \rightarrow N = \frac{V \times D}{B_{\mathrm{f}}} = \frac{2{,}000 \times 20}{0.35} = 114{,}000$ fish in a batch on the final day. The survival rate at the final stage was 70 %, and in a 7,7,12 layout, a batch is divided into 1.71 parts; therefore, the original batch of purchased fingerlings should comprise $\frac{114k}{0.7} \times 1.71 = 280{,}000$ fish. In a 5,5,16 layout, a batch is divided into two parts; therefore, the original batch of purchased fingerlings should comprise $\frac{114k}{0.7} \times 2 = 326{,}000$ fish.

9.3.3.4 Fine Tuning Calculating Grading (Sorting) Criteria

The optimal growth periods, i.e., 147, 147, and 147 days for the 7,7,12 layout, determine the grading criteria as follows: from (9.1), $Y_{\mathrm{ave}}(t = 147) = 46$ g for moving a fish from phase I to phase II at the age of 147 days. $Y_{\text{ave}}(t = 294) = 183$ g for moving it from phase II to phase III at the age of $147 \times 2 = 294$ days. However, other growth functions lead to different grading criteria: $Y_{\min}$ (9.1) leads to 20 g and 126 g, respectively, and $Y_{\max}$ (9.1) leads to 71 g and 239 g, respectively.

9.3.4 1,000-ton/year Recirculating Aquaculture System

Based on (9.8)–(9.11) and converting the measurement units, the optimization problem was:

$$\max\ T = \lambda \times N_{\mathrm{f}} \times B_{\mathrm{f}}$$

with
$$\lambda = \frac{1}{\sum Si}\left(\frac{c_1}{p_1} + \frac{c_2}{p_1 p_2} + \frac{c_3}{p_1 p_2 p_3}\right) \tag{9.18}$$

S.T.:

$pt\frac{1}{\Sigma Si}\left(\frac{c_1}{p_1} + \frac{c_2}{p_1 p_2} + \frac{c_3}{p_1 p_2 p_3}\right) - 52 \leq 0$ is the arrival rate, $\lambda \leq 52$ fish batch arrivals/year (once per a week); $c_1 + c_2 + c_3 - 50 \leq 0$ is the total number of culture tanks that can fit into the given building space ≤ 50; $\frac{\left(B_{\text{fingerings}} + Gr \times S_1\right) \times 0.9\ \mathrm{N}}{V_1 \times p_1} - 30 \leq 0$, seabream biomass density in first growth phase $\leq 30\ \frac{kg}{m^3}$; $\frac{\left(B_{\text{fingerings}} + Gr \times (S_1 + S_2)\right) \times 0.9\ \ \mathrm{N}}{V_2 \times p_1 p_2}$ $-55 \leq 0$ is the seabream biomass density in second growth phase $\leq 55\ \frac{kg}{m^3}$; $\frac{\left(B_{\text{fingerings}} + Gr \times (S_1 + S_2 + S_3)\right) \times 0.9\ \ \mathrm{N}}{V_3 \times p_1 p_2 p_3} - 55 \leq 0$ is the seabream biomass density in third growth phase $\leq 55\ \frac{kg}{m^3}$; $\frac{273}{365} - S_1 - S_2 - S_3 \leq 0$ is the total growth period $S_1 + S_2 + S_3 = \frac{273}{365}\text{years} = 273\,\text{days}$; $V_i - 350 \leq 0$ is the off-the-shelf round tank, 2-m height and volume up to 350 m^3.

$$B_{\mathrm{f}} = B_{\text{fingering}} + \mathrm{Gr}\sum Si \quad \text{fish body weight} \tag{9.19}$$

$$\mathrm{Gr} = 1.7\mathrm{e} - 3 \times 365\,\mathrm{kg/year}; \quad B_{\text{fingering}} = 2\mathrm{e} - 3\,\mathrm{kg} \tag{9.20}$$

$c_1 = \lambda \times S_1 \times p_1$ is the number of culture tanks in the first growth phase,

$c_2 = \lambda \times\ S_2 \times p_1 p_2$ is the number of culture tanks in the second growth phase,

$c_3 = \lambda \times S_3 \times p_1 p_2 p_3$ is the number of culture tanks in the third growth phase, $p_1 = p_2 = p_3 = 1$. In our case, the farmer preferred only one single batch throughout the entire rearing process, i.e., no batch splits were allowed, and, additional lower bound was $0 \leq x_i$ (all decision parameters are above zero).

The solution vector $X = [c_1, c_2, c_3, S_1, S_2, S_3\ V_1, V_2, V_3, N]$ can be obtained by use of Matlab's *fmincon* function, with the "*active-set*" algorithm (Coleman et al. 1999).

Solution. The optimal layout was found to comprise 13 culture tanks in each of the three growth phases, which sums to 39 culture tanks. Optimal parameters included: arrival frequency—a single fish batch into the system every 7 days, then 91 days in each phase; growth up to 77, 233, and 468 g in the successive growth phases (the so-called *three growing phases*, *13*, *13*, *13 culture tanks and 91*, *91*, *91 days*, *respectively*). The optimal values satisfied the criteria of biomass density below 50 kg/m^3 and culture tank utilization above 99 %. Expected production was 1,000 ton/year. The proposed layout can accommodate different fish species with different growth rates, e.g., seabream and grouper can be reared under the same proposed layout, culture volume, density, and schedule. Increasing the desired biomass density from 50 to 60 kg/m^3 advances the expected production to 1,335 ton/year.

9.3.5 Simulation

9.3.5.1 Constructing the Simulation Model

The computer program, ARENA, simulates the flow of entities, i.e., fish batches, through the growth phases in the culture tanks (a *process-orientation* method rather than an *event-sequencing* method—Banks 1998). An entity can "request" use of a resource (culture tank): if its request is denied, it joins a virtual queue and changes its color value to red to signal an alarm; if its request is "approved," it is permitted to capture a resource, which it retains until the entity has grown to its predetermined size, at which point it releases the resource.

The resources are the culture tanks, each of which has its own capacity, its own cleaning schedule, and its own state variables, comprise fish population, growth function, and mortality rate. The resource (tank) status may be idle, busy, or blocked, i.e., suffering a technical problem or being cleaned.

The model incorporated two types of input variables. Discrete-event variables comprise number of fish, number of culture tanks, time between arrivals of fingerlings, and frequency of fish sorting. Continuous-time variables were: fish body weight, O_2 demand, feed intake, and excretions of TAN and TSS.

For detailed for description of the simulation model building blocks, refer to Halachmi et al. (2014).

Model inputs:

(i) Number of tanks in each of the three growth phases
(j) Stocking frequency, i.e., number of days between successive stocking events
(k) Number of fingerlings per stocking event, i.e., batch size
(l) Initial fingerling size or weight
(m) Final product size

(n) Desired stocking density at each time point in the fishes' life in each culture tank
(o) Fish-growth functions and mortality rate in each culture tank

Model outputs (i.e., the simulation responses):

(p) Monthly and annual sales, expressed as tons and number of fish per year, costs, and revenues
(q) Stocking density (kg/m^3) and standing stock biomass (tons) in each culture tank at any given time, and feed load on the filters
(r) Utilization of each culture tank and overall facility utilization at any given time
(s) Sorting frequency and criteria at various points along the production line

For detailed for description of a model verification and validation, refer to Halachmi (2000), Halachmi et al. (2001), and Kleijnen et al. (1998)

9.3.6 DOE and RSM

DOE. Use of the D-optimal (Anonymous 2008) reduced the number of input combinations that were actually simulated to 100. Matlab's D-optimal design maximizes the determinant of Fisher's information matrix XTX. Maximizing det (XTX) is equivalent to minimizing the determinant of the covariance of the estimated parameters. There are several Matlab functions that generate D-optimal designs: *cordexch*, *daugment*, *dcovary*, and *rowexch*; we used *cordexch*. *Cordexch* (n factors, n runs) uses a coordinate-exchange algorithm to generate a D-optimal design with n runs (the rows) for a linear additive model with n factors (the columns). The DOE followed the procedure (Kleijnen and Standridge 1988) and (Kleijnen 2008).

The DOE outcome was 100 input combinations. The inputs were fed to the simulation software. Then, after running the simulation 100 times, an input–output function (*I/O transformation*) was fitted using stepwise regression. The regression model itself was found to be significant ($p = 0.000$), and the adjusted R^2 was above 0.90 in both cases.

The *I/O transformation* which is defined as a model of the underlying simulation experiments and is formulated by means of RSM, uses regression models, and steepest ascent (gradient, unique maximum, saddle point, etc.). In our case study, this meta-model was a first- or second-order polynomial, which meant that Kuhn–Tucker conditions were both necessary and sufficient for a global optimum point. The global optimum was found with ordinary projection methods (Coleman et al. 1999). Thus, the meta-model enabled a global optimum to be found and the integrated design methodology to be completed.

9.3.7 Proposed Integration Optimization and Simulation Throughout the Life-Time of a Project

Given the above considerations, a typical simulation–optimization aquaculture project involved the following steps:

i. Clear formulation of the project aims and phases, with farm staff participation (1 week)
ii. Clear formulation of the simulation model goals, with farm staff participation (1 week)
iii. Gaining an understanding of the Inputs and Outputs ("I/O") of the intended simulation model, in parallel with real-life data measurements (1/2 year, given 3 years data exist in the farm management computers)
iv. Formulation of a *conceptual model* of the system, and obtaining farm staff confirmation of its validity, before starting any programming work (2 months)
v. Translation of the conceptual model to "Arena" modeling software (1 month)
vi. Software verification (2 months)
vii. Model validation, including statistical validation, visual analysis, etc. (1/2 year)
viii. Designing simulation experiments (DOE, 2 months)
ix. Formulation of the meta-model using RSM (1 month)
x. Formulation of the optimization problem, objective function, and constraints (1 month)
xi. Simulation runs, scenario analysis, interpretation, and solving the optimization problem while brainstorming/discussion sessions with the farm staff (2 months, two meetings)
xii. Project documentation and delivery

The iterative process described in Fig. 9.3 (mentioned above) is performed in step xi.

9.4 Applications

The real-life data are described in (Halachmi 2012b, c, d). Based on the simulation optimization, the flowing results were obtained:

In the 2,500 ton/year netcages case-study, the model optimizes: (1) facility allocation, i.e., number and volume of netcages in each growing phase; (2) fish-batch arrival frequency; (3) number of fingerlings in a batch; (4) number of days in each culture netcage; and (5) grading criteria along the production lines.

Compared with today existing management (200,000 fish per batch, 4 g fingerlings, 441 days in sea, 7,7,12 layout), the 5,5,16 layout was superior, giving 1,687 vs. 981 ton/year).

For the new system that is now under construction, the 4,8,16 layout was selected. Optimal arrival frequency is a batch every month, and optimal retention times are 122 days in each successive growing phase (up to 62, 196, and 382 g, respectively). Use of these parameters did not violate the biomass-density criterion (15, 20, and 25 kg/m^3, respectively) or the netcage utilization criterion (never below 99 %), which suggests that it is not feasible to have fewer culture netcages.

The above numerical values reflect local conditions, but the concept is applicable anywhere. It is recommended that every aquaculture enterprise apply this concept in its design stage. The onus now lies with the industry.

9.5 Discussion

The main imperfection of simulations lies in their heuristic character: outputs are obtained only in response to the selected inputs, so that there is no guarantee of having obtained the best possible solution (Halachmi et al. 2002, 2012; Halachmi 2000; Banks 1998; Law and Kelton 1991). Classic simulation–optimization (Fig. 9.1) made attempts to overcome this imperfection. However, classical simulation–optimization does not pretend to guarantee global optimal point. If the commercial simulation software could have grant access to the exact optimization algorithm and would allow controllable $F(x)$ and $G(x)$ functions, then the entire proposed methodology has no necessity anymore. The onus is now on the simulation industry to improve their software.

The proposed method (Fig. 9.2) cuts the automatic linkage between the simulation kernel and the optimization engine and replaces that linkage with a set "global-optimization-enabled" objective function (up to third-order polynomial form, KKT conditions, etc.). However, one should verify that his fitted functions reached enough degree of goodness-of-fit. Otherwise, no proof of optimality. The responsibility is on the modeler to verify that fitted functions well adequately mimic the simulation and the real system.

In the future, more applications in other aquaculture farms are foreseen and highly recommended. The authors encourage practitioners also in other industries to adopt the proposed simulation–optimization and to send us their conclusions.

9.6 Conclusion

A simulation–optimization model aimed at best possible managerial parameters of marine netcages and recirculating aquaculture system was developed and applied.

Acknowledgements Many thanks go to the fish farmers Mr. Nir Tzohari and Dubi Helman (Ardag) who collected the data and actively seeks out professional management tools. Thanks to Dudi Gada (Ardag) and Roni (Ardag) who participated in the scenario running events and the brainstorming at the face-validity meeting in Newe Yaar.

This study was also financed by grants no. 459-4255-05 and 459-0265-06 from the Israel Ministry of Agriculture's chief scientist fund. This chapter is contribution no. 705/06 from the Institute of Agricultural Engineering, ARO. Israel.

References

Anonymous (1976) Report of the National Commission on Agriculture, 1976, Pt. VIII. Fisheries. Government of India, Ministry of Agriculture, New Delhi

Anonymous (2008) Matlab statistics toolbox. The MathWorks, Natick

Banks J (1998) Handbook of simulation. Wiley, New York

Better M, Glover F, Kochenberger G, Wang H (2008) Simulation optimization: applications in risk management. Int J Inf Technol Decis Mak 7(4):571–587

Brune DE, Schwartz G, Eversole AG, Collier JA, Schwedler TE (2003) Intensification of pond aquaculture and high rate photosynthetic systems. Aquacult Eng 28(1–2):65–86

Caillouet CW (1972) Rotatable cage for high density aquaculture. Prog Fish Cult 34(1):8

Chua TE, Teng SK (1978) Effect of stocking density on the growth of estuary grouper, Epinephelus salmoides Maxwell, cultured in floating net-cages. Aquaculture 15:273–287

Coleman TF, Branch MA, Grace A (1999) Optimization toolbox—for use with MATLAB®, user's guide, 2nd edn. The MathWorks, Inc., Natick

Forsberg OI (1996) Optimal stocking and harvesting of size-structured farmed fish: a multi-period linear programming approach. Math Comput Simul 42:299–305

Fu MC (1994) Optimization via simulation: a review. Ann Oper Res 53:199–248

Fu MC (2002) Optimization for simulation: theory vs. practice. INFORMS J Comput 14 (3):192–215

Fu MC, Andradóttir S, Carson JS, Glover F, Harrell CR, Ho YC, Kelly JP, Robinson SM (2000) Integrating optimization and simulation: research and practice. In: Proceedings of the winter simulation conference, pp 610–616. http://www.informscs.org/wsc00papers/082.PDF

Fujiya M (1979) Coastal culture of yellowtail (Seriola quinqueradiata) and sea breams (Sparus major) in Japan. In: Advances in aquaculture. Fishing News Books, Oxford

Gates JM, MacDonald CR, Pollard J (1980) Salmon culture in water reuse systems: an economical analysis, vol 78. Technical report. University of Rhode Island Marine

Halachmi I (2000) Designing the optimal robotic barn, Part 2: Behaviour-based simulation. J Agric Eng Res 77(1):67–79

Halachmi I (2006) Systems engineering for ornamental fish production in a recirculating aquaculture system. Aquaculture 259(1–4):300–314

Halachmi I (2007) Biomass management in re-circulating aquaculture systems using queuing networks. Aquaculture 262:514–520

Halachmi I (2012) Mathematical principles of production management and robust layout design: Part I. 250-ton/year recirculating aquaculture system (RAS) Aquacultural Engineering 50:1–10

Halachmi I (2012b) Mathematical principles of production management and robust layout design: Part I. 250-ton/year recirculating aquaculture system (RAS). Aquacult Eng 50:1–10

Halachmi I (2012c) Mathematical principles of production management and robust layout design: Part II. Upscaling to a 1000-ton/year recirculating aquaculture system (RAS). Aquacult Eng 50:11–19

Halachmi I (2012d) Mathematical principles of production management and robust layout design: Part III. 2500-ton/year fish farming in marine net cages. Aquacult Eng 54:110–117
Halachmi I, Dzidic A, Metz JHM, Speelman L, Dijkhuizen AA, Kleijnen JPC (2001) Validation of simulation model for robotic milking barn design: case study. Eur J Oper Res 134(3):677–688
Halachmi I, Metz JHM, van't Land A, Halachmi S, Kleijnen JPC (2002) Optimal facility allocation in a robotic milking barn. Trans ASAE 45(5):1539–1546
Halachmi I, Simon Y, Guetta R, Hallerman EM (2005) A novel computer simulation model for design and management of recirculating aquaculture systems. Aquacult Eng 32(3–4):443–464
Halachmi I, Simon Y, Mozes N (2014) Simulation of the shift from marine netcages to inland recirculating aquaculture systems. Annals of Operations Research 219(1):85–99
Hamlin HJ, Michaels JT, Beaulaton CM, Graham WF, Dutt W, Steinbach P, Losordo TM, Schrader KK, Main KL (2008) Comparing denitrification rates and carbon sources in commercial scale upflow denitrification biological filters in aquaculture. Aquacult Eng 38(2):79–92
Harada T (1970) The present status of marine fish cultivation research in Japan. Helgoländer Wiss Meeresunters 20:594–601
Jahncke LM, Schwarz MH (2000) Application of hazard analysis and critical control point (HACCP) principles as a risk management approach for re-circulating aquaculture systems (RAS). In: Proceedings of the third international symposium on recirculating aquaculture systems, Roanoke, VA
Johnson FC (1974) Hatch—a model for fish hatchery analysis. Report NBSIR. U.S. National Bureau of Standards, Washington, DC
Kennedy WA (1975) An experimental fishfarm for salmon at the Pacific Biological Station. Tech Rep Fish Mar Serv Can 543, 33 p
Kleijnen JPC (2008) Design and analysis of simulation experiments. Springer, New York
Kleijnen JPC, Standridge CR (1988) Experiment design and regression analysis in simulation: an FMS case study. Eur J Oper Res 33:257–261
Kleijnen JPC, Bettonvil B, van Groenendaal W (1998) Validation of trace-driven simulation models: a novel regression test statistical perspective. Manag Sci 44:812–819
Law AM (1990) How to build valid, credible, and appropriately detailed simulation models. J Ind Eng 22:16–17
Law AM, Kelton WD (1991) Simulation modelling and analysis. McGraw-Hill, New York
Leung PS, Shang YC (1989) Modelling prawn production management system: a dynamic Markov decision approach. Agr Syst 29:5–20
Leung PS, Hochman E, Rowland LW, Wyban JA (1990) Modelling shrimp production and harvesting schedules. Agr Syst 32(1990):233–249
Libey GS, Timmons MB (eds) (1996) Successes and failures in commercial re-circulating aquaculture. In: Proceedings of an international workshop. Northeast Regional Aquacultural Engineering Service (NRAES-98). Cornell University, Ithaca, NY
Lipschultz E, Krantz GE (1980) Production optimisation and economic analysis of an oyster (Crassostera virginica) hatchery on the Chesapeake Bay. In: Proceedings of the World Mariculture Society, vol 11, MD, USA, pp 580–591
Milne PH (1979) Selection of sites and design of cage, fishpens and net enclosures for aquaculture. In: Advances in aquaculture. Fishing News Books, Ltd., Oxford
Møller D (1979) Recent developments in cage and enclosure aquaculture in Norway. In: Advances in aquaculture. Fishing News Books, Ltd., Oxford
O'Rourke K (1996) The economics of re-circulating aquaculture systems. In: Libey GS, Timmons MB (eds) Successes and failures in commercial re-circulating aquaculture. Proceedings of an international workshop, Roanoke, VA, USA, July 19–21. Northeast Regional Aquacultural Engineering Service (NRAES-98). Cornell University, Ithaca, NY, pp 61–78
Seginer I, Halachmi I (2008) Optimal stocking in intensive aquaculture under sinusoidal temperature, price and marketing conditions. Aquacult Eng 39(2–3):103–112
Shaftel TL, Wilson BM (1990) A mixed-integer linear programming decision model for aquaculture. Manag Decis Econ 11:11–31

Smith SA (1996) HACCP program for disease and therapeutics for intensive culture of food fish. In: Libey GS, Timmons MB (eds) Proceedings of the third international symposium on recirculating aquaculture systems, Roanoke, VA, 19–21 July 1996. University of Virginia, pp 344–345

Sparre P (1976) A Markovian decision process applied to optimisation of production planning in fish fanning. Medd Dan Fisk Havunders 7:111–197

Summerfelt ST (1996) Engineering design of modular and scalable RAS containing circular tanks, micro screen filtering, fluidised sand bio filter, cascade aeration, and low-head or U-tube oxygenation. In: Libey GS, Timmons MB (eds) Proceedings of the third international symposium on recirculating aquaculture systems, Roanoke, VA, 19–21 July, pp 217–244

Summerfelt ST, Hankins JA, Summerfelt SR, Heinen JM (1993) Modelling continuous culture with periodic stocking and selective harvesting to measure the effect on productivity and biomass capacity to fish culture systems. In: Wang JK (ed) Techniques for modern aquaculture. Proceedings of an aquacultural engineering conference at Spokane, Washington, Feb 1993. ASAE Publication—American Society of Agriculture Engineers, St. Joseph, MI, pp 581–595

Summerfelt ST, Davidson J, Wilson G, Waldrop T (2009) Advances in fish harvest technologies for circular tanks. Aquacult Eng 40(2):62–71

Tatum WM (1973) Comparative growth of pompano (Trachinotus carolinus) in suspended cages receiving diets of a floating trout chow and then receiving a mixture of 50 percent trout chow and 50 percent sinking ration. Proc Annu Meet World Maricult Soc 4:125–141

Timmons MB, Ebeling JM, Wheaton FW, Summerfelt ST, Vinci BJ (2001) Recirculating aquaculture systems, 2nd edn. Cayuga Aqua Ventures, Ithaca, NY

Varvarigos P, Home MT (1986) Application of linear programming to individual fish farm production planning. In: Proceedings of IFAC automation and data processing in aquaculture, Trondheim, Norway, pp 243–254

Weintraub A, Romero C, Bjorndal T, Epstein R (2007) Handbook of operations research in natural resources. Springer, London

Wilson BM (1991) Schaftel TL (1991) A mathematical programming approach to production decisions in the emerging aquaculture industry. Decis Sci 22:194–205

Yu R, Leung P (2005) Optimal harvesting strategies for a multi-cycle and multi-pondshrimp operation: a practical network model. Math Comput Simul 68:339–354

Yu OK, Vizcarra AT (1979) Development of circular floating cages for milkfish broodstock at the SEAFDEC Aquaculture Department. In: International workshop on pen and cage culture of fish. SEAFDEC Aquaculture Department, Tigbauan, Philippines

Yu R, Leung PS, Bienfang P (2006) Optimal production schedule in commercial shrimp culture. Aquaculture 254(254):426–441

Chapter 10
Swarm Intelligence in Optimal Management of Aquaculture Farms

A. Cobo, I. Llorente, and L. Luna

10.1 Introduction

Optimization problems are of high importance and practical application both for the industry and for the scientific world. Of course, the field of aquaculture farm management is no exception to the applicability of optimization techniques. Many strategic and operational decisions in the management of such farms can be optimized by mathematical modeling. In particular, seeding strategies, feeding policies, feed selection, and harvest times determination are examples of decisions that can be optimized using mathematical methodologies. An optimization algorithm is a search method where the goal is that a given quantity is optimized (maximized or minimized), possibly subject to a set of constraints. In general, any classical optimization problem has the three following elements: an objective function which represents the quantity to be optimized, a set of decision variables, and a feasible region defined by a set of constraints that restrict the values that can be assigned to the decision variables. Due to the practical importance of optimization problems, many algorithms to tackle them have been developed. However, in many practical problems, it can arise various problems or difficulties. Firstly, the complexity of the problems can make it difficult to find the optimal solution or an approximation in reasonable computation times. This happens, for example, in many problems of combinatorial nature, where there are a finite number of feasible solutions but classical optimization techniques are not applicable or lead to computation times too high for practical purposes. Moreover, in many real-life problems, the goal is often to optimize several objective functions at the same time; in these cases we need to use multicriteria optimization techniques that can be applied

A. Cobo (✉) • I. Llorente • L. Luna
Research Group in Economic Management for the Primary Sector Sustainability,
University of Cantabria, Cantabria, Spain
e-mail: angel.cobo@unican.es

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_10

for solving problems that require an evaluation and measuring in which different and very often opposed criteria intervene. Finally, sometimes decisions are affected by nonformalizable criteria or subjective judgments of the decision-makers. In such cases it is advantageous to have techniques that generate sets of possible solutions, thereby facilitating the process of decision-making. These are some of the reasons that the use of metaheuristic methods to solve optimization problems has received more and more attention in the last 30 years, especially the population-based metaheuristics. In these methods we sacrifice the guarantee of finding optimal solutions for the sake of getting good solutions in a significantly reduced amount of time (Blum and Roli 2003).

The term metaheuristic, first introduced in Glover (1986), refers to a new kind of algorithm which basically tries to combine basic heuristic methods in higher-level frameworks aimed at efficiently and effectively exploring a search space. In short, metaheuristics are high-level strategies for exploring search spaces by using different methods. There is no commonly accepted definition for the term metaheuristic, but in the following we present one of the first definitions proposed in the specialized literature: "A metaheuristic is formally defined as an iterative generation process which guides a subordinate heuristic by combining intelligently different concepts for exploring and exploiting the search space, learning strategies are used to structure information in order to find efficiently near-optimal solutions" (Osman and Laporte 1996). This definition highlights one of the key features of metaheuristic techniques: the dynamic balance between diversification (exploration of the search space) and intensification (exploitation of the accumulated search experience). Metaheuristic algorithms form an important part of contemporary global optimization algorithms, computational intelligence, and soft computing (Yang 2012).

Blum and Roli (2003) outline fundamental properties which characterize metaheuristics: efficiently exploration of the search space in order to find near-optimal solutions, approximate and usually nondeterministic algorithms, use of learning processes, not problem specific, incorporation of mechanisms to avoid getting trapped in confined areas of the search space, use of domain-specific knowledge in the form of heuristics that are controlled by the upper-level strategy, and use of search experience to guide the search.

There are different criteria to classify metaheuristic algorithms (Blum and Roli 2003). Based on the origins of the algorithm, we can distinguish between nature-inspired and non-nature-inspired. The first ones try to mimic some successful characteristics in nature and have attracted a great deal of attention in engineering and industry. Example of nature-inspired metaheuristic is the genetic algorithms, inspired by evolution and natural selection principles. Another characteristic that can be used for the classification of metaheuristics is the number of solutions used at the same time. Algorithms working on single solutions are called trajectory methods. Population-based metaheuristics, on the contrary, perform search processes which describe the evolution of a set of points in the search space, that is, a population of potential solutions. In general, population-based metaheuristics are more complex to use than trajectory methods; they require mechanisms to manage population of solutions, and they need effective operators for the combination of solutions. It is clear that the use of a population is an appropriate way to achieve search space diversification. Thus population-based methods are desirable if their complexity is not overwhelming (Kochenberger 2003).

A subset of population-based metaheuristics is often referred to as swarm intelligence (SI) algorithms. These algorithms have been developed by mimicking the so-called swarm intelligence characteristics of biological agents such as birds, fish, insects, and others (Yang 2012). The increasing popularity of these computational techniques is due to their flexibility, versatility, and efficiency to deal with very complex optimization problems. In the following sections the theoretical foundations of the two main SI techniques will be presented, and then we will expose a practical application in the context of optimal planning of aquaculture farms.

10.2 Swarm Intelligence

Swarm intelligence, also referred to as collective intelligence, is an innovative distributed artificial intelligence paradigm for solving optimization problems that originally took its inspiration from the collective behavior of social insects such as ants, termites, bees, and wasps, as well as from other animal societies such as flocks of birds or schools of fish. These algorithms use multiple interacting agents in order to exploit the benefits of cooperation in situations where you do not have global knowledge of an environment. In these situations individuals within the group (agents) interact to solve the global objective, exchanging locally available information, which in the end propagates through the entire group such the problem is solved more efficiently than can be done by a single individual (Engelbrecht 2005). The term swarm intelligence was first used by Beni (1998) in the context of cellular robotic systems where simple agents organize themselves through nearest-neighbor interaction. In short, SI refers to various techniques based on the idea that groups of extremely simple agents with little or no organization can exhibit complex and intelligent behavior by using simple local rules and communication mechanisms (Bonabeau et al. 1999). Thanks to this intelligent behavior, a group of social agents can carry out actions on a complex level and form decentralized and self-organizational systems. These groups of agents are known as swarms; formally a swarm can be defined as a group of generally mobile agents which communicate with each other (either directly or indirectly) by acting on their local environment (Hoffmeyer 1994).

The roots of SI are deeply embedded in the biological study of self-organized behaviors in social insects (Bonabeau et al. 1999). Colonies of social insects have fascinated researchers for many years, and the mechanisms that govern their behavior remained unknown for a long time. Biological swarm systems that have inspired computational models include ants, termites, bees, and spiders. Even though the single members of these colonies are non-sophisticated individuals, they are able to achieve complex tasks in cooperation. Examples of collaborative work are nest building, task allocation, recruitment of colony members for prey retrieval, foraging behaviors, larvae organization, clustering task of corpses... Many aspects of these collective activities are self-organized and work without a central control (Blum and Li 2008). Other examples of collective intelligence and emergent behavior from nature are the self-organization in optimal spatial patterns of birds in a flock and fish in a school, hunting strategies of predators, communication using molecules in bacteria, and behaviors of some simple cellular organisms in times of food shortage (Engelbrecht 2005).

Different mathematical models inspired by such behaviors have been successfully applied for solving a wide range of real problems. Among the most well-known SI models are the ACO (ant colony optimization) and PSO (particle swarm optimization) techniques. Both are population-based metaheuristic algorithms that can be applied to optimum solution-seeking problems and represent an interesting alternative to problems of a combinatorial nature that are difficult to solve using classic techniques.

10.2.1 Ant Colony Optimization

Ant colony optimization (ACO) was one of the first techniques for approximate optimization inspired by SI. It was introduced as a technique for combinatorial optimization in the early 1990s (Dorigo 1992). The inspiring source of ant colony optimization is the foraging behavior of real ant colonies. The basic ACO algorithm mimics the way real ants find shortest route between a food source and their nest. The ants communicate with one another by means of chemical pheromone trails, which enables them to find short paths between their nest and food sources. When searching for food, ants initially explore the area surrounding their nest in a random manner. While moving, ants leave a chemical pheromone trail on the ground; and wherever an ant has a choice of direction, it chooses the branch to follow with a probability that is dependent on the pheromone concentration. With each successful round-trip to the food source by an ant, the trail beacon comes stronger. The deposited pheromone is subject to evaporation over time, and then the pheromone concentration will be higher in the shortest paths and more ants get attracted toward the source using these routes. This indirect communication between the ants via pheromone trails enables them to find shortest paths between their nest and food sources.

In ACO, principles of communicative behavior occurring in real ant colonies are used, and several generations of artificial ants search for good solutions. An artificial ant is a stochastic constructive procedure that incrementally builds a solution by adding opportunely defined solution components to a partial solution under construction. Every ant of a generation builds up a solution step by step using information provided by the previous ants (pheromone trails) and heuristic information that represents a priori information about the problem. Once a solution is completed, pheromone trails are updated according to the quality of the solution built, so that the following ants are attracted by the pheromone and will likely search in the solution space near good solutions (Dorigo and Stützle 2004). ACO has been applied successfully to solve various optimization problems (Dorigo and Stützle 2004). The first ACO algorithm, called ant system (AS), was applied to the traveling salesman problem (TSP) (Dorigo 1992), but in general ACO can be applied to any combinatorial optimization problem that can be represented by a graph, consisting of a finite number of nodes and links between nodes. In the case of continuous optimization problems, the main problem is to determine a way of mapping the continuous space problem to a graph search problem. Different approaches have been used for this purpose; a simple solution is to encode floating-point variables using binary string representations (Engelbrecht 2005).

In order to apply the basic ACO approach to solve an optimization problem, we need to proceed as follows. First, we need to derive a finite set $C = \{c_1, c_2, \ldots, c_n\}$ of solution components which will be used to assemble solutions to the combinatorial optimization problem and a finite set of possible transitions among these components. Each component is associated to one of the nodes of the graph, and each link in the graph represents one possible transition. We also need a cost function, which associates a cost to each solution generated by the search process. Each component added to a solution contributes to the total cost of the solution. Given these basic elements, the optimization algorithm aims to construct a feasible sequence of components such that the cost of the solution is minimized, that is, to construct a minimum cost path in the graph.

One of the central elements of ACO is the pheromone model, defined as a set of pheromone values T. Each link between solution components i and j has associated a pheromone value $\tau_{ij} \in T$. The pheromone model is used to probabilistically generate solutions to the problem by assembling them from the set of solution components, that is, by constructing incrementally a path in the graph. Normally a random constant value is used to initialize the pheromone values $\tau_{ij} = \tau_0$, and then the candidate solutions will modify these values in order to concentrate the search in regions of the search space containing high-quality solutions.

The success of ACO algorithm is in the cooperative behavior, so a set of software agents called artificial ants search for good solutions. The number of ants is a parameter that must be specified; the fewer ants used, the less the exploration ability of the algorithm. Small values of this parameter may cause suboptimal solutions. Moreover, if the parameter value is high, the computational complexity grows. Then we need to find the right balance between complexity and effectiveness.

Each artificial ant stars the solution construction with an empty path S. At each construction step the path is extended by adding a solution component (node) from a set of feasible components $N(s)$. This set of feasible components includes the unvisited nodes that are connected with the ant's current node. The choice of a node from $N(s)$ is at each construction step performed probabilistically with respect to the pheromone model and heuristic information. The heuristic information is represented by a new matrix (η_{ij}) whose values can be seen as a priori measure of the quality of inclusion in the path of node c_j after node c_i, i.e., the attractiveness, or desirability, of the move from c_i to c_j. For example, in the case of TSP, these values can be defined as a function of the distance between nodes (cities). For most of ACO algorithms, the transition probabilities from node c_i to node c_j are defined as follows:

$$p_{ij} = \begin{cases} \dfrac{(\tau_{ij})^{\alpha}(\eta_{ij})^{\beta}}{\sum\limits_{q \in N(s)} (\tau_{iq})^{\alpha}(\eta_{iq})^{\beta}} & \text{if } j \in N(s) \\ 0 \quad \text{if } j \notin N(s) \end{cases} \tag{10.1}$$

where α and β are two positive parameters whose values determine the relative influence of pheromone and heuristic information. This transition probability balances the exploration-exploitation trade-off; the best balance is achieved through proper selection of the parameters α and β, so their values must be selected carefully. If $\alpha = 0$, no pheromone information is used, previous search experience is neglected, and the search can degrade to a stochastic greedy search. Moreover, large values of α give excessive importance to the pheromone information which may lead to a rapid convergence to suboptimal solutions, that is, all ants follow the same nonoptimal path.

After completion of a path by each ant, the pheromone values must be updated. Different update strategies can be used; in the following we outline a general rule in order to provide the basic idea. This rule includes a pheromone evaporation, which uniformly decreases all the pheromone values, and an update process based on the quality of the solution found by each ant. According to the ideas, the pheromone on each link is updated as

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{k=1}^{n_a} \Delta\tau_{ij}^{k}(t) \tag{10.2}$$

where ρ in $[0, 1]$ is a parameter called evaporation rate that controls the influence of search history; n_a denotes the number of ants in the colony, i.e., the number of solutions that are used for the update (each ant constructs its solution); and $\Delta\tau_{ij}^{k}(t)$ is the amount of pheromone deposited by ant k on the link (i, j) at time step t. This amount must be calculated using an increasing function of the quality of the solution. The value of the evaporation rate must be selected carefully. For large evaporation rates, pheromone evaporates rapidly and more random the search becomes.

10.3 Particle Swarm Optimization

Particle swarm optimization (PSO) is a stochastic technique based on the evolution of populations for problem solving. PSO was developed by Kennedy and Eberhart in 1995 and has been successfully applied in a great variety of contexts

(Kennedy and Eberhart 1995, Poli et al. 2007). PSO is a kind of SI that simulates the social behavior of a flock of birds or fish schooling when moving all together following a common tendency in their displacements.

A PSO algorithm is initialized with a population (swarm) of "particles." Each particle represents a single candidate solution of the optimization problem and makes use of its individual memory and knowledge gained by the swarm as a whole to try to find the best solution to the problem. In PSO the particle swarm simulates the social optimization commonly found in communities with a high degree of organization. For a given problem, some fitness function is needed to evaluate the proposed solution. In order to get a good one, PSO methods incorporate both a global tendency for the movement of the set of individuals and local influences from neighbors (Eberhart and Kennedy 1995, Kennedy and Eberhart 1995).

PSO procedures start by choosing a swarm of random candidate solutions in the search space. Then they are displaced throughout their domain looking for an optimum, taking into account global and local influences, the latest coming from the neighborhood of each particle. To this purpose, all particles have a position $X_i(t)$ and a velocity $V_i(t)$ that allows updating the particle's position in each iteration.

The initial position vectors $X_i(0)$ and velocity vectors $V_i(0)$ are randomly selected over the search space. Then these particles evolve all through the space according to two essential reasoning capabilities: a memory of their own best position and knowledge of the global or their neighborhood's best. The meaning of the "best" must be understood in the context of the problem to be solved. In a minimization problem that means the position with the smallest value for the target function.

The dynamics of the particle swarm is considered along successive iterations, like time instances. Each particle modifies its position $X_i(t)$ along the iterations, keeping track of its best position in the variable domain implied in the problem. This is made by storing for each particle the coordinates ${P^b}_i$ associated with the best solution (fitness) it has achieved so far along with the corresponding fitness value, ${f^b}_i$. These values account for the memory of the best particle position. In addition, members of a swarm can communicate good positions to each other, so they can adjust their own position and velocity according to this information. To this purpose, we also collect the best fitness value among all the particles in a neighborhood of each particle i, ${f^{bg}}_i$, and its position ${P^{bg}}_i$ from the initial iteration. This is a social information for modifying the position of each particle. Different PSO algorithms have been developed which differ in the size and topology of their neighborhoods (Engelbrecht 2005). In the simplest implementation of PSO, known as global best PSO, the neighborhood for each particle is the entire swarm, that is, the social information reflects information obtained from all the particles in the swarm. Furthermore, the local best PSO uses a ring social network topology where smaller neighborhoods are defined for each particle; in this case the social information reflects local knowledge of the environment.

The evolution for each particle i is given by

$$V_i(t+1) = wV_i(t) + \alpha r_1\left(P_i^{bg}(t) - X_i(t)\right) + \beta r_2\left(P_i^b(t) - X_i(t)\right) \tag{10.3}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{10.4}$$

where $X_i(t)$ and $V_i(t)$ are the position and the velocity of particle i at time t, respectively; w is called inertia weight and it decides how much the old velocity will affect the new one; and coefficients α and β are constant values called acceleration coefficients, which decide the degree of affection of $P^{bg}{}_i$ and $P^b{}_i$. In particular, α is a weight that accounts for the "social" component, while β represents the "cognitive" component, accounting for the memory of an individual particle along the time. The acceleration coefficients are also referred to as trust parameters, where α expresses how much confidence a particle has in its neighbors, while β expresses how much confidence a particle has in itself. Finally, two random numbers, r_1 and r_2, with uniform distribution on $[0, 1]$ are included to enrich the searching space. Figure 10.1 shows the three components of the velocity vectors—inertia, social component, and cognitive component—and their influence in the new position of each particle.

After the update process of positions and velocities, a fitness function must be given to evaluate the quality of the new positions, and the local and global best positions of each particle must also be updated. This procedure is repeated several times (thus yielding successive generations) until a termination condition is reached. Common terminating criteria are that a solution is found that satisfies a lower threshold value, or that a fixed number of generations has been reached, or that successive iterations no longer produce better results. The final PSO procedure is briefly sketched in Table 10.1.

As happens in other metaheuristics, the basic PSO algorithm is influenced by a number of control parameters: swarm size, neighborhood size, inertia weight,

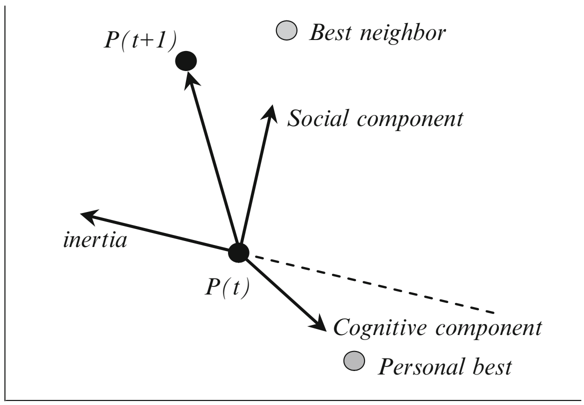


Fig. 10.1 Components of the velocity vector in a PSO algorithm

Table 10.1 General structure of a PSO algorithm

begin
  t=0
  random initialization of individual positions X_i and velocities V_i in *Pop(t)*
  fitness evaluation of *Pop(t)*
    while (not termination condition) **do**
      for each particle i in *Pop(t)* **do**
        Calculate best fitness particle in the neighborhood $P^{bg}{}_i(t)$
        Calculate particle position $P^b{}_i(t)$ with best fitness
        Calculate velocity $V_i(t+1)$ for particle i according to (3)
        while not feasible $X_i(t)+V_i(t+1)$ **do**
          Apply scale factor to $V_i(t+1)$
        end
        Update position X_i according to (4)
      end
      $t = t + 1$
    end
end

acceleration coefficients, and number of iterations. Implementing a PSO algorithm requires a careful selection of these parameters.

10.4 Metaheuristics for Management Aquaculture Farms

The complexity of the problems that arise in the management of aquaculture farms makes metaheuristic techniques to be particularly suitable for the solution of many optimization problems of practical importance. In this section we will examine the research contributions to the application of metaheuristic techniques and swarm intelligence in different problems related to optimal management of aquaculture farms. In particular, several research works that can be found in the literature use population-based metaheuristics; perhaps the most commonly used in this area are genetic algorithms (GA), which is one of the modern optimization techniques because of its evolutionary nature; it can handle any kind of objective function and constraint. Recent works of application of GA in aquaculture are cited below.

Atia et al. (2012) use genetic algorithm to the optimal design of solar water heating systems; the work presents a model of a forced circulation solar water heating system for supplying a hot water at a certain temperature for an aquaculture system. In this context Liu et al. (2011) address the water quality prediction in aquaculture management using a hybrid approach. In the work, a prediction model based on support vector regression is proposed, and genetic algorithms are used for the optimal selection of parameter values. Gutiérrez-Estrada et al. (2012) evaluate several linear and nonlinear models for modeling the water exchange process in gilthead sea bream semi-intensive aquaculture systems; in particular, they use genetic algorithms to find the optimal values of the parameters of a fuzzy logic model. A very different application is the use of genetic algorithms in environmental models to predict the potential distribution of invader species. In Therriault and Herborg (2008), a genetic algorithm for rule-set prediction environmental model is presented; the model analyzes the potential distribution of a tunicate specie in Canadian waters. Chen et al. (2007) use the same methodology to model the niches of silver carp and bighead carp in their native ranges using hydrologic and general environmental parameters in concert with native distributional data. The model suggests that both species have the potential of spreading throughout the eastern USA and selected areas of the West Coast. Genetic algorithms have also been used to design an optimal diet composition with the objective to ensure the maximal survival in the culture of common octopus (Hormiga et al. 2010). Other applications of genetic algorithms in aquaculture are fish stock-recruitment processes (Chen et al. 2000), predicting fish distributions (D'Angelo et al. 1995), optimization of multi-objective fisheries bioeconomic models (Mardle et al. 2000), or optimization of harvest management strategies of many species, considering interactions between and within species (Stafford 2008).

ACO is not a metaheuristic that has been widely used in practical problems of aquaculture management; however, it is also possible to find some practical applications. For example, Liao et al. (2012) develop four metaheuristic algorithms for the solution of grouping problem in high-throughput cryopreservation operations of fish sperm; one of the algorithms proposed is based on ant colony optimization. Another area of application of optimization techniques is transport and logistics; in the case of the distribution of aquaculture products, it is possible to find works that apply other types of metaheuristics, such as simulated annealing (Wang et al. 2012).

Although there are fewer works making use of PSO techniques, it is possible to find interesting practical applications that have been addressed with these techniques. For example, in Deng et al. (2006), a PSO algorithm was proposed to train a fuzzy neural network that defines a prediction model for dissolved oxygen concentration in fishponds. This work presents experimental results showing that the proposed method is effective and more accurate than other conventional approaches as back-propagation method. Xuemei et al. (2011) address the same problem of prediction of dissolved oxygen by using a neural network but introduce an adaptive genetic algorithm to optimize the network and make it faster convergence. Chau (2005) uses PSO for training perceptrons to forecast real-time algal bloom dynamics on the basis of several input hydrodynamic and/or water quality variables. It is shown that when compared with the benchmark back-propagation algorithm, its results can be attained both more accurately and speedily.

10.4.1 Example of Application: A PSO Model for Optimal Management of Aquaculture Farms

Similar to other animal breeding industries, aquaculture production is based in the daily growth of individuals. Modeling this biological process is complex due to the broad range of factors that influence fish growth, which are both internal, such as feeding rates or fingerling weight, and external, such as water temperature or the social behavior of individuals. In this complex scenario, bioenergetic models measure the daily growth of fish as the energy gain from the difference between the energy input and output. By modeling the complex interactions between economic and biological systems, it is possible to make the most efficient decisions about aspects such as the diet composition, feeding rates, and harvesting time. Thus, bioeconomic models have played a crucial role in OR methods (Bjørndal et al. 2004).

Although these studies have contributed significantly to the optimization of production planning, the complex interactions of technical, biological, and economic aspects in fish farming limit the application of classical optimization methods. A useful method when the optimization problem is complex, stochastic, and nonlinear is the particle swarm optimization (PSO). As an example of application, we present in this section a PSO model to address problems of production planning in aquaculture farms. This model aims to provide aquaculture producers with a methodology and a tool to facilitate their decision-making process, in particular, the determination of optimal time of planting and harvesting. Determining the optimal planting and harvesting strategy allows managers to maximize results and eliminate operational risk, and swarm intelligence techniques are particularly suited to address such problems and eliminate the uncertainty of aquaculture enterprises due to the complexity and the large number of factors and constraints involved in the process.

The model consists of a biological submodel of the process of farming in sea cages, which is interrelated with an economic submodel that quantifies the process so as to continue the economic implications of any change in the parameters of farming and market. Different factors are considered in the proposed model: abiotic factors affecting the growth process; feeding rate and daily growth, which are a function of water temperature and the average weight of fish; and survival rates depending on the average weight of specimens and environmental conditions. The proposed model considers weights in harvest time, fingerling weights, and delay periods between harvests as decision variables. The objective function to optimize is the farming gross margin obtained in a period of time. To calculate this function the cost of stocking and the cost of feed are considered.

The use of a population-based metaheuristic like PSO provides additional advantages. Firstly, it provides a set of possible schedules close to the optimal and allows to perform a complete analysis of the temporal evolution of the process of fattening and bioeconomic analysis of production location.

10.5 Particular Case: Production of Sea Bream in Floating Cages

Traditionally, gilthead sea bream were cultured extensively in coastal lagoons and saltwater ponds until intensive rearing systems were developed during the 1980s. Artificial breeding was successfully achieved in Italy in 1981–1982, and large-scale

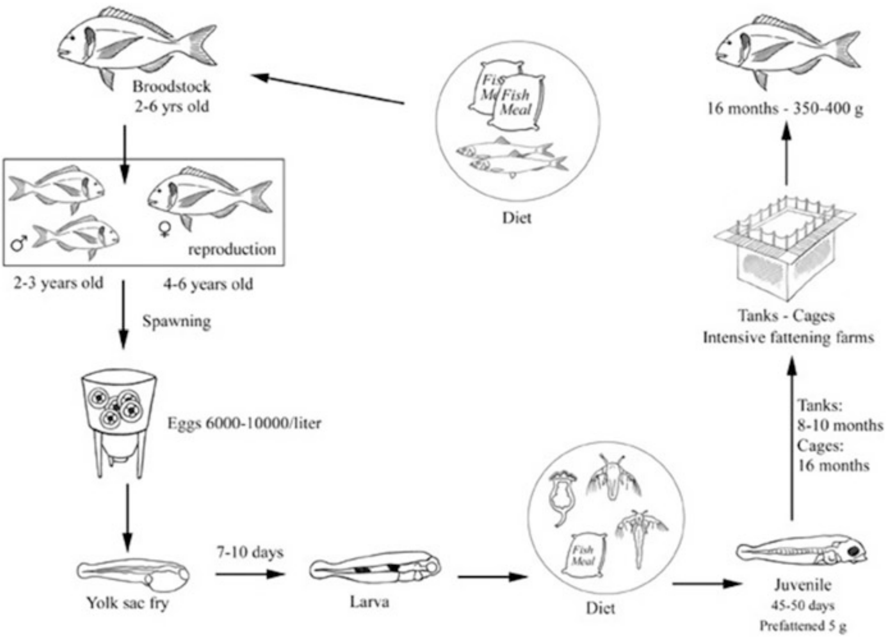


Fig. 10.2 Production cycle of *Sparus aurata*—intensive system (*Source*: Food and Agriculture Organization of the United Nations http://www.fao.org/fishery/culturedspecies/Sparus_aurata/en)

production of gilthead sea bream juveniles was definitively achieved in 1988–1989 in Spain, Italy, and Greece.

Gilthead sea bream are very suitable species for extensive aquaculture in the Mediterranean due to their good market price, high survival rate, and feeding habits (which are relatively low in the food chain). This species very quickly demonstrated a high adaptability to intensive rearing conditions, both in ponds and cages. Ongrowing in sea cages is simple and economical, and it is the fattening system normally used in the Mediterranean basin. In sea cages biomass densities are lower than in tanks (10–20 kg/m^3), but there are great advantages that make cage farming more profitable: no energy costs for pumping, aeration, or post-rearing water treatment. However, it is not possible to control temperature in cage rearing, it is necessary to stock larger juveniles, and longer rearing periods to market size are needed. On average, larger pre-fattened gilthead sea bream (10 g) reach first commercial size (350–400 g) in about 1 year, while smaller juveniles (5 g) reach the same size in about 16 months. The production cycle of an intensive system of production of *Sparus aurata* is shown in Fig. 10.2.

10.6 Objective Function and Decision Variables

The proposed approach aims to determine the optimal planting and harvesting strategy in order to maximize the present value of farming gross margin obtained in a period of time and eliminate the uncertainty of aquaculture enterprises. To calculate the gross margin for farming, the only costs considered are the cost of stocking and the cost of feed.

Production costs do not fit a linear function in aquaculture. There are two issues that directly affect gross margin: location, which determines the environmental variables, logistic costs, etc., and decision of operational strategy, that is, the sequencing of the strategy of planting and harvesting. In this example of application, only variables associated with operational strategy are considered. Variables associated with environmental conditions are considered as input parameters in the model.

A planted and harvesting strategy in a time period is defined by the following decisions:

- Number of seed-harvest processes in the time period.
- Time of planting.
- Juvenile weight.
- Market size, that is, weight in harvest time. This size determines the time of harvesting.
- Diet scheduling.

Stocking costs are calculated by estimating the number of fingerlings in order to reach the maximum biomass density in the cage at the time of harvesting and considering the price of the juvenile as a function of its weight.

In order to simplify the model, we assume that the production process uses a single floating cage and only one type of feed.

We have developed a bioeconomic model to evaluate the production of gilthead sea bream in floating cages based on location. The biologic submodel contains a growth model based on the specie physiology and using the following functions:

- $T(t)$: water temperature at time t
- $R(T(t), w(t))$: ration size as function of water temperature and fish weight
- $GR(T(t), w(t))$: specific fish growing rate as function of water temperature and fish weight, that is, a function that relates feed consumption to weight gain
- $M(T(t), w(t))$: probability of mortality depending on the temperature and fish weight

In the model, the time is considered as a discrete variable that counts the number of days from an initial date t_0. The function $T(\,)$ is determined by the environmental conditions at the location of the farm; whereas the functions $R(\,)$ and $GR(\,)$ are determined by the technical specifications provided by the manufacturer of the feed. The daily mortality rate is a function with a larger value for fingerlings than for individuals with more of 50 g.

Table 10.2 Input parameters: characteristics of the farm and environmental conditions

Location	Canary islands
Maximum biomass density	20 kg/m^3
Volume of the cage	100 m^3
Maximum number of days between harvesting and seeding	60 days
Feasible sale sizes	[300, 700] g
Available juvenile weight	[3, 10] g
Temporal horizon	5 years
Maximum number of seed-harvest processes	8

If N_t represents the number of individuals in the cage at time t, N_o can be initialized with the number of fingerlings that allows to reach the maximum biomass density in the cage at the time of harvesting, and then the following rule is used:

$$N_t = N_0 \prod_{k=0}^{t} \left(1 - M\left(T(k), w(k)\right)\right). \tag{10.5}$$

And the weight function is defined as follows:

$$w(t) = w(0) \prod_{k=0}^{t} \left(1 + GR\left(T(k), w(k)\right)\right) \tag{10.6}$$

where $w(0)$ is the weight of the seed juveniles.

The economic submodel aims to maximize the gross margin of breeding. This gross margin must be calculated considering juvenile prices as function of its weight $w(0)$, sale price of the final product, minimum weights for sale, seasonality of prices, and feeding prices. All prices are also weight dependent.

The model also considers a set of parameters that define the characteristics of the aquaculture farm, the environmental conditions, and several constraints that limit the feasible production strategies. In the example to be presented below, the values of these parameters are shown in Table 10.2.

10.7 Application of a PSO Algorithm

In order to apply a PSO strategy for planning the optimal production of a farm with the characteristics described in Table 10.2, we must follow the general structure defined in Table 10.1. The algorithm starts with random initialization of particle position and velocity. Each position in the particle represents a production planning:

$$P_i(t) = (r_{i1}(t), hw_{i1}(t), jw_{i1}(t);\ \ r_{i2}(t), hw_{i2}(t), jw_{i2}(t);\ \ \ldots;\ \ r_{in}(t), hw_{in}(t), jw_{in}(t)) \tag{10.7}$$

where n is the maximum number of planting-harvesting processes to perform in the period under planning; $r_{ik}(t)$ represents the number of days of delay with respect to the previous process of harvest; $hw_{ik}(t)$ represents the harvest weight of the fish; and $jw_{ik}(t)$ represents the juvenile weight. The velocity vector has the same structure. Both vectors are initialized randomly in the feasible ranges.

The PSO is influenced by a number of control parameters; Table 10.3 shows the values of these parameters used in the example.

The basic PSO algorithm was implemented in a Web-based application that can be used in decision-making processes by managers of aquaculture farms. Figure 10.3 shows a screenshot of this application with a set of potential near-optimal planting-harvesting strategies obtained using a PSO algorithm.

Table 10.3 Control parameters of the PSO algorithm

Number of particles	80
Neighborhood topology	Ring
Neighborhood size	20
Inertia weight	1.0
Weight of the social component (α)	1.0
Weight of the cognitive component (β)	0.4
Termination criteria (number of consecutive iterations without improving)	15

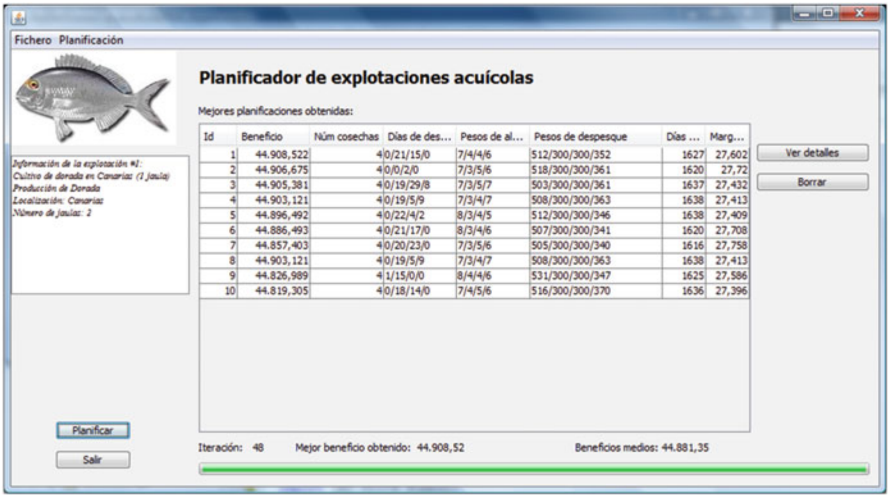


Fig. 10.3 Web-based application for optimal management of planting-harvesting processes in aquaculture farms

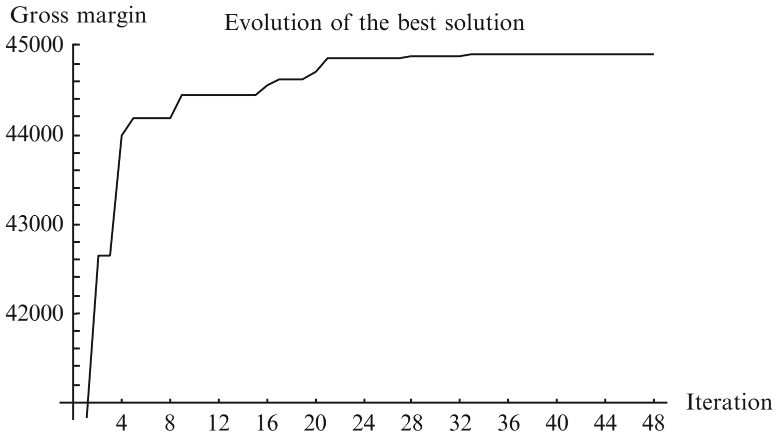


Fig. 10.4 Evolution of best gross margin for each generation along the iterations of the PSO algorithm

After applying the algorithm with the parameter values defined in Tables 10.2 and 10.3, a set of 80 potential solutions is obtained. Figure 10.3 shows only the ten best solutions. The fact of offering to the decision-maker a set of good alternatives or near-optimal schedules facilitates the final decision and allows considering new factors not initially included in the model. The implemented tool also provides decision-makers with detailed information on the evolution of the product (number of fish available, average size of fish, daily feeding ration, etc.); all this information is extremely useful in the control processes.

Figure 10.4 displays the evolution of best gross margin for each generation along the iterations. As can be seen, a total of 48 iterations were required to complete the algorithm. Also observe how as it advances the gross margin of the schedules found also improves.

Although the maximum number of permitted harvests in the 5-year temporal horizon was 8, all of the schedules obtained in the final iteration could fit only 4 harvests. Specifically, the best position found by the particles determined the following planning:

$$P_{best} = (0, 512, 7;\ \ 21, 300, 4;\ \ 15, 300, 4;\ \ 0, 352, 6).$$

As can be observed, the days of delay with respect to the previous process of harvest are minimized ($r_1 = 0$; $r_2 = 21$; $r_3 = 15$; and $r_4 = 0$); different juvenile weights are used ($jw_1 = 7$; $jw_2 = 4$; $jw_3 = 4$; and $jw_4 = 6$), and in most of harvest, the harvest weight is close to the minimum commercial size (300 g). However, in the first harvest, the algorithm recommends delaying time to market of the product. Given fattening rates, survival probabilities, and farm parameters, it is possible to determine the exact times in which to proceed to planting and harvesting each harvest. Specifically, the algorithm recommends to the producer the following strategy:

- Day 1: Initial planting with 8,232 juveniles of 7 g
- Day 514: Harvesting with 512 g of fish weight. Number of surviving fish, 7,812
- Day 535: Second planting with 13,866 juveniles of 4 g
- Day 884: Harvesting with 300 g of fish weight. Number of surviving fish, 13,320
- Day 899: Third planting with 13,832 juveniles of 4 g
- Day 1,248: Harvesting with 300 g of fish weight. Number of surviving fish, 13,287
- Day 1,248: Fourth planting with 11,830 juveniles of 6 g
- Day 1,627: Harvesting with 352 g of fish weight. Number of surviving fish, 11,356

The gross margin associated with this best production plan is 44908.52 €. As happens in any metaheuristics, there is no guarantee of optimality. However, planning obtained makes rational use of resources and generates a near-optimal strategy.

10.8 Conclusions

Swarm intelligence offers a new and powerful approach to the optimization problems, and their use is made possible by the increasing availability of high-performance computers at relatively low costs. These algorithms have recently found extensive applications in solving global optimization searching problems when the classical optimization techniques cannot be applied.

In spite of their wide applicability and simplicity, metaheuristic techniques are not being used in the management of aquaculture farms as widely as in other contexts. In particular PSO techniques have proven to be more efficient than genetic algorithms, making them well suited to address complex problems that arise in the management of such farms. It is demonstrated that PSO gets better results in a faster, cheaper way compared with other methods. In this work we have presented a PSO approach for optimal planning of production of sea bream in floating cages as an example of the applicability of swarm intelligence to find the optimal operational strategies that maximizes gross margin and to delimit operational risk. In general, swarm intelligence provides aquaculture managers with methodologies and useful tools that can be used in decision-making processes in aquaculture systems.

References

Atia DM, Fahmy FH, Ahmed NM, Dorrah HT (2012) Optimal sizing of a solar water heating system based on a genetic algorithm for an aquaculture system. Math Comput Model 55 (3–4):1436–1449

Beni G (1998) The concept of cellular robotic systems. In: Proceedings of the IEEE international symposium on intelligent systems. IEEE, Piscataway, pp 57–62

Bjørndal T, Lane DE, Weintraub A (2004) Operational research models and the management of fisheries and aquaculture: a review. Eur J Oper Res 156:533–540

Blum C, Li X (2008) Swarm intelligence in optimization. In: Blum C, Merkle D (eds) Swarm intelligence: introduction and applications. Springer, Berlin

Blum C, Roli A (2003) Metaheuristics in combinatorial optimization: overview and conceptual comparison. ACM Comput Surv 35(3):268–308

Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm intelligence: from natural to artificial systems. Oxford University Press, New York

Chau KW (2005) Algal bloom prediction with particle swarm optimization algorithm. Lect Notes Comput Sci 3801:645–650

Chen DG, Hargreaves NB, Ware DM, Liu Y (2000) A fuzzy logic model with genetic algorithm for analyzing fish stock-recruitment relationships. Can J Fish Aquat Sci 57(9):1878–1887

Chen P, Wiley EO, Mcnyset KM (2007) Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biol Invasions 9(1):43–51

D'Angelo DJ, Howard LM, Meyer JL, Gregory SV, Ashkenas LR (1995) Ecological uses for genetic algorithms: predicting fish distributions in complex physical habitats. Can J Fish Aquat Sci 52(9):1893–1908

Deng C, Wei X, Guo L (2006) Application of neural network based on PSO algorithm in prediction model for dissolved oxygen in fishpond. In: Proceedings of the World Congress on Intelligent Control and Automation (WCICA), pp 9401–9405

Dorigo M (1992) Optimization, learning and natural algorithms. Ph.D. Thesis, Politecnico di Milano

Dorigo M, Stützle T (2004) Ant colony optimization. Bradford/MIT, Cambridge

Eberhart RC, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp 39–43

Engelbrecht AP (2005) Fundamental of computational swarm intelligence. Wiley, Chichester

Glover F (1986) Future paths for integer programming and links to artificial intelligence. Comput Oper Res 13:533–549

Gutiérrez-Estrada JC, Pulido-Calvo I, De la Rosa I, Marchini B (2012) Modeling inflow rates for the water exchange management in semi-intensive aquaculture ponds. Aquac Eng 48:19–30

Hoffmeyer J (1994) The swarming body. In: Rauch I, Carr GF (eds) Semiotics around the world. In: Proceedings of the Fifth Congress of the International Association of Semiotic Studies, pp 937–940

Hormiga JA, Almansa E, Sykes AV, Torres NV (2010) Model based optimization of feeding regimens in aquaculture: application to the improvement of Octopus vulgaris viability in captivity. J Biotechnol 149(3):209–214

Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks. Perth, Australia, pp 1942–1948

Kochenberger GA (2003) Handbook of metaheuristics. Springer, Berlin

Liao TW, Hu E, Tiersch TR (2012) Metaheuristic approaches to grouping problems in high-throughput cryopreservation operations for fish sperm. Appl Soft Comput 12(8):2040–2052

Liu S, Tai H, Ding Q, Li D, Xu L, Wei Y (2011) A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Math Comput Model. doi:10.1016/j.mcm.2011.11.021

Mardle SJ, Pascoe S, Tamiz M (2000) An investigation of genetic algorithms for the optimization of multi-objective fisheries bioeconomic models. Int Trans Oper Res 7:33–49

Osman IH, Laporte G (1996) Metaheuristics: a bibliography. Ann Oper Res 63:513–623

Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization: an overview. Swarm Intell 1(1):33–57

Stafford R (2008) A computational approach to ecological and economic sustainable harvest management strategies in a multi-species context, with implications for cod recovery plans. Ecol Informat 3(1):105–110

Therriault TW, Herborg LM (2008) Predicting the potential distribution of the vase tunicate Ciona intestinalis in Canadian waters: informing a risk assessment. ICES J Mar Sci 65(5):788–794

Wang EJ, Tsai DM, Su TS, Lin KY (2012) Simulated annealing for cost-effective transport of live aquaculture products. Aquac Econ Manag 6(1):68–95

Xuemei H, Yingzhan H, Xingzhi Y (2011) The soft measure model of dissolved oxygen based on RBF network in ponds. In: Proceedings 4th International Conference on Information and Computing, ICIC 2011, pp 38–41

Yang XS (2012) Nature-inspired metaheuristic algorithms: success and new challenges. J Comp Eng Inform Technol 1(1). doi: 10.4172/jceit.1000e101

Chapter 11
Multi-objective Optimization for Improved Agricultural Water and Nitrogen Management in Selected Regions of Africa

M. Pastori, A. Udías, F. Bouraoui, A. Aloe, and G. Bidoglio

11.1 Introduction

African agriculture has enormous potential for growth thanks to its natural resources, including water and land, that in many cases are only partially used: for example, only 10 % of the crop-suitable land in the Guinean savannah is actually cropped (Morris et al. 2009).

Nitrogen pollution is recognized as a crucial threat to our planet together with biodiversity loss and climate change (Giles 2005). It can be considered as the main factor for increasing crop production and has become an issue for the environment after 1970 when the amount of global reactive N increased rapidly (Zavattaro et al. 2012).

Research shows that in most African countries, the main limiting factor to crop production is nitrogen, while water limitation is more restricted in only a few countries (Pastori et al. 2011). Furthermore, it can be estimated that irrigation may substantially increase yield in water-rich regions, but the lack of infrastructure does not allow these countries to reach higher production levels. On the other hand, some North African countries are already mining water resources (Pastori et al. 2011), and an improved sustainable use of water resources will be needed, in particular, to cope with climate change (drier and hotter climate). In this context, the problem of nitrogen and, more generally, of nonpoint source pollutants (NPS), such as phosphorus (P), sediment, and pesticides, shall increase in the future, and it will require the availability of tools and methods to optimize the use of fertilizer and irrigation.

M. Pastori • A. Udías (✉) • F. Bouraoui • A. Aloe • G. Bidoglio
European Commission Joint Research Centre (JRC),
Institute for Environment and Sustainability, Ispra, Italy
e-mail: angel.udias-moinelo@jrc.ec.europa.eu

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_11

African agriculture will need to invest capital and technology to adapt to the new situation: higher yields required multi-cropping systems, new or more intensive irrigation systems, increased fertilizer application, changes in sow rates and livestock types, etc. (Easterling and Apps 2005; FAO 2011). However, the inadequate implementation of these measures could increase the impacts on the environment and lead to new conflicts between users of ecosystems (Schröter et al. 2005, IPCC 2007). For example, increased use of water for irrigation could conflict with the demands of water for domestic or industrial use, leading to adverse ecological effects (Bates et al. 2008). In addition, loss of soil by erosion may increase due to climate change; this effect could be exacerbated by changes in land management (Lee et al. 1999).

Fortunately, researchers are developing models to evaluate the economic impacts of environmental effects to achieve sustainable agriculture. They usually use a crop-simulator model to predict the environmental effects of various management practices and then link to an optimization model to determine trade-offs between economic returns and environmental impacts. The final output is usually a sustainable compromise between economic achievements and environmental quality (Weintraub and Romero 2006).

Multi-objective optimization methods in connection with biophysical models have shown great potential for addressing such issues of opposing management goals. Johnson et al. (1991) linked CERES, a crop-simulator model, to a dynamic optimization model to determine the optimum applications of water and fertilizers needed to maximize gross margin. Zekri and Herruzo (1994) combined NTRM, a crop-simulator model, and a mathematical mixed multi-objective programming model to assess the effects of an increase in nitrogen prices and a reduction in drainage irrigated water, thus inducing the adoption of best management practices. Finally, Teague et al. (1995) used the EPIC-PST simulation model to predict the environmental risks of using pesticides and nitrates. Sadeghi et al. (2009) applied an optimization approach to maximize profits from land use, while minimizing erosion risk. Meyer et al. (2009) coupled SWAT (soil and water assessment tool) with an optimization routine to determine optimum farming system patterns to reduce nitrogen leaching while maintaining income. Similarly, Whittaker et al. (Whittaker et al. 2009) applied SWAT in connection with a Pareto-optimization approach considering profits from land use and chemical pollution from farm production. Latinopoulos (2009) applied optimization to a problem of water and land resource allocation in irrigated agriculture with respect to a series of socioeconomic and environmental objectives. Thus, the multi-objective modeling of joint production processes that combine private goods sold on the market place and public goods without established markets, such as environmental protection, is an important line of research (Nalle et al. 2004).

The overall goal of this paper is to propose a multi-objective methodological tool, able to incorporate conflicting elements such as environmental objectives and economical issues, to identify optimum crop and land management patterns in different African countries, demonstrating the ability to provide trade-off Pareto solutions which simultaneously minimize total nitrate pollution through runoff and leaching and at the same time maximize the exploitation benefits by choosing the adequate crop, fertilization, and irrigation management sequences. Knowledge of these sets helps the decision-maker to choose optimum alternative patterns of agricultural management adapted to countries with multiple soil types and different climate, soil composition, and crops.

The methodological tool integrates the multi-objective evolutionary algorithm presented in Udias et al. (2011) with the GIS-EPIC modeling tool developed at the African continental scale and described in Pastori et al. (2011). Results for different crops and African countries are presented to illustrate the method which is shown to be powerful and fully operational in making management decisions. This methodology may become an important support for policy makers, farmers, and water managers, providing information about cost-effectiveness of different agricultural practices in African countries.

11.2 Multi-objective Evolutionary Optimization Model

The starting point for handling multi-objective optimization problems (MOPs) is to consider a set of best alternatives or solutions that represent optimal criterion trade-offs. When a scenario involves an arbitrary optimization problem with M objectives, all of them to be maximized, a general multi-objective problem can be formulated as follows:

$$\begin{aligned} &\text{maximize} \quad f_m(x), \quad m = 1, 2, \ldots., M \\ &\text{subject to}: g_j(x) \geq 0, \quad j = 1, 2, \ldots, J \\ &\qquad\qquad h_k(x) = 0, \quad k = 1, 2, \ldots, K \\ &\qquad\qquad x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1, 2, \ldots, n \end{aligned} \tag{11.1}$$

where x is a vector of n decision variables $x = (x_1, x_2, \ldots, x_n)^T$. The terms $g_j(x)$ and $h_k(x)$ are called constraint functions and $f_m(x)$ is the multi-objective function. J inequality and K equality constraints are associated with the problem. The last subsets of constraints are called variable bounds, which restrict each decision variable x_i to take a value within an interval with a lower $x_i^{(L)}$ and an upper $x_i^{(U)}$ bound. All of these constraints define the decision variable space D or simply the decision space. In this case, a Pareto-optimal objective vector $f^* = (f_1^*, f_2^*, \ldots, f_M^*)$ is such that it does not exist any feasible solution x' and corresponding objective vector $f' = (f_1', f_2', \ldots, f_M') = (f_1(x'), f_2(x'), \ldots, f_M(x'))$ such that $f_m^* \leq f_m'$ for each $m = 1, 2, \ldots., M$ and $f_j^* < f_j'$ for at least one $1 \leq j \leq M$.

Many applications of multi-criteria methods conclude that their main value does not lie in providing the "answer," but in endowing such process with an improved transparency, setting a better structuring of the problems, and facilitating decision-maker learning (Ananda and Herath 2003; Prato 1999; Mills et al. 1996). Even if decision-makers disagree with MCA's output, it can still provide a valuable input to the decision procedure (RAC Resource Assessment Commission 1992). The notion of multi-criteria methods as a "glass box," rather than a "black box," suggests that those using multi-criteria techniques can understand in a better way the implicit trade-offs and appreciate the consequences of alternative preference positions.

The variety of techniques for solving continuous and nonlinear multi-objective optimization methods has grown rapidly over recent decades. Usually the methods are divided into three major categories: methods with a priori articulation of preferences (which implies that the user indicates the relative importance of the objective functions or desired goals before running the optimization algorithm), methods with a posteriori articulation of preferences (which entail selecting a single solution from a set of mathematically equivalent solutions), and methods that require no articulation of preferences are addressed. A complete survey of all them is included in Marler and Arora (2004).

Multi-objective genetic algorithms provide an approach for a posteriori articulation of preferences; they are intended for depicting the complete Pareto-optimal set. In this sense, they provide an alternative to the a posteriori methods. There, the methods determine one Pareto point at a time, and each point requires the solution of a single-objective optimization problem. Alternatively, multi-objective genetic algorithms do not require solving a sequence of single-objective problems; it has the ability to converge on the Pareto-optimal set as a whole. In addition, these algorithms are relatively robust, which has led some researchers to combine them with gradient-based methods (Poloni et al. 2000; Coverstone-Carroll et al. 2000).

In our approach for identifying optimum patterns of agricultural management considering multiple ecosystems of countries with different climate, soil composition, and crops, we applied a multi-objective evolutionary algorithm (MOEA) routine which had proven to be highly suitable for addressing complex nonlinear and combinatorial problems in many previous applications (Udías et al. 2007, 2009, 2011; Galbiati et al. 2007) and provides the Pareto cost-efficient management strategies which usually helps the decision-maker to choose the best alternative.

MOEA is an iterative search algorithm that is based on the principles of evolution (Goldberg 1989). A solution is represented as a "genome." The optimization starts with an initial "population" of "genomes." In each iterative step, the "genomes" of the "population" are evaluated with a defined objective function and the "fittest genomes" are chosen to be recombined. The newly generated solutions or "offspring genomes" also are evaluated, and the least "fit genomes" are excluded from the population to maintain the original population size. The MOEA algorithm applies the usual procedures of selection (tournament), crossover (multipoint), and mutation (uniform) to generate the new population. The Pareto fitness function and the tournament selection approach can be relatively efficient methods of incorporating into fitness a point's non-dominated tendencies. It also introduces elitism by searching for and storing Pareto-optimal points (separate from the general population) that surface with each one.

This process is continually repeated for a given number of iterations known as generations: a big population and a higher value of iterations are usually a guarantee of a better GA performance, but it requires longer computation time to reach optimal solutions. The output of the optimization process is a range of non-dominated solutions know as Pareto-optimal solutions (Deb 2001) that can be visualized in a two-or-more-dimensional plot to see the optimal solution trade-off between the different objective functions.

11.3 Methodology

The approach proposed in this work combines a biophysical model to predict the effects of candidate agricultural management practices, an economic model to estimate the benefit of the candidates, and a multi-objective evolutionary algorithm (MOEA) to search for best management practices. A flow chart of this approach is shown in Fig. 11.1.

The geodatabase includes all the African data required for a biophysical model (meteorological, soil, land use, and agricultural management). The biophysical model is coupled with a geodatabase covering the entire African continent and has proven to have a good performance in simulating actual yields in most of the African countries (Pastori et al. 2011).

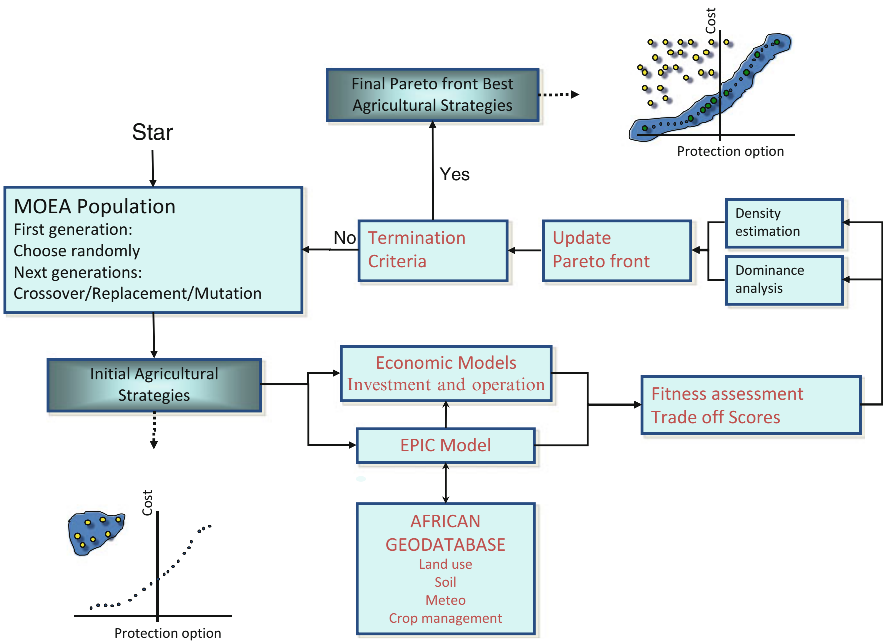


Fig. 11.1 Flow chart of the integrative MOEA with EPIC and GIS methodology

The MOEA generates candidate strategies according to the considered decision variables (regarding the amount and performance of the fertilization and irrigation processes); the biophysical model estimates the outputs (production, fertilization, irrigation, etc.) for the MOEA proposed strategies. These outputs are used by the economic model for determining the gross income, production costs, and net benefits for each crop and country. This iterative process continues until convergence is reached.

The final result is a set of efficient trade-off (Pareto) alternatives in relation to the exploitation of both resources are generated by the integrated tool, allowing comparisons between optimal and inefficient management of the agricultural production and water pollution. In addition, the proposed approach provides a platform for public discussion of alternative measures or management decisions between stakeholders with often opposing interests. It could also provide information, at the level of regions or watersheds, about the environmental impacts of real and simulated management practices to decision-makers.

11.3.1 The Biophysical Model and the African GIS

The use of biophysical models makes it possible to estimate, among other factors, crop nutrient uptake, the leached fraction of fertilizers, and soil erosion. These estimates provide decision-makers with information about the environmental impacts of current and modified farm management practices.

The biophysical model used in this study is the environmental policy integrated climate (EPIC) model. It is a biophysical, field-scale agriculture management model. It simulates crop water requirements and the fate of nutrients and pesticides as affected by farming activities such as timing of agrochemical application, tillage, crop rotation, irrigation strategies, etc., while providing a basic farm economic account at the same time. The main components can be divided in the following items: hydrology, weather, erosion, nutrients, soil temperature, and plant growth. For a detailed and complete description of the model and the simulated processes, see Williams (1984).

The model EPIC was chosen as it simulates crop production under different farming practices and operations including fertilization and irrigation application rates and timing and therefore considers nutrient losses to the environment. In addition, it has been thoroughly evaluated and applied from local to continental scale (Gassman et al. 2005) and used in global assessment (Liu et al. 2008, 2009; Pastori et al. 2011). The model has been applied for irrigation scheduling assessment (Rinaldi 2001; Wriedt et al. 2009), climate change studies (Mearns et al. 1999), and biofuel production and assessment (Velde et al. 2009).

The capability of the EPIC biophysical model to simulate cereal crop yields in the African continent was assessed in a previous study (Pastori et al. 2011) comparing simulated yields with the reported: maize, millet, sorghum, barley, and wheat showed a good correlation between simulated and measured values (at country level, R^2 value is around 0.8–0.9 for maize and sorghum and 0.6–0.7 for sorghum, wheat, millet, and barley).

11.3.2 Economic Model

The economic analysis was assessed taking various factors into consideration and the potential income deriving from the sales of the crop on the local market and the difference in costs of management in relation to the different management practices hence adopted (fertilization and/or irrigation levels). More specifically, for each crop, the selling price is considered homogenous at a country level, and the average from FAO statistical data was chosen as the reference value (see Table 11.1).

The value of nitrogen fertilizer is quite variable year by year depending on market trends possibly on local matters as well. For this study, we considered the average price of two important fertilizers commonly used and available on the African market: DAP and urea that result respectively at an average cost of 1.7 and 0.5 US$ per kg N ha^{-1}. In the simulations, a generic cost of 1.1 US$ per kg N ha^{-1} was used (Table 11.1). Having an accurate estimate of the water cost is an important issue when making a trade-off assessment of cost benefit of alternative management practices. Water cost is variable year by year depending on many variables such as the availability of water, the type of irrigation applied, and the investment cost required to set up the irrigation plant.

For this study, we decided to consider only the variable water cost for irrigation, not taking into account the fixed cost (e.g., linked with investments) as we wanted to focus on the assessment of different management practices also, considering the investment costs in developing countries can be paid or supported by local authorities or external organizations. The operating water cost was estimated by considering the use of a water pump to draw up and/or distribute water in the field, and for each country, this cost depends on the cost of diesel and the depth of the well. The method and the equation used are described in detail in Hogan et al. (2007).

The total income is computed according to (11.2) based on the EPIC output values of the yield, water consumption, and fertilization consumption:

Table 11.1 Crop values and water and fertilizer costs in each African country assumed in the analysis

Country	Crop value [US$ t^{-1}]				Water irrigation cost [US$ 100 mm^{-1}]	Fertilizer cost [US$ kg^{-1}]
	Maize	Wheat	Barley	Sorghum		
Algeria	270	340	210	160	7	1.1
Congo DEM	370	n.a.	n.a.	n.a.	9	1.1
Ethiopia	180	280	250	230	12	1.1
Libya	n.a.	280	250	n.a.	5	1.1
Morocco	250	280	220	300	28	1.1
Mozambique	150	n.a.	n.a.	140	9	1.1
South Africa	140	210	200	150	7	1.1
Tunisia	n.a.	250	160	n.a.	29	1.1

$$B = Y \times SP - WC \times WP - FC \times FP \tag{11.2}$$

where B represents benefits (US$/ha), Y yield (t/ha), SP crop selling price (US$/Tm), WC water consumption (m3),WP water price (US$/m3), FC fertilizer consumption (kg/ha) and FP fertilizer price (US$/kg).

11.3.3 Optimization (Genome Definition)

The optimization starts with an initial "population" of "genomes." With each iterative step, the "genomes" of the "population" are evaluated with the defined objective functions and the "fittest genomes" (non-dominated) are chosen to be recombined. The newly generated solutions or "offspring genomes" also are evaluated and the dominated genomes are excluded from the elitist population.

11.3.3.1 Genome Definition

The variables used to build the population of different management strategies are related mainly with the fertilization and irrigation practices that can be considered, in this analysis, to be the most important factors in the hands of farmers to control crop production.

The EPIC model was set with the auto-fertilization and auto-irrigation options. In this case, the model estimates the amounts and the number of operations required according to crop-specific characteristics and to user-defined parameters. The decision variables considered in this study are described in Table 11.2, including the ranges, which are the management elements on which it is possible to act. EPIC determines the daily amounts of fertilizer and water that should be applied to each crop, cell, and day in the under consideration period, according to the value of these variables: the climatic conditions, the soil characteristics, and the crops presented in each cell.

For example, the fertilization is calculated on the basis of a stressing factor linked to nitrogen availability that can range from minimum of 0 (no stress and no fertilization is applied) to a maximum of 1 (as soon as the crop encounters a nitrogen stress, the model applies the fertilizer).

Table 11.2 Range of the decision variables considered in all the analyses

Parameter	Description	Minimum	Maximum
BIR	Water stress (irrigation trigger)	0	1
VIMX	Maximum annual irrigation	20	900
ARMX	Maximum irrigation in single application	50	80
ARMN	Minimum irrigation in single application	10	60
BFT0	Fertilizer stress (fertilization trigger)	0	1

The final water or fertilizer amounts calculated are also dependent on local specific (for each cell of 15 km × 15 km) characteristics that influence crop growth such as the meteorological conditions and the soil type and fertility.

11.3.3.2 Objective Functions

The objectives simultaneously considered are:

(a) Maximization of the net income dependent on the crop yield production combined with the general cost of water and fertilizer and computed according to (11.2).
(b) Minimization of the environmental impact of the practice calculated as nitrogen leaching through the soil profile, potential indicator of soil, and subsurface and groundwater body quality.

The figures of nitrates are directly provided by the EPIC model. The EPIC model allows the determination of these figures, for each crop, country, and management practice, required to estimate total income by applying (11.2).

11.3.4 Case Study

To illustrate the implementation of the presented approach and its possible outcomes, we conducted a model at country scale in Africa for the most frequent cereal crops (see Table 11.1): maize, sorghum, millet, wheat, and barley. The maize is the most important one in Central and Southern Africa; wheat is also the most frequent crop in Northern Africa. For this analysis, we applied the optimization tool for each crop in five countries representative of different environments in the continent where the studied crop is dominant.

For the optimization study, the model was applied at a spatial resolution of $15 \times 15\ km^2$ individual cells characterized by uniform topography, soil, and climate. Each cell could account for up to five different crops/rotations. The simulation period is 20 years with measured data. The Harmonized World Soil Database (HWSD) with a resolution of about 1 km (30 arc sec) was used to characterize the soils. A detailed database of global land use (SAGE) describing the area (harvested) and yield of 175 distinct crops for the year 2000 on a 5 min by 5 min (approximately 10 km × 10 km) grid was used for crop area and management. A climate database with a resolution of 10 min was developed including daily precipitation, minimum and maximum temperature, wind speed, and solar radiation for the period 1961–2006. The geodatabase is linked with the EPIC model and was developed to support a quick application of the model for the entire African continent. The geodatabase includes all the data required for EPIC modeling (meteorological daily data, soil profile data, land use data with crop distribution, and agriculture management data) and all necessary sets of attributes required to simulate different strategies, management, and scenarios.

For a complete description of the geodatabase structure, the methodology adopted, and the input data used, refer to the European report (Pastori et al. 2011).

The integrated optimization tool was executed for each considered African country and crop. The MOEA population size was 10 chromosomes per generation, for 200 generations. The number of evaluations of the objective function (around 2000) is quite low compared to the requirements of similar problems. Experimentally, it was observed that in most cases, this small amount of evaluations was enough to achieve the convergence, most likely due to working with the option of auto-fertilization and auto-irrigation of EPIC, as it speeds the process of convergence to the Pareto trade-off. In any case, one of the advantages of our approach is the low number of evaluations of the biophysical model required, as each evaluation of one strategy scenario by EPIC requires high computational times. For most African countries, executing the EPIC model 2000 times requires more than two, one of the computations on a standard PC.

11.4 Results and Discussions

Tables 11.3 and 11.4 respectively reported the trade-off optimal strategy solutions for barley and maize. Included in both tables are averages of average yield, fertilizer, irrigation, N leaching, benefits, water cost, and fertilizer cost from the execution of the multi-criteria optimization tool.

One of the results of the analysis is that each country shows a particular pattern that is closely related to the specific characteristics of the country itself such as the climate, the soil, and the management (see Tables 11.3 and 11.4). Logically, average benefit also varies highly according to the country, the crop type, and obviously the crop selling price and management costs (the range is between 145 and 2,580 US$ ha^{-1}). Nitrate losses for optimized solutions range from 2 to 9 kg ha^{-1}, with the highest value corresponding to the maximization of gross margin. Nitrogen pollution varies widely between countries and it is so strictly linked with local climatic and soil condition that can facilitate nitrogen losses, especially under new high productive management practices.

Table 11.3 Average values of main outputs derived from the optimization process for barley

	Yield t ha^{-1}	Fertilizer kg ha^{-1}	Irrigation mm ha^{-1}	N leaching kg ha^{-1}	Benefits US$ ha^{-1}	Water cost US$ ha^{-1}	Fertilizer cost US$ ha^{-1}
Algeria	0.93	43	50	4.38	414	8	64
Libya	1.42	86	187	1.99	280	9	95
Morocco	1.91	73	149	2.59	304	41	73
South Africa	1.68	49	314	3.36	260	21	54
Tunisia	1.31	76	29	7.55	194	8	84

Table 11.4 Average values of main outputs derived from the optimization process for maize

	Yield t ha^{-1}	Fertilizer kg ha^{-1}	Irrigation mm ha^{-1}	N leaching kg ha^{-1}	Benefits US\$ ha^{-1}	Water cost US\$ ha^{-1}	Fertilizer cost US\$ ha^{-1}
Congo	6.56	155	46	8.29	2,251	4	171
Ethiopia	4.50	113	0	6.79	684	0	125
Mozambique	4.66	99	0	9.17	599	0	99
Nigeria	7.26	166	18	4.87	2,663	1	163
South Africa	1.68	49	314	3.36	260	21	54

Results of these tables have a fairly limited utility and only could be used to give orders of magnitude of the results for each crop and country. However, one of the main advantages of applying an evolutionary multi-criteria methodology approach is to provide decision-makers information about all the efficient trade-off between the objectives they have previously selected. This is the Pareto frontier curves which allow simultaneous comparisons of all efficient strategies and show the optimal solutions for best management practices that can be applied in the studied area.

The Pareto curves generated with the integrated tool are the result of an optimization process that considers concurrently the different objectives identified: in our case of the crop production, the management cost (in this case summarized with use of water and nitrogen fertilizer) and the impact on the environment (simplified for this analysis with the nitrogen leaching into the soil). All the solutions presented in the graph are optimized under the perspective of the objectives chosen, but, at this point, a further analysis or decision that should consider a wider context is required.

The integrated tool may apply up to five objectives simultaneously, for example, looking at how to maximize the production of a crop minimizing water consumption and fertilizer application as well as minimizing nitrate contamination. However, to simplify the understanding, in this chapter, we only present results obtained by simultaneously considering two objectives: net profit and nitrate percolation.

Figures 11.2 and 11.3 show the barley and maize values for the efficient trade-off strategies (Pareto solution) with net profit and minimization of nitrate percolation as objectives. That is, the values of Tables 11.3 and 11.4 are the average of the values obtained by each strategy of each country, shown in Figs. 11.2 and 11.3, respectively.

For each country, it is also possible to identify a range of optimal solutions where the increase on the benefit for the farmers is linearly correlated with the environmental impact. After a specific point, the relation benefit/nitrogen losses starts to increase much more rapidly and this implies that a small increase in the benefit corresponds to a high increase in the environmental impact. It is quite clear that the shape of the optimized solutions is strictly linked with the relation between

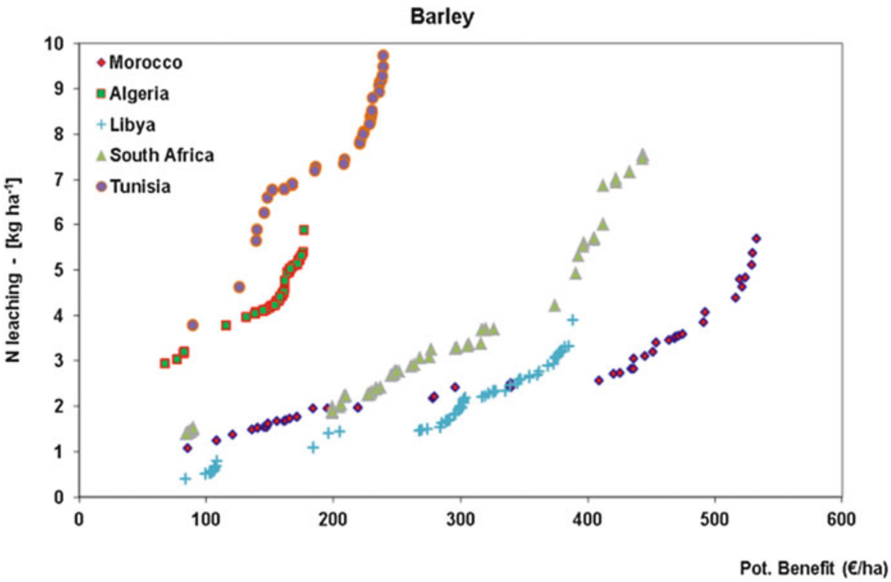


Fig. 11.2 Pareto-optimal front solutions obtained by the application of the multi-objective optimization tool for barley in different countries in Africa

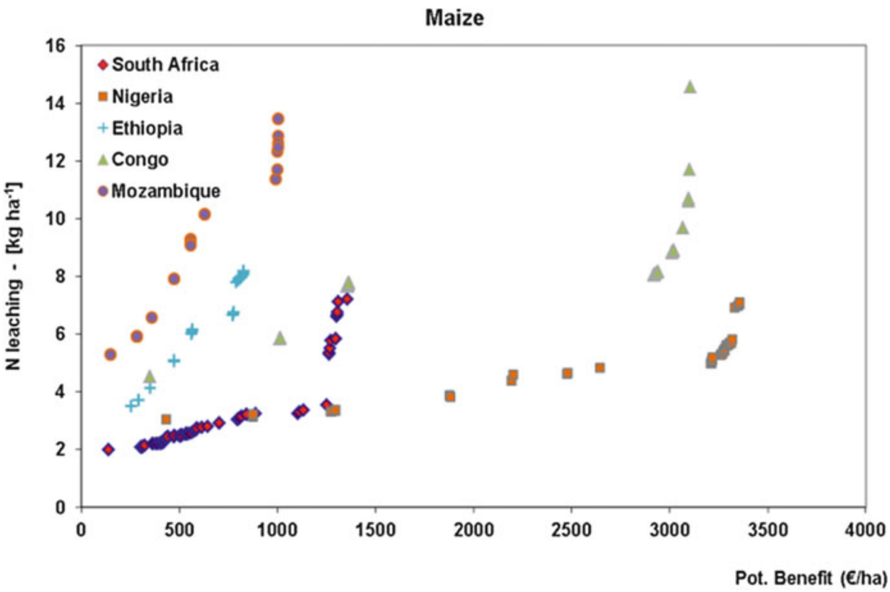


Fig. 11.3 Pareto-optimal front solutions obtained by the application of the multi-objective optimization tool for maize in different countries in Africa

the economic components and the environmental issues. A greater slope in the curve indicates that we can gain less net benefit for the same environmental impact.

In the case of maize (Fig. 11.3), we applied the tool in some central and southern countries. Nigeria shows a higher potential for benefit increase: this partially depends on the maize price that is quite high on the local market (380 US$ t in the period 2000–2009 according to FAO statistics), but it is also linked with the high water availability (no or very limited irrigation is necessary in this region) that limits the management costs required to increase the yield production and also reduces the risk for nitrogen leaching associated with high irrigation inputs.

In all the countries, it is possible to identify the benefit value beyond which the risk for water contamination increases very rapidly: in the case of South Africa, for example, above the value of benefit of 1,300 $\$\,\mathrm{ha}^{-1}$, the Pareto curve tends almost to be vertical. In the case of Mozambique, the nitrogen availability in the soil is quite high and part of the nitrogen fertilizer added to raise the production is partially lost in the water percolating into the soil resulting in a leaning Pareto front.

Barley is another important cereal crop for the Northern and Southern Africa. The optimization tool points out a similar trend for all the countries also varying the crop type. It is quite interesting to compare the results for the Northwest African countries: in the case of Morocco, the benefit is maximum for barley because the nitrogen leaching is generally low, also with high input of water and fertilizer. More generally, Morocco, Libya, and South Africa show a similar trend while Tunisia and Algeria point out a higher impact to the environment, an aspect that can be very useful in the case of an integrated management of crop production in these countries.

Another advantage of applying multi-criteria analysis is that it allows for performing the sensitivity analysis of the influence of certain parameters of the model in the final results almost automatically.

Figure 11.4 presents for Morocco and wheat the effect of variations in the economic model parameter. In this case, there is an increase and reduction of 30 % in the water and fertilizer costs. In the linear part of the Pareto frontiers, the slopes are quite similar in just all the cases, suggesting that changes in water and fertilizer costs lower than 30 % generally have no great impact on the net income which may be achieved without causing an increase in percolation of nitrates. However, an exponential increase is observed when the cost benefit made from the fertilizer does not exceed 1,550 $/ha allowing to achieve a reduction of 1,750 $\$\,\mathrm{ha}^{-1}$, both having a similar environmental impact. The case of the increased cost of water (+30 % WC) is the only one in which the slope of the linear region is slightly higher and it is apparently hard to find strategies that achieve net profit to be between 600 and 1,300 $\$\,\mathrm{ha}^{-1}$.

If we consider the case of wheat (Fig. 11.4) in Morocco, as an example, we may observe that the actual crop management stands in the lower left part of the graph that corresponds to the lower impact on the environment and also on the lower production and finally on the lower benefit.

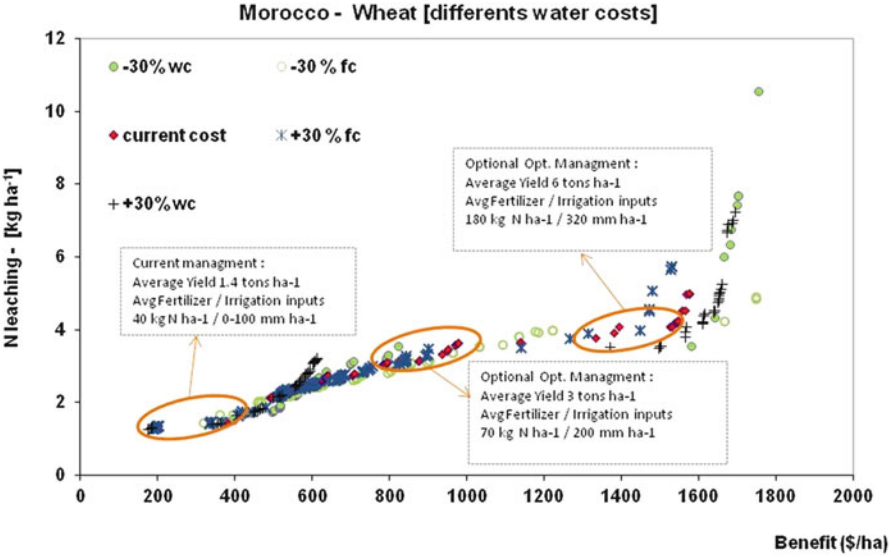


Fig. 11.4 Wheat in Morocco Pareto-optimal front solutions resulted by the application of the multi-objective optimization tool varying water and fertilizer cost

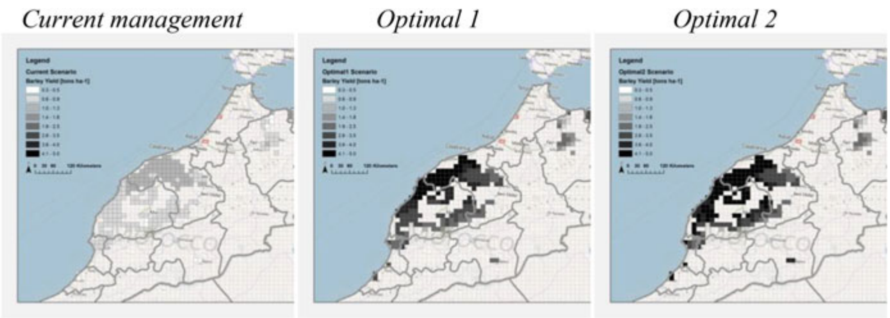


Fig. 11.5 Crop yield maps. Crop yield maps of GIS-EPIC Africa system set up with current and two optimal set of management configuration resulted from the application of MOO; the three maps correspond to the circled areas of Fig. 11.4

Figure 11.5 shows the spatial distribution of crop production for the current management compared with two of the optimal management solutions identified with MOO tool.

Wheat is one of the major crops in Morocco and counts for 40 % of the total agricultural harvested area according to FAO statistics (2009). Nevertheless, the final output is not yet enough to satisfy the full local demand over the past few years. As a result, the country has to import high quantities of wheat. The integrated optimization tool points out the possibility to move toward a more profitable yield production by preserving the environment at the same time: the second highlighted area in the graph show a set of optimal solutions that correspond to a very sensible increase of the yield reaching an average level of production in the country of 3 t ha^{-1} y^{-1}. These optimized and more productive solutions correspond to an increase of the fertilization from 40 to 70 kg ha^{-1} and to an increase of the irrigation amounts to 150–200 mm. The increase of nitrogen input according to the model is sustainable and it demonstrates real transferability to the farmers linked with the market price of fertilizer and crops. In the case of the new water management strategies, the analysis points out that the sustainability is more complex because it depends on the current water availability for the irrigation withdrawals and also on the balance with other sectors that could have alternative (and more profitable) water requirements.

11.5 Conclusions

For this study, we have coupled a multi-objective optimization model with a GIS crop management simulation model (EPIC) in order to identify best management practices for different crops and countries in Africa. The optimization tool is a multi-objective genetic algorithm that controls biophysical model variables related with the level of fertilizer and irrigation application for each crop and cell (15 km $\times$ 15 km). Depending on these parameter values and climatic and soil characteristics, the biophysical model estimates, among other things, the crop production, the fertilizer and water consumption, and the potential effects on the environment (nitrates percolation and runoff, erosion, etc.).

Thus, after a reasonable number of evaluations of the EPIC model outputs (less than 1,000), the optimizer finds the efficient trade-off strategies according to the simultaneous considered criteria. The integrated tool has been used taking up to five simultaneous objectives into consideration, for example, looking at how to maximize the production of a crop minimizing water consumption and fertilizer application as well as minimizing nitrate contamination. Moreover, the methodological approach presented in this study also includes a simplified economic model. It allows to select the best compromise solution by taking into account at the same time the gross margin potential for farmers and the environmental impact (in this case, nitrate losses as a simplified indicator of the environmental impact on soil and water) of the agricultural production.

We applied this methodology for most common crops in African countries, showing that actual management practices are in general inefficient solutions or correspond to the lowest values of potential production and benefits that farmers can use. This implies that the African agriculture farmers can improve their gross margin by maintaining low nitrate losses that are already limited under current management due to the low input (both for water and fertilizer). On the other hand, it has been shown that in some countries the nitrate leaching can increase considerably by moving to a more productive and profitable production, thus becoming a potential source of contamination for water resources.

The tool can be easily used on a more detailed scale if required by decision-makers to develop new Pareto-optimal fronts specifically defined for particular regions and/or conditions. The final proposed solutions provide a wide range of local management strategies to optimize nitrogen leaching and the farmers benefit. The final decision can be obviously more focused on one of the two objectives and can be taken by managers considering other socioeconomic aspects.

Finally, it is shown that the coupling of multi-objective programming models with crop simulation models and GIS spatial information is a powerful tool to address the agricultural–environmental issue. This methodology is based on objective quantitative data and gives valuable information to decision-makers.

Acknowledgements This research has been partially supported by the projects MTM2009-14039-C06-03 and MTM2012-36163-C06-06 from the Spanish Government, RIESGOS-CM from Comunidad de Madrid (Spain).

References

Ananda J, Herath G (2003) Incorporating stakeholder values into regional forest planning: a value function approach. Ecol Econ 45:75–90. doi:10.1016/S0921-8009(03)00004-1

Bates B, Kundzewicz ZW, Wu S, Arnell N, Burkett V et al (2008) Climate change and water. Technical paper of the intergovernmental panel on climate change. IPCC Secretariat, Geneva

Coverstone-Carroll V, Hartmann JW, Mason WJ (2000) Optimal multi-objective low-thrust spacecraft trajectories. Comput Methods Appl Mech Eng 186:387–402. doi:10.1016/S0045-7825(99)00393-X

Deb K (2001) Multi-objective optimization using evolutionary algorithms. Wiley, New York

Easterling W, Apps M (2005) Assessing the consequences of climate change for food and forest resources: a view from the IPCC. Clim Change 70:165–189. doi:10.1007/s10584-005-5941-0

FAO (2009) The special challenge for sub-Saharan Africa "How to feed the world 2050", Rome. http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Africa.pdf. Accessed Jan 2013

FAO (2011) The state of food and agriculture—women in agriculture, closing the gender gap for development. Office of Knowledge Exchange, Research and Extension, Rome. http://www.fao.org/docrep/013/i2050e/i2050e.pdf. Accessed Jan 2013

Galbiati L, Elorza FJ, Udías A, Bouraoui F (2007) Multiobjective optimization for river basin management plan. In: Candela L, Vadillo I, Aagaard P et al (eds) Water pollution in natural porous media at different scales. Assessment of fate, impact and indicators "WAPO2". Instituto Geológico y Minero de España, Madrid, pp 627–631

Gassman PW, Williams JR, Benson VW, et al (2005) Historical development and applications of the EPIC and APEX models. Working paper 05-WP 397, Center for Agricultural and Rural Development, Iowa State University, Ames. http://www.card.iastate.edu/publications/synopsis.aspx?id=763. Accessed 1 Nov 2006

Giles J (2005) Nitrogen study fertilizes fears of pollution. Nature 433:791. doi:10.1038/433791a

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Boston

Hogan R, Stiles S, Tacker P, Vories E, Bryant KJ (2007) Estimating irrigation costs. Little Rock: University of Arkansas Cooperative Extension Service, Fact Sheet No. FSA28

IPCC (2007) Climate change 2007: mitigation. Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

Johnson SL, Adams RM, Perry GM (1991) The on-farm costs of reducing groundwater pollution. Am J Agric Econ 73:1063–1073. doi:10.2307/1242434

Latinopoulos D (2009) Multicriteria decision-making for efficient water and land resources allocation in irrigated agriculture. Environ Dev Sustain 11:329–343. doi:10.1007/s10668-007-9115-2

Lee JJ, Phillips DL, Benson VW (1999) Soil erosion and climate change: assessing potential impacts and adaptation practices. J Soil Water Conserv 54:529–536

Liu J, Fritz S, van Wesenbeeck CFA, Fuchs M et al (2008) A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change. Global Planet Chang 64:222–235. doi:10.1016/j.gloplacha.2008.09.007

Liu J, Zehnder AJB, Yang H (2009) Global consumptive water use for crop production: the importance of green water and virtual water. Water Resour Res 45, W05428. doi:10.1029/2007WR006051

Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. Struct Multidiscip Optim 26:369–395. doi:10.1007/s00158-003-0368-6

Mearns LO, Mavromatis T, Tsvetsinskaya E et al (1999) Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. J Geophys Res 104:6623–6646

Meyer B, Lescot J-M, Laplana R (2009) Comparison of two spatial optimization techniques: a framework to solve multiobjective land use distribution problems. Environ Manage 43:264–281. doi:10.1007/s00267-008-9225-0

Mills D, Vlacic L, Lowe I (1996) Improving electricity planning—use of a multicriteria decision making model. Int Trans Oper Res 3:293–304. doi:10.1016/S0969-6016(96)00023-8

Morris M, Binswanger-Mkhize HP, Byerlee D (2009) Awakening Africa's sleeping giant: prospects for commercial agriculture in the Guinea Savannah zone and beyond. The World Bank, Washington, DC

Nalle DJ, Montgomery CA, Arthur JL et al (2004) Modeling joint production of wildlife and timber. J Environ Econ Manag 48:997–1017. doi:10.1016/j.jeem.2004.01.001

Pastori M, Bouraoui F, Aloe A, Bidoglio G (2011) GISEPIC AFRICA: a modeling tool for assessing impacts of nutrient and water use in African agriculture—JRC63230. Publications Office of the European Union, Luxembourg

Poloni C, Giurgevich A, Onesti L, Pediroda V (2000) Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics. Comput Methods Appl Mech Eng 186:403–420. doi:10.1016/S0045-7825(99)00394-1

Prato T (1999) Multiple attribute decision analysis for ecosystem management. Ecol Econ 30:207–222. doi:10.1016/S0921-8009(99)00002-6

RAC Resource Assessment Commission (1992) Multi-criteria analysis as a resource assessment tool (RAC Research Paper No. 6). Australian Government Publishing Service, Canberra

Rinaldi M (2001) Application of EPIC model for irrigation scheduling of sunflower in southern Italy. Agric Water Manag 49:185–196. doi:10.1016/S0378-3774(00)00148-7

Sadeghi SHR, Jalili K, Nikkami D (2009) Land use optimization in watershed scale. Land Use Policy 26:186–193. doi:10.1016/j.landusepol.2008.02.007

Schröter D, Cramer W, Leemans R et al (2005) Ecosystem service supply and vulnerability to global change in Europe. Science 310:1333–1337. doi:10.1126/science.1115233

Teague ML, Bernardo DJ, Mapp HP (1995) Farm-level economic analysis incorporating stochastic environmental risk assessment. Am J Agric Econ 77:8–19. doi:10.2307/1243884

Udías A, Galbiati L, Bouraoui F, Elorza FJ (2007) Mejora de la sostenibilidad de la actividad agrícola mediante un algoritmo evolutivo multiobjetivo, V Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB'07

Udías A, Galbiati L, Elorza FJ, Efremov R, Gómez A, Chiang G, Arrosa M, Lejarraga T (2009) Algoritmos genéticos para la selección de medidas de restauración de cuencas. MAEB—VI Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados:139–146

Udías A, Galbiati L, Elorza FJ, Efremov R, Pons J, Borras G (2011) Framework for multi-criteria decision management in watershed restoration. J Hydroinf 14(2):395–411. doi:10.2166/hydro.2011.107

Velde MVD, Bouraoui F, Aloe A (2009) Pan-European regional-scale modelling of water and N efficiencies of rapeseed cultivation for biodiesel production. Glob Chang Biol 15:24–37. doi:10.1111/j.1365-2486.2008.01706.x

Weintraub A, Romero C (2006) Operations research models and the management of agricultural and forestry resources: a review and comparison. Interfaces 36:446–457. doi:10.1287/inte.1060.0222

Whittaker G, Confesor R Jr, Griffith SM et al (2009) A hybrid genetic algorithm for multiobjective problems with activity analysis-based local search. Eur J Oper Res 193:195–203. doi:10.1016/j.ejor.2007.10.050

Williams JR, Jones CA, Dyke PT (1984) The EPIC model and its applications. In: Proceedings of International Symposium for Agrotechnology Transfer, ICRISAT, Patancheru

Wriedt G, Van der Velde M, Aloe A, Bouraoui F (2009) Estimating irrigation water requirements in Europe. J Hydrol 373:527–544. doi:10.1016/j.jhydrol.2009.05.018

Zavattaro L, Monaco S, Sacco D, Grignani C (2012) Options to reduce N loss from maize in intensive cropping systems in Northern Italy. Agr Ecosyst Environ 147:24–35. doi:10.1016/j.agee.2011.05.020

Zekri S, Casimiro Herruzo A (1994) Complementary instruments to EEC nitrogen policy in non-sensitive areas: a case study in Southern Spain. Agr Syst 46:245–255. doi:10.1016/0308-521X(94)90001-V

Chapter 12
Modelling of Catastrophic Farm Risks Using Sparse Data

V.A. Ogurtsov, M.A.P.M. van Asseldonk, and R.B.M. Huirne

12.1 Introduction

Farmers often face risky events in agriculture. Risk means the possibility of a loss of income or property resulting from some event (Pritchet et al. 1996). Catastrophic risks are infrequent events but can cause large losses to farmers. In deriving the optimal farm plan, alternative options should be addressed to cope with catastrophe events. Those options are on-farm risk management strategies (e.g. diversification) as well as risk transfer strategies (e.g. insurance) (Ogurtsov et al. 2009). Thus, for a proper risk assessment of a catastrophe event, its probability and magnitude need to be taken into account. The assessment should ideally be based on a long-term and reliable farm-level history. But, in practice, farm-level data is often very sparse to provide a good and reliable basis for such a risk assessment (Hardaker et al. 2004), and this is certainly the case when focusing on catastrophe events. The reliability can be enhanced by eliciting subjective probability judgements, in addition to the available data (Hardaker and Lien 2005). Furthermore, it is advised to smooth the sparse data (i.e. interpolating between observations and extrapolating outside observations) by fitting a parametric or empirical distribution (Shlaifer 1959; Anderson et al. 1997, pp. 42–44). However, by smoothing the data in such a way, the risk analyst might face the problem of overrepresenting the middle part of the distribution and underestimating one or both tails. Catastrophes cause a serious downside risk, and therefore it is important to analyse the tail of the distribution very carefully by investigating alternative tail estimations. Before the smoothing

V.A. Ogurtsov
Business Economics, Wageningen University, Wageningen, The Netherlands

M.A.P.M. van Asseldonk (✉) • R.B.M. Huirne
Agricultural Economic Research Institute, Wageningen University, Wageningen, The Netherlands

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_12

procedure, a realistic assumption should be made about the upper and lower bounds, ensuring that the distribution will be a reasonable approach to include the downside and upside tails.

One of the ways to smooth data and to include the downside and upside tails is to fit the sparse data to a (parametric) normal distribution. There is a continuous discussion about the applicability of normal distribution assumptions of yields in agriculture (i.e. Antwood et al. 2003; Galagher 1987; Just and Weninger 1999; Ramirez et al. 2003; Swinton and King 1991). The problem arises because it is difficult to reject normality assumptions, especially when data is sparse. For farm outcomes, it is generally hard to obtain 10 relevant observations under the same economic policy, management regime, farm programme or trade policy (Just and Weninger 1999; Richardson 2006). At least 20 or more observations are usually required to test with any accuracy whether a distribution is normally distributed or not (Richardson 2006). Non-normality might therefore be masqueraded as normality, simply because of the misspecification of the test (Just and Weninger 1999).

Normality is not likely because the upward potential of yields is biologically bounded and there is a risk of (complete) crop failure because of, for example, adverse meteorological circumstances (Galagher 1987). Many studies stated that crop yields are skewed and do not follow normality (Just and Weninger 1999; Galagher 1987; Antwood et al. 2003; Swinton and King 1991; Ramirez et al. 2003). However, Just and Weninger (1999) argued that many studies that rejected normality are typically cited as the basis for making non-normality assumptions but are no better individually justified than normality.

Alternatively, to a parametric normal distribution, the technique of kernel density estimation (KDE) can be used to generate unobserved data to supplement sparse data. The KDE procedure is a non-parametric approach of smoothing data by hand. Instead of minimising the sum of squared residuals, the KDE method weights observations on relative proximity to estimate the probability. The estimation of the probability at a given point depends on a preselected probability density. In that way the kernel is analogous to the principle of local averaging, by smoothing, using evaluations of the function at neighbouring observations (Yatchew 1998). Therefore, the probabilities in the tails depend largely on the choice of kernel. The kernel density smoothing procedure is popular in many fields, but it is not widely used in agriculture (Richardson et al. 2006), and only a limited number of agricultural studies were conducted (i.e. Hardaker et al. 2006; Richardson et al. 2000, 2006). In the early work of Richardson et al. (2000), analysis of simulated statistics showed that KDE gives acceptable results for simulating sparse data. Hardaker et al. (2006) suggested that in case of sparse historical data, additional information and judgements need to be incorporated about the tails of the distribution when applying the KDE approach to improve the confidence of the results. Richardson et al. (2006) found that KDE provided better results than parametric distributions and a linear interpolation of the empirical distribution.

In complex systems with more than one activity, like farming, the stochastic dependency needs to be accounted for (Hardaker et al. 2004). Ignoring stochastic dependency between risky prospects in farm planning can be seriously misleading

(Richardson et al. 2000). For example, crop yields tend to be positively correlated in that a good year for one crop also often suits other crops and vice versa. Similarly, prices for several kinds of farm products tend to move together, depending on the general economic conditions (Hardaker and Lien 2005). Therefore, the univariate normality versus the univariate KDE debate needs to be upscaled to multivariate normality (MVN) versus multivariate kernel density estimation (MVKDE).

This paper compares the alternative ways of conducting a farm risk analysis using sparse data with special reference to catastrophe events. For this purpose MVKDE and MVN procedures are applied to simulate the joint distributions of crop yields and prices. Six case farms were chosen to reflect the conditions of typical Dutch arable farms. The risk analysis focuses on the impact of the functional form chosen to generate the joint distribution on the density in the downside tail. Subsequently, the result of incorporating the downside tail alternatively on the optimised net farm income and farm plan is addressed by applying utility-efficient programming (UEP).

12.2 Methods to Characterise Catastrophe Events in Farm Planning Models

In this section the methods of MVN and MVKDE procedures are described, which can be used to generate the required probability distributions that then can be incorporated into UEP models for obtaining the optimal farm plans.

12.2.1 Simulation Procedure for the Multivariate Normal Distribution

An MVN distribution for some random variables (crop yields and prices) is specified by three components: a (deterministic) component capturing the mean (i.e. the expected value of the observations), a (stochastic) component based on the variance and a multivariate component based on the covariances of the observations. The steps for constructing an MVN distribution are the following (Richardson 2006):

1. Calculate the best possible model to predict each variable, whether this is simply the arithmetic mean or based on a trend regression, a multiple regression or a time-series model.
2. Calculate the residuals based on the prediction for each random variable.
3. Calculate variances for each variable using their residuals.
4. Calculate covariances using their residuals.

12.2.2 Simulation Procedure for the MVKDE

Besides the MVN procedure, also the MVKDE simulation procedure will be applied to simulate the joint distribution of random variables (crop yields and prices) alternatively. The procedure in specifying MVKDE distribution consists of the following steps (Richardson et al. 2006):

1. From the matrix of observations (yields and prices, usually historical de-trended), the covariance matrix R_{kxk} is estimated and then factored by Cholesky decomposition so that $P = RR^T$, where P is identity matrix, k is a set of variables (yields and prices) and T is used to transpose matrix R into R^T.
2. The minimum $X_{\mathrm{Min},j}$ and maximum $X_{\mathrm{Max},j}$ bounds for each variable k are then determined. The cumulative probabilities for these values are assumed $F(X_{\mathrm{Min},j}) = 0$ and $F(X_{\mathrm{Max},j}) = 1$, where $j = 1, \ldots, k$ is one of the k variables.
3. For each variable k, a new vector of X^A_{sj}, of dimension S ($s = 2, \ldots, S$) is created with a given minimum $X^A_{sj} = X_{\mathrm{Min},\,j}$ (i.e. $s = 1$ for the minimum observation) and maximum $X^A_{sj} = X_{\mathrm{Max},\,j}$ by the formula:

$$X^A_{sj} = \left(\frac{1}{S-1}\right)\left(X_{\mathrm{Max},\,j} - X_{\mathrm{Min},\,j}\right) + X^A_{(S-1)j} \tag{12.1}$$

4. The smoothed percentiles for each X^A_{sj} between the extreme points $F(X_{\mathrm{Min},j}) = 0$ and $F(X_{\mathrm{Max},j}) = 1$ are calculated based on KDE (Silverman 1986; Scott 1992). For each variable j, the smoothed percentile is evaluated at a given point X^A_{sj} as

$$\hat{F}\left(X^A_{sj}\right) = \frac{1}{nh_j}\sum_{i=1}^{n} K\left[\frac{\left(X^A_{sj} - X_{ij}\right)}{h_j}\right] \tag{12.2}$$

where $K(\cdot)$ is the cumulative kernel function associated with a symmetric continuous kernel density $k(\cdot)$ such that $K(x) = \int_{-\infty}^{x} k(t)dt$, and h_j is the bandwidth of the variable j.

With a specific kernel function, the value of bandwidth, called a smoothing parameter, determines the degree of averaging in the estimate of the density function. Bandwidth is also called the standard deviation of the kernel density function. It is important to choose the most appropriate bandwidth because a value that is too small leads to under-smoothed data, or if too large to over-smoothed data. When a bandwidth decreases towards zero, the number of modes increases and the KDE is very noisy. As bandwidth increases to infinity, the number of modes drops to one, so that the KDE displays a unimodal pattern.

The best criterion to select a kernel is the smallest root mean square (RMSE) of residuals between the historical kernel cumulative probabilities and probabilities for the kernel function.

Formula (12.2) can be used for a univariate KDE. If the interest is in a multivariate distribution, covariances of the underlying random variables have to be taken into account. In this way the MVKDE procedure can be used to incorporate the stochastic dependency (Richardson et al. 2006). Then, the simulation of MVKDE would take the following steps:

1. Generate correlated uniform standard deviates (CUSDs) from the observed random variables; the result will be a value between 0 and 1.
2. Given the $CUSD_j$, along with respective vectors X_{sj}^A and smoothed percentiles $\hat{F}\left(X_{sj}^A\right)$ with a scale (including $F(X_{\mathrm{Min},j})=0$ and $F(X_{\mathrm{Max},j})=1$) between the nearest lower $\hat{F}_L\left(X_{Lj}^A\right)$ and nearest upper $\hat{F}_L\left(X_{Uj}^A\right)$ percentiles, interpolate among the X_{sj}^A random vector of $\widetilde{X}_j$ is generated.

The final formula of the generated MVKDE vector is the following:

$$\widetilde{X}_j = X_{Lj}^A + \left(X_{Uj}^A - X_{Lj}^A\right) * \frac{\left(CUSD_j - \hat{F}_L\left(X_{Lj}^A\right)\right)}{\left(\hat{F}_U\left(X_{Uj}^A\right) - \hat{F}_L\left(X_{Lj}^A\right)\right)} \tag{12.3}$$

Goodness-of-fit tests can be conducted whether the simulated joint MVN and MVKDE distributions of yields and prices are appropriate.

12.2.3 Utility-Efficient Programming (UEP)

UEP is a mathematical programming method and can be used for optimising farm plans (Van Asseldonk et al. 2005, Lien et al. 2009, 2011). In UEP the expected utility of the farm plan is maximised (Ogurtsov et al. 2008). UEP is a non-parametric method, which implies that it is free of distribution assumptions and includes the joint distribution by means of the so-called states of nature (i.e. specific combinations and probabilities of possible outcomes). UEP takes the following form (Hardaker et al. 2004)

$$\text{Maximise } E[U] = p\,U(z,\, r), \quad r \text{ is varied} \tag{12.4}$$

and is subject to

$$Ax \leq b \tag{12.5}$$

$$Cx - Iz = U(z, r) \tag{12.6}$$

$$\text{And} \quad x \geq 0 \tag{12.7}$$

where

$E[U]$ is expected utility.
z is a vector of farm incomes by state of nature.
r is a coefficient of risk aversion.
p is a probability of each state of nature.
$U(z, r)$ is a vector of utilities of farm incomes by state of nature with risk aversion level r.
A is a vector of technical-economic coefficients per each activity.
x is a vector of activities.
b is a vector of available resources (constraints).
C is a vector of the state of nature matrix of activity incomes.
I is an identity matrix.

In most cases, r represents a coefficient of absolute risk aversion. As long as the risk aversion coefficient of a farmer is not known, a range of risk aversion coefficients can be considered for modelling. Hardaker et al. (2004) developed a method called SERF, where alternative farm plans can be provided in terms of certainty equivalents as a measure of risk aversion over a definite range, developed by Anderson and Dillon (1992). For a risk-averse farmer, the coefficient of relative risk aversion of wealth $r_r(W)$[1] varies from 0.5 to 4, typically about 1, with the following interpretation: 0.5, hardly risk averse at all; 1.0, somewhat risk averse (normal); 2.0, rather risk averse; 3.0, very risk averse; and 4.0, almost paranoid about risk.

12.2.4 Available Sparse Data and Optimisation Constraints

For the current analysis, six Dutch arable farms were selected from the Farm Accountancy Data Network (FADN) database. The FADN data is an official European Union dataset, which includes detailed farm-specific data of, among other things, yields per unit per crop. A prerequisite for the selection of the arable farms was that at least 10 consecutive years with observations was available for a farm to be selected. The corresponding number of states of nature ranged from 11 to 13 for the farms under study (Table 12.1). The main crops in the production plan constituted consumption potato, wheat, rye and sugar beet.

[1] Absolute risk aversion coefficient is usually calculated as a proportion of the relative risk aversion coefficient to wealth.

Table 12.1 Summary of characteristics of the farms selected

Farm number	Number of observations	Cultivated area (ha)	Main activities in production plan
I	13	17	Potato, wheat, rye, sugar beet
II	11	80	Potato, wheat, sugar beet
III	11	101	Wheat, rye, sugar beet
IV	11	37	Potato, wheat, sugar beet
V	11	205	Wheat, rye, sugar beet
VI	11	155	Potato, wheat, sugar beet

The farm-specific yields observed in the states of nature were de-trended by a linear function (formula 12.8):

$$y_{qit} = \alpha_{qi} + \beta_{qi1} t + \varepsilon, \quad \varepsilon \sim N\left(0, \ \sigma^2\right) \tag{12.8}$$

where y_{qit} is the yield unit of activity q on farm i in year t ($t = 1, \ldots, T$); α_{qi} is the regression constant for activity q on farm i; β_{qi} is the systematic change per activity q on farm i (it is assumed that the trend caused by technological change among other things will continue in the future); and ε is a normally distributed random error term (Murdoch 1966, p. 34).

Farm gate prices and costs of production were assumed to be identical for all farms considered. The average annual crop prices were de-trended by the Paasche equation with the consumer price index (CPI) as deflator (Mas-Colell et al. 1995, p. 37):

$$I(p)_{qt} = \frac{p_{qt}}{p_{qy}} \tag{12.9}$$

where $I(p)_{qt}$ is a deflator price of activity q in year t ($t = 1, \ldots, T$), p_{qt} is the volume of price of activity q in year t and p_{qy} is the fixed volume of price of activity q in basic year y.

Crop-specific production costs were obtained from norms (see Dekkers 2002) and were equivalent to prices deflated. Following the usual crop-rotation rules, cereal crops (e.g. wheat and rye) were restricted to a maximum of two-thirds of the cultivated area. Tuberous crops (consumption potato and sugar beet) were restricted to a maximum of one-third of the cultivated area. Each crop was also restricted to the maximum observed area in its past (i.e. 11–13 years). Moreover, for sugar beet, the maximum quota limitations were accounted for.

12.2.5 Expanding the States of Nature Matrix for MVN and MVKDE to Account for Catastrophe Events

MVN and MVKDE approaches can be applied to generate a more enhanced sample than the observed sparse data as explained before. By doing so, it will make them more relevant and reliable to the uncertainty to be faced in the future farm planning period to date, having been accounted for, among other things, catastrophe events. The densities in the downside tails are predefined when applying the MVN approach and root from the specified means, variances and covariances. The MVN distribution can be truncated to prevent the anomalies occurring (e.g. negative yields and prices). Given the MVKDE procedure, subjective maximums and minimums need be added prior to the sampling.

Catastrophe events in arable farming, resulting into high losses, stem from numerous risks (i.e. perils), for example, weather-related perils as hail, storm and drought. However, the different catastrophic risks are generated simultaneously, since the applied MVN and MVKDE approaches do not discriminate if a downside outcome originates from one peril or another (no separate distributions are generated for different perils). Note that catastrophe events correspond to extreme unfavourable outcomes, not necessarily the minimum value that is specified for each KDE. For instance, a 50 % reduction of the expected level is often regarded as a catastrophe event.

12.2.6 Computations

We used the Simetar software to compare the MVN procedure with the MVKDE procedures (Richardson 2006). The following kernel density functions were applied: Cauchy, cosine, double exponential, Epanechnikov, Gaussian, Parzen, quartic, triangle, triweight and uniform (see Richardson 2006). On the basis of the available historical yields, prices and corresponding covariance matrix, the MVN distribution and each MVKDE alternative were parameterised, and subsequently 500 states of nature (of yields and prices) were derived by the Latin hypercube (LH) sampling procedure. In this way, the impact of the functional form on the joint distribution and the density in the downside tail could be studied. The LH procedure was taken in favour of Monte Carlo simulation (MCS), because it divides the distribution in an equal number of intervals so that tails with a downside risk and upside potential are taken into account (Richardson 2006). On the contrary, MCS randomly selects points, so that the tails can be underestimated even with a higher number of replications. The minimum values, for both MVN distribution and MVKDE, equalled zero. The change of the maximum affects the shape of the distribution, and the maximum values imposed arbitrarily were calculated as the observed (from the limited sparse data) maximum value plus one standard deviation.[2]

[2] Different assumptions in defining the maximum value were considered: ‘maximum plus one standard deviation’, ‘maximum plus two standard deviations’ and ‘maximum plus three standard deviations’. The choice was made in favour of the ‘maximum plus one standard deviation’ because it accommodates a more dense tail.

Subsequently, the impact of incorporating the downside tail alternatively when optimising net farm income was addressed by applying UEP. Hereto, the 500 generated samples per alternative were incorporated as states of nature in UEP. Detailed results are presented for farm II whereas only the aggregated results for the other five farms.

12.3 Results

12.3.1 Probability Distributions of Random Variables

12.3.1.1 Graphical Representation

The kernel functions under study were parameterised with the available states of nature, as discussed before. The appropriate approach is to select subsequently the kernel function with the smallest RMSE between the kernel itself and the historical observations (derived from the available states of nature complemented with the specified bandwidth). It was observed, however, that the density in the downside tail was underestimated for the majority of the kernels. The only kernel function that encompassed a denser downside tail, inherent to catastrophic risks and imposed by an extremely lower bound, was the Cauchy kernel. The remainder kernels definitely overestimated the middle section of the distribution and were equivalent to each other with respect to the downside tail. The double exponential and the Parzen kernel functions are typical representatives of kernels that overestimate the middle part and underestimate the downside tail. The remainder of this paper focuses therefore on the normal distribution as well as the Cauchy, the double exponential and the Parzen kernel functions.

For only farm II, we elaborate on the generated cumulative distribution functions (CDFs) and the corresponding test characteristics. Then, the general results for all farms will be presented. In Fig. 12.1, the CDFs of yields and prices for consumption potato, wheat and sugar beet are shown. For both yields and prices, it can be seen that the Cauchy kernel matched the downside tail better (e.g. entire crop failure).

Since the Cauchy kernel captured the downside tail best, the crop yield distributions simulated by the Cauchy kernel for all six farms were compared in Fig. 12.2. As presented before, identical prices were assumed for all the farms and are therefore not presented.

As can be seen, the Cauchy kernels of the several farms had a similar pattern, but there were significant differences between the yield levels of the farms. The probability of an entire potato failure was almost 5 % for farms I, IV and VI, while for farm II the most extreme event was a potato yield of 5 t per hectare with a probability of 2 %. Note that the observed crop plans of farms III and V did not comprise potatoes.

In general, more extreme unfavourable wheat and sugar beet yields were generated for farm II than for the other five farms. For example, given farm II, the probability of an entire wheat or sugar beet failure was approximately 3 % and 5 % respectively.

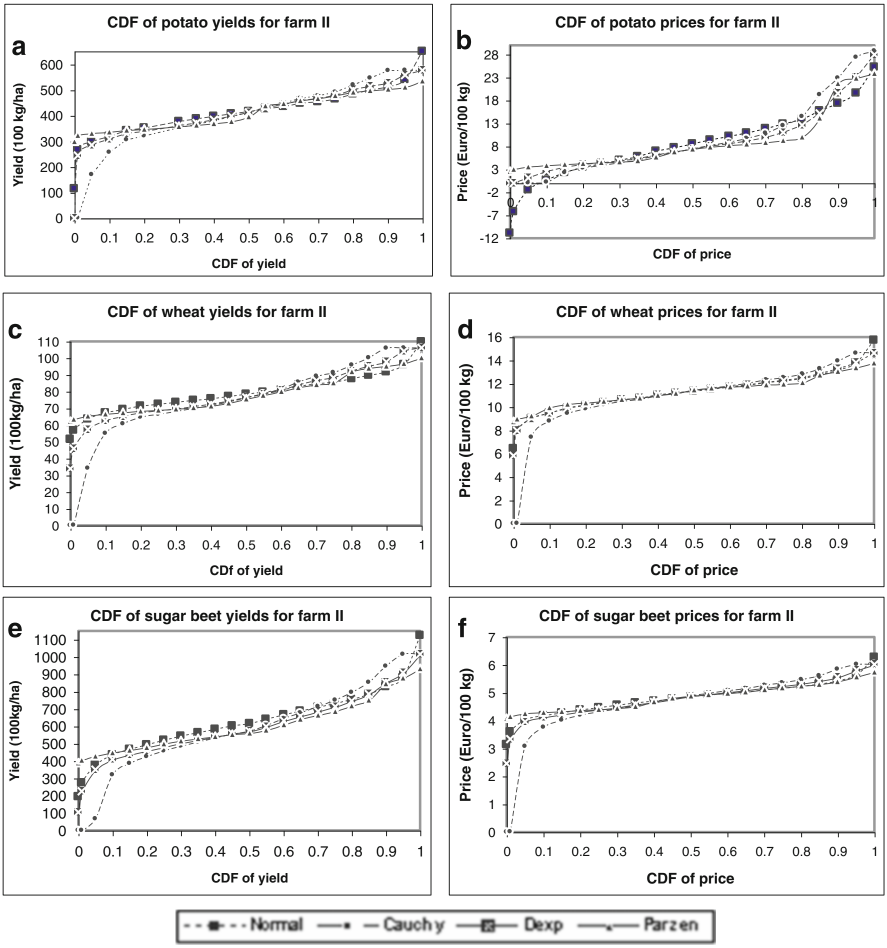


Fig. 12.1 Cumulative distributions of yields and prices for farm II

12.3.1.2 Test Statistics

Several statistical tests were performed to validate whether the structure of the simulated data adequately captured the structure present in the available sparse dataset. In Table 12.2, test values and critical values for normality tests, two-sample Hotelling T^2, Box's M test and complete homogeneity test, were summarised at the 95 % confidence level. If the test value does not exceed its critical value, then the null hypothesis is not rejected for the test under consideration. The critical values for farms II–VI were identical and are shown in the last column (equal number of degrees of freedom given three activities). The preceding column depicts the critical values for farm I (number of degrees of freedom given four activities).

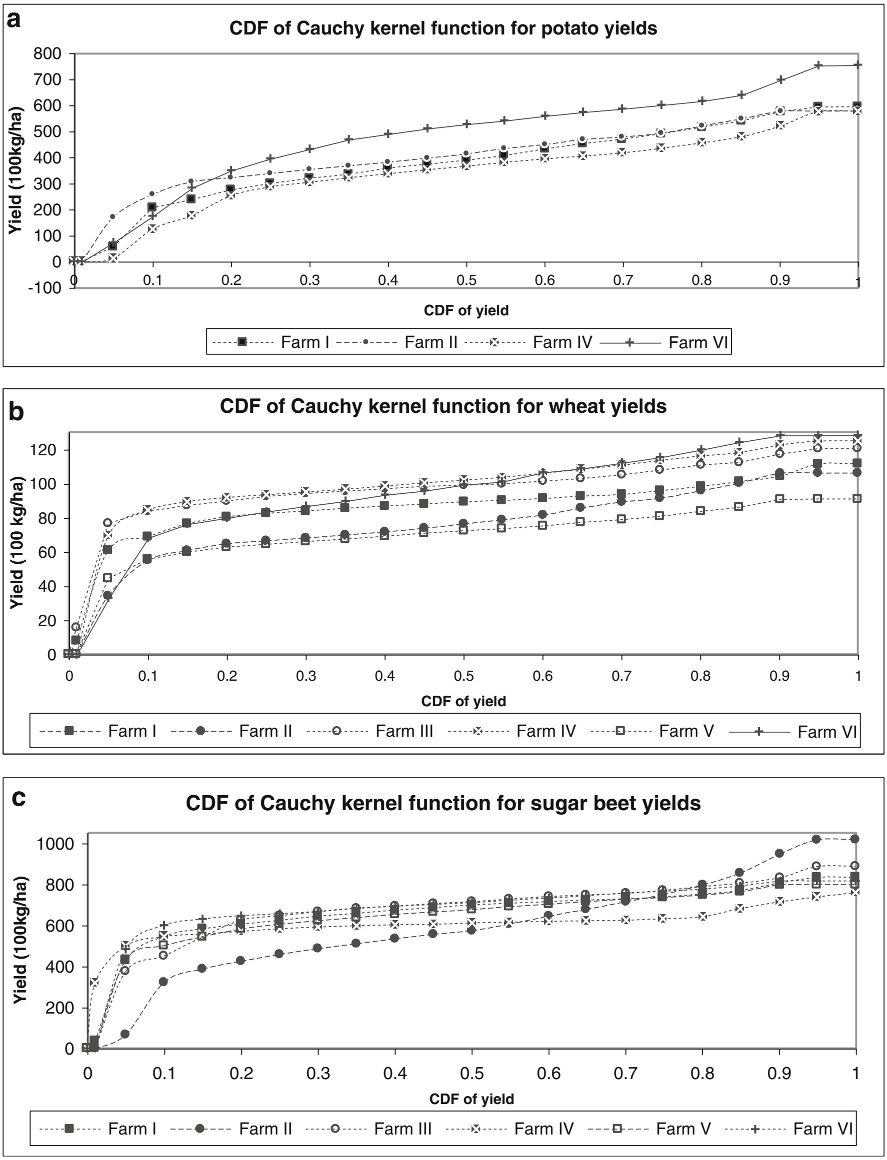


Fig. 12.2 CDFs of Cauchy kernel function yield distributions

The skewness and kurtosis criterion of the MVN distribution showed that the hypothesis that the data are multivariate and normally distributed was not rejected (Table 12.2). However, this finding can illustrate that the model with a limited number of states of nature can be misspecified as in the study by Just and Weninger (1999).

Table 12.2 Tests statistics of different distribution assumptions

	Farm I	Farm II	Farm III	Farm IV	Farm V	Farm VI	Critical values for farms	
Distributions	Test value						I	II, III, IV, V and VI
Normal	Skewness criterion							
	93.75	48.4	81.2	57.49	63.73	49.11	146.57	74.47
	Kurtosis criterion							
	1.02	0.06	0.96	0.06	−0.98	−1.02	1.96	1.96
	Two-sample Hotelling T^2 test							
Normal	2.40E−05	0.277	2.50E−05	3.20E−05	1.70E−05	1.40E−05	15.87	12.83
Cauchy	1.623	0.534	0.588	0.545	0.829	0.658	15.87	12.83
Double exp.	0.022	0.01	0.005	0.005	0.023	0.02	15.87	12.83
Parzen	0.043	0.003	0.006	0.018	0.03	0.03	15.87	12.83
	Box's M test							
Normal	31.65	25.24	21.3	22.9	32.96	23.23	51	32.67
Cauchy	61.9	47.36	41.07	48.23	60.5	43.8	51	32.67
Double exp.	16.61	27.79	23.43	24.92	35.85	26.23	51	32.67
Parzen	11.91	24.74	22.07	22.7	33.59	33.59	51	32.67
	Complete homogeneity test							
Normal	45.14	37.7	33	33.51	48.25	33.96	60.48	40.11
Cauchy	96.6	73.15	67.13	74.38	92.48	67.01	60.48	40.11
Double exp.	27.01	42.51	37.43	37.9	53.94	40.06	60.48	40.11
Parzen	17.88	35.75	33.42	32.84	48.7	48.7	60.48	40.11

The two-sample Hotelling T^2 test was applied to test whether the mean vectors of the simulated data and available sparse data were different. The hypothesis that the mean vectors are equal was not rejected for all four distributions for each farm (at 95 % confidence level).

The Box's *M* test was used to test whether the covariance matrices were equivalent. The simulated and the historical covariance matrices were not statistically different at the 95 % confidence level for the multivariate normal distribution for almost all farms (except in the case of farm V, where the test value of 32.96 was slightly higher than the critical value of 32.67). The hypothesis that the covariance matrices obtained from the Cauchy kernels are equal to the historical covariance matrices was persistently rejected. For the double exponential kernel, the hypothesis of maintaining the covariance structure was accepted for 5 farms out of 6 (except farm V), while the Parzen kernel was appropriate four times (farms I up to IV).

To test simultaneously whether both simulated mean vectors and covariance matrices were equal to the historical ones, the complete homogeneity test was used. The test failed to reject (at 95 % confidence level) that both simulated mean vectors and covariance matrices are statistically equivalent to the historical ones for the normal distribution (except farm V). Maintaining of the mean and covariance structure simulated by means of Cauchy kernels was always rejected. The results from the double exponential and Parzen kernels were rather mixed.

The test results differ from the study by Richardson et al. (2006), where the hypothesis of the appropriate covariate structure between sparse and simulated data was preserved. This might be explained by the fact that in their state of nature matrix very low yields were observed, close to our extreme subjective minimums, whereas in this study the observed states of nature did not represent observations in the downside tail.

12.3.2 Impact of Input Distributions on Optimal Farm Plan

The optimal farm plans resulting in the maximal expected utility were obtained in GAMS on the basis of a negative exponential utility function. The absolute risk aversion coefficients (*Ra*) were calculated as the proportion of the relative risk aversion (*Rr*) coefficients (on a scale from 0.5 to 4) to the permanent income (for details see Hardaker et al. 2004). The permanent income was obtained for each farm with a separate linear programming model. Then, for each level of risk aversion, the optimal farm plan with corresponding certainty equivalents (*CEs*), expected monetary values (*EMV*) of net farm income and risk premiums (*RP*) were calculated.[3]

Table 12.3 presents the results obtained from UEP for farm II on the basis of MVN distribution and MVKDE (Cauchy, double exponential and Parzen kernels) of inputs. In general, it can be seen that if a farmer was more risk averse, he was

[3] The risk premium is defined as the difference between EMV and CE and is expressed as a percentage; it is calculated as RP % = Risk premium/EMV.

Table 12.3 UEP results for farm II

	Ra	*Rr*	EMV, euro	*CE*, euro	Risk premium (*RP*), euro	*RP* (%)	Activities		
							Potato	Wheat	Sugar beet
Normality									
Ra1 = *Ra* min	5E−06	≈ 0.5	94,941	81,122	13,819	14.6	26.4	37.6	16
Ra2	1E−05	≈ 1	83,656	70,826	12,830	15.3	17.8	46.2	16
Ra3	2E−05	≈ 2	81,356	61,047	20,309	25	16	48	16
Ra4	3E−05	≈ 3	81,356	52,406	28,950	35.6	16	48	16
Ra5 = *Ra* max	4E−05	≈ 4	81,356	44,753	36,604	45	16	48	16
Cauchy									
Ra1 = *Ra* min	5E−06	≈ 0.5	94,422	74,662	19,760	20.9	26.4	37.6	16
Ra2	1E−05	≈ 1	82,443	62,262	20,181	24.5	18.7	45.3	16
Ra3	2E−05	≈ 2	78,243	50,519	27,724	35.4	16	48	16
Ra4	3E−05	≈ 3	78,243	41,702	36,541	46.7	16	48	16
Ra5 = *Ra* max	4E−05	≈ 4	78,243	34,617	43,625	55.8	16	48	16
Double exponential									
Ra1 = *Ra* min	5E−06	≈ 0.5	93,886	81,586	12,300	13.1	26.4	37.6	16
Ra2	1E−05	≈ 1	93,458	72,202	21,257	22.7	26.1	37.9	16
Ra3	2E−05	≈ 2	79,884	63,047	16,838	21.1	16	48	16
Ra4	3E−05	≈ 3	79,884	57,352	22,532	28.2	16	48	16
Ra5 = *Ra* max	4E−05	≈ 4	79,884	52,669	27,216	34.1	16	48	16
Parzen									
Ra1 = *Ra* min	5E−06	≈ 0.5	93,656	84,150	9,505	10.1	26.4	37.6	16
Ra2	1E−05	≈ 1	93,656	77,292	16,364	17.5	26.4	37.6	16
Ra3	2E−05	≈ 2	89,239	68,300	20,939	23.5	23.1	40.9	16
Ra4	3E−05	≈ 3	79,897	63,781	16,116	20.2	16	48	16
Ra5 = *Ra* max	4E−05	≈ 4	79,897	60,779	19,118	23.9	16	48	16

more prone to choose a production plan comprising more less profitable lower-variance crops (wheat instead of potato) compared to the optimal plan achieved with *Ra1* (implying that the decision-maker is almost risk neutral). The changes in the production plan correspondingly resulted into changes in the net farm income. With an increase of risk aversion, the farmer was willing to pay a higher risk premium.

The impacts of alternatively specified input distributions on the optimal farm plan (i.e. level of activities) were mixed. The allotted acreage in the farm plan of sugar beet, which was the most profitable cropping activity, always corresponded to the maximum quota allowed. The changes in production plans between potato and

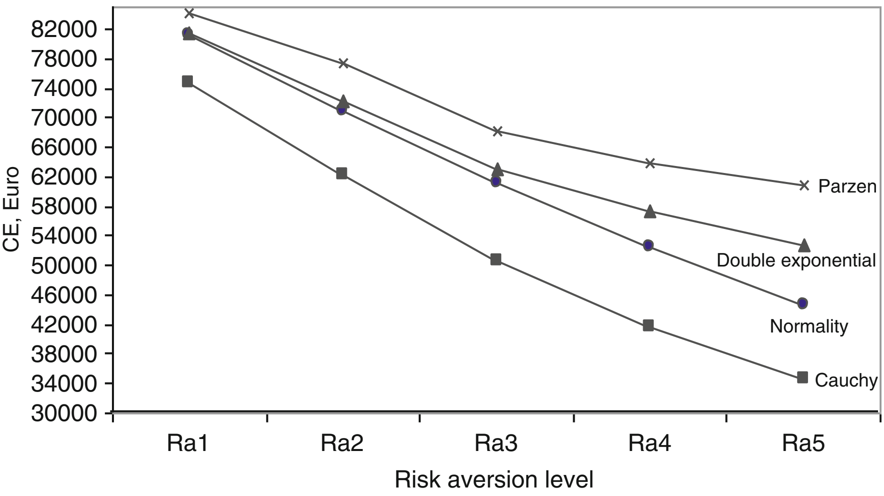


Fig. 12.3 CEs for farm II

sugar beet between the distribution alternatives were considerable for a 'somewhat risk-averse' (*Ra2*) and 'rather risk-averse' (*Ra3*) farmer. For a 'very risk-averse' (*Ra4*) and 'almost paranoid about risk' (*Ra5*) farmer, the production plan did not alter despite the differences in input distributions. The net farm incomes (*EMV*) were not much different between the models based on a normal distribution, Cauchy, double exponential and Parzen kernels. For farm II, with Cauchy kernel distribution, which better incorporates the lower tail, the net farm income was lower than for other distribution assumptions.

Substantial changes in the size of *CEs* were observed (Fig. 12.3). As in theory, the *CEs* are decreasing as a cost of paying for increasing risk aversion (Hardaker et al. 2004). The decrease of *CEs* was steeper for Cauchy kernel, which better incorporates the downside tail.

The conclusions drawn from farm II were also valid for the other farms under study. The risk premiums increased if the level of risk aversion increased. It corresponded to the decrease in *CEs*, due to worse optimal plans and increased levels of risk aversion.

12.4 Conclusions and Discussion

Initially, the sample of historical data comprising 11–13 observations of annual returns for an individual farm situation, which is already difficult to obtain, was not appropriate to analyse the impact of catastrophe events. However, the available sparse data was then used to generate data by applying MVN and MVKDE procedures to incorporate the downside tail. The analysis showed that the functional form chosen to generate the joint distribution substantially impacted the density in tail, although they were parameterised with the same observations. The differences in the optimal farm plan obtained (i.e. activity levels) generated by UEP were less profound.

To specify kernel density functions, usually expert opinions are elicited to define subjectively the minimum and the maximum values. If, on the basis of these subjective judgements, it is believed that catastrophe losses do occur (such as an entire crop failure), one might be inclined to specify the lower bound accordingly (equal or close to zero). It was observed that the normal distribution and all kernels, except the Cauchy kernel function, underestimated the impact of these beliefs, thereby neglecting the downside tail of the distribution. Note that the upper bound was arbitrary, augmented to the value of the mean plus one standard deviation. Limiting the upside potential will definitely have its impact of the density over the whole distribution, thus also the downside tail.

The statistical tests showed that the simulated mean vectors from the Cauchy kernel were not statistically different from the mean vectors of the sparse data. Furthermore, the covariance structure was statistically different. However, it was not logical to expect that on the basis of the available sparse data, in which catastrophe states of nature were absent, the covariance structure of the Cauchy kernel distribution would not change. Sensitivity analysis, by altering the minimum and maximum values, consequently rejected the hypothesis that the covariance structures of sparse and simulated data were approximately identical. The limited available observations were only positioned in the mode part of the kernel density, and therefore it was not possible to simulate the appropriate tail data on the basis of the observed data (under the assumption that catastrophe events do occur).

In the statistical field, there is extensive discussion about the choice of bandwidth. For this paper we used the standard bandwidth settings in Simetar. However, changing of bandwidth parameters could result in different estimates of the low tail. Thus, there is a need to explore the effect of bandwidth choice in farm-level catastrophe simulation models.

Contrary to the asset integration assumptions, in which the decision-maker views gains and losses as a change in wealth position, this paper applied the measure of permanent income for UEP on the basis of constant absolute risk aversion properties of the expected utility function. According to these assumptions, farmers make their decisions on the basis of the annual incomes that are permanent in the long term. By doing so, relatively high risk premiums to avoid downside risks are expected. Alternatively, if wealth measures were taken as the basis of rational decision-making, differences in the optimal farm plans would be limited between alternatively generated joint distributions. However, when a more simplistic utility function containing the target minimum level of net farm income is the basis for decision-making, the approach on how the tail is included does certainly affect the optimal farm plan.

References

Anderson JR, Dillon JL (1992) Risk analysis in dryland farming systems. FAO, Rome

Anderson JR, Dillon JL, Hardaker JB (1997) Agricultural decision analysis. Iowa State University Press, Ames

Antwood J, Shaik S, Watts M (2003) Are crop yields normally distributed? A reexamination. Am J Agric Econ 85(4):888–901

Dekkers DA (2002) Quantitative information: arable and vegetable production. PPO, Lelystad

Galagher P (1987) U.S. soybean yields: estimation and forecasting with nonsymmetric disturbances. Am J Agric Econ 69(4):796–803

Hardaker JB, Lien G (2005) Towards some principles of good practise for decision analysis in agriculture. Norwegian Agricultural Economics Research Institute, Oslo, Working paper 2005-1.

Hardaker JB, Huirne RBM, Anderson JR, Lien G (2004) Coping with risk in agriculture. CABI, Wallingford

Hardaker JB, Lien G, Richardson JW (2006) Expanding the states of nature for risk programming. Contributed paper, AARES 51st conference, Queenstown, New Zeeland, 13–15 February 2007.

Just RE, Weninger Q (1999) Are crop yields normally distributed? Am J Agric Econ 81:287–304

Lien G, Hardaker JB, van Asseldonk MAPM, Richardson JW (2009) Risk programming and sparse data: how to get more reliable results. Agr Syst 101(1–2):42–48

Lien G, Hardaker JB, Van Asseldonk MAPM, Richardson JW (2011) Risk programming analysis with imperfect information. Ann Oper Res 190(1):311–323

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory. Oxford University Press, New York

Murdoch DC (1966) Linear algebra for undergraduates. Wiley, New York

Ogurtsov VA, Van Asseldonk MAPM, Huirne RBM (2008) Assessing and modelling catastrophic risk perceptions and attitudes in agriculture: a review. NJAS—Wageningen J Life Sci 56 (1–2):39–58

Ogurtsov VA, Van Asseldonk MAPM, Huirne RBM (2009) Purchase of catastrophe insurance by Dutch dairy and arable farmers. Rev Agric Econ 31(1):143–162

Pritchet ST, Schmit JT, Doerpinghaus HI, Athearn JL (1996) Risk management and insurance. West, St. Paul

Ramirez OA, Mishra S, Field S (2003) Crop-yield distributions revisited. Am J Agric Econ 85 (1):108–120

Richardson JW (2006) Simulation for applied risk management with an introduction to Simetar. Texas A&M University, College Station

Richardson JW, Klose SL, Gray AW (2000) An applied procedure for estimating and simulating multivariate empirical (MVE) distribution in farm-level risk assessment and policy analysis. J Agric Appl Econ 32(2):299–315

Richardson JW, Lien G, Hardaker JB (2006) Simulating multivariate distributions with sparse data: a kernel density smoothing procedure. IAAEA Annual Meeting, Gold Coast, Australia

Scott DW (1992) Multivariate density estimation: theory, practice, and visualisation. Wiley, New York

Shlaifer R (1959) Probability and statistics for business decisions. McGraw-Hill, New York

Silverman BW (1986) Density estimation and data analysis. Chapman and Hall, New York

Swinton SM, King RP (1991) Evaluating robust regression techniques for detrending crop yield data with nonnormal errors. Am J Agric Econ 73(2):446–451

Van Asseldonk MAPM, Kobzar O, Meuwissen MPM, Huirne RBM, Hardaker JB (2005) Market-based crop insurance appraisal using whole-farm planning. Proceedings of the 86th EAAE seminar on agricultural income stabilisation: what role should public policy play? Anacapri, Italy.

Yatchew A (1998) Non-parametric regression techniques in economics. J Econ Lit 36:669–721

Chapter 13
Forecasting Grape Maturation Under Heat Stress Using MatPred

Leorey Marquez, Geoff Robinson, and Simon Dunstall

13.1 Introduction

Recent episodes of extreme heat in Australia have highlighted the vulnerability of the wine industry to heatwave events. For many of its grape-growing regions, Australia is projected to experience shifts in annual average temperature between 2006 and 2030 in the order of 0.2–1.1 °C. By 2050, the projected increase in annual average temperature in grape-growing regions is 0.4–2.6 °C (Webb 2006). The best estimate of warming over Australia by 2030 relative to the climate of 1990 is approximately 1 °C with warmings of around 0.7–0.9 °C in coastal areas and 1–1.2 °C inland. Mean warming in winter is a little less than in the other seasons, as low as 0.5 °C in the far south (CSIRO 2007).

13.1.1 *Impact of Extreme Heat on Grapes*

The effects of a heatwave on winegrapes will vary depending on the location of the vineyard and on the timing of the heat event relative to the developmental stage (or phenology) of the grapevine. Webb et al. (2008) described the climate sensitivity of winegrape quality and provided a model to quantitatively inform the Australian wine industry of the impacts of projected climatic changes.

L. Marquez (✉)
CSIRO Computational Informatics, Gate 5, Normanby Road, Clayton, VIC 3168, Australia
e-mail: Leorey.Marquez@csiro.au

G. Robinson • S. Dunstall
CSIRO Mathematics, Informatics & Statistics, Clayton, VIC 3168, Australia
e-mail: Geoff.Robinson@csiro.au; Simon.Dunstall@csiro.au

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_13

Extreme heat causes winegrapes to ripen at a faster rate, reducing the number of days available for optimised harvest conditions (Webb et al. 2008). Thus, grape quality becomes increasingly compromised in hot regions (White et al. 2006) as quality may be inversely related to the degree and duration of warm temperatures (Jackson and Lombard 1993). In addition, the crushing of the grapes is more likely to take place at higher temperatures, increasing the likelihood of oxidation faults (Coombe 1987).

Excessively high temperatures are detrimental to grape development as they inhibit berry growth, delay sugar accumulation, impede fruit coloration, cause fruit to shrivel, and may cause abnormal pigmentation of white fruit (Hashim-Buckey 2006). Grapes are highly susceptible to heat, wind, and water stress during the flowering stage and any exposure to extreme weather events may result in yield loss and poor fruitset (Hayman et al. 2012). Excessive and prolonged heat can impair photosynthesis by causing plants to close their stomata and shut down the photosynthetic process.

Grapes are most vulnerable during the veraison or ripening phase. This is the period when grape berries resume growth (through cell expansion), become soft and accumulate sugar. Levels of acids decline and color appears in red or purple fruit. Studies have documented that very high temperatures during the ripening phase reduce or completely inhibit key enzymes that are responsible for the synthesis of anthocyanins. This results in poor coloration of fruit which reduces the amount of marketable fruit at harvest (Hashim-Buckey 2006).

Heat-induced shriveling of grapes is often referred to as "sunburn" or heat injury. This type of damage generally occurs after a sudden rise in temperature and may occur at any time from fruit set to harvest. The type and extent of damage varies; single berries, parts or whole clusters may wilt, shrivel, and dry. In some cases, damage occurs only to fruit that is directly exposed to sunlight. However, shaded fruit may also become damaged when temperatures exceed 40 °C, but susceptibility to such heat damage is usually variety-dependent (Hashim-Buckey 2006).

13.1.2 Recent Extreme Heat Events in Australia

Most of the climate warming in the last 100 years has been attributed to increases in greenhouse gases from human activities. Global temperatures from 11 of the last 12 years (1995–2006) have been ranked among the 12 warmest years in the instrumental record of global surface temperature (IPCC 2007). The projected changes in maximum and minimum temperature are associated with a projected strong increase in frequency of hot days and warm nights and a moderate decrease in frost. A substantial increase in fire weather risk is also likely at most sites in southeastern Australia (CSIRO 2007).

As the world becomes warmer, the frequency of heatwaves is likely to increase, but there will still be variability on a year-by-year basis. In the past 4 years alone, five significant extreme heat events have occurred namely:

1. March 2008 Autumn Heatwave—An exceptionally prolonged heatwave affected much of southern Australia in the first half of March 2008 (BOM

2008). This event was unusually late in the season and became a major concern for the southern regions of Victoria, South Australia, and Western Australia. Affecting mostly late ripening grape varieties, vines were defoliated and suffered from sunburn and heat damage. Ripening was delayed causing harvest schedules to be thrown into disarray (Hayman et al. 2012).

2. November 2009 Late Spring Heatwave—An exceptionally prolonged heatwave affected large parts of central and south-eastern Australia in November 2009, resulting in the month being the hottest November on record for many areas (BOM 2009a). This event came unusually early in the season and affected vineyards mostly in South Australia along with several areas in Victoria and New South Wales. Grape stock flowering during this event suffered from low yield, including Granache in the Barossa Valley and Merlot in the Limestone Coast (Hayman et al. 2012).
3. 2011 Summer heatwave—Between January 30 and February 6, 2011, the Hunter Valley in New South Wales experienced a heatwave that was both exceptionally hot and humid (BOM 2011).
4. December 2011 WA Record—Roebourne, in the Pilbara, recorded a maximum temperature of 49.4 °C on 21 December 2011, breaking Western Australia's previous December record of 48.8 °C at Mardie in the Pilbara on 26 December 1986. This was also the second-warmest December day on record for Australia, just 0.1 °C behind the all-Australia record 49.5 °C, observed at Birdsville on the 24 December 1972, and the fifth warmest day ever recorded in WA, for any month (BOM 2012).
5. January–February 2009 Heatwave—An exceptional heatwave affected south-eastern Australia during late January and early February 2009. The most extreme conditions occurred in northern and eastern Tasmania, most of Victoria and adjacent border areas of New South Wales, and southern South Australia, with many records set both for high day and night time temperatures as well as for the duration of extreme heat (BOM 2009b). A survey of 92 winegrowers across ten selected regions affected by the heatwave showed unprecedented impacts of the heatwave on vineyards with significant heat-related crop losses at some sites (Webb et al. 2009). The most obvious effects on grape vines included stalled development, leaf burn, leaf drop, berry sunburn, berry 'bagging', and berry shrivel. On February 7 (Black Saturday), catastrophic bushfires engulfed many of the heat-affected areas causing 173 fatalities, destroying more than 2,100 homes, and destroying a number of townships including Marysville and Kinglake (Egan 2012).

Having established the critical impact of extreme heat events on wine growers, the rest of the paper presents an innovative maturation forecasting methodology as implemented in the CSIRO decision tool MatPred. The next subsections describe the VitiForecaster package of which MatPred is a component and explains the logic behind the calculation of the harvest dates. Section 13.3 then describes the major data inputs required by MatPred followed by a discussion of the model estimation approach in Sect. 13.4. Finally, Sect. 13.5 presents the fitting of various regression models and the selection of the recommended regression function.

13.2 Maturation Forecasting with MatPred

As noted earlier, extreme heat accelerates the ripening process for grape berries resulting in significantly reduced opportunities for harvesting grapes at their desired level of maturity. With the increase in the frequency, intensity, and duration of extreme heat events, there have been stronger calls for more research into managing these events. The challenge will be for grape growers to appropriately manage the risk and conditions once informed of impending extreme heat events (Hayman et al. 2012).

The implementation of best-practice methods for managing extreme heat events will become critical for continued profitable wine production in Australia and globally (Webb 2006). This paper presents the forecasting tool MatPred for monitoring the maturation of grapevines with or without heat stress and for identifying the optimal period for harvest, thereby minimising the risks associated with extreme heat events. This paper also describes the process of fitting statistical models to provide a means of forecasting the maturation of grapes for a given vintage based on block and grower attributes, vintage characteristics, sampling measurements, and weather data. Using analysis of data from Vintage 2009, the results will show that statistical models that use meteorological data generally perform better than ones that do not (Marquez et al. 2009b). Based on analysis of the sample data, regression estimates for a certain model, referred to as “Model 5”, are recommended for use in maturation forecasting in Australia.

13.2.1 VitiForecaster and Components

Grapes are a highly perishable product, so timing is crucial. To assist the wine industry, Australia’s CSIRO in collaboration with Pernot-Ricard Pacific has developed VitiForecaster, a decision support system for planning and managing grape intake logistics. MatPred is part of the VitiForecaster package. VitiForecaster assists viticulturists and winemakers by providing analysis on crucial decision issues concerning winery intake, such as (CSIRO 2012):

1. Ripened grapes must be harvested as close as possible to the ideal time
2. Harvested grapes must be transported to a winery before significant berry deterioration occurs
3. Transport of unfermented juice and incomplete wine between wineries needs to be integrated with grape logistics and other winemaking considerations

Knowing when grapes are likely to be ready to harvest not only improves the quality of the wine they produce, but also helps with scheduling and planning for the rest of the supply chain. Harvest logistics must be planned about a week in advance in order to schedule harvesters and transport. Grape processing equipment is expensive and requires careful scheduling to maximise its use and efficiency. VitiForecaster allows customers to predict numbers of grape blocks that will be ready for harvest and therefore make much better decisions about harvesting and cartage (CSIRO 2012).

With VitiForecaster, CSIRO aims to make the intake planning process less intermittent, and instead institute a continuous process where plans and schedules (vintage-plans, crushing-plans, bookings, etc.) are updated whenever new information becomes available. VitiForecaster provides the following tools and components to support the intake planning process (Dunstall and Owens 2008):

1. Maturation forecasting—MatPred is used to predict the maturity time (time of reaching preferred Brix) for blocks by investigating the grape maturation patterns that are evolving in the current vintage.
2. Vintage projection—A view of vintage can be formed early in seasons when there is little or no grape sample data. VitiForecaster can provide indications of likely harvest dates, as well as giving a whole-of-vintage view from which decisions on harvesting and winemaking resourcing can be made.
3. Vintage planning—Microsoft Excel spreadsheets provide harvest date predictions to viticulturists, helping them to maximise overall vintage intake quality and to construct individual block plans taking into consideration block maturation, winery capacities, parcelling, and various other factors. In particular, the predictions can affect decisions about harvest dates and about which grape blocks send their grapes to which wineries.
4. Parcel planning—Spreadsheets are again used to facilitate the process of converting the vintage plans for blocks to individual parcels of grapes that will be scheduled for physical crushing at a winery.
5. Intake scheduling—Bid Sheet Workbooks allow limited crushing capacity to be allocated and crushing and pressing equipment to be scheduled.

For more details on the individual modules of VitiForecaster, please refer to Dunstall and Owens (2008).

13.2.2 Harvest Prediction Overview

The main function of MatPred is to generate estimates of the best harvest dates for blocks of grapevines given the desired level of maturity assigned to each block. As grapes mature, their sugar content increases. Thus, the level of maturity of a block is indicated by the average sugar content of the berries, measured in Brix, the number of grams of sucrose per 100 g of grapes. In technical terms, the Brix of grapes is the percentage of sugar by mass in an aqueous solution which has the same specific gravity. Higher Brix levels generally indicate higher potential levels of alcohol in the wine produced. Another commonly used measure of sugar content is Baume where 1 Baume = 1.8 Brix (DeGaris 2004). For consistency, this paper will use Brix to indicate sugar content. However, MatPred allows users to use either Brix or Baume for sugar content as shown in forthcoming examples.

Given the target or preferred Brix value for a block of grapevines, the process of predicting the best harvest date generally involves the following steps:

1. Obtain sample measurements of sugar level from the block during the entire ripening period.
2. From the series of sample measurements recorded, calculate the average daily increase in Brix for the block giving more weight to the most recent Brix measurements.
3. Update the predicted harvest date using the latest estimate of the average daily increase in Brix.
4. Repeat steps 1–3, preferably with samples taken more frequently as the predicted harvest date draws nearer.

MatPred's modelling approach uses the rate of increase in Brix per day as the response variable in a regression model estimated using explanatory variables that include time of year, block location, grape type, Brix sampling measurements and the weather. Other factors such as altitude, block orientation, and soil type will be included in the regression estimate in future versions. Details of the input data records describing block characteristics, grower attributes, sampling measurements, and weather are presented in Sects. 13.3.1, 13.3.2, 13.3.4, and 13.3.5, respectively.

The predicted date of harvest, referred in MatPred as "P-dates", is calculated by applying the derived daily rate of increase in Brix to the latest sampling date and projecting forward until the target Brix is reached. MatPred can also provide harvest dates for blocks with missing or incomplete sampling measurements. For these blocks, MatPred provides an "inferred" date of harvest ("I-date") based on the "P-dates" of blocks with similar grape characteristics, location, and weather data.

The flowchart in Fig. 13.1 summarises the process in MatPred of calculating predicted harvest dates (P-dates) or inferred harvest dates (I-dates). Given the target sugar content ("preferred Baume" in Fig. 13.1), a "P-date" can be obtained if there are sample measurements for the block. If there are none, a sufficient number of similar blocks having the same preferred Baume are then used to compute the inferred harvest date (I-date) for the block.

13.2.3 Harvest Dates

Aside from P-dates and I-dates, MatPred can also provide other types of harvest dates depending on available data. Single-letter codes are used to indicate the type of harvest date. Harvest dates fall into one of three categories, described as follows (Dunstall and Owens 2009):

1. *Maturity dates*. These are harvest dates that are related to grape maturity. These dates estimate when an average grape on a block will reach preferred target sugar level. There can only be one maturity-type harvest date for a block. There are four types of maturity dates described as follows:

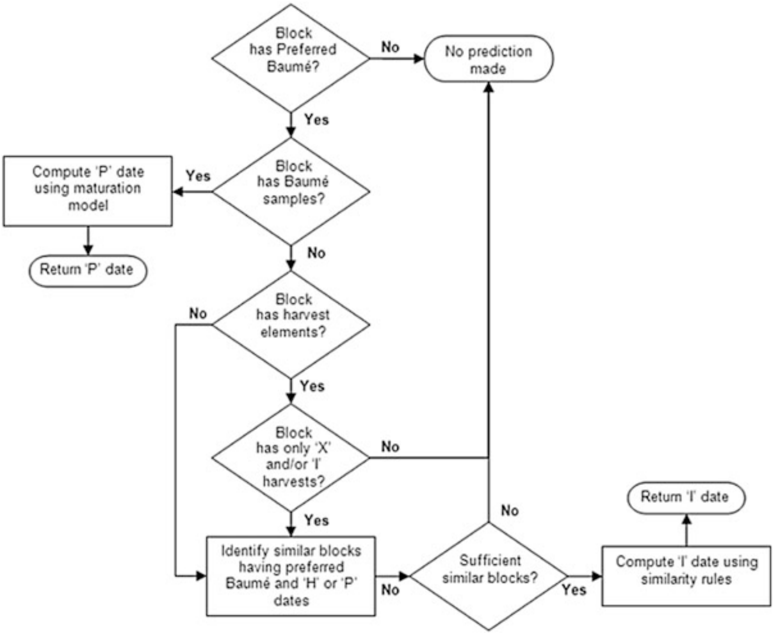


Fig. 13.1 Logic applied by MatPred to calculate a P (predicted) or I (inferred) harvest date for a block (*Source*: Dunstall and Owens 2009)

- An **X** date indicates that nothing is known about a block's history or future maturation profile.
- An **H** date (*historical date*) provides an expected harvest date this vintage, based solely on what the corresponding date was in a preceding vintage.
- An **I** date (*inferred historical date*) is computed by the CSIRO system when an H date is not known for the block and there are no samples for the block. The date is computed using date information relating to "similar" blocks.
- A **P** date (*predicted date*) is the most useful maturity type of harvest date, and it is computed by MatPred using Brix sample information for the block. A **P**-type date is most preferred, and an **X** date is least preferred.

2. *Bookings and actual dates*. These dates relate to actual harvests and crushes, and there can be several of these types current for a block at any given time (showing a series of bookings or crushes). This group consists of three date types, namely:

- A **B** date (*booked date*) corresponds to a booking that has been recorded in the system
- An **A** date (*actual date*) corresponds to a crush that has actually occurred (including the weight as recorded on a weighbridge)
- A **U** date (*uncrush date*) is computed when a block has some actual crushes, no more bookings in the system, and further tonnes are expected. The *date* part of the U date is equal to the time of the last crush and the tonnage part is equal to the estimated tonnes remaining to be harvested

3. *Planning dates*. Planning dates provide additional control for harvest management purposes. The date types in this category are:
   - A **V** date (*vintage-planned date*) is determined by vintage planning staff and typically updated twice-weekly during vintage. A block can have different vintage-planned dates for different intended-uses (e.g. if it is planned that grapes will go to two different wineries, to one as a red and the other as a rosé).
   - An **R** date (*recommended date*) indicates that a block has been marked as to-be-harvested in the short term (e.g. next fortnight) and is ready to be booked.
   - An **L** date (*latest-view date*) is a special case and is manually set by users. Sometimes an L date acts as a maturity date, sometimes as a planning date, and at other times as both, depending on the circumstances.

Table 13.1 presents a summary of the various harvest-date types and their relationship with each other.

13.3 MATPRED Input Data

MatPred requires historical and current data from three principal groups:

1. Grower data—Attributes of the blocks and area where the grapes are grown and managed
2. Maturation data—Brix, Ph, and acidity measurements from samples of the grapes to be harvested
3. Meteorological data—historical and forecast data on the weather conditions surrounding the vintage

13.3.1 Block Attributes

The basic unit of grape area used is the grape block, or simply block. Each block designates an area where grapes of the same variety from one area are being grown for the same intended use and will be harvested at the same time.

The fields in the block records consist of:

1. VendorCode—the identification code for the grower.
2. BlockRef—the identification code for the blocks within a given grower.
3. VinYr—the year of vintage, coded as 0 for 2000, 1 for 2001, 2 for 2002, etc.
4. IntendedUse—the identification code for the product type that will be produced with the grapes from the block. For example, “LILINC” refers to Lindauer Cuvee.
5. Grapetype—the type of wine that will be produced from this block. The value is one of “Red” (R), “White” (W), “Sparkling” (S), or “Fortified” (F).

Table 13.1 Harvest `date_type` definitions and update protocols (*Source*: Dunstall and Owens 2009)

Date type	Meaning	Member of set	Constraints	Invalidated by arrival of	Invalidates, on arrival
X	No history	Maturity	Only one of (XHIP) can be current (for a given (block, destination code) tuple)	XHIP	XHIP
H	Historical date	Maturity	Only one of (XHIP) can be current(for a given (block, destination code) tuple)	XHIP	XHIP
I	Inferred historical date	Maturity	Only one of (XHIP) can be current (for a given (block, destination code) tuple)	XHIP	XHIP
P	Predicted date	Maturity	Only one of (XHIP) can be current (for a given (block, destination code) tuple)	XHIP	XHIP
L	Latest view	Maturity	Only one L can be current (for a given (block, destination code) tuple)	L	L
B	Booking	Bookings and actuals	Multiple B can be current, and multiple harvests from (B,A,U) can be current, for any (block, destination code, winery) tuple	New set of (B,A,U) harvest elements for same block	Existing (B,A,U) harvest elements for same block
A	Actual crush	Bookings and actuals	Multiple A can be current, and multiple harvests from (B,A,U) can be current, for any (block, destination code, winery) tuple	New set of (B,A,U) harvest elements for same block	Existing (B,A,U) harvest elements for same block
U	Uncrush	Bookings and actuals	Only one U can be current, but multiple harvests from (B,A,U) can be current, for any block (even when a block streams to multiple products/wineries, only one U can be computed)	New set of (B, A,U) harvest elements for same block	Any existing (B,A,U) harvest elements for same block
V	Vintage plan date	Planning dates	Multiple V can be current for any given (block, destination code, winery)	Any V for same block (assuming V dates pass through the `vp2pp` service first)	

13.3.2 Grower Attributes

Each grower data record consists of:

1. VendorCode—a three-character identification code for the grower. Also referred to as GrowerCode.
2. Location—the label for the principal geographical location of the grower's activities. This is usually the suburb or local government area where the blocks are located.
3. Longitude—the longitude coordinate of the centroid denoting the centrepoint of the growers' activities.
4. Latitude—the latitude coordinate of the centroid denoting the centrepoint of the growers' activities.
5. Altitude—the altitude (in meters) of the centroid denoting the centrepoint of the growers' activities.

When altitude values are available for both weather stations and blocks, the MATPRED software adjusts temperature for altitude differences by subtracting 2 °C per 1,000 ft rise. In a setup phase, block coordinates are also helpful for determining which weather stations to use for meteorological data.

13.3.3 Maturation data

The term "maturation data" includes measurements for levels of pH (PH), titratable acidity (TA), and sugar concentration (Brix) either prior to harvest (referred to as sampling data) or at harvest (referred to as weightag data). For Australia, there were 9 years of sampling data (2000–2008) with each record consisting of:

1. VendorCode—the identification code for a grower.
2. BlockRef—the identification code for a block.
3. Vintage—the year of vintage for grapevines.
4. SampleDate—the day, month, and year when the sample was taken.
5. AnalysisCode—Code for the grape property being measured. This is denoted by BX for Brix, PH for pH value, or TA for titratable acidity.
6. AnalysisResult—the value observed for the AnalysisCode.

There were also 9 years of weightag data (2000–2008). The weightag data records consist of:

1. VendorCode—the identification code for the grower.
2. BlockRef—the identification code for the blocks within a given grower.
3. VinYr—the year of vintage for grapevines, denoted by 0 for 2000, 1 for 2001, 2 for 2002, etc.
4. WeightagDate—the day, month, and year when the weightag data was taken.
5. WeightagBr—the value observed for Brix.

Table 13.2 Numbers of blocks for each grape type for each vintage

Vintage	Fortified	O = Rosé	Red	Sparkling	White	Unknown
2002	30	0	1,644	161	864	0
2003	21	0	1,606	226	820	0
2004	20	16	774	136	557	297
2005	24	104	1,638	291	1,042	0
2006	17	167	1,552	307	957	0
2007	21	189	1,541	300	981	0
2008	13	54	1,566	306	963	0

The blocks of unknown grape were not used for fitting of models (Marquez et al. 2009b)

6. WeightagpH—the value observed for pH.
7. WeightagTA—the value observed for titratable acidity.
8. WeightagTonnes—total volume of grapes harvested in tonnes.

Table 13.2 shows the number of blocks per year per grape type for which maturation data was provided.

13.3.4 *Meteorological Data*

The MatPred software requires the following meteorological data for each day of vintage (DOV):

- Minimum (ambient) daily temperature (°C)
- Maximum (ambient) daily temperature (°C)
- Daily rainfall (mm)

These data are required as either observations (past) or forecasts (future), for the set of weather stations that are relevant to the vineyard blocks of interest.

In Australia there are two major types of weather stations, (1) automated weather stations (AWS) and (2) rainfall-only stations. Observations of many meteorological variables are available for an AWS, and near-term forecasts are also available for AWS. Forecasts are not readily available for non-AWS stations.

Figure 13.2 shows the locations of rainfall-measuring stations in Australia, while Fig. 13.3 shows the distribution of stations that measure temperature. Notice that weather stations are clustered closer to the coasts particularly in the southeast and southwest.

Required Australian meteorological data can be obtained from the Bureau of Meteorology (BOM). The BOM website (http://www.bom.gov.au/) provides access to various types of weather data, including:

- *Historical data*. There is historical data available from which long-term average rainfall and temperature values can be calculated.
- *Recent daily rainfall observations*. Recent rainfall data for many stations including non-AWS stations is typically available for the daily period up to 9 am local

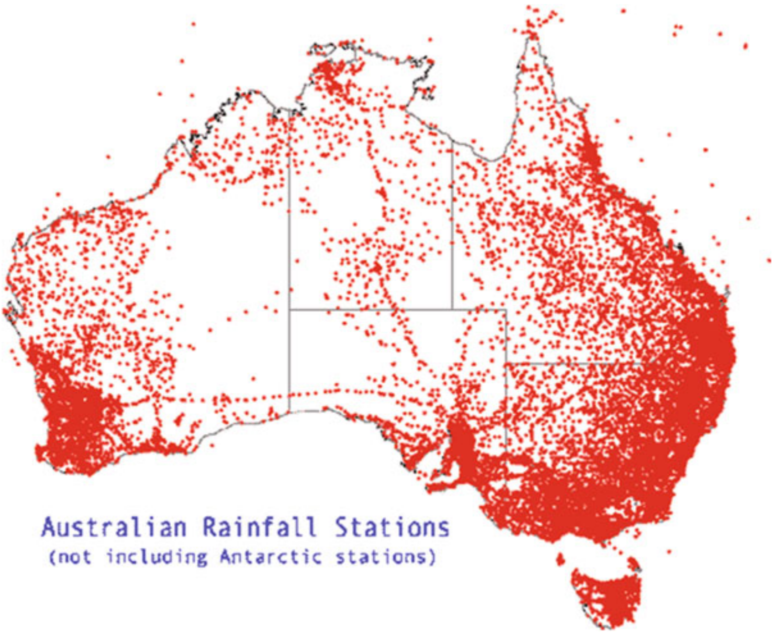


Fig. 13.2 Locations of rainfall-measuring stations in Australia (from Bureau of Meteorology)

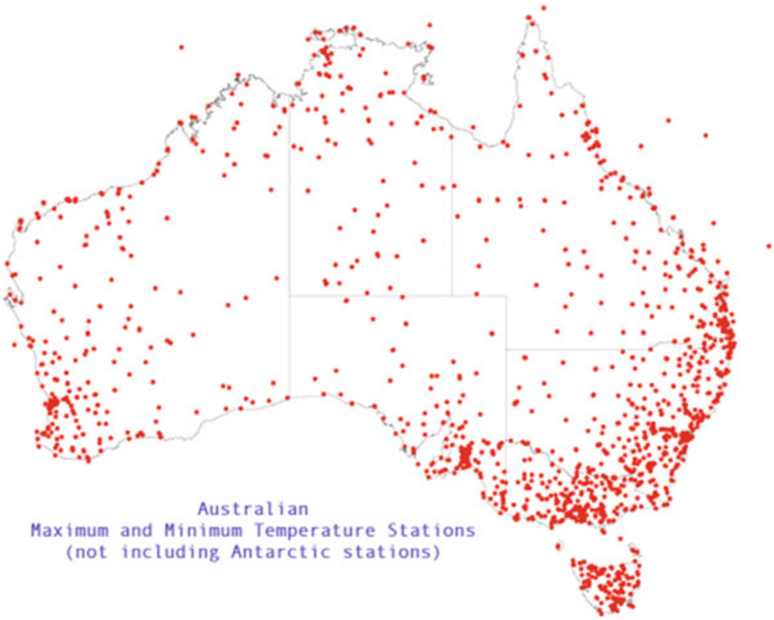


Fig. 13.3 Locations of temperature-measuring stations in Australia (from Bureau of Meteorology)

Table 13.3 Summary of Australian meteorological data used in analysis of Vintage 2009

Data set	Weather stations	BOM data product	URL (as of 21 Nov 2008)	Minimum update frequency
Weather station definitions	All	–	ftp://ftp.bom.gov.au/anon2/home/ncc/metadata/sitelists/stations_20081120.txt or at least the subset relating to the required IDCL* and IDY* products	Annually/service start-up only
Long-term averages	AWS only?	n/a	n/a	Annually/service start-up only
Daily rainfall observations	>3,000 stations (i.e. AWS and non-AWS)	IDCLRD00001	ftp://ftp.bom.gov.au//anon/home/ncc/www/rainfall/totals/daily/data/latest.data	Daily (approx 3 pm AEDST)
Daily temperature observations	~865 stations (i.e. most are AWS?)	IDCLCD0002	ftp://ftp.bom.gov.au//anon/home/ncc/www/temperature/silo/daily/data/latest.dc	Daily (approx 3 pm AEDST)
Temperature and rainfall forecasts (OCF)	~783 stations (i.e. most are AWS?)	IDY02122 and IDY02123	ftp://ftp.bom.gov.au/anon/gen/fwo/IDY02122.dat (0:00Z) ftp://ftp.bom.gov.au/anon/gen/fwo/IDY02123.dat (12:00Z)	Twice daily (approx 3 am and 3 pm AEDST)

time in the preceding day (data to 9 am today is posted at some number of hours after 9 am). It relates to one day only and is updated daily.

- *Recent hourly AWS observations*. There is data for near-real-time observations at AWS. These are recent to at least the last 3 h, if not better. The data has a 1-h frequency and covers a range of measures including ambient temperatures and rainfall.
- *Daily forecasts*. BOM provides daily forecasts up to 8 days into the future for various weather-related measures including ambient temperatures and rainfall.

Table 13.3 summarises the meteorological data used for the 2009 vintage.

13.3.5 Data Lookup Protocol

The meteorological data tables used by MatPred contain the following fields:

1. StationCode—the three-character code used to identify a weather station
2. CompDate—the day, month, and year when the data was observed
3. TotRain—the total amount of rain (in mm) observed for the entire day

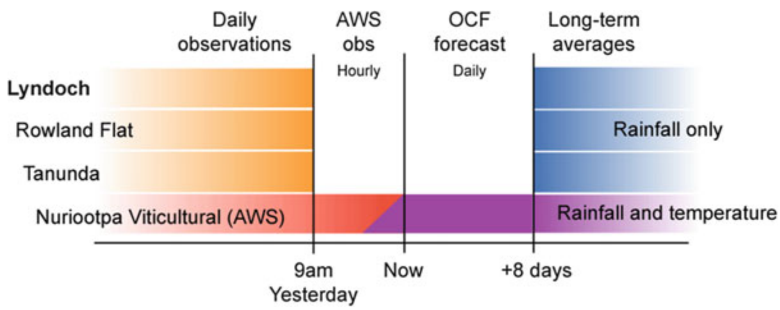


Fig. 13.4 General data lookup protocol

4. MinTemp—the minimum temperature (in degrees Celsius) observed for the entire day
5. MaxTemp—the maximum temperature (in Celsius) observed for the entire day

When assembling meteorological records for a particular place and time, MatPred will (in general) access different kinds of meteorological data depending on the availability of data and the meteorological measure(s) sought. MatPred applies the following lookup protocol when searching for temperature and/or rainfall values from stations for a particular date or time (Dunstall et al. 2009):

- If the sought date is far in the past, prior to the time at which daily observations began to be accumulated, MatPred uses the long-term averages.
- If the sought date is in the past, between the time that the observations started and around 36 h prior to now, MatPred can usually look up the daily observations (in Australia: AWS for temperatures, and AWS or non-AWS for rainfall).
- If the sought date-and-time is in the last 36 h, MatPred will typically use AWS observations or forecasts received in the past few hours, depending on when the last set of observations was downloaded.
- If the sought date-and-time is in the near future (i.e. in the forecast horizon), MatPred will use forecasts for the AWS.
- If the sought date is further into the future, MatPred will use long-term averages (in Australia, these apply at AWS only).

Figure 13.4 illustrates the lookup protocol used when searching for representative rainfall and temperature values for a given time period "Now". If "Now" is too far in the past or in the future, then long-term averages are used for the location of interest. For more recent times, daily observations, hourly observations, or daily forecasts may be used.

13.3.6 Assigning Blocks to BOM stations

In order to find the meteorological conditions corresponding to the maturation data, MatPred assigns the grape blocks to a set of Key BOM stations. For each wine-growing area, MatPred uses a two-step procedure to assign a Key BOM station to the area. This is performed for data-validation and data gap-filling purposes. As a first step, MatPred identifies the Key BOM stations most commonly assigned to blocks in the wine-growing areas. Key BOM stations are chosen with the following criteria in mind:

- Historical data on daily maximum temperature and daily rainfall are constructed for each Key BOM station. Where such data is missing for a Key BOM station, values are estimated using historical data from neighbouring BOM stations.
- Forecasts of daily maximum temperature and daily rainfall are constructed for each Key BOM station. Where such forecasts are unavailable for a Key BOM station, values are estimated using forecast data from neighbouring BOM stations.
- The Key BOM stations cover the regions in which most wine is being grown, so for most blocks it is reasonable to expect that their weather will be similar to that at the nearest Key BOM station, allowing for adjustments such as those for altitude. (A commonly used figure for the atmospheric lapse rate is 2 °C per thousand feet.)
- The number of Key BOM stations is optimised so that the computer resources and processing time required to perform the estimation of parameters from historical data and the construction of predictions are minimised while providing cover for the maximum number of grape blocks.

In the second step, the closest Key BOM station is assigned to each block, with the second closest Key BOM station as alternate. For each growing area, maps were used to select a Key BOM station that is close to the middle of that area. These are recorded in a BOM-station-to-area table (`Area2BOMStn.csv`). For blocks with no GPS data, block-to-area information is used to assign a Key BOM station, using the BOM-station-to-area table.

13.4 Forecasting Methodology

13.4.1 Modelling Brix Changes as Brownian Motion

The modelling approach essentially consists of fitting a regression model to the rate of change of Brix over a given period. This rate of change is expected to be affected by factors such as the time of year, the starting Brix, rainfall, temperature, and grape type. In building the regression equations for Vintage 2009, the statistical analysis observed the following guidelines:

1. Sugar concentration (measured in Brix) varies as a function of a number of factors including time, grape characteristics, block location attributes, weather, and grower activities.
2. For each block, the pattern of Brix changes from season to season.
3. Measurement of Brix is subject to sampling and testing errors.

4. The starting Brix value may vary from block to block and from vintage to vintage.
5. The pattern of Brix over time can be modelled as Brownian motion.

The variance of the change in Brix over a period of h days is expected to be approximately linear as a function of h. The variance for $h = 0$ is the sum of the measurement error variances for the initial and final Brix. The slope of the variance as a function of h may be interpreted as a Brownian motion variance per unit time. Hence, the variance of the change in Brix may be written as

$$2\sigma^2 + V_{\mathrm{B}}h$$

where

- σ^2 is the variance of a Brix measurement
- V_{B} is the Brownian motion variance per day

Therefore, the variance of the rate of change of Brix over a period of h days is given by

$$2\sigma^2 h^{-2} + V_{\mathrm{B}}h^{-1}$$

When fitting a model to data items which are each predicting rates of change of Brix, we use weighted linear regression, with weights given by the reciprocal of this variance.

In practice, the parameters σ^2 and V_{B} are estimated by an iterative model-fitting technique as follows:

1. The ratio $2\sigma^2/V_{\mathrm{B}}$ is initially set to a typical value, say 10.
2. The model is then fitted to data on the rate of change of Brix. The residuals from this model are multiplied by h, so that they can be interpreted as residuals from a model for the change in Brix over a period.
3. Then a simple linear regression model for the squares of these residuals using h as a predictor enables the parameters σ^2 and V_{B} to be estimated. These estimated values are then used for fitting the model to the rate of change of Brix.
4. Step 3 is repeated until it converges (which usually takes about seven iterations).

One complication to this model-fitting procedure arises when there are outliers among the data. The solution adopted for dealing with this problem is to fit the model to the rate of change of Brix twice. First the model is fitted to the raw data. Then scaled residuals are computed by multiplying the residuals by the square root of the weighting factor. The median of the absolute values of the scaled residual is computed and divided by the quartile of a standard normal distribution (the median of absolute values of standard normals) in order to provide an estimate of the standard deviation of the scaled residuals.

Several regression models for the daily change in Brix for Vintage 2009 were estimated using different sets of explanatory factors and their interactions. These models were then analysed and compared in terms of robustness and accuracy.

13.4.2 Regression Estimation Without Heat Stress

Marquez et al. (2009b) describes the series of linear models estimated and analysed to plan the 2009 Vintage for a major Australian wine company. Information from 2002 to 2008 were provided to CSIRO on the attributes, grape samples, and harvests for all blocks supplying grapes to this wine company and its contract processors. These data were then used to estimate the daily rate of change in Brix as a function of sampling date, starting Brix, rainfall, temperature, and other factors. This initial set of explanatory variables did not include heat stress. Table 13.4 lists some of the coefficients used in the regression models for Vintage 2009 and their corresponding variables when implemented in MatPred.

Nine regression models were estimated for Vintage 2009 using weighted linear regression with the length of the interval between Brix measurements as the weighting variable (Marquez et al. 2009b). The models were intended to predict Brix levels 1 day at a time so only rates of change for intervals of up to 10 days were used for fitting. Analysis of the weightag data provided showed that the weightag values were subject to different sources of noise and errors. As a result, the weightag data were discarded and only sampling data were used for the fitting of the models for Vintage 2009.

Analysis and comparison of the nine regression models for Vintage 2009 showed that (Marquez et al. 2009b):

- The effect of **DOV** is positive, which is different from the effect found for New Zealand data (Marquez et al. 2009a). Other factors being equal, grapes ripen slightly more slowly late in the season than early in the season. The effect of being 60 days later in the season is approximately an increase of 0.08 Brix per day.
- The rate of increase in Brix decreases as the initial Brix (**StartBX**) increases. For instance, it is about 0.25 Brix per day slower at a starting Brix of 23 than at a starting Brix of 15.
- Rain during the period tends to decrease Brix. However, previous rain (rain more than 4 days before the sampling date) has a smaller effect and tends to increase the rate of change of Brix.
- Grapes ripen at about 0.01 Brix per day faster for each degree Celsius increase in maximum temperature.
- Grapes for red, fortified, and rosé wines ripen slightly faster than white wine grapes, whereas grapes for sparkling wine ripen more slowly.
- The model with the smallest estimated Brownian motion variance was judged as the most robust and selected as the model to be used for predicting harvest dates. This model combines fortified wines with white wines and aggregates the rainfall over the 4 days before the start of the interval into a single variable, **rainp**. The square of that 4-day-aggregate rainfall, **rainpSq,** is also used in this model.

Table 13.4 Model variables for Vintage 2009 and their corresponding MatPred implementation

Variable name in data analysis	Variable name in software implementation	Meaning	Definition and units
DoV	YEAR_DAY	Day of vintage/day number in the year, at the midpoint of the time interval	Integer days elapsed since (and including) Jan 1: at Jan 1 YEAR_DAY is equal to 1
DoV2	YEAR_DAY_SQ100	Square term for day number in the year	Equal to (YEAR_DAY)(YEAR_DAY − 100)
DoV3	YEAR_DAY_CU100	Cubic term for day number in the year	Equal to YEAR_DAY_SQ100 × YEAR_DAY
BX_ini	START_BRIX	Initial Brix value in the time interval	Units of Brix
BxSq	START_BRIX_SQ	Square term for initial Brix	Equal to START_BRIX squared
AvTotRain	PRECIP_DAILY	Total rainfall in the immediately preceding daily period	Units of mm
Rainp	PRECIP_4DAILY	Total rainfall in the immediately preceding 4-day period	Units of mm
rainpSq	PRECIP_4DAILY_SQ		Equal to PRECIP_4DAILY squared
AvMaxTemp	TEMP_MAX	The maximum temperature in the immediately preceding daily period	Units of °C
rainp: AvMaxTemp	PRECIP_4DAILY_m_TEMP_MAX	An interaction term	Equal to PRECIP_4DAILY × TEMP_MAX
Intercept	INTERCEPT	A constant term	Units of Brix per day

13.4.3 Estimating Maturation Under Heat Stress

As noted earlier, the extreme heat events in January and February 2009 caused unprecedented impacts on the affected winegrape-growing regions both in terms of yield loss as well as the scale of damage noted on the grape bunches (Webb et al. 2009). This event rendered the then-current maturation forecasting model inadequate as the meteorological conditions have completely altered the regression factors.

In response, CSIRO incorporated several revisions into the regression modeling to account for the impact of extreme heat events. The principal changes introduced in the analysis are summarized as follows:

- Sampling measurements from 2009 were added to the maturation data. However, the matching of grape blocks to BOM weather stations was not as meticulously followed as previous efforts where weather predictions were available. For the 2009 data, we only used grape blocks such that the BOM station could be found from earlier matching of grape blocks to BOM stations.
- The earlier analyses used time-of-year effects (**ToY** and quadratic, cubic and quartic terms in **ToY**) when fitting models but these were not expected to be equally appropriate in all latitudes. Hence, time-of-year effects have largely been replaced by a measure of the amount of sunshine which is computed from latitude and the time of year. It was expected that models fitted using terms involving the amount of sunshine will be useful over a wide range of latitudes.
- In direct response to the observation that many blocks of grapes were adversely affected by very high temperatures during the 2009 season, a term representing heat stress was added to the model. This term does not consider the nature of the effects of heat on plant physiology, but simply estimates how the rate of ripening of grapes is affected by high temperatures. The mathematical form of the heat stress term was developed in conjunction with Angus Davidson of Pernot-Ricard Pacific.

The next two sections provide details on the derivation of measures for sunshine and heat stress.

13.4.4 Modelling the Amount of Sunshine

Previous models fitted used the daily rate of change in Brix as the response variable and the day of the year (DOY) as an explanatory variable. With the introduction of sunshine and heat stress into the regression formulation, a procedure was formulated whereby sunshine can be calculated from DOY which represents the position of the earth relative to its orbit, and latitude which represents the position of a point on the earth's surface relative to the sun.

The calculation of the amount of sunshine per day impacting on a horizontal surface is described as follows.

First, the DOY is expressed as an angle ϕ in relation to the earth's orbit. To obtain ϕ (in radians), we use

$$\phi = (\mathrm{DOY}/(365 + 97/400))^{*}2\pi$$

where

- 365 + 97/400 is the average number of days in a year (since there are 97 leap years every 400 years)
- 2π is the total radians in a complete orbit

If ψ is the apparent latitude of the sun at DOY, then its tangent is approximated by

$$\tan(\psi) = 0.0712\sin(\phi) - 0.3999\cos(\phi) + 0.0006\sin(2\phi) - 0.0067\cos(2\phi) \\ +0.0013\sin(3\phi) + -0.0026\cos(3\phi)$$

The apparent latitude of the sun can then be computed by taking the arctangent. Note that positive angles are conventionally used to mean that the sun is north of the equator.

The length of daylight **D** at the given point depends on time of the year (DOY) and the latitude λ of the point. Half the length of daylight, **D/2**, expressed as an angle in radians, at the specified latitude λ is then given by

$$D/2 = \arccos(-\tan(\psi)^{*}\tan(\lambda))$$

If latitudes within the Arctic or Antarctic circles are used, then the quantity whose arccosine is computed must be constrained to be between −1 and +1.

A measure of the amount of sunshine **S** is then given by

$$S = \cos(\psi)^{*}\cos(\lambda)^{*}\sin(D/2) + \sin(\psi)^{*}\sin(\lambda)^{*}(D/2)$$

This quantity is plotted in Fig. 13.5. The scale of this measure is calculated such that the amount of sunshine is unity at the equator at the equinoxes.

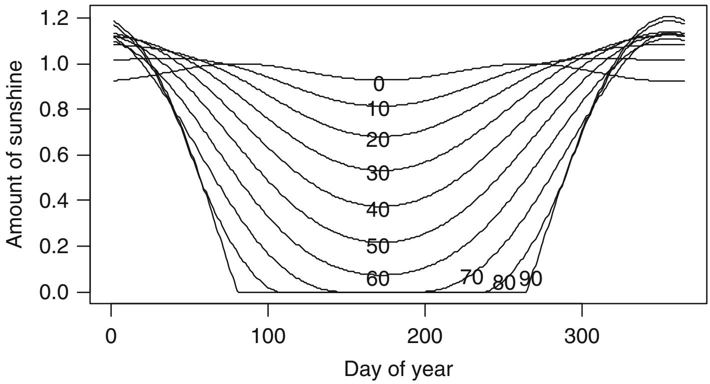


Fig. 13.5 Amount of sunshine as a function of day of year and degrees south latitude (written over lines). Unity is taken to be the daily amount of sunshine at the equator at the time of an equinox. Cloud cover and shade are ignored

13.4.5 Estimating Heat Stress

The calculation of the term for heat stress consists of four steps.

1. An inquiry into the temperature at which heat stress starts to occur was conducted with initial values for these parameters chosen based on advice from Angus Davidson from Pernot-Ricard-Pacific. Alternative values were tested for goodness-of-fit with observed data for vintages 2003–2009. The resulting temperature thresholds obtained were:

   (a) 42.5 °C if the previous overnight minimum temperature was 20 °C or less
   (b) 39.5 °C if the previous overnight minimum temperature was 30 °C or more
   (c) 42.5 °C – 0.35× (previous overnight minimum temperature −20 °C) if the previous overnight minimum temperature was between 20 and 30 °C.

2. If the daily maximum temperature is greater than the temperature at which heat stress starts to occur, then the daily heat stress is taken to be equal to the temperature excess in degrees celsius. Daily heat stresses are accumulated and the cumulative sum of daily heat stresses is the number which is assumed to be related to the rate of change of Brix.
3. If the daily maximum temperature is at least 5 °C lower than the temperature at which heat stress starts to occur, then the cumulative sum of daily heat stresses is reduced by 2 or to its minimum possible value of zero.
4. If the daily maximum temperature is lower than the temperature at which heat stress starts by between 0 and 5 °C, then the reduction in the cumulative sum of daily heat stresses is proportionately smaller.

For example, Fig. 13.6 shows a pattern of daily temperatures where the maximum temperature on the fourth day was 44 °C and there was a series of 4 very hot days from day 12 to day 15 with maximum temperatures of 44, 46, 45, and 42 °C

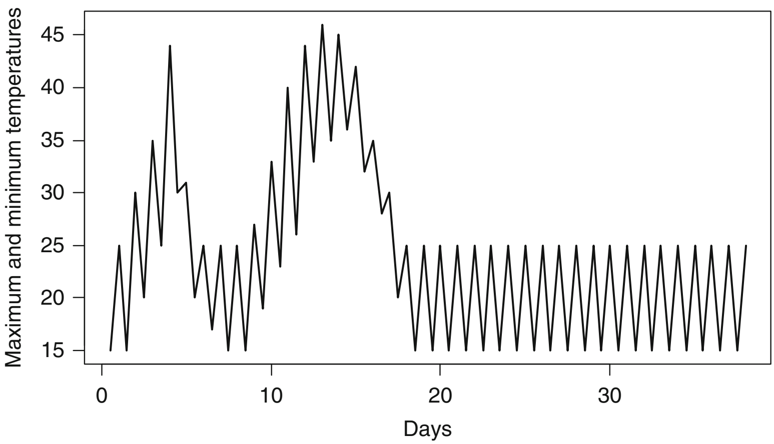


Fig. 13.6 Pattern of maximum and minimum temperatures to illustrate calculation of heat stress

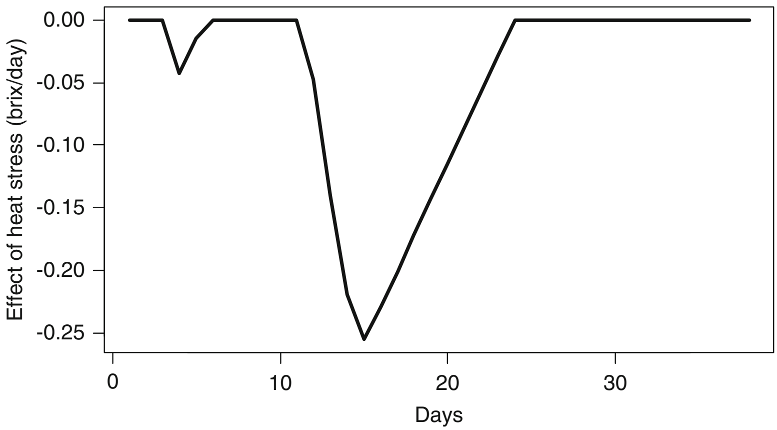


Fig. 13.7 Example of amount of heat stress according to model

and high overnight minimum temperatures. Figure 13.7 presents the calculated effect of heat stress for the pattern of daily temperatures shown in Fig. 13.6.

Note that the isolated very hot day causes a small amount of heat stress, but the series of very hot days causes much more substantial heat stress according to the model, with the rate of increase of Brix being much smaller than it would otherwise have been. There are other terms in the model which tend to increase Brix more rapidly at higher temperatures, but this heat stress term can be larger than them in magnitude so that the overall model predicts that Brix will decrease.

13.5 Fitting Models

Several linear regression models were fitted to the daily rates of change in Brix (Y), largely following the models used in previous data analyses for Vintage 2009, but now incorporating sunshine and heat stress as explanatory variables (Marquez et al. 2009a, b). Table 13.5 presents some of the principal explanatory variables used in obtaining the regression estimates. Only rates of change for intervals of up to 10 days were used for fitting these models. It is intended that the models will be used for predicting one day at a time. Again, only sampling data was used for the fitting of these models.

Table 13.5 Regression variables used in estimating daily rate of change in Brix

Regression variable	Description
DoV	Day of vintage which is 1 for March first, 32 for April first, etc. (The midpoint of the interval was used)
DoV2	Quadratic term in day of vintage, defined as DoV * (DoV – 100)
DoV3	Cubic term in day of vintage, defined as DoV * DoV2
DoV4	Quartic term in day of vintage, defined as DoV2 * DoV2
Sunshine	Amount of sunshine computed from latitude and day of vintage
sunshineSq	Quadratic term in sunshine, defined as (sunshine – 1) squared
StartBx	Initial Brix
BxSq	Square of initial Brix
GrapeTypeRed	Difference of slope for red wine grapes from the slope for white wine grapes
AvTotRain	Average rainfall in millimetres over interval
heatStress$reg	Calculated heat stress for the given day and location
rainpN	Total rainfall for the 24-h period *N* days before the start of the interval
AvTotRain: AvMaxTemp	Interaction term for average total rain and average maximum temperature

13.5.1 Model 1

This is the simplest model fitted, estimating the daily rate of change in Brix (Y) as a constant value of 0.17 for all grape types. For Vintage 2009, Model 1 estimates that:

$Y = 0.17$

13.5.2 Model 2

Model 2 expands Model 1 by providing a separate estimated daily rate of change in Brix (Y) for each wine type. These estimates are only slightly different from those reported in Marquez et al. (2009b). For Vintage 2009, Model 2 estimates that:

White: $Y = 0.24$
Red: $Y = 0.19$
Sparkling: $Y = 0.30$
Fortified: $Y = 0.17$
Rose: $Y = 0.24$

13.5.3 Model 3

This model expands Model 2 by making the daily rate of change in Brix (Y) a function of the initial Brix value (**StartBX**). This adds the feature that the daily rate of change in Brix is smaller when the starting Brix is larger. The estimated coefficients for this fitted model are similar to those reported in Marquez et al. (2009b). For Vintage 2009, Model 3 estimates that:

White: Y = 0.74 – 0.03 * **StartBX**
Red: Y = 0.77 – 0.03 * **StartBX**
Sparkling: Y = 0.81 – 0.03 * **StartBX**
Fortified: Y = 0.43 – 0.01 * **StartBX**
Rose: Y = 0.68 – 0.02 * **StartBX**

Note that the coefficients from this model can be used to find estimates of the maximum Brix, by looking at the Brix for which the fitted rate of change is zero. These estimated maxima are 26.39, 28.61, 24.07, 39.63, and 29.18 for white, red, sparkling, fortified, and rosé wine grapes, respectively. These maxima should not be thought of as real biological limits. However, the data do suggest that the rate of increase of Brix becomes smaller as these values are approached.

13.5.4 Model 4

This model included a large number of predictor variables and interactions, in order to obtain the maximum fit that can be expected. The intention was to trim this large model into a more compact model that can be adopted for routine use. To illustrate, the Model 4 estimate for White wines for Vintage 2009 is given by:

Y = 1.16 + 0.036 * **StartBX** – 0.0023 * **BxSq** +
0.094 * **AvTotRain** – 0.004 * **AvTotrainSq**
–1.26 * **sunshine** – 0.75 * **sunshineSq+**
0.0036 * **rainp1** + 0.0031 * **rainp2+**
0.0033 * **rainp3** + 0.0026 * **rainp4+**
–0.0079 * **AvMinTemp** + 0.019 * **AvMaxTemp+**
–0.0053 * **StartBX** * **interval+**
0.0002 * interval * **BxSq+**
–0.0013 * **StartBX** * **AvTotRain+**
–0.0029 * **AvTotRain** * **AvMaxTemp+**
–0.0002 * **AvTotRainSq** * **AvMaxTemp**

Model 4 estimates for Red, Sparkling, Fortified, and Rose wines follow a similar format.

Many of the fitted coefficients are similar to those found in previous reports. Both sunshine and day-of-vintage (**DoV**) variables have been included in this model even though they are alternative ways of explaining the same types of effects, so it would not be reasonable to expect the coefficients for the **DoV** terms to be similar to those found previously.

The residual standard error from this model is intended to be used as a benchmark for choosing a simpler model. We expect to be able to find a simpler model with residual standard error not very much larger than for this model.

Table 13.6 Numbers of intervals affected by heat stress

Vintage	Number of intervals affected by heat stress	Average size of heat stress effects
2003	926	0.0066
2004	864	0.0282
2005	118	0.0036
2006	1,098	0.051
2007	724	0.0042
2008	481	0.0024
2009	5,500	0.1144

13.5.5 Model 5

This model aggregates the rainfall over the 4 days before the start of the interval (**rainp1, rainp2, rainp3, rainp4**) into a single variable, **rainp**. The square of that 4-day-aggregate rainfall, **rainpSq**, is also used in this model. To illustrate, the Model 5 estimate for White wines for Vintage 2009 is given by:

Y = 0.98 + 0.033 * **StartBX** − 0.0004 * **BxSq** −
0.008 * **AvTotRain** + 0.053 * **interval**
−0.37 * **sunshine** − 0.41 * **sunshineSq+**
0.0039 * **rainp** + 0.00002 * **rainpSq+**
−0.014 * **heatStress$reg** + 0.014 * **AvMaxTemp+**
0.0021 * **StartBX** * **interval+**
–0.00006 * **rainp** * **AvMaxTemp**

Again, Model 5 estimates for Red, Sparkling, Fortified, and Rose wines follow a similar format.

The heat stress term in this model was zero most of the time. The number of intervals between successive maturity samples for which the heat stress term was not zero is given in Table 13.6. The average estimated size of the non-zero heat stress effects is also given in Table 13.6. We can see that heat stress was much more of a problem in 2009 than in previous vintages.

Similar information is displayed in Fig. 13.8. The heat stress effects are generally zero. However, for 2009 there were a much larger number of non-zero heat stress effects than for other vintages and these effects were generally larger in magnitude.

13.5.6 Model 6

This model used day of year (**DoY**) and its square and cube (**DoY2**, **DoY3**) as substitutes for the amount of sunshine and its square. The results show that these alternatives are equally good as Model 5 in terms of measurement variance, but not as robust when considering the Brownian motion variance, as shown in Table 13.7. To illustrate, the Model 6 estimate for White wines for Vintage 2009 is given by:

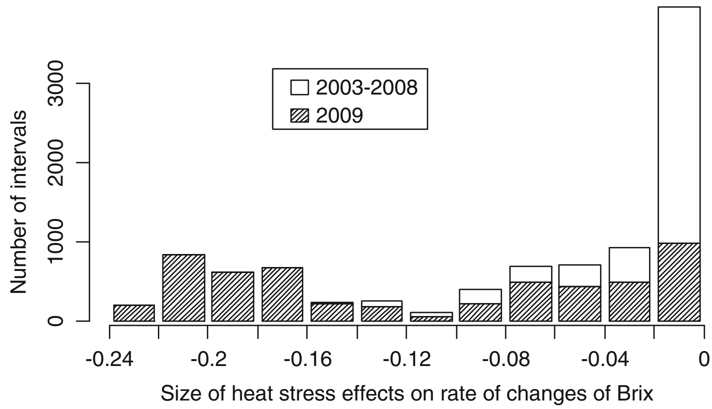


Fig. 13.8 *Stacked bar chart* showing frequencies of non-zero heat stress effects for 2009 and for all other vintages combined

Table 13.7 Comparison of models for the components of variation

Derived model	Measurement variance	Brownian motion variance	Number of outliers
Model 1	0.2311	0.1711	531
Model 2	0.2358	0.1558	544
Model 3	0.2743	0.1184	618
Model 4	0.2806	0.0922	627
Model 5	0.2868	0.0917	611
Model 5 red wine only	0.2995	0.0781	384
Model 5 without 2008 data	0.2918	0.0770	512
Model 6	0.2849	0.1103	609
Model 7	0.2830	0.0953	618

$$
\begin{aligned}
Y = {} & 0.62 - 0.032 * \mathbf{StartBX} - 0.0004 * \mathbf{BxSq} - \\
& 0.008 * \mathbf{AvTotRain} - 0.053 * \mathbf{interval} \\
& -0.0013 * \mathbf{DoV} - 0.00003 * \mathbf{DoV2} + \\
& 0.004 * \mathbf{rainp} - 0.00002 * \mathbf{rainpSq} + \\
& -0.014 * \mathbf{heatStress\$reg} + 0.014 * \mathbf{AvMaxTemp} + \\
& 0.0021 * \mathbf{StartBX} * \mathbf{interval} + \\
& -0.00006 * \mathbf{rainp} * \mathbf{AvMaxTemp}
\end{aligned}
$$

Again, estimates for Red, Sparkling, Fortified, and Rose wines follow a similar format.

13.5.7 Model 7

This model uses meteorological information like Model 5, but only uses linear terms in an attempt to produce a model which is easy to interpret. To illustrate, the Model 7 estimate for White wines for Vintage 2009 is given by:

$$Y = 0.76 - 0.033 * \mathbf{StartBX} + 0.008 * \mathbf{AvTotRain} - 0.012 * \mathbf{interval} - 0.30 * \mathbf{sunshine} + 0.0014 * \mathbf{rainp} - 0.014 * \mathbf{heatStress\$reg} + 0.014 * \mathbf{AvMaxTemp}$$

As in the previous models, estimates for Red, Sparkling, Fortified, and Rose wines follow a similar format.

13.5.8 Selection of a "Best" Model

13.5.8.1 Comparing Estimates of Variance Components

For each of the models fitted we find estimates of the measurement variance and the Brownian motion variance per day. These estimates are given in Table 13.7. As in Marquez et al. (2009b), Model 5 is preferred largely because it has the smallest estimate of Brownian motion variance. When the model for the rate of change of Brix is better, the estimated Brownian motion variance is smaller because more of the variation in the slope data is explained by the factors included in the model. Model 5 gives the smallest estimate and appears to be a sensible choice. In this model, sunshine, heatstress, total rain, and maximum temperature are all significant variables in explaining the daily change in Brix.

The Brownian motion variance is much larger than those obtained for New Zealand by Marquez et al. (2009a). For Model 5, the estimate of Brownian motion variance is 0.0917 compared to 0.0189 for New Zealand. This suggests that grape maturity is much harder to predict in Australia, perhaps because its weather is more variable. This component of variance would be expected to be smaller if more accurate weather data were used for fitting the models.

The measurement variance is slightly larger than for New Zealand (0.2868 compared to 0.2123 for Model 5). One possible explanation for this difference is that there is more bunch-to-bunch variation in Australia.

13.5.8.2 Comparing Predictions of Models

The amount of difference in fitted values between Model 5 and some of the other models is summarised in Table 13.8. Most of the differences are small, compared to the average slope of about 0.15 Brix per day. This suggests that Model 5 is fairly robust. In particular, it is substantial proof that fitting the model leaving out the 2008 data does not change the predictions much.

Table 13.8 Comparing differences in predictions between models

Derived model	Min	1 %	5 %	25 %	50 %	75 %	95 %	99 %	Max
Model 1	−0.598	−0.274	−0.187	−0.070	0.005	0.077	0.179	0.252	0.806
Model 2	−0.595	−0.259	−0.179	−0.069	0.004	0.074	0.175	0.248	0.787
Model 3	−0.266	−0.164	−0.124	−0.056	−0.009	0.037	0.107	0.165	0.549
Model 4	−0.334	−0.066	−0.039	−0.013	0.002	0.014	0.034	0.054	0.482
Model 5 red wine only	−0.087	−0.035	−0.021	−0.008	0.001	0.010	0.025	0.038	0.224
Model 5 without 2008 data	−0.263	−0.053	−0.037	−0.018	−0.006	0.004	0.019	0.032	0.075
Model 6	−0.234	−0.144	−0.106	−0.045	0.000	0.041	0.107	0.167	0.543
Model 7	−0.171	−0.053	−0.028	−0.009	−0.001	0.009	0.033	0.063	0.276

13.5.9 Conclusion

Recent extreme heat events have shown the need to update MatPred's maturation forecasting capabilities to take into account the impact of heat stress on wine grapes. As a result, the estimation procedure has been extended to include measures for sunshine and heat stress and the examination of several new regression estimates. These variables not only maintained the prediction capabilities of earlier models, but have enhanced the robustness of the new models by accounting for extreme variations in weather events. A comparison of these new estimates has established a recommended process for formulating regression relationships for future vintages that will enable analysis of the impact of future extreme heat events on the Australian wine industry to be achieved and corresponding risks to be managed effectively.

References

BOM (2008) An exceptional and prolonged heatwave in southern Australia, Australian Bureau of Meteorology. http://www.bom.gov.au/climate/current/statements/scs15b.pdf

BOM (2009a) A prolonged spring heatwave in central and south-eastern Australia, Australian Bureau of Meteorology. http://www.bom.gov.au/climate/current/statements/scs19b.pdf

BOM (2009b) The exceptional January-February 2009 heatwave in south-eastern Australia, Australian Bureau of Meteorology. http://www.bom.gov.au/climate/current/statements/scs17d.pdf

BOM (2011) An exceptional summer heatwave in greater Sydney and the Hunter Valley. Special Climate Statement 27, Australian Bureau of Meteorology. http://www.bom.gov.au/climate/current/statements/scs27.pdf

BOM (2012) New high temperature record set for Western Australia. Australian Bureau of Meteorology. http://www.bom.gov.au/climate/current/statements/scs37.pdf

Coombe BG (1987) Influence of temperature on composition and quality of grapes. Acta Hortic 206:23–35

CSIRO (2007) Climate change in Australia. http://www.climatechangeinaustralia.gov.au/technical_report.php. Chap. 5, p 53.

CSIRO (2012) Research into wine supply networks: maturation forecasting. http://www.csiro.au/Outcomes/Food-and-Agriculture/wine-supply-networks/Maturation-forecasting.aspx

Degaris K (2004) Measuring Baume, Victoria Department of Primary Industries. http://www.crcv.com.au/resources/Grape%20and%20Wine%20Quality/Workshop%20Notes/Measuring%20Baume.pdf

Dunstall S, Owens B (2008) ASN Software for Orlando Wines—Vintage 2008. CMIS Report 07/165, CSIRO Mathematics, Informatics and Statistics, June 2008.

Dunstall S, Owens B (2009) ASN Viticulture Software—Vintage 2009. CMIS Report 09/45, CSIRO Mathematics, Informatics and Statistics, Apr 2009.

Dunstall S, Marquez L, Robinson G (2009) Maturation forecasting using meteorological data. CMIS Report 08/148, CSIRO Mathematics, Informatics and Statistics, July 2009.

Egan J (2012) Perfect storm led to Black Saturday bushfires, Australian Geographic, 9 Aug 2012. http://www.australiangeographic.com.au/journal/perfect-storm-of-weather-led-to-black-saturday-bushfires.htm

Hashim-Buckey J (2006) Grapes suffering from heat wave. Western Farm Press, 28 July 2006. http://westernfarmpress.com/grapes-suffering-heat-wave

Hayman P, Longbottom M, McCarthy M, Tomas D (2012) Managing grapevines during heatwaves, Grape and Wine Research Development Corporation. http://www.gwrdc.com.au/webdata/resources/files/Manging_grapevines_during_heatwaves.pdf

IPCC (2007) Climate change 2007: the physical science basis, summary for policymakers. Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Jackson DI, Lombard PB (1993) Environmental and management practices affecting grape composition and wine quality—a review. Am J Enol Vitic 44:409–430

Marquez L, Robinson G, Dunstall S (2009a) An analysis of grape maturation data for Pernod Ricard NZ. CMIS Report 08/146, CSIRO Mathematics, Informatics and Statistics, June 2009

Marquez L, Robinson G, Dunstall S (2009b) An analysis of grape maturation data for Orlando Wines. CMIS Report 08/147, CSIRO Mathematics, Informatics and Statistics, June 2009

Webb L (2006) The impact of projected greenhouse gas-induced climate change on the Australian wine industry. PhD Dissertation, University of Melbourne, Australia

Webb L, Whetton P, Barlow E (2008) Modelling the relationship between climate, winegrape price and winegrape quality in Australia. Clim Res 36:89–98

Webb L, Watt A, Hill T, Whiting J, Wigg F, Dunn G, Needs S, Barlow S (2009) Extreme heat: managing grapevine response. Report to Grape and Wine Research Development Corporation (GWRDC), University of Melbourne. http://www.landfood.unimelb.edu.au/vitum/Heatwave.pdf

White MA, Diffenbaugh NS, Jones GV, Pal JS, Giorgi F (2006) Extreme heat reduces and shifts United States premium wine production in the 21st century. Proc Natl Acad Sci USA 103:11217–11222

Chapter 14
Technical Efficiency of Sow Farms: A Parametric and Non-parametric Approach

Xavier Ezcurra and Lluís M. Plà-Aragonés

14.1 Introduction

The importance of pig production for the Spanish economy is reflected by recent agricultural statistics for Spain's swine industry. The sector contributes 15 % of the Final Agricultural Product and accounts for 35 % of the total economic value of the country's livestock production. Pork is the main meat consumed in Spain (60 kg/person/year); 55 % of total meat consumption. Moreover, after Germany, Spain is the second largest pig producer in the European Union (EU).

Due to recent EU regulation of pig farms and continuous growth of the census, there has been increasing concern about the measurement and comparison of the technical efficiency of different Spanish sow farms. Vertical integration is more and more common in the sector, concentrating production in few hands. Private companies and cooperatives play the role of the so-called integrators (Rodriguez et al. 2014). This integration leads to base production on different farms owned by the same integrator. Hence, identifying the best practices among farms to increase technical efficiency is crucial for either farmers or integrators. The future of swine producers, integrated or not, will depend on their ability to enhance their economic performance by improving productive efficiency rather than increasing farm size. The current literature on livestock production contains several studies of

X. Ezcurra (✉)
Department of Mathematics, University of Lleida, Lleida, Spain
e-mail: xezcurra@matematica.udl.es

L.M. Plà-Aragonés
Department of Mathematics, University of Lleida, Lleida, Spain

Agrotecnio Research Center, University of Lleida, Lleida, Spain
e-mail: lmpla@matematica.udl.es

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_14

the efficiency of dairy farms (Cloutier and Rowley 1993; Reinhard et al. 1999; Jaforullah and Whiteman 1999; Hansson and Öhlmér 2008), sheep farms (Gaspar et al. 2008; Ripoll-Bosch et al. 2012; Theodoridis et al. 2012), and extensive livestock farming (Gaspar et al. 2009), but fewer for sow farms (Galanopoulos et al. 2006). Moreover, hardly any economic studies have been undertaken on Spanish swine farms, which is strange given the importance of the sector in Spain.

Pig farming in Spain could be divided into three different phases according to final product and different economic activities. The first one relates to farms producing piglets (FPP), the second one to producing feeder pigs, and the third one to producing fattened pigs. Integrators own more than one sow farm and also several rearing and fattening farms. However, it is common to host the second phase in a sow farm generating two types of sow farms: those producing piglets or producing feeder pigs. Less and less common are the farrowing-to-finish farms embracing all the phases. The foundation of the economic activity relies on good herd management practices in sow farms which are much more complicated compared to the management of the other pig farms (Rodriguez et al. 2014). In this context it is reasonable that companies owning several sow farms are wondering about the efficiency of their farms and detecting the ones more efficient to be taken as a reference. Hence, for the purposes of this study, we consider a sow farm to be a farm that houses sows and that produce as output either weaned piglets or feeder pigs. Inputs include reproductive sows, concentrates and labour, etc. Different farms tend to organise their operations in different ways, so consequently values for individual outputs will also tend to differ, even if they are integrated under the same company. There is, therefore, a special interest in comparing different sow farms and highlighting efficient practices, in order to identify a best-practice sow farm group. This group of farms could then be used as a point of reference for less efficient units and for benchmarking performance. As observed by Weersink et al. (1990), identifying possibilities for improving efficiency should help to enhance the profitability of farms and make the pig industry more competitive. The existence of an official record keeping system (the BD-porc® 2013), which registers the main controllable variables on a Spanish farm, allows us to select the inputs and outputs registered by farm basis to calculate efficiency and perform subsequent improvements.

The simplest way of measuring technical efficiency, the pure relationship between input and output as such, is often inadequate due to the existence of multiple inputs and outputs relating to different resources, activities, and environmental factors. A variety of techniques have been proposed to study the efficiency overcoming this inconvenience. For instance, the measurement of relative efficiency where there are multiple possibly incommensurate inputs and outputs was early addressed by Farrell (1957) and developed by others in the 1960s and early 1970s. The method is focusing on the construction of a hypothetical efficiency frontier of a firm to compute efficiency measures relative to this reference firm. Most of the papers related to the measurement of productive efficiency have based their analysis either on parametric or non-parametric methods. The choice of estimation method has been an issue of debate, and some researchers prefer parametric approach (e.g. Berger 1993) and other the non-parametric approach (Banker et al. 2004). Parametric frontier functions require the definition of a specific functional form for the function of production, meanwhile DEA does not distinguish between technical efficiency and statistical noise effects avoiding the need to assume functional relationship between inputs and outputs. The aim of this chapter is to analyse the technical efficiency of Spanish sow farms comparing parametric and non-parametric approaches. In addition, several technical indexes used regularly for sow herd management will be explored as explanatory variables for efficiency scores. Therefore, the present chapter is structured as follows; in the next section an overview of both approaches is presented. Sow farm data used in this analysis are presented in Sect. 14.3. This is followed by some results and conclusions, in Sects. 14.4 and 14.5, respectively. Finally, the chapter concludes with a brief outlook of the subject in Sect. 14.6.

14.2 Methodology

14.2.1 Parametric Approach

The parametric approach requires the definition of a specific functional form for the technology and for the inefficiency error term, using mathematical programming or econometric techniques. It can be subdivided into deterministic and stochastic models. Deterministic models envelope all the observations, identifying the distance between the observed production and the maximum functions, defined by the frontier and the available technology, as technical inefficiency. On the other hand, stochastic approaches allow distinguishing between technical efficiency and statistical noise.

Farrell (1957) suggested the use of functional forms in the estimation of production functions. Aigner and Chu (1968) were the first ones to estimate a parametric frontier, adjusting a Cobb–Douglass function and imposing the non-negativity of the error terms. The model was:

$$Y_i = \alpha + \sum_{j=1}^{r} \beta_j X_{j,i} + V_i - U_i \tag{14.1}$$

where $i = 1, \ldots, N$ indicates the units and $j = 1, \ldots, r$ indicates de inputs, Y_i is output of the ith firm, $X_{j,i}$ are productive factors used by the ith firm, β is a vector of parameters to be estimated, and $V_i - U_i$ is the composed error term where V_i represents randomness (or statistical noise) and U_i represents technical efficiency. V_i are assumed to be independently and identically distributed $N(0, \sigma_i^2)$ random errors, independent of U_i, and U_i are non-negative random variables associated with technical inefficiency production, which are assumed to be independent and identically distributes and truncates (at zero) the normal distribution with mean, μ and variance σ_u^2.

It allows the definition of the likelihood functions and it gets estimators for β and variance parameters, $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2/\sigma^2$. Subtracting V_i from both sides of (14.1) yields

$$\widetilde{Y}_i = Y_i - V_i = \alpha + \sum_{k=1}^{r} \beta_k X_{k,i} - U_i \tag{14.2}$$

where $\widetilde{Y}_i$ is the observed output of the ith firm adjusted for the stochastic noise captured. For a given level of output $\widetilde{Y}_i$, the technically efficiency input vector for the ith firm, X_i^t is derived by simultaneously solving (14.2) and the input ratios $X_1/X_i = K_i$ ($i > 1$), where K_i *is* the ratio-observed inputs X_1 and X_i.

The measures of technical efficiency relative to the production frontier are defined as:

$$\mathrm{EFF}_i = E\left(Y_i^* \middle| U_i, X_i\right) / E\left(Y_i^* \middle| U_i = 0, X_i\right) \tag{14.3}$$

where Y_i^* is the production of the ith firm, which will be equal to Y_i when the dependent variable is in original units and will be equal to $\exp(Y_i)$ when the dependent variable is in logs. EFF_i will take a value between zero and one. The efficiency measures can be shown to be defined as (Jondrow et al. 1982; Battese and Coelli 1988):

Logged dependent variable	Efficiency (EFF_i)
Yes	$\exp(-U_i)$
No	$(x_i\beta - U_i)/x_i\beta$

14.2.2 Non-parametric Approach

Non-parametric approach doesn't require the specification of any particular functional form to describe the efficient frontier. In these circumstances, suppose that we have observations of n farms, each one transforming inputs into m outputs, efficiency of a target farm j can be expressed as:

$$\mathrm{rE}_j = \frac{a_1 y_{1j} + a_2 y_{2j} + \cdots + a_m y_{mj}}{u_1 x_{1j} + u_2 x_{2j} + \cdots + u_s x_{sj}} = \frac{A^T \cdot y_j}{U^T \cdot x_j} \tag{14.4}$$

where

rE_j is the relative efficiency for farm j
a_i is the weight given to output i, $i = 1, 2, \ldots, m$
y_{ij} is the amount of output i from farm j
u_i is the weight of input i from farm j, $i = 1, 2, \ldots, s$
x_{ij} is the amount of input i from farm j

The initial assumption is that this measure of efficiency requires a common set of weights to be applied across all sow farms. This immediately raises the problem of how such an agreed common set of weights can be obtained. It could be possible that a farm might value inputs and outputs differently and therefore adopt different weights, and consequently each farm should be allowed to adopt a set of weights which shows it in the most favourable light in comparison to the other farms. In that case relative efficiency of farm j respect to the set of k farms can be obtained by solving the following model:

$$\begin{aligned} &\text{maximise} \quad && \mathrm{rE}_j = \frac{A^T \cdot y_j}{U^T \cdot x_j} \\ &\text{subject to}: && \frac{A^T \cdot y_k}{U^T \cdot x_k} \leq 1; \quad k = 1, \ldots, n \end{aligned} \tag{14.5}$$

where k represents the total number of farms involved in the analysis and the weights, a's and u's components of the vectors A and U, are treated as the decision variables of the problem. They could be constrained to be greater than or equal to some small positive quantity in order to avoid any input or output being totally ignored in determining the efficiency. The solution produces the weights most favourable to farm j and also produces a measure of efficiency, rE_j. If $\mathrm{rE}_j = 1$ then farm j is efficient relative to the others, but if rE_j turns out to be less than 1 then some other farm is more efficient than farm j, even when the weights are chosen to maximise efficiency of farm j. These farms constitute the peer group for farm j. A peer group is a group of efficient farms that act as a reference for an inefficient one. Thus, an inefficient farm can identify and eliminate their less efficient practices by comparing to its peer group.

The model presented is a fractional linear program with infinite solutions when there exist. To solve the model, it is first necessary to convert it into an equivalent linear form as Charnes et al. (1978) proposed. They were the first to develop the DEA approach based on the concept of technical efficiency of Farrell (1957). Hence, DEA is a linear programming technique that converts multiple inputs and outputs into a scalar measure of efficiency and it is extensively used in Economics and Operations Research (Seiford 1996). The transformation of (14.5) into a linear model provides (Charnes et al. 1978):

$$\begin{aligned} &\text{maximise} \quad && A^T \cdot y_j \\ &\text{subject to}: && A^T \cdot y_j - U^T \cdot x_k \leq 0; \quad k = 1, \ldots, n \\ & && U^T \cdot x_k = 1 \\ & && A, U \geq 0 \end{aligned} \tag{14.6}$$

About the linear transformation applied in (14.5), we can remark that in maximising a fraction or ratio it is the relative magnitude of the numerator and denominator that are of interest and not their individual values. It is thus possible to achieve the same effect by setting the denominator equal to a constant and maximising the numerator.

For linear programs in general the more constraints the more difficult a problem is to solve. Hence, the dual DEA model involves fewer constraints and uses to be simpler than primal and it is usual to solve it rather than the primal. Following mathematical formulation corresponds to the dual model of the linear version of (14.6). Let η_j be the output-oriented efficiency associated to farm j. Let $Y = (y_j)$ be an $(m \times n)$ matrix of outputs for n Spanish sow farms with y_j representing the $(m \times 1)$ vector of outputs for the jth farm. Let $X = (x_j)$ be an $(s \times n)$ matrix of inputs with x_j representing the $(s \times 1)$ vector of inputs for the jth farm and μ an $(n \times 1)$ vector of weights to be defined. The linear version of the model is as follows:

$$\begin{aligned} &\eta_j = \max_{\eta,\mu} \begin{pmatrix} 1 & 0_n^T \end{pmatrix} \begin{pmatrix} \eta \\ \mu \end{pmatrix} \\ &\text{subject to}: \begin{pmatrix} 0_m & -X \\ -y_j & Y \end{pmatrix} \begin{pmatrix} \eta \\ \mu \end{pmatrix} \geq \begin{pmatrix} -x_j \\ 0_s \end{pmatrix} \\ &\eta, \mu \geq 0 \end{aligned} \tag{14.7}$$

which assumes the existence of constant returns to scale (CRS). This assumption of the original model may be relaxed following Banker et al. (1984) by adding any of the constraints $\sum \mu_i = 1$ for variable returns to scale (VRS) or $\sum \mu_i \leq 1$ for non-decreasing returns to scale (NDRS) (Banker et al. 1984; Färe et al. 1985; Lovell 1994). Apart from output-oriented relative technical efficiency measure defined as in (14.7), input-oriented measures can be also obtained:

$$\begin{aligned} &\theta_j = \min_{\theta,\lambda} \begin{pmatrix} 1 & 0_n^T \end{pmatrix} \begin{pmatrix} \theta \\ \lambda \end{pmatrix} \\ &\text{subject to}: \begin{pmatrix} 0_s & Y \\ x_j & -X \end{pmatrix} \begin{pmatrix} \theta \\ \lambda \end{pmatrix} \geq \begin{pmatrix} y_j \\ 0_m \end{pmatrix} \\ &\theta, \lambda \geq 0 \end{aligned} \tag{14.8}$$

where θ_j represents the input-oriented efficiency associated to farm j and (14.8) assumes the existence of CRS. This assumption of the original model may be relaxed like in (14.7) by adding any of the constraints $\sum \mu_i = 1$ for VRS or $\sum \mu_i \leq 1$ for non-increasing returns to scale (NIRS). Input-oriented measures of inefficiency measure the potential reduction in inputs holding outputs constant. Alternatively, output-oriented measures of inefficiency measure the potential increase of outputs, holding inputs constant. We understand inefficiency as the complementary to one of the efficiency. Efficiencies are usually expressed in percentage terms.

In this work we will focus on input-oriented measures, then the θ represents a proportional reduction in all inputs ($0 \leq \theta \leq 1$) and θ_j is the minimum value of θ for farm j. Maximum value for θ is one and represents the farm operating at best-practice (given the existing set of observations). We will consider θ_j^c, θ_j^v and θ_j^n solutions for DEA models assuming CRS, VRS, and NIRS, respectively.

There are different methods of testing a farm's return to scale nature (Banker et al. 1984; Färe et al. 1985; Seiford and Zhu 1999). We will use the scale efficiency index method provided by Färe et al. (1985) because it is robust and simple. We assume no inefficiency due to input congestion, i.e. farms are subject to strong input disposability. The scale efficiency index measure for farm j can be calculated as:

$$S_j = \theta_j^c / \theta_j^v \tag{14.9}$$

If the value of the ratio is equal to unity, then farm j is scale-efficient. This means that the farm is operating at its optimum size and hence the productivity of inputs cannot be improved by increasing or decreasing the size of the sow farm. The VRS model ensures that a farm is only compared to other farms of a similar size (Fraser and Cordina 1999).

If not and $\theta_j^c = \theta_j^n$, then the results suggest that scale inefficiency is due to increasing returns to scale. This means that the farmer can improve the productivity of inputs by increasing the farm size. Or when $\theta_j^c < \theta_j^n$, the results suggest that scale inefficiency is due to decreasing returns to scale. This means that the farm is bigger than its optimum size.

14.2.3 Explanatory Variables

Following the two-step approach (Coelli 1998), different explanatory variables were proposed to estimate inefficiency scores. First estimates of relative efficiencies using the inputs and outputs are calculated. Second the effect of different variables on efficiency is analysed. Apart from all the inputs and output, other exogenous variables as piglet mortality, culling rate, litter size, piglets alive, and farrowings per sow per year were considered. Since the inefficiency scores are censored, values between zero and one, a Tobit model is proposed:

$$\mathrm{Ineff}_k^* = \alpha + \beta z_k + \varepsilon_k$$

$$\mathrm{Ineff}_k = \begin{cases} \mathrm{Ineff}_k^* & \text{if } \mathrm{Ineff}_k^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where Ineff_k^* represents the latent variable related to the inefficiency scores and is a dependent variable not censored. Ineff_k is the censored variable defined by the DEA efficiency scores; z is a vector of independent explanatory variables related to the k-farm, α is the constant term; β is the vector of parameters to be estimated and ε is the statistical noise, normally distributed with mean zero.

14.3 Sow Farm Data

An initial sample of 193 sow farms from the north-east of Spain was considered. It is the main pig-producing area of Spain (around 2 % of the national surface area concentrates 15 % of total national production). The farms belonged to the same pig supply chain and data is recorded in the main swine data bank of Spain (BD-porc®), which is promoting a new extension program to encourage pig industry economists to complement economic analysis with efficiency studies. The farms were classified into two groups according to their final product: piglets or feeder pig. Homogeneity was considered from the perspective of both sow farms (belonging to the same company) and the common environment. This meant that observed differences in technical efficiency would be the result of managerial ability. A filtering process was performed, and several farms were rejected because of problems associated with previous healthcare problems (e.g. classical swine fever and Aujersky disease), different production systems, geographical situation, and recent initiation in the activity or outliers detected by statistical analysis. If a farm reported unreasonable values, or values more than two standard deviations from the mean, it was eliminated from the data set. Hadi's (1992, 1994) method for identifying and removing multiple outliers was also used. Hence, only 96 farms from the initial sample were finally considered. The reasons for rejecting the other 97 farms were: incomplete economic data (64), inconsistencies in data (9) and outliers (24). The farms used in the study included 45 producing piglets (average weight 5.8 kg per piglet sold) and 51 producing feeder pigs (average weight 18.7 kg per feeder pig sold), respectively. Data relating to these farms are shown in Tables 14.1 and 14.2. The period analysed was 1st January to 31st December 2006.

The choice of variables was constrained by the availability of data registered with the BD-porc® data-bank, economic data provided by the company, and

Table 14.1 Summary of variables for DMUs producing FPP

Variables	Mean	SD	Median	Minimum	Maximum
Output (t) {O}	63.96	28.65	56.42	23.03	141.61
Labour (#) {I}	2.374	1.10	2.12	0.77	5.23
# Sows {I}	520.99	242.24	420.26	174.15	1,131.60
Feed (t) {I}	554.68	242.98	492.42	161.96	1,189.45
Veterinary (€) {I}	397.44	221.66	308.60	127.12	981.12
Expenses (€) {I}	10,657	4,874	9,290	3,349	22,501
FeedWP (kg) {I}	108.90	71.95	88.81	20.72	292.65
AI (#) {I}	3,911	1,755	3,318	1,609	8,344

Table 14.2 Summary of variables for DMUs producing feeder pigs

Variables	Mean	SD	Median	Minimum	Maximum
Output (t) {O}	109.71	63.23	89.11	31.81	321.65
Labour (#) {I}	1.36	0.79	1.12	0.36	3.95
# Sows {I}	301.43	176.23	248.58	81.08	885.58
Feed (t) {I}	325.01	190.92	264.18	79.56	923.601
Veterinary (€) {I}	501.70	274.98	444.93	141.93	1,433.09
Expenses (€) {I}	9,463	5,466	7,494	2,662	27,825
FeedS (kg) {I}	5,830	3,575	4,575	1,352	18,497
FeedWP (kg) {I}	588.71	345.19	493.22	0.00	1,627.14
AI (#) {I}	2,262	1,466	1,812	361	7,950

protocols suggested by Dyson et al. (2001) for avoiding pitfalls in the use of DEA. It was assumed that sow farms produced one output: weaned piglets or feeder pigs. The two outputs are different in age and weight. Weaned piglets are from 3 to 4 weeks old with a weight of around 7 kg, while feeder pigs are from 4 to 6 weeks older and weighting between 15 and 20 kg. Depending on the activity, seven or eight inputs were considered: labour working in the farm, both salaried and family workers, average number of breeding sows, feed consumed by sows, veterinary expenses, other expenses (water, fuel, electricity, repairs, etc.), feed consumed by piglets (FeedWP) and/or feeder pigs (FeedS), and number of inseminations (AI). It was difficult to measure labour because the registered data was not introduced in the same way for all farms. Labour was therefore expressed in terms of equivalent workers (1,920 h/year). Tables 14.1 and 14.2 summarise statistics on inputs and outputs for each group of farms considered. Sow FPP produced more units, i.e. piglets, than farms producing feeder pigs (FPFP). The size of the farm, in terms of its number of sows, was also bigger. This seems logical considering both the shorter productive cycle for FPP and the greater value per unit produced by the second as opposed to the first group. Data presented in Tables 14.1 and 14.2 make it possible to calculate several sow-related ratios. Some of these ratios are similar for both groups of farms, for instance kg of feed consumed per sow (1,065 kg for FPP and 1,080 kg for FPFP) and the number of inseminations per sow (7.51 for FPP and 7.52 for FPFP).

However, other ratios calculated from Tables 14.1 and 14.2 revealed differences between the two groups and were more useful for characterising piglet and feeder pig production. For instance, the output produced for each type of farm (112 kg vs. 364 kg), feed consumed by suckling piglets (109 kg vs. 589 kg), or the veterinary expenses per sow or per unit produced (0.76€ vs. 1.67€) were also greater for producers of feeder pigs than piglets. All of these reveal that the longer productive cycle and lifespan of feeder pigs imply an increase in both feed consumption and veterinary expenses by young pigs respect to farms producing only piglets. In addition, the average size of farms (571 vs. 301) was different, being farms producing piglets bigger than those producing feeder pigs.

14.4 Empirical Results and Discussion

14.4.1 Parametric Approach

The maximum likelihood estimated of the parameters of the stochastic production frontier was obtained for feeder pigs and weaned piglets using the program, FRONTIER 4.1 (2013). The stochastic frontier production for the cross-section of feeder pigs and weaned piglets is specified as follows:

For weaned piglets:

$$\begin{aligned}\ln \text{Output}_i = \beta_0 + \beta_1 \ln \text{Labour}_i + \beta_2 \ln \text{Sows}_i + \beta_3 \ln \text{Feed}_i \\ + \beta_4 \ln \text{Veterinary}_i + \beta_5 \ln \text{Expenses}_i + \beta_6 \ln \text{FeedWP}_i + \beta_7 \ln \text{AI}_i \\ + V_i - U_i\end{aligned}$$

For feeder pigs:

$$\begin{aligned}\ln \text{Output}_i = \beta_0 + \beta_1 \ln \text{Labour}_i + \beta_2 \ln \text{Sows}_i + \beta_3 \ln \text{Feed}_i \\ + \beta_4 \ln \text{Veterinary}_i + \beta_5 \ln \text{Expense}s_i + \beta_6 \ln \text{FeedWP}_i + \beta_7 \ln \text{AI}_i \\ + \beta_8 \ln \text{Feed}S_i + V_i - U_i\end{aligned}$$

where i refers to the ith DMU in the sample; and V_i and U_i are the random variables as defined in Sect. 14.2. The variance parameters were estimated in terms of $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\gamma = \sigma_u^2/\sigma^2$.

A Cobb–Douglas production function with the forward-selection technique and the stepwise method using PROC REG of SAS was estimated for each group of farms (Table 14.3). The estimated regression coefficients for input variables were

Table 14.3 Estimated Cobb–Douglas production frontiers

	FPP		FPFP	
Estimates	Value	SE	Value	SE
β_0	0.53	1.00	4.46	2.27
β_1	–	–	–	–
β_2	0.13	1.00	0.58	0.36
β_3	−0.49	1.00	0.18	0.32
β_4	–	–	0.18	0.55
β_5	0.20	1.0	–	–
β_6	–	–	–	–
β_7	−0.10	1.0	–	–
β_8	–	–	–	–
Σ	0.0064	1.0	0.0360	0.0100
Γ	0.0500	1.0	0.9190**	0.0860
Log (likelihood)	50.189**		37.4630**	

**Significant at the 5 % level

Table 14.4 Frequency distributions of technical efficiency estimates from the stochastic frontier method

Efficiency score	FPP	Efficiency score	FPFP
Mean	0.9858	Mean	0.8720
Minimum	0.9820	Minimum	0.5970
Maximum	0.9890	Maximum	0.9745
Standard deviation	0.0015	Standard deviation	0.0797

different depending on the activity analysed and not significant (at the 5 % level) in any case. However, parameter γ was not significantly different from zero for FPP (i.e. the inefficiency effects are not significant in determining the level and variability of the output) and significant for FPFP (at the 5 % level).

The parameters β_i represent the elasticity of output with respect to each input i. For instance, those with the greatest elasticity were number of sows (FPP) and feed consumption of sows (FPFP). These results are meaningful because both variables are important components in the production cost of piglets. However, the signs of the slope coefficients in FPP had different signs, positive and negative mixed. In particular, the sign of the slope coefficients of β_3 (Feed) was not consistent: negative for FPP and positive for FPFP. Finally, the Log (likelihood) was significant at the 5 % level in both groups of farms and so there is a significant relationship between the dependent variable and the set of independent variables. These results revealed the existence of different production functions for each group of sow farms.

Once the production function was estimated, the technical efficiency for each farm was calculated. Some statistics of the estimated technical efficiencies are presented in Table 14.4. The mean technical efficiency estimated was 0.98577 and 0.87198 for weaned piglets and feeder pigs, respectively. The main implication of these results is that the set of farms analysed are very efficient, being on average FPP more efficient than FPFP. Surprisingly, the efficiency of FPP is very high, with a capability of less than a 2 % of technical efficiency improvement and with a low standard deviation. The fact that FPFP are less efficient than FPP can be interpreted as production in FPP is more complicated, controlled, and delicate than in FPFP. A reason for that is the longer production cycle deployed in FPFP, introducing variability in the final output as revealed the higher coefficient of variability (0.45 vs. 0.58) in FPFP. On the other hand, this situation may partially explain the bad estimates of the stochastic production frontier for FPP and suspecting of problems related to multicolinearity.

14.4.2 Non-parametric Approach

As stated before, the solution to the DEA model (14.8) provides a measure of the relative efficiency of each DMU and the weights leading to the efficiency. To solve the different DEA models, the DEAP software was utilised (Coelli 1996). The CRS, VRS, NIRS, and scale (Scale) input-oriented DEA frontiers were estimated for the same number of farms as for the stochastic frontier depending on the final product: total weight of weaned piglets or feeder pigs. This meant solving three linear

Table 14.5 Statistics of DEA models

	FPP			FPFP		
	Mean	SD	Minimum	Mean	SD	Minimum
CRS	0.8903	0.0870	0.7143	0.88681	0.10793	0.5355
VRS	0.93320	0.0744	0.7473	0.91066	0.10595	0.5370
NIRS	0.9009	0.0858	0.7143	0.89934	0.10832	0.5370
Scale	0.9546	0.0590	0.7143	0.97427	0.04103	0.7687

Table 14.6 Summary of relative efficiency (CRS and VRS) for FPP

	CRS		VRS	
Variables	Efficients	Inefficients	Efficients	Inefficients
Mean	1.00	0.86	1.00	0.89
% farms	17.77 %	82.23 %	40.00 %	60.00 %
Labour{I}	2.41	2.36	2.35	2.38
Sows{I}	535.64	517.82	520.20	521.52
Feed{I}	565.67	552.31	543.58	562.08
Veterinary{I}	347.79	408.17	370.65	415.30
Expenses{I}	10,473	21,631	10,366	10,850
FeedWP{I}	82.31	114.65	86.41	123.90
AI{I}	3,610	4,976	3,671	4,072
Output{O}	71.17	82.32	66.23	62.44

mathematical programs for every farm. The scale efficiency for every farm was obtained by (14.9). The summarised statistics for the four estimated measures for the two groups of farms are presented in Table 14.5.

In both cases the average efficiency index is very high showing a strong market competition in agreement with the trend observed from the parametric approach similar to other European studies. For instance, the average efficiency under CRS and VRS is very similar to the figures reported by Lansink and Reinhard (2004) in the Netherlands, but slightly higher than those reported by Galanopoulos et al. (2006) in Greece or Sharma et al. (1999) in Hawaii. However, the stochastic frontier and DEA approach showed similar values, though with lower minimum values in the DEA approach which derived in different standard deviations. As the DEA approach is not stochastic, it interprets noise as inefficiency and so we can consider the different estimates consistent. This comparison agrees with the findings of Sharma et al. (1999) who obtained similar conclusions from both approaches. Furthermore, Banker et al. (2004) considered DEA-based estimator of efficient input better than stochastic-based ones even under heteroscedasticity. The mean efficiencies for the VRS, CRS, NIRS, and Scale DEA frontiers range from 0.88681 to 0.97427. Thus, the DEA analyses reveal substantial productive efficiency in all the cases.

Summarising apart the number of relative efficient and inefficient farms (Tables 14.6 and 14.7), we observe the mean technical inefficiency is quite high

Table 14.7 Summary of relative efficiency (CRS and VRS) for FPFP

	CRS		VRS	
Variables	Efficients	Inefficients	Efficients	Inefficients
Mean	1.00	0.85	1.00	0.86
% farms	25.49 %	74.51 %	35.29 %	64.71 %
Labour{I}	1.27	1.40	1.51	1.29
Sows{I}	280.60	308.56	333.31	284.05
Feed{I}	309.11	330.44	359.51	306.19
Veterinary{I}	400.41	536.34	484.13	511.28
Expenses{I}	8,589	9,762	10,284	9,016
FeedS{I}	5,440	5,963	6,573	5,425
FeedWP{I}	459.99	632.75	568.22	599.89
AI{I}	2,005	2,350	2,409	2,182
Output{O}	114.26	108.15	129.77	98.77

in both FPP (0.86 and 0.89) and FPFP (0.85 and 0.86) under CRS and VRS assumptions. The percentage of efficient DMUs producing feeder pigs under the CRS (25.49 %) is greater than the percentage in producing weaned piglets (17.77 %). In terms of the VRS model the percentage of efficient DMUs producing weaned piglets (40.00 %) is greater than the percentage in feeder pigs (35.29 %). For comparison reasons, we have also included in Tables 14.6 and 14.7 the average inputs and output of efficient and inefficient farms under CRS and VRS. The comparison of efficient and inefficient farms within each group has no clear outcome. It seems reasonable that the results of inefficient FPP under VRS and FPFP under CRS with more inputs produce less output with respect to efficient farms. More difficult to explain is the behaviour of the other inefficiencies like FPP under CRS where a reduction in several inputs leads to an increment in output, but maybe the noticeable increment of expenses (from 10,474 to 21,631) in this group of farms compromises the efficiency. On the other hand, inefficient FPFP under VRS exhibit less input in general (only FeedWP increases) and less output than corresponding efficient FPFP under VRS.

Table 14.8 presents the scale efficiency scores complementing the mean scaled efficiency showed in Table 14.5 (0.95 for FPP and 0.97 for FPFP). Although previous mean values implied that the average size of farms is not far from the optimal size, most of the farms are characterised by increasing returns to scale (58 % of FPP farms and 45 % of FPFP). According to the efficiency analysis theory, these farms are small farms and efficiency gains would be expected by increasing the size and achieving cost savings, assuming no other constraining factor. Again, the variability in FPFP is higher than those in FPP as shown in Table 14.8. Lansink and Reinhard (2004) reported similar scale efficiency for pig farms in the Netherlands, while Galanopoulos et al. (2006) presented lower scores for Greek farms.

Table 14.8 Optimal, sub-optimal, and super-optimal distribution of DMUs producing weaned piglets and feeder pigs

Sow farm	Scale efficiency	*N*	%	Output		
				Mean	SD	CV
FPP	Sub-optimal	26	57.78	46.14	12.95	0.28
	Supra-optimal	11	24.44	100.83	23.00	0.23
	Optimal	8	17.78	71.17	17.04	0.24
FPFP	Sub-optimal	23	45.09	74.77	23.75	0.32
	Supra-optimal	14	27.45	165.60	69.68	0.42
	Optimal	14	27.45	111.21	58.00	0.52

Table 14.9 Technical efficiency for FPP by output

	Output			
	<44.00	44.00–56.40	56.41–78.70	>78.70
Number DMUs	11	11	12	11
Labour{I}	1.32	1.74	2.42	3.99
Sows{I}	285.10	386.68	531.95	879.23
Feed{I}	307.57	414.37	583.76	910.38
Veterinary{I}	258.86	262.24	404.07	663.97
Expenses{I}	6,008	7,743	11,144	17,689
FeedWP{I}	76.41	113.54	110.07	135.48
AI{I}	2,329	2,828	4,127	6,342
Scale (mean)	0.89	0.98	0.99	0.96

Table 14.10 Technical efficiency for FPFP by Output

	Output			
	<67.35	67.35–89.11	89.12–138.47	>138.47
Number DMUs	14	12	12	13
Labour{I}	0.67	0.99	1.40	2.43
Sows{I}	148.26	217.33	305.98	539.83
Feed{I}	156.24	230.64	337.57	582.27
Veterinary{I}	275.33	400.16	598.68	749.68
Expenses{I}	4,767	6,721	9,565	16,957
FeedWP{I}	303.02	455.12	579.31	1,028.38
FeedS(I)	2,843	4,211	5,570	10,781
AI{I}	998	1,653	2,215	4,232
Scale (mean)	0.97	0.99	0.99	0.96

To discuss further the possible link between efficiency and input variables depending on the production level of the DMU, and also to draw an approach to the relationship between efficiency and optimal dimension, we have considered four groups of DMUs by production level (i.e. output). In Tables 14.9 and 14.10 we summarise these results by group. Differences in efficiency by output do not seem very important in any of both groups of DMUs. The results are similar to those reported by Sharma et al. (1999). Medium-size farms (532 sows in FPP and 306 sows in FPFP) were more scale-efficient, but with small differences. Small FPP are the least scale-efficient showing a remarkable difference according to the rest of scores, so we can assert the size of a farm plays a more important role in FPP than in FPFP regarding the scale efficiency. Most of the DMUs are characterised by increasing returns to scale. However, the results differ with respect to returns to scales properties with Sharma et al. (1997).

As Galanopoulos et al. (2006) recognise, the DEA analysis can neither fully explain the underlying differences in efficiencies in the use of a particular input, nor assess the constraints to changes in operational practices that would improve efficiency. In part this is why we considered in the next section additional explanatory variables to explain variations in efficiency scores and identifying places to make improvements in pig production systems.

14.4.3 Explanatory Variables

To explain some variations the inefficiency scores were regressed on the DMU-level characteristics, using a Tobit model, since the inefficiencies vary from zero to unity. The objective is to identify the common features in the most efficient farms. Authors as Hansson and Öhlmér (2008) had used the same approach to investigate how operational managerial practices can contribute to improved dairy farm efficiency. Apart from the input–output variables already considered, additional variables selected by livestock experts from Bdporc databank were included in the analysis to explain variations in efficiency scores. These five exogenous variables are: piglet mortality, culling rate, litter size, piglets alive, and farrowings per sow per year. Some other different explanatory variables related to Greek managerial practices had been considered by Galanopoulos et al. (2006) in a similar analysis. The inefficiency scores were regressed on these 14 variables (inputs + output + exogenous) and are presented in Tables 14.11 and 14.12. For the results presented, the independent variable is the inefficiency score, so a positive (negative) sign of a coefficient reflects a negative (positive) effect on efficiency levels. Recall that the estimated coefficients in Tobit regression models do not have a direct interpretation as a true marginal effect, but rather a two-scale effect: effect on the mean of the dependent variable and on the probability of the dependent variable being observed.

The Tobit results for FPP (Table 14.11) and FPFP (Table 14.12) indicate that no exogenous variable related to technical indexes considered in the analysis was significant explaining the inefficiencies of farms. Only the output, the size of the farm (Sows), the feed consumption of sows (Feed), suckling piglets (FeedWP), and the number of artificial inseminations (AI) are significant. As expected, Output has negative effects on inefficiency scores, while Sows has positive effects. In a similar study, Galanopoulos et al. (2006) found that several managerial practices such as

Table 14.11 Tobit model of FPP

CRS			VRS		
Variables	Estimate	SE	Variables	Estimate	SE
Intercept	0.7675	0.2359	Intercept	−0.6335	0.3349
Output**	−0.0000	0.0000	Output**	−0.0000	0.0000
Labour	−0.0005	0.0004	Labour	−0.0009	0.0007
Sows**	0.0010	0.0005	Sows	0.0014	0.0007
Feed	0.0000	0.0000	Feed**	0.0000	0.0000
Veterinary	−0.0000	0.0001	Veterinary	−0.0001	0.0001
Expenses	0.0000	0.0000	Expenses	−0.0000	0.0000
FeedWP**	0.0002	0.0001	FeedWP**	0.0006	0.0001
AI**	0.0000	0.0000	AI**	0.0001	0.0000
FeedS	−0.0045	0.0093	FeedS	−0.0281	0.0145
Piglet mortality	0.0034	0.0032	Piglet mortality	−0.0043	0.0050
Culling rate	0.0006	0.0004	Culling rate	0.0008	0.0006
Litter size	−0.0113	0.0265	Litter size	0.0240	0.0400
Piglets alive	−0.0614	0.0332	Piglets alive	0.0036	0.0494
Farrowings pspy	0.0229	0.0552	Farrowings pspy	0.1495	0.0836

Estimated parameters for CRS and VRS inefficiencies (14 variables)
**Significant coefficients at the 5 % level

Table 14.12 Tobit model of FPFP

CRS			VRS		
Variables	Estimate	SE	Variables	Estimate	SE
Intercept	0.6297	0.2661	Intercept	0.3812	0.3594
Output**	−0.0000	0.0000	Output**	−0.0000	0.0000
Labour	−0.0004	0.0013	Labour	0.0012	0.0017
Sows	0.0024	0.0015	Sows	−0.0002	0.0020
Feed	−0.0000	0.0000	Feed	−0.0000	0.0000
Veterinary	0.0001	0.0001	Veterinary	0.0002	0.0001
Expenses	−0.0000	0.0000	Expenses	−0.0000	0.0000
FeedWP	0.0001	0.0001	FeedWP	0.0001	0.0001
AI**	0.0000	0.0000	AI**	0.0001	0.0000
FeedS**	0.0000	0.0000	FeedS	0.0000	0.0000
Piglet mortality	0.0014	0.0032	Piglet mortality	−0.0020	0.0046
Culling rate	−0.0008	0.0007	Culling rate	−0.0011	0.0010
Litter size	0.0492	0.0372	Litter size	−0.0088	0.0524
Piglets alive	−0.0739	0.0426	Piglets alive	−0.0123	0.0590
Farrowings pspy	−0.1305	0.0875	Farrowings pspy	0.0056	0.1188

Estimated parameters for CRS and VRS inefficiencies (14 variables)
**Significant coefficients at the 5 % level

insemination method, origin of genotype, and how the feed was prepared significantly influenced the technical efficiency of Greek pig farms. Although we didn't consider genotype because it was the same for all farms belonging to the same company and AI and Feed were considered as inputs, these variables showed similar significant influence on the efficiency of farms.

The results for FPFP (Table 14.12), only the output, the number of inseminations (AI), and the feed consumed by feeder pigs (FeedS), were significant. That is, less variables than for FPP. However, again, as expected, output had negative effects on inefficiency scores, while AI and FeedS had positive effects under CRS. Under VRS only output and AI were significant. Overall, in terms of signs of the regression coefficients, these results are quite consistent for both activities (FPP and FPFP). However, it seems a little bit strange to observe some negative signs pointing out interesting conclusions. For instance, under VRS the effect of piglet mortality is negative either in FPP or FPFP. From the analysis of these specific farms, it is not mortality itself which explains inefficiency, but a higher prolificity correlated to more efficient farms and hence more susceptible to suffering more casualties of piglets. Even though the different signs shown by litter size and piglets alive in FPP and FPFP suggest again the importance of litter size in FPP to produce many piglets, but perhaps, larger litter sizes are not so suitable for FPFP. Less piglets, but weightier, might be more interesting and profitable for FPFP. This argument can be reinforced observing the sign of the culling rate that also differs between FPP and FPFP. In FPP inefficiency is associated with higher culling rates, while in FPFP it is in the contrary. This agrees with the general idea of managers about high culling rates in FPP with younger sows, more productive in the number of piglets than older populations, putting the emphasis more on the quantity than the quality of piglets.

The fact that artificial insemination appeared as significant in all the analyses may suggest the importance of the reproductive performance in the technical efficiency of sow farms, either producing piglets or feeder pigs, regardless of the number of sow or the feed consumption of concentrates in many cases. The importance of artificial insemination as explanatory variable was already detected by Galanopoulos et al. (2006).

14.5 Conclusions

Pig farming is a biological activity with many uncertainties, so in view of efficiency measurements the choice of stochastic frontier analysis allowing for a correction of stochastic events would seem obvious. However, the parametric specification of the production technology can be problematic and may not always provide suitable results and are more difficult to interpret as we have shown. For example, the results for the stochastic frontier production function for feeder pigs and weaned piglets exhibit problems related to multicolinearity. However, the observed trends of technical efficiencies calculated from the parametric approach were consistent with those calculated with the non-parametric method for the same set of DMUs.

The sow farms analysed were highly technically efficient, FPP being slightly more efficient than FPFP. With respect to scale efficiency, scores were also high, FPFP being more scale-efficient than FPP. However, the important percentage of farms operating at below their optimal scales suggests that the current trend towards larger farm sizes could have a beneficial impact upon the efficiency of sow farms (either they produce piglets or feeder pigs) in future. Considering that current levels of efficiency are already quite high, a lot of effort is expected to achieve further improvements. Mean technical efficiency and the percentages of efficient DMUs are higher in this study compared with other published results. This fact can suggest a more homogeneous and competitive DMUs in the Spanish context dominated by vertically integrated companies. On the other hand, it has been seen how the increase of number of inseminations leads to a higher level of technical efficiency.

The strength of the DEA methodology lies in the fact that it focuses on individual farms (microeconomic agents) and can be used by advisers, specialist, or extension service agents to promote and diffuse best practices in farm management. It may therefore facilitate local action to combat relative inefficiency and become an important feature of programmes aimed at raising overall performance standards in the pig farming sector.

The computational and interpretative simplicity of DEA face stochastic methods make it a practical tool for individual agents such as small companies. Furthermore, the structure of the Spanish pig sector, with production concentrated in a relatively small number of companies and cooperatives, may benefit from such efficiency studies. However, DEA analysis should be only considered a starting point for identifying places to make improvements in farm production rather than an ending point.

14.6 Outlook

Although it was not in the scope of the study, other applications of technical efficiency should be pointed out as important future trends in this kind of studies. Sustainable development is a matter of concern with increasing attention from policy-makers and academics. For instance, the concept of environmental efficiency is gaining importance recently, mainly in Europe, but also in other countries. Manure management issues and GHG emissions are concerns that also have a rising interest (Piot-Lepetit 2014).

Many times, the consideration of environmental aspects is related to the existence of undesirable inputs (Piot-Lepetit 2014) and the way they can be dealt and interpreted by the methodology, either stochastic frontier or DEA models. Hence, Piot-Lepetit and Vermersch (1998) used DEA to measure the efficiency of French pig farms and derived a shadow price of organic nitrogen. Similarly, Lansink and Reinhard (2004) investigate the possibility of improving the environmental performance of Dutch pig farms reducing ammonia emissions, while Asmild and Hougaard (2006) employed the same DEA methodology to evaluate the environmental improvement potential of Danish pig farms. Environmental efficiency was also considered by Yang (2009) in pig farming in Taiwan. Other proposals beyond pig farming are also published concerning the eco-efficiency of farms (Picazo-Tadeo et al. 2011). Eco-efficiency benefits public expenditure in agri-environmental programs, although the cost-benefit balance is disputable. Finally, Yang (2009) and Picazo-Tadeo et al. (2011) emphasised also the benefits of training farmers to promote the integration between farming and environment, and hence, achieving a more efficient and sustainable production.

References

Aigner D, Chu SF (1968) On estimating the industry production function. Am Econ Rev 58:826–839

Asmild M, Hougaard J (2006) Economic and environmental efficiency of Danish pig farms. Agric Econ 35:171–181

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in DEA. Manag Sci 30:1078–1092

Banker RD, Chang H, Cooper WW (2004) A simulation study of DEA and parametric frontier models in the presence of heteroscedasticity. Eur J Oper Res 153:624–640

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. J Econ 38:387–399

BD-porc® (2013) http://www.irta.es/bdporc/

Berger AN (1993) Distribution-free estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. J Prod Anal 4:261–292

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Cloutier LM, Rowley R (1993) Relative technical efficiency: data envelopment analysis and Quebec's dairy farms. Can J Agric Econ 41:169–176

Coelli T (1996) A guide to DEAP version 2.1: a data envelopment analysis (computer) program. CEPA working paper 96/08. Centre for Efficiency and Productivity Analysis. Department of Econometrics, University of New England, Australia

Coelli TJ (1998) A multi-stage methodology for the solution of oriented DEA models. Oper Res Lett 23:143–149

Dyson RG, Allen R, Camacho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. Eur J Oper Res 132:245–259

Färe R, Grosskopf A, Cooper WW (1985) The measurement of efficiency of production. Kluwer Nijhoff, Boston

Farrell MJ (1957) The measurement of productive efficiency. J R Stat Soc Ser A 120:253–281

Fraser I, Cordina D (1999) An application of data envelopment analysis to irrigated dairy farms in Northern Victoria, Australia. Agr Syst 59:267–282

FRONTIER 4.1 (2013) A computer program for stochastic frontier production and cost function estimation. http://www.uq.edu.au/economics/cepa/frontier.php

Galanopoulos K, Aggelopoulos S, Kamenidou I, Mattas K (2006) Assessing the effects of managerial and production practices on the efficiency of commercial pig farming. Agric Syst 88:125–141

Gaspar P, Escribano M, Mesías FJ, Rodriguez de Ledesma A, Pulido F (2008) Sheep farms in the Spanish rangelands (dehesas): typologies according to livestock management and economic indicators. Small Rumin Res 74:52–63

Gaspar P, Mesías FJ, Escribano M, Pulido F (2009) Assessing the technical efficiency of extensive livestock farming systems in Extremadura, Spain. Livest Sci 121:7–14

Hadi AS (1992) Identifying multiple outliers in multivariate data. J R Stat Soc Ser B 54:761–777

Hadi AS (1994) A modification of method for the detection of outliers in multivariate samples. J R Stat Soc Ser B 56:393–396

Hansson H, Öhlmér B (2008) The effect of operational managerial practices on economic, technical and allocative efficiency at Swedish dairy farms. Livest Sci 118:34–43

Jaforullah M, Whiteman J (1999) Scale efficiency in the New Zealand dairy industry: a non-parametric approach. Aust J Agric Resour Econ 43:523–541

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econometrics 19:233–238

Lansink AO, Reinhard S (2004) Investigating technical efficiency and potential technological change in Dutch pig farming. Agr Syst 79:353–367

Lovell CAK (1994) Linear programming approaches to the measurement and analysis of productive efficiency. TOP 2:175–223

Picazo-Tadeo AJ, Gómez-Limón JA, Reig-Martínez E (2011) Assessing farming eco-efficiency: a data envelopment analysis approach. J Environ Manage 92:1154–1164

Piot-Lepetit I (2014) Technological externalities and environmental policy. How to simulate manure management regulation within a DEA framework. Ann Oper Res 214:31–48

Piot-Lepetit I, Vermersch D (1998) Pricing organic nitrogen under the weak disposability assumption: an application to the French pig sector. J Agric Econ 49:85–99

Reinhard S, Lovell CAK, Thijssen G (1999) Econometric estimation of technical and environmental efficiency: an application to Dutch dairy farms. Am J Agric Econ 81:44–60

Ripoll-Bosch R, Díez-Unquera B, Ruiz R, Villalba D, Molina E, Joy M, Olaizola A, Bernués A (2012) An integrated sustainability assessment of Mediterranean sheep farms with different degrees of intensification. Agr Syst 105:46–56

Rodriguez S, Plà LM, Faulin J (2014) Piglet production inside the pork supply chain management. a linear optimization model in sow farms. Ann Oper Res 219:5–23

Seiford L (1996) Data envelopment analysis: the evolution and state of the art. J Prod Anal 7:99–137

Seiford L, Zhu J (1999) An investigation of return to scale in data envelopment analysis. Omega 27:1–11

Sharma KR, Leung P, Zaleski HM (1997) Productive efficiency of the swine industry in Hawaii: stochastic frontier vs. data envelopment analysis. J Prod Anal 8:447–459

Sharma KR, Leung P, Zalesky HM (1999) Technical allocative and economic efficiencies in swine production in Hawaii: a comparison of parametric and nonparametric approaches. Agric Econ 20:23–35

Theodoridis A, Ragkos A, Roustemis D, Galanopoulos K, Abas Z, Sinapis E (2012) Assessing technical efficiency of Chios sheep farms with data envelopment analysis. Small Rumin Res 107:85–91

Weersink A, Turvey CG, Godah A (1990) Decomposition measures of technical efficiency for Ontario dairy farms. Can J Agric Econ 38:439–456

Yang C-C (2009) Productive efficiency, environmental efficiency and their determinants in arrow-to-finish pig farming in Taiwan. Livest Sci 126:195–205

Chapter 15
Multicriteria Analysis of Olive Farms Sustainability: An Application of TOPSIS Models

Laura Riesgo and Jordi Gallego-Ayala

15.1 Introduction

Spain is the first world producer and exporter of olive oil and table olives. Olive grove is grown in 2,032,290 ha, considering rainfed and irrigated area (MAGRAMA 2010). Spanish olive production yields 43 % of the total olive oil world production in 2007/2008 (1.2 million tonnes) which comprises a gross production of 1,990 million Euro (MAGRAMA 2010, 2012).

Andalusia region located in southern Spain is the major olive production area worldwide with a total area of 1.5 million hectares (19 % of the total olive grove area in the world, 30 % of the total olive grove area in the EU and 59 % in the Spanish territory) (CAyP 2008). In macroeconomic terms, olive production is the second most important agricultural sector in Andalusia, creating an overall income of 2,660 million Euro in 2007 (26 % of total agricultural production of Andalusia which accounts for 10,227 million Euros). Olive groves are identified as a 'social crop' as it is one of the agricultural activities that creates the most jobs per hectare (CAyP 2008). Indeed, the olive industry creates 32 % of agricultural employment in Andalusia (91,327 direct jobs), more than any other dynamic agricultural sector (like, e.g. horticulture). In summary, olive grove production is an important socio-economic activity, which is particularly relevant in rural municipalities in Andalusia, where olive farming is almost the only source of income for the population.

The environmental relevance of olive groves is also worth highlighting. The olive groves in Andalusia traditionally were associated to high biodiversity, being an example of a 'high natural value' agricultural system. This was possible due to

L. Riesgo (✉)
Department of Economics, University Pablo de Olavide, Seville, Spain
e-mail: laurariesgo@upo.es

J. Gallego-Ayala
Water Regulatory Council of Mozambique, Maputo, Mozambique

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_15

low intensity olive farming (minimum use of agrochemicals), old olive trees with semi-natural herbaceous vegetation and their location in areas with different land uses (Beaufoy and Cooper 2009). However, in recent years this ecological value has diminished due to the 'modernization' of olive groves. This modernization has been based on the expansion (new farms that have led to single-crop systems in large areas of Andalusia) and intensification of the crops (intensive use of fertilizers, pesticides, machinery and a large number of farms with uncovered soil). In spite of this modernization process, many olive grove systems are still associated with specific natural ecosystems, and 138,536 ha of olive groves (10 % of the olive grove area in Andalusia) are included in the Natura 2000 Networking Programme.

Therefore, olive grove systems provide a wide array of goods and services to the Andalusian society. Some of these goods and services are 'commodity outputs' since they are commercialized by the market, for instance olive oil. Alternatively other goods and services are 'non-commodity outputs' or 'public goods' since they have no market to be commercialized (e.g. the contribution of olive grove systems to support rural areas). Due to the lack of markets for public goods, olive growers do not receive any monetary compensation for their provision. The concurrence of production systems that provides both commodities and non-commodities to the society and the possibility of 'market failure' (unsuitable supply of public goods due to the lack of incentives—remuneration—for a suitable supply) makes olive farming a perfect example of multifunctional agricultural system.

15.1.1 Recent Development of the Olive Groves and Sustainability Problems in Andalusia

Spain's accession to the EU and the implementation of specific mechanism of support to the olive sector within the Common Agricultural Policy has encouraged the expansion and intensification of olive grove systems in Andalusia over the last two decades (DG ENV 2010). However, this rapid expansion and intensification caused several negative environmental impacts (Gómez-Limón and Riesgo 2012):

(a) *Soil erosion*. This environmental impact has been exacerbated in recent years due to the expansion of olive groves towards soils with unfavourable conditions for agricultural production (steep slopes, lands particularly sensitive to erosion or with frequent torrential rain). These adverse conditions and poor management of soils by farmers has damaged the natural vegetation cover (farms with uncovered soil). The Regional Andalusian Government reported that 29.7 % of olive farms had moderate soil erosion problems (12–50 $t \cdot ha^{-1} \cdot year$), 11.8 % showed high soil erosion (50–100 $t \cdot ha^{-1} \cdot year$) and 11.2 % very high soil erosion (more than 100 $t \cdot ha^{-1} \cdot year$).

(b) *Overexploitation of water resources*. The massive intensification of the olive sector is also reflected in the expansion of irrigated systems. Thirty years ago olive farming was almost non-irrigated, but nowadays the irrigated olive groves surface is estimated to be 546,425 ha, representing 35.3 % of the olive grove area in Andalusia. Despite the low water requirements of olive trees and the highly efficient irrigation systems used (water extractions range between 1,500 and 2,000 $m^3 \cdot ha^{-1} \cdot year$), the pressure on water resources is high. Increasing water extraction causes not only the overexploitation of water resources but jeopardizes the satisfaction of other water demands.

(c) *Non-point source water pollution*. Olive grove systems have contributed to the worsening of water quality due to the use of agrochemical products (mainly herbicides and fertilizers). Non-point source water pollution in rivers, dams and aquifers has produced several sanitary alarms in the last few years in Andalusia, such as the prohibition of drinking water from dams surrounded by olive trees.

(d) *Biodiversity loss*. One of the main characteristics of the olive grove in the 1980s (traditional farming) was the high biodiversity associated with this crop. The presence of trees and scrubland provided a habitat similar to meadows (*dehesa*), where a large number of insects, birds, reptiles and mammals lived. However, the intensification of olive farms changed this situation (disappearance of vegetable cover, water pollution, and high insecticide use and soil erosion) and diminished both the number and the diversity of animal species in the olive grove systems.

(e) *Damages in traditional agricultural landscapes*. The olive grove systems coexisted in the past with other crops such as pastures, vineyards or cereals. However, this diversity disappears with the intensification of olive grove systems and the olive grove is now often the only crop on farms.

What ‘sustainable agriculture’ actually means is a difficult question to answer. There is a scientific debate on how it is possible to reconcile the preservation of natural resources with production growth to satisfy food and fibre requirements as the world’s human population expands. Examining this issue, several definitions and alternative approaches can be found. Notwithstanding, there is a broad consensus that agricultural sustainability meets the following requirements (Raman 2006): (a) enhance food security, (b) protect natural resources and prevent environmental degradation, (c) be economically viable and (d) be socially acceptable. Taking these requirements into consideration, agricultural sustainability can be defined by the ‘mosaic’ approach, as a concept that encompasses three main dimensions (Yunlong and Smit 1994; Raman 2006):

Economic sustainability. To be sustainable, agriculture must be economically viable, ensuring not only adequate profitability for farmers (microeconomic approach), but also a positive contribution to national/regional income (macroeconomic approach).

Sociocultural sustainability. Agriculture must be socially and culturally relevant, i.e. it should ensure food security and equitable income distribution, as well as contributing to the viability of rural communities.

Environmental sustainability. Sustaining the preservation of biological productivity and ecosystem services is fundamental to achieve sustainable agriculture. Indeed, agricultural sustainability can be defined as the ability to ensure greater agricultural productivity while simultaneously conserving natural resources and preventing the depreciation of ecosystems.

The analysis of agricultural sustainability requires some geographic bounds. The farms (i.e. Andalusian olive farms) are considered the basic unit for this analysis of agricultural sustainability, like in other related works in the literature (Bockstaller et al. 1997; Girardin et al. 2000; Andreoli and Tellarini 2000; van Passel et al. 2007; Andersen et al. 2007; Russillo and Pintér 2009; Gómez-Limón and Riesgo 2009).

15.1.2 Objective

This chapter develops a theoretical framework and a methodology to evaluate the sustainability of olive farms. To this end, we used a hybrid Multicriteria Decision Making (MCDM) model integrating the Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to rank the olive farms in terms of their performance with respect to a set of socio-economic and environmental attributes. The objective of this study is to evaluate the sustainability of olive farms in Andalusia through the construction of composite indicators using an MCDM approach. The three dimensions of sustainability (economic, social and environmental) as stated previously are considered in the analysis and a set of indicators was selected in order to obtain a precise diagnosis of the sustainability of olive farms in this region. This result can be useful for public decision-making to encourage policy actions to promote the most sustainable farms.

Notwithstanding, it should be highlighted that the most common MCDM approach to construct composite indicators is based on the weighted sum method (Zhou and Ang 2009; Rowley et al. 2012). However, one of the main disadvantages of analysing the sustainability of agriculture through the implementation of the weighted sum method is that allows compensability between the different dimensions of sustainability (i.e. compensability amongst the indicators considered to measure sustainability). Some authors (Hansen 1996; Bockstaller et al. 1997; Morse et al. 2001; Ebert and Welsch 2004; Munda 2008) have criticized this approach because it considers that trade-offs between attributes (commensurability) are incompatible with the concept of sustainability. In order to test the results obtained by the hybrid TOPSIS model, we also calculate the weighted sum method to analyse the sustainability of olive farms. This analysis will allow us to compare results of two MCDM methods: TOPSIS and weighted sum.

The chapter is organized as follows. Section 15.2 presents the materials, with a brief description of the dataset of indicators utilized for this research. Section 15.3 includes an explanation of the AHP technique, the TOPSIS method used to calculate the sustainability indicator and the weighted sum approach. Section 15.4 presents the results obtained and aims to determine which the most sustainable farm types are. In addition, this section includes a comparison between the results obtained by TOPSIS and the weighted sum methods. Finally, Sect. 15.5 contains a discussion of the results and the conclusions drawn.

15.2 Materials

15.2.1 Economic, Social and Environmental Indicators

With the aim to achieve the objective proposed in this research, it is necessary to select a set of criteria that allow quantifying the multidimensional performance of the olive farms in Andalusia from a holistic perspective (economic, social and environmental). We use the comprehensive theoretical framework proposed by Sauvenier et al. (2006) and van Cauwenbergh et al. (2007) known as SAFE (Sustainability Assessment of Farming and the Environment Framework) to select a set of indicators which cover the multifunctional aspects of the olive grove agricultural systems. Thus, a total of 27 indicators were selected to quantify the sustainability of olive grove. See below a brief explanation regarding each sustainability indicator (interpretation and numeric calculation). For more information on indicators' selection and calculation, see Gómez-Limón and Riesgo (2012).

15.2.1.1 Economic Sustainability Indicators

Olive grove economic sustainability encompasses two principles: (a) farmers' economic sustainability (i.e. the economic viability of olive farms) and (b) public economic sustainability (i.e. food security and wealth creation in the society as a whole). According to these principles, seven indicators were included:

1. *Olive farmers' profit* (*PROFITOLIV*, in $€ \cdot \mathrm{ha}^{-1} \cdot \mathrm{year}$). Net profit is defined as gross income less total expenses in a given period, including depreciation on capital goods. Only those olive farms that record positive PROFITOLIV scores will be sustainable in the long run. The sustainability of olive farms would increase as PROFITOLIV records higher positive values.
2. *Variation in farmers' profits* (PROFITVAR, dimensionless and bounded [0,1]). The variation in farmers' profit over a period of time may be quantified through measures of dispersion from time series of annual profits. This variation was calculated by a coefficient of variation of the indicator PROFITOLIV over the

last 8 years. Farmers' income stability over a period of time (low values of PROFITVAR) would result in olive farms being more economically sustainable.

3. *Adaptation index* (ADAPTIND, dimensionless and bounded [0,1]). Olive farm viability depends not only on income and costs (profits and their stability over time), but also on how farmers adapt to changes in technology, policy reforms, changes in agricultural outputs or inputs, market or environmental changes (climate change). An ad hoc index is developed as a proxy to quantify farmers' capacity to face changes. This indicator is defined as a mathematical function of a set of variables such as (a) average slope of the land as a shaping factor of the technologies applied on the farm, (b) irrigation water availability as a factor needed for a potential irrigation transformation of the farm, (c) farmers' age, as young farmers are usually more willing to confront changes and (d) farmers' education, as educated farmers are usually more willing to confront changes. ADAPTIND values were bounded between 0 and 1. Farms that recorded high scores for this indicator are viable in the long run and more sustainable from an economic perspective.
4. *Production value* (PRODVAL, in $€ \cdot ha^{-1} \cdot year$). The contribution of olive farms to food security can be approached by the value of olive production. The higher the value of this indicator, the greater the economic sustainability of the olive farm.
5. *Changes in farm sales* (SALESVAR, dimensionless). Changes in farm sales over a period of time may be quantified through measures of dispersion. These changes were calculated using a coefficient of variation of the indicator PRODVAL over the last 8 years. Changes in PRODVAL due to changes in yields or prices imply a reduction in agricultural sustainability. Production stability implies a steady olive oil supply chain, as it minimizes the risk of olive supply being insufficient to meet demand. Therefore, the higher the SALESVAR score, the less economically sustainable the olive farm.
6. *Contribution to agricultural added value* (CONTRAAV, in $€ \cdot ha^{-1} \cdot year$). The contribution of olive farms to regional wealth can be assessed through gross value added (GVA). GVA is defined as income from output sales less expenses due to intermediate consumption goods. This indicator is a proxy to quantify olive farms' contribution to regional gross domestic product (GDP), as it shows the value added in the olive oil supply chain by olive farms. Positive values of CONTRAAV imply a positive contribution to regional wealth (i.e. high economic sustainability of olive farms from a public perspective).
7. *Percentage of income from subsidies* (PERCSUBV, bounded [0,1]). The economic viability of olive farms, excluding subsidies received by farmers, helps to achieve acceptable levels of economic sustainability from a public perspective. A zero value of the indicator PERCSUBV means the highest sustainability, as olive farm viability does not depend on public support (i.e. public subsidies). By contrast, higher values of this indicator represent lower economic sustainability.

15.2.1.2 Sociocultural Sustainability Indicators

The sociocultural sustainability of olive farms is based on two principles: (a) social sustainability due to the contribution of olive farms to rural development and (b) cultural sustainability as olive farms contribute to the conservation of cultural heritage. According to these principles, nine indicators were included.

8. *Total labour* (TOTLAB, in agricultural labour unit $\cdot$ ha^{-1}). Job creation in rural areas is one of the most important social roles of agriculture. Total labour in olive farms was selected as an indicator to quantify the social implications of olive farms in rural areas. Higher values of TOTLAB show labour-demanding olive farms and thus more sustainable farms from a social perspective.
9. *Apparent labour productivity* (PRODLAB, in € $\cdot$ agricultural labour unit^{-1}). Fulfilling a social role requires not only creating jobs but also generating income to guarantee proper remuneration of jobs. Apparent labour productivity is considered as an indicator to quantify the capacity of olive farms to remunerate jobs. Apparent labour productivity is defined as value added per person employed. The higher the value of PRODLAB, the more sustainable farms are from a social perspective, because olive farms help job creation in the long run.
10. *Risk of agricultural and rural abandonment* (ABANDON, bounded [0,1]). Agricultural and rural abandonment is a consequence of a number of factors, such as low profitability of agriculture in less favoured areas (i.e. the presence of environmental handicaps), perceived lack of opportunities for young people in rural areas and well-paid jobs in neighbouring territories. ABANDON is bounded between 0 when nobody manages the farm after the farmer's retirement and 1 when the farm transfer is guaranteed.
11. *Percentage of family and permanent labour* (FAMPERLAB, bounded [0,1]). Olive farming shows seasonal employment as labour is mainly demanded during harvesting (around 45–60 % of total labour in olive farms is required during harvesting). Seasonal employment neither increases population density in rural areas nor contributes to rural development in olive grove systems. The indicator FAMPERLAB quantifies the percentage of family and permanent labour of total labour in olive farms. As family and permanent workers usually live close to the farm, values of FAMPERLAB close to 1 imply more socially sustainable olive farms.
12. *Guarantee of origin membership* (ORIGIN, dimensionless and bounded [0,1]). Agriculture must provide high-quality food. This indicator varies between 1 if the olive farm is a member of a Designations of Origin and 0 if not. A value of 1 denotes the highest cultural sustainability of an olive farm.
13. *Percentage of olive oil classified as extra virgin olive oil* (VIRGINOIL, bounded [0,1]). Extra virgin olive oil satisfies the high-quality criteria on olive oil production. Values close to 1 show that most of the olive oil produced on the farm is extra virgin and consequently the sustainability of the olive farm is higher.

14. *Percentage of farm planted with crops other than olive trees* (OTHERCROP, bounded [0,1]). One of the cultural sustainability criteria is to protect the visual quality of agricultural landscape. OTHERCROP is defined as the percentage of land covered by crops other than olive trees. As the visual quality of olive grove landscapes in Andalusia includes contrasting colours and textures due to a mixture of olive trees and other crops (Arriaza et al. 2004), breaking single-crop farming contributes to enhancing the visual quality of the landscape. The indicator OTHERCROP ranges from 0 to 1. A value of 0 means a farm solely consists of olive trees, which does not enhance the visual quality of the landscape (lowest cultural sustainability), whereas a value of 1 indicates a multiple-crop farm with higher quality agricultural landscape (highest cultural sustainability).
15. *Soil cover* (COVER, bounded [0,1]). Since soil cover contributes to enhancing landscape valuation, soil cover was selected as an indicator to quantify the visual quality of agricultural landscape. This indicator is actually defined as the percentage of days during the year in which vegetation covers the soil. In this case, a value of zero value implies uncovered soil and low-valued olive grove landscape, whereas soils with vegetation denote high-valued landscape (higher sustainability).
16. *Index of protection of olive heritage* (HERITAGE, dimensionless and bounded [0,1]). Agricultural landscape includes the protection of a number of anthropogenic elements such as 100-year-old olive trees, ranches (haciendas), old mills for making olive oil, stone walls and hedges. The protection of olive heritage is considered an intangible factor and consequently an ad hoc index was built to quantify HERITAGE. This indicator is defined as a mathematical function of a set of variables such as the presence of 100-year-old olive trees on the farm, the presence of ranches or old mills for making olive oil on the farm, the presence of stone walls, terraces, hedges or similar heritage elements on the farm and the presence of rural tourism activities (rural houses, guide tours, etc.) on the farm. The indicator HERITAGE is bounded between 0 and 1. The higher the values of this indicator, the higher the sociocultural sustainability of olive farms.

15.2.1.3 Environmental Sustainability Indicators

The environmental sustainability of olive groves addresses three criteria: (a1) guarantee of olive grove genetic diversity, (a2) protection of biological diversity and (a3) protection of habitat diversity (ecosystem). In order to quantify the achievement of each criterion, 11 indicators were selected.

17. *Number of olive grove varieties* (NUMVAR, in number of crops). The genetic diversity of olive groves is a natural heritage that should be protected for future generations. A new indicator is included in the analysis to quantify the contribution of olive farms to the protection of the phylogenetic resources of olive farms. NUMVAR calculates the number of olive grove varieties on the

farm. The minimum value of NUMVAR is 1, denoting the least sustainable olive farm (e.g. 1 olive grove variety on the farm).

18. *Index of biological diversity* (DIVERSIND, dimensionless and bounded [0,1]). Biological diversity in the ecosystems of olive groves includes several living beings. Quantifying species at farm level goes beyond the scope of this research, but an ad hoc indicator has been built to analyse biological diversity on olive farms. DIVERSIND is defined as a mathematical function of a number of variables: (a) the presence of vegetation cover (flora and fauna protection), (b) weed control through sheep grazing (the least harmful soil management method), (c) placement of branches from pruning into piles on the borders of the farm (refuge areas for some species), (d) olives left on olive trees after harvesting (olives for fauna feeding) and (e) removal of ferti-irrigation or subsurface drip irrigation (minimizing animal poisoning). The indicator DIVERSIND is bounded between 0 and 1. A value of 1 indicates optimum biodiversity on the farm and the highest environmental sustainability.
19. *Pesticide risk* (PESTRISK, in rat $\cdot$ kg $\cdot$ ha^{-1} $\cdot$ year). Pesticides help control pests but may also reduce the population of non-target species. The lowest value of this indicator is zero, denoting organic olive farms. These production systems are the most sustainable from an environmental perspective because they have the highest value of biodiversity protection.
20. *Percentage of land with crops other than olive groves* (OTHERCROP, bounded [0,1]). This indicator achieves two criteria as it contributes to the visual quality of agricultural landscape and biodiversity, i.e. as a proxy of heterogeneity of land use and diversity of the ecosystem. High values of the indicator represent the presence of several crops or land uses on the olive farm and thus the existence of several ecosystems and higher environmental sustainability.
21. *Percentage of non-arable land* (NONARABLE, bounded [0,1]). This indicator assesses the value of non-agricultural ecosystems in olive farms such as river flows, and rocky out-crops. The higher the value of NONARABLE, the more environmentally sustainable the olive grove is.
22. S*oil erosion* (EROSION, in t $\cdot$ ha^{-1} $\cdot$ year). Soil erosion is one of the main environmental problems in olive grove systems (Gómez Calero and Giráldez 2009; Gómez Calero et al. 2009). High values for the indicator EROSION (high soil loss) denote olive farms with a limited capacity to protect soil and which are consequently less sustainable from an environmental perspective.
23. *Soil organic matter* (ORGMAT, dimensionless and bounded [0,1]). As one of the main determinants of soil quality is soil organic matter, an ad hoc indicator was built to analyse the soil organic matter of olive farms considering a set of variables such as tillage activities to maintain vegetation cover, the vegetation cover and the milling of pruning rests. The indicator ORGMAT is bounded between 0 and 1. The highest value of the indicator shows the most sustainable olive farms in terms of maintaining soil fertility.

24. *Nitrogen balance* (NITROGENBAL, in N $\text{kg} \cdot \text{ha}^{-1} \cdot \text{year}$). This indicator is defined as the physical difference (excess/shortage) between the nitrogen content of inputs (fertilizers) and outputs (harvesting). The difference between both quantities is the nitrogen liberated into the environment. Positive values of NITROGENBAL imply less environmentally sustainable olive farms.
25. *Residual herbicide use* (RESHERB, in active ingredient $\text{kg} \cdot \text{ha}^{-1} \cdot \text{year}$). This indicator measures the active ingredient content of residual herbicides used in olive farming. The lowest value of this indicator is 0 indicating that no residual herbicides are used on the farm. This value suggests organic olive farming and consequently no damage is caused to the environment.
26. *Irrigation water use* (WATERUSE, in $\text{m}^3 \cdot \text{ha}^{-1} \cdot \text{year}$). Irrigated olive farms account for 47 % of irrigated land in Andalusia (CHG 2008) and are the main irrigation water users (864 hm^3/year or 26 % of water demand). This high consumption leads to problems of over-extraction and environmental damage and, consequently, an indicator measuring the water actually extracted from the ecosystems (irrigation) was chosen. The indicator WATERUSE takes a value of zero in non-irrigated olive farms. These farms are the most environmentally sustainable as water is not used for irrigation purposes.
27. *Energy balance* (ENERGYBAL, in $\text{kcal} \cdot \text{ha}^{-1} \cdot \text{year}$). This balance is defined as the difference between the energy contained in the output (agricultural production) and the energy contained in agricultural inputs (input use and tillage practices). Positive values of ENERGYBAL mean that olive farms are using less energy than that produced by photosynthesis. The higher the positive values of this indicator, the higher the environmental sustainability.

15.2.2 Dataset: Olive Grove in Andalusia

A total of 410 olive farmers were interviewed face-to-face in Andalusia during May and September 2010. To carry out the survey, prior stratification was undertaken with respect to growing areas in Andalusia, and five agricultural districts were selected. Together, these districts account for 29.3 % of the total olive grove area in Andalusia (429,156 ha). In each district, 82 olive farmers were randomly interviewed.

A questionnaire was designed to collect information for the calculation of the above-mentioned indicators. The questionnaire consisted on three sections. The first section aimed to gather data on farm characteristics such as location, farm size or hired and family workers. The second section included questions on the management of olive groves, such as variety of olive grove, tillage system, use of herbicides and pesticides, prices and production). Finally, the last section compiled socio-economic information from the respondents. In order to calculate some economic and environmental indicators, secondary information was collated to complement the primary information supplied by the survey.

15.3 Methodology

The methodological framework proposed to evaluate the sustainability of olive farms using two MCDM approaches is based on four steps: (1) selection of the socio-economic and environmental attributes and indicators, (2) identification of the relative importance of the selected attributes and indicators via AHP, (3) ranking the olive farms in terms of sustainability by using a TOPSIS model and the weighted sum method and (4) classification of the farms by their sustainability performance using cluster analysis and description of the main factors that influence the sustainability performance of the olive farms. Therefore, this study is rooted in the combination of different MCDM methods, first the combination of the AHP and TOPSIS and secondly the combination of the AHP and the weighted sum. The methodological framework followed in this research is summarized in Fig. 15.1.

15.3.1 The Analytic Hierarchy Process

The AHP method developed by Saaty 1980 is an MCDM technique based on arranging decision-making problems in a hierarchical structure which allows the

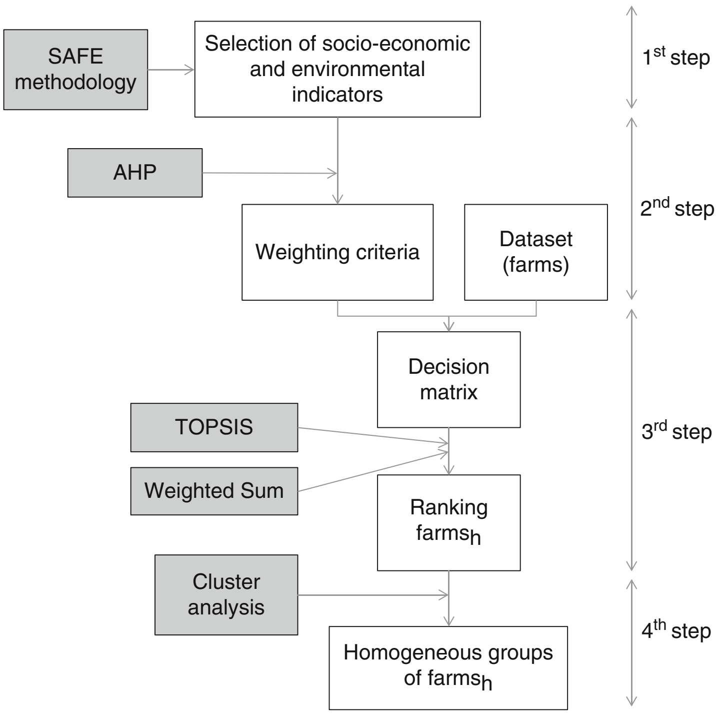


Fig. 15.1 Outline of the methodological approach

identification of the relative importance of each element under analysis by using a pairwise comparison system. Scores of these comparisons are used to build the Saaty matrices ($A = a_{kl}$), which are employed to determine the vector of priorities or weights ($w_1, \ldots w_k, \ldots w_n$). Although various procedures to estimate the weights of the criteria under evaluation have been proposed, for this research we select the simplest one: the geometric mean method (Aguarón and Moreno-Jiménez 2000).

Although the AHP technique was originally developed for individual decision-makers, overtime groups began using it in decision-making sessions (Easley et al. 2000), such as our case study. Thus, in order to determine the weights attached to each criterion we need to consider the judgements of a group of people (p), each with his/her own pairwise comparison matrix ($A = a_{klp}$) and its related weights (w_{kp}). This individual information is suitably treated in order to obtain a synthesis of aggregated weights (w_k). For this purpose, Forman and Peniwati (1998) suggest that group decision-making should be performed by aggregating individual priorities using the geometric mean:

$$w_k = \sqrt[m]{\prod_{p=1}^{p=m} w_{kp}} \tag{15.1}$$

The AHP technique was first applied to a representative sample of the population of Andalusia region (survey of 503 individuals), with the aim of obtaining the weights of the three criteria dimensions (economic, social and environmental) considered for the study, in order to calculate the relative importance of the set of economic (w_{eco}), social (w_{soc}) and environmental (w_{env}) criteria. Thus, we have: $w_{\text{eco}} = 58.6\ \%$; $w_{\text{soc}} = 14.3\ \%$; $w_{\text{env}} = 27.1\ \%$. For more information on this survey, see Arriaza and Gómez-Limón (2011).

AHP was afterwards applied to a panel of 18 experts, with the aim of obtaining the weights of the socio-economic and environmental criteria contained in each of the three criteria dimensions. This experts panel comprised eight scientific experts in different fields such as agricultural economics, sociology and rural development, ecology and environmental management and olive agronomy, together with ten experts from the olive industry (two experts from the Regional Ministry of Agriculture and the Environment, two experts from agricultural professional organizations, three technical managers, one representative from the Spanish Association of Olive Municipalities and two olive growers).

Taken into consideration the weights given by the population of Andalusia and the expert panel, the normalized weight given to each indicator is display in Fig. 15.2. For more information on the weights resulting from the AHP methodology, see Arriaza and Gómez-Limón (2011).

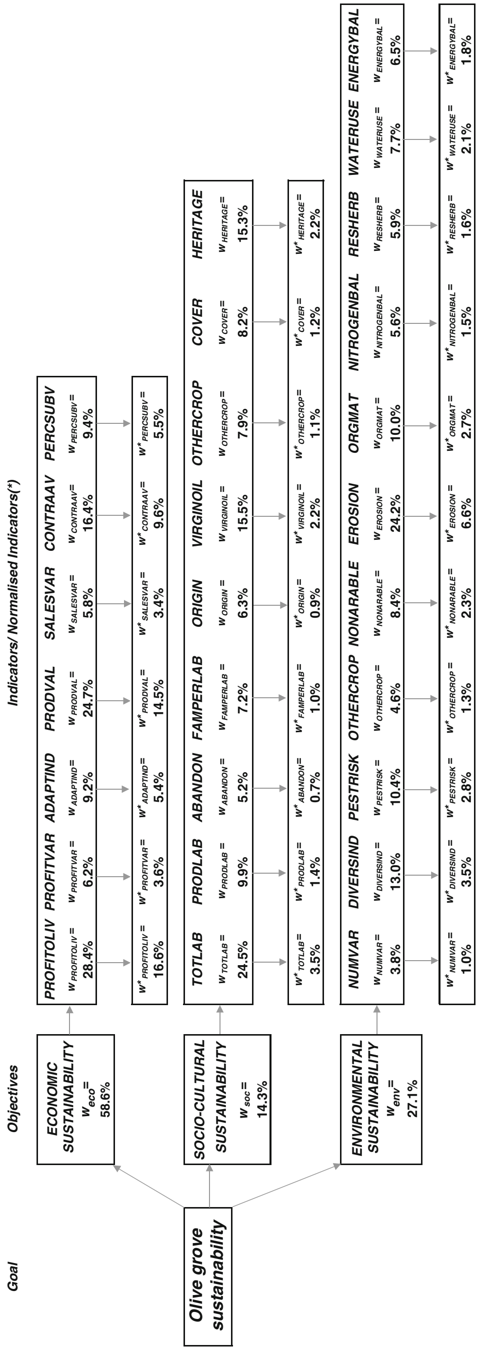


Fig. 15.2 Normalized weights for sustainability indicators

15.3.2 The TOPSIS Method

The TOPSIS analysis developed by Hwang and Yoon (1981) is a widely used MCDM tool to support the selection of the best compromise solution between a finite set of alternatives (Olson 2004), resulting in a rank of alternatives by using a distance measures framework. This procedure is based on the premise that the best alternative should have the closest distance to the positive ideal solution and the farthest distance from the negative ideal solution (Lai et al. 1994; Zanakis et al. 1998).

The TOPSIS model is becoming one of the most common multicriteria tools used to solve real-life problems (Behzadian et al. 2012). The main advantages of the TOPSIS models are: (a) the idea behind this model is rational and comprehensible for alternative selection, (b) low rank reversal and (c) the approach allows to identify the best alternatives in a simple mathematical and computational form (Kim et al. 1997; Deng et al. 2000; Olson 2004; Shih et al. 2007; Wu et al. 2010). The TOPSIS models have been used in the fields of supply chain management and logistics (Awasthi et al. 2011), energy management (Yan et al. 2011), water resources management (Srdjevic et al. 2004) and irrigated agriculture (Gallego-Ayala 2012) amongst others.

This technique was initially developed as a decision-support tool for alternatives selection, but can be adapted perfectly well to the composite indicator construction (Zhou and Ang 2009; Rowley et al. 2012). In fact, the multicriteria aggregation procedure followed in the TOPSIS model allows evaluating in a comprehensive way the sustainability performance of a system (Rowley et al. 2012). Thus, the TOPSIS approach has been used to evaluate the sustainability performance of the olive groves farms. The TOPSIS is applied in seven steps as listed below:

Step 1. Obtain the decision matrix (D)

Build the decision matrix (D) for ranking farms, by using the data obtained for each indicator (F) by farm (A). The decision matrix for evaluating the farms sustainability is as follows:

$$D = \begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_i \\ \vdots \\ A_m \end{array} \begin{array}{c} \begin{array}{cccccc} F_1 & F_2 & \cdots & F_j & \cdots & F_n \end{array} \\ \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1j} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2j} & \cdots & f_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{i1} & f_{i2} & \cdots & f_{ij} & \cdots & f_{in} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mj} & \cdots & f_{mn} \end{bmatrix} \end{array} \quad (15.2)$$

where f_{ij} denotes the performance value of farm A_i with respect to each indicator F_j.

Step 2. Calculate the normalized decision matrix (R)
The normalization has been carried out by using vector normalization, following the next expression:

$$v_{ij} = \frac{f_{ij}}{\sqrt{\sum_{j=1}^{n} f_{ij}^2}} \tag{15.3}$$

where v_{ij} denotes the normalized value. Thus $V = [v_{ij}]_{m \times n}$.

Step 3. Weight the normalized decision matrix (R) is obtained as follows:

$$r_{ij} = v_{ij} \cdot w_j \tag{15.4}$$

where w_j denotes the associated weight to each attribute. Weighting makes it possible to differentiate the relative importance of the various socio-economic and environmental attributes considered in this research. The weights calculated using the AHP technique have been integrated in this research to determine the weights of the criteria evaluation when developing the TOPSIS model to rank the olive farms. Thus, $R = [r_{ij}]_{m \times n}$.

Step 4. Determine the positive (T+) and negative (T−) ideal solution using the following formulation:

$$T^{+} = \{r_1^{+}, r_2^{+}, \ldots, r_n^{+}\} = \left\{\left(\max_i r_{ij} \middle| j \in J'\right), \left(\min_i r_{ij} \middle| j \in J''\right)\right\} \tag{15.5}$$

$$T^{-} = \{r_1^{-}, r_2^{-}, \ldots, r_n^{-}\} = \left\{\left(\min_i r_{ij} \middle| j \in J'\right), \left(\max_i r_{ij} \middle| j \in J''\right)\right\} \tag{15.6}$$

where J' and J'' are linked to the indicators with positive polarity (more is better) and the indicators with negative polarity (less is better), respectively.

Step 5. Calculate the separation distance of each alternative from the positive and negative ideal solution.
Separation distance is calculated by using the n-dimensional Euclidean distance. The separation of each farm from the positive-ideal solution (S_i^{+}) is given by the expression:

$$S_i^{+} = \sqrt{\sum_{j=1}^{n} \left(r_{ij} - r_j^{+}\right)^2} \tag{15.7}$$

Similarly, the separation of each farm from the negative-ideal solution (S_i^{-}) is as follows:

$$S_i^{-} = \sqrt{\sum_{j=1}^{n} \left(r_{ij} - r_j^{-}\right)^2} \tag{15.8}$$

Step 6. Calculate the relative closeness to the ideal solution, for each farm. The relative closeness of the farm A_i *to the ideal solution is given by:*

$$C_{i\mathrm{TOPSIS}} = \frac{S_i^-}{S_i^+ + S_i^-}, \quad i = 1, \ldots, m \tag{15.9}$$

where $C_{i\mathrm{TOPSIS}}$ is an index with values ranging between 0 and 1, where 0 corresponds to the worst possible performance of the farm and 1 to the best.

Step 7. Rank the farms, according to the descending order of $C_{i\mathrm{TOPSIS}}$ *index values.*

15.3.3 The Weighted Sum Approach

In order to compare the TOPSIS model with other MCDM approaches, we applied a conventional MCDM methodology to calculate agricultural sustainability, the weighted sum method combined with the AHP methodology. Once the normalized weights (w_k^*) have been obtained using the methodology mentioned in Sect. 15.3.1, resolving problems by means of the AHP technique is equivalent to optimizing a multi-attribute utility function, as has been demonstrated by Zahedi (1987). Adjusting this formula to our case study, the sustainability of each olive farm can be obtained through the following expression:

$$C_{i\mathrm{WS}} = \sum_{k=1}^{k=27} w_k^* \cdot I_{ki} \tag{15.10}$$

where $C_{i\mathrm{WS}}$ is the sustainability by olive farm i using the weighted sum method, w_k^* is the normalized weight to indicator k (using the weights obtained from the AHP technique), and I_{ki} is the normalized outcome of indicator k in the olive farm i.[1]

Finally, the olive farms will be ranked according to the descending order of $C_{i\mathrm{WS}}$ values.

15.4 Results

15.4.1 TOPSIS Method

Following TOPSIS methodology, we obtained a ranking of farms. A cluster analysis was carried out to classify the olive farms into homogeneous groups regarding

[1] Instead of vector normalization to express the base indicators in homogeneous units as in the TOPSIS approach, we used linear normalization (re-scaling method), to normalize the indicators in the weighted sum approach.

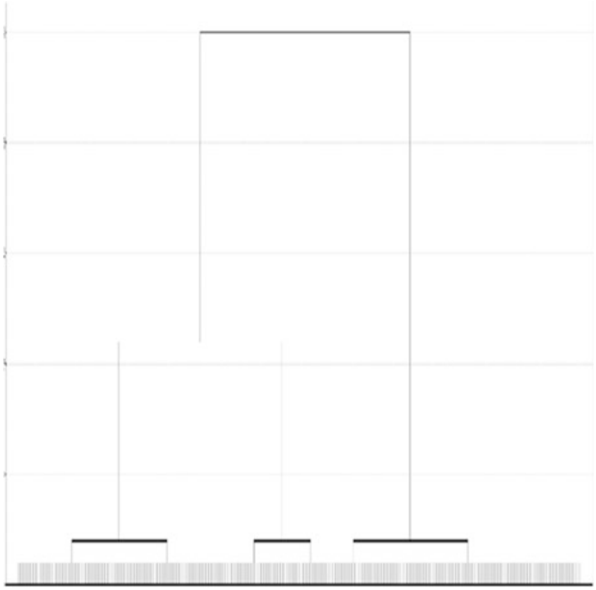


Fig. 15.3 Dendrogram considering $C_{i\mathrm{TOPSIS}}$ values

its sustainability level ($C_{i\mathrm{TOPSIS}}$). To develop this classification, the $C_{i\mathrm{TOPSIS}}$ values were considered as classifying variables. Furthermore, a hierarchical aggregation method was applied (the Ward or minimum distance method), defining the distance between elements as the Euclidean square distance. This aggregation procedure results in a dendrogram (see Fig. 15.3). We considered appropriate to cut the dendrogram in order to group the olive farms into three homogeneous groups or clusters.

Statistical tests were conducted to ensure that groups obtained from the cluster analysis differ on average from each other (see Table 15.1). In particularly, and before checking that the distribution is normal, we use the ANOVA test for equality of means not assuming equal variances (Games-Howell test).

Taking into account olive farm characteristics (see Table 15.2), these groups can be characterized as follows:

- *Cluster 1*. This group comprises traditional mountain rainfed olive farms (91 % of the farm size is non-irrigated on average) in high sloping lands (17 % on average) and with low olive production (2,714 kg ha^{-1}). Around one third of the olive farms (31 %) are located in a Denomination of Origin (DO) area and eco-friendly production systems (organic and integrated production farming) are used in 19 % of farms. These features led us to denominate this group as 'Traditional mountain rainfed olive farms'.
- *Cluster 2*. This group includes both rainfed (70 % of the farm size) and irrigated olive plots (30 % of the farm size) in moderate sloping lands (9 % on average) and moderate olive production (around of 5,000 kg ha^{-1}). Production systems are mainly conventional (92 %) and only 20 % of the olive farms are located in a DO area. This group is therefore labelled as 'Traditional plain olive farms'.

Table 15.1 Means of olive grove sustainability for each olive farm type

	N	Mean $C_{i\text{TOPSIS}}$	S.D.	F ANOVA (p value) Post hoc tests Games-Howell
Cluster 1	180	0.3871	0.039	1,022.938
Cluster 2	157	0.4950	0.037	(0.000)
Cluster 3	72	0.6347	0.050	$C_{\text{cluster3TOPSIS}} > C_{\text{cluster2TOPSIS}} > C_{\text{cluster1TOPSIS}}$

Table 15.2 Olive farm characteristics per olive farm group

	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	F ANOVA (p value) Post hoc tests Games-Howell
Rainfed land (% of farm size)	91	69	53	35.280 (0.000) c_1*** > c_2*** > c_3***
Irrigated land (% of farm size)	8	30	47	35.530 (0.000) c_3*** > c_2*** > c_1***
Average olive production (kg ha^{-1})	2,714	4,994	7,722	712.421 (0.000) c_3*** > c_2*** > c_1***
Slope land (%)	17.11	8.85	6.39	39.438 (0.000) c_1*** > c_2*** > c_3***
Denomination of origin (% of farm size)	31.11	19.75	15.28	4.859 (0.008) c_1** > c_2**; c_1*** > c_3***
Eco-friendly production system (% of farms)	18.89	7.64	0	11.588 (0.000) c_1*** > c_2*** > c_3***

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level

- *Cluster 3*. This group includes both rainfed and irrigated olive plots (50 % of the farm on average) in low sloping lands (6 % on average) and with high olive production (7,722 kg ha^{-1}). Production system is conventional (100 %) and only 15 % of the farms are located in a DO area. This group is labelled as 'intensive olive farms'.

Once this classification of farms has been established, the differences in the sustainability values of the three groups have been analysed (see Table 15.1). The results (mean $C_{i\text{TOPSIS}}$ values) show that 'intensive olive farms' (cluster 3) are significantly more sustainable than 'Traditional mountain rainfed olive farms' (cluster 1) and 'Traditional plain olive farms' (cluster 2). Thus, agricultural sustainability of olive farms is correlated to the farm profile.

When analysing separately the dimensions of olive grove sustainability (i.e. economic, social and environmental dimensions) for each of the three clusters, we obtained the results in Table 15.3.

Results show that *intensive olive farms* (cluster 3) are significantly more economic sustainable than the other two olive farm types (*Traditional mountain rainfed olive farms*—cluster 1—and *Traditional plain olive farms*—cluster 2). When analysing sociocultural sustainability, results show no statistically significant

Table 15.3 ANOVA tests for comparison of economic, social and environmental sustainability between olive farm groups

	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	F ANOVA (p value) Post hoc tests Games-Howell
Economic $C_{iTOPSIS}$ (S.D.)	0.2785 (0.0599)	0.4388 (0.0556)	0.6445 (0.0756)	952.498 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
Sociocultural $C_{iTOPSIS}$ (S.D.)	0.3123 (0.0629)	0.3160 (0.0627)	0.3124 (0.0554)	0.177 (0.838)
Environmental $C_{iTOPSIS}$ (S.D.)	0.6300 (0.0526)	0.6499 (0.0290)	0.6462 (0.0445)	9.556 (0.000) $c_2{}^{***} > c_1{}^{***}$; $c_3{}^{**} > c_1{}^{**}$

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level

Table 15.4 ANOVA tests for comparison of economic indicators between olive farm groups

Indicators	Weights in economic dimension (%)	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	F ANOVA (p value) Post hoc tests Games-Howell
PROFITOLIV (S.D.)	28.4	802,498 (683,087)	4,133,155 (1,862,871)	**13,893,060** (5,894,452)	579.710 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
PROFITVAR (S.D.)	*6.2*	***0.0387*** *(0.0074)*	*0.0366 (0.0046)*	*0.0358 (0.0039)*	*8.393 (0.000)* $c_1{}^{***} > c_2{}^{***}$; $c_1{}^{***} > c_3{}^{***}$
ADAPTIND (S.D.)	9.2	0.1079 (0.1278)	0.2073 (0.1511)	**0.2784** (0.1780)	40.511 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
PRODVAL (S.D.)	24.7	1,431,941 (847,047)	4,572,452 (1,532,731)	**11,031,886** (3,304,415)	759.930 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
SALESVAR (S.D.)	*5.8*	*0.0344 (0.0054)*	***0.0349*** *(0.0043)*	*0.0347 (0.0038)*	*0.398 (0.672)*
CONTRAAV (S.D.)	16.4	948,266 (581,069)	3,321,626 (1,133,276)	**8,778,744** (3,066,206)	689.974 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
PERCSUBV (S.D.)	*9.4*	***0.1236*** *(0.0505)*	*0.1100 (0.0053)*	*0.1090 (0.0037)*	*8.586 (0.000)* $c_1{}^{***} > c_2{}^{***} > c_3{}^{***}$

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level
See the highest values of each indicator in bold
See indicators with negative contribution to economic sustainability in italics

differences amongst the three olive farm types. Regarding environmental sustainability, results show that the least sustainable olive farms in Andalusia are the *Traditional mountain rainfed olive farms*.

In order to analyse these results, we examine the mean values of each indicator included in each sustainability dimension (economic, sociocultural and environmental dimensions) per cluster (see Tables 15.4, 15.5 and 15.6).

Intensive olive farms show the highest mean values in those indicators with the greatest importance in the economic dimension of olive grove sustainability (see Table 15.4). Olive farmers' profit (PROFITOLIV), production value (PRODVAL) and contribution to agricultural added value (CONTRAAV) account for 69.5 % of

Table 15.5 ANOVA tests for comparison of sociocultural indicators between olive farm groups

Indicators	Weights in social dimension (%)	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	*F* ANOVA (*p* value) Post hoc tests Games-Howell
TOTLAB (S.D.)	24.5	0.0223 (0.0043)	0.0275 (0.0058)	**0.0289** (0.0038)	70.074 (0.006) c_3*** > c_2*** > c_1***
PRODLAB (S.D.)	9.9	68,475 (58,167)	**77,629** (48,442)	55,493 (40,695)	4.844 (0.008) c_2** > c_3**; c_1* > c_3*
ABANDON (S.D.)	*5.2*	*0.8792 (0.2281)*	*0.8726 (0.2258)*	***0.9028** (0.2079)*	*0.457 (0.633)*
FAMPERLAB (S.D.)	7.2	0.6689 (0.3230)	0.7269 (0.2866)	**0.8098** (0.2584)	5.893 (0.003) c_3*** > c_1***; c_3* > c_2*
ORIGIN (S.D.)	6.3	**0.35** (0.478)	0.20 (0.399)	0.06 (0.231)	14.324 (0.000) c_1** > c_2** > c_3**
VIRGINOIL (S.D.)	15.5	**88.34** (11.67)	73.89 (20.83)	76.56 (16.74)	34.501 (0.000) c_1*** > c_2***; c_1*** > c_3***
OTHERCROP (S.D.)	7.9	**0.04** (0.131)	0.03 (0.108)	0.02 (0.105)	1.195 (0.304)
COVER (S.D.)	8.2	0.59 (0.365)	0.51 (0.344)	**0.95** (0.173)	44.000 (0.000) c_3*** > c_1***; c_3*** > c_2***
HERITAGE (S.D.)	15.3	**0.1583** (0.1432)	0.1047 (0.0785)	0.1261 (0.0996)	9.296 (0.000) c_1*** > c_2***

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level and * at 0.10-level
See the highest values of each indicator in bold
See indicators with negative contribution to sociocultural sustainability in italics

the total value of economic sustainability. These indicators have a positive contribution to the economic sustainability since the higher the values of these indicators the higher the economic sustainability. In contrast, *Traditional mountain rainfed olive farms* show the lowest mean values for these indicators and the highest values for indicators such as changes in farm sales (SALESVAR) and percentage of income from subsidies (PERCSUBV). These latter indicators have a negative contribution to the economic sustainability since the higher the values of these indicators the lower the economic sustainability.

When analysing sociocultural sustainability, we can conclude that *intensive olive farms* show the highest values for total labour (TOTLAB), percentage of family and permanent labour (FAMPERLAB) and soil cover (COVER). Since this olive farm type accounts for the highest olive production (7,722 kg ha^{-1}), this is as well the most labour demanding (see Table 15.5). In contrast, *Traditional mountain rainfed olive farms* mountain show the lowest mean values for these indicators but the highest for cultural indicators (guarantee of origin membership—ORIGIN, percentage of olive oil classified as extra virgin olive oil—VIRGINOIL, percentage

Table 15.6 ANOVA tests for comparison of environmental indicators between olive farm groups

Indicators	Weights in environmental dimension (%)	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	F ANOVA (p value) Post hoc tests Games-Howell
NUMVAR (S.D.)	3.8	**6.322** (5.923)	4.955 (4.784)	4.972 (5.352)	3.192 (0.042) $c_1{}^{*} > c_2{}^{*}$
DIVERSIND (S.D.)	13.0	**0.3893** (0.1224)	0.3690 (0.1336)	0.3339 (0.1997)	3.913 (0.021) $c_1{}^{*} > c_3{}^{*}$
PESTRISK (S.D.)	*10.4*	*17,916,009 (47,499,675)*	*28,834,701 (109,782,782)*	***48,601,773*** *(79,166,418)*	*3.636 (0.027)* $c_3{}^{***} > c_1{}^{***}$
OTHERCROP (S.D.)	4.6	**0.0162** (0.0757)	0.0151 (0.0659)	0.0124 (0.0411)	0.084 (0.919)
NONARABLE (S.D.)	8.4	**0.0434** (0.1280)	0.0113 (0.0621)	0.0135 (0.1089)	4.631 (0.010) $c_1{}^{**} > c_3{}^{**} > c_2{}^{**}$
EROSION (S.D.)	*24.2*	***1,063*** *(1,679)*	*297 (488)*	*379 (1,138)*	*17.788 (0.000)* $c_1{}^{***} > c_2{}^{***}$; $c_1{}^{***} > c_3{}^{***}$
ORGMAT (S.D.)	10.0	0.5043 (0.3249)	0.5515 (0.3402)	**0.5743** (0.3892)	1.375 (0.254)
NITROGENBAL (S.D.)	*5.6*	*1,362 (3,114)*	*1,744 (4,505)*	***3,296*** *(3,872)*	*6.635 (0.001)* $c_3{}^{***} > c_1{}^{***}$; $c_3{}^{**} > c_2{}^{**}$
RESHERB (S.D.)	*5.9*	***973,582*** *(1,878,860)*	*953,594 (1,831,441)*	*941,391 (1,383,052)*	*0.010 (0.990)*
WATERUSE (S.D.)	*7.7*	*34,396 (144,201)*	*176,922 (391,245)*	***309,698*** *(610,977)*	*16.105 (0.000)* $c_3{}^{***} > c_1{}^{***}$ $c_3{}^{***} > c_2{}^{***}$
ENERGYBAL (S.D.)	6.5	37,435,092 (27,131,785)	88,409,585 (50,044,877)	**196,425,993** (96,990,380)	222.206 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level and * at 0.10-level
See the highest values of each indicator in bold
See indicators with negative contribution to environmental sustainability in italics

of farm planted with crops other than olive trees—OTHERCROP, and index of protection of olive heritage—HERITAGE).

As it is mentioned above, *traditional mountain olive farms* are less environmental sustainable than the other two olive farms types. When analysing the indicators included in the environmental dimension, soil organic matter (ORGMAT) and energy balance (ENERGYBAL) show the lowest values in *traditional mountain olive farms* (see Table 15.6). These indicators have a positive contribution to the environmental sustainability since the higher the values of these indicators the higher the environmental sustainability. In addition, residual herbicide use (RESHERB) and soil erosion (EROSION) show the highest value, but they have a negative contribution to the environmental sustainability since the higher the values of these indicators the lower the environmental sustainability. The weight of these four indicators in the environmental sustainability of olive grove in Andalusia is 46.6 % explaining the low score of *traditional mountain olive farms*.

In contrast, we cannot correlate the characteristics of the highest environmental sustainable with the olive farm type, since no statistically significant differences were found between *traditional plain olive farms* and *intensive olive farms* (see Table 15.3). However, *intensive olive farms* show the highest mean values of negative indicators such as pesticide risk (PESTRISK), nitrogen balance (NITROGENBAL) and irrigation water use (WATERUSE).

15.4.2 The Weighted Sum Approach

Following the weighted sum methodology combined with AHP, we obtained a ranking of farms. A cluster analysis was carried out to classify the olive farms into homogeneous groups regarding its sustainability level ($C_{i\mathrm{WS}}$). The resulting dendrogram produced three olive farm types (see Fig. 15.4).

Table 15.7 shows how the homogeneous groups obtained from the cluster analysis differ on the agricultural sustainability $C_{i\mathrm{WS}}$ mean values from each other.

There is a high degree of similarity between the farm groups obtained using the weighted sum method, and those obtained via TOPSIS technique (see Table 15.8). Thus, we label the first group as *traditional mountain olive farms* (cluster 1), the second group as *traditional plain olive farms* (cluster 2) and the third group as *intensive olive farms* (cluster 3).

Despite sustainability scores per cluster in the case of the weighted sum approach (mean $C_{i\mathrm{WS}}$ values) are lower than the sustainability values obtained through the TOPSIS methodology (mean $C_{i\mathrm{TOPSIS}}$ values), results shows that 'intensive olive farms' (cluster 3) are significantly more sustainable than the other two olive farm types ('Traditional mountain rainfed olive farms'—cluster 1—and 'Traditional plain olive farms'—cluster 2).

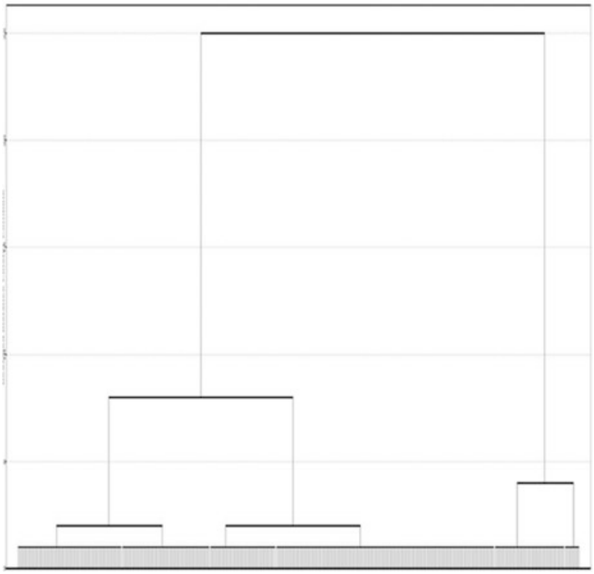


Fig. 15.4 Dendrogram considering $C_{i\mathrm{WS}}$ values

Table 15.7 Means of olive grove sustainability for each olive farm type

	N	Mean $C_{i\mathrm{WS}}$	S.D.	F ANOVA (p value) Post hoc tests Games-Howell
Cluster 1	207	0.2740	0.022	743.308
Cluster 2	140	0.3380	0.023	(0.000)
Cluster 3	62	0.4434	0.059	$C_{\mathrm{cluster3WS}} > C_{\mathrm{cluster2WS}} > C_{\mathrm{cluster1WS}}$

Table 15.8 Olive farm characteristics per olive farm group

	Cluster 1 c_1	Cluster 2 c_2	Cluster 3 c_3	F ANOVA (p value) Post hoc tests Games-Howell
Rainfed land (% of farm size)	93	64	48	55.319 (0.000) $c_1{}^{***} > c_2{}^{***} > c_3{}^{***}$
Irrigated land (% of farm size)	7	34	52	56.630 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
Average olive production (kg ha^{-1})	3,018	5,152	7,782	423.993 (0.000) $c_3{}^{***} > c_2{}^{***} > c_1{}^{***}$
Slope land (%)	16.49	8.37	5.55	39.874 (0.000) $c_1{}^{***} > c_2{}^{***} > c_3{}^{***}$
Denomination of origin (% of farm size)	25	25	18	0.773 (0.462)
Eco-friendly production system (% of farms)	20	8	0	12.386 (0.001) $c_1{}^{*} > c_2{}^{*} > c_3{}^{***}$

***Mean differences are statistically significant at the 0.01 level of significance, ** at 0.05-level

An in-depth analysis of the olive farms comprised in each cluster shows that 81.17 % of farms belong to the same cluster using both methodologies (i.e. TOPSIS and weighted sum methods). Consequently, results obtained under the weighted sum methodology seem to be equivalent to those obtained under the TOPSIS methodology (see Sect. 15.4.1).

15.5 Discussion and Conclusions

This chapter contributes to the literature on the analysis of agricultural sustainability by using composite indicators rooted in MCDM techniques. Sustainability's views expressed through a survey to the population in Andalusia and a panel of olive grove experts allowed us to calculate the weights of each sustainability dimension and indicator, respectively. Three different olive farm types were identified in Andalusia according to their sustainability index calculated by using TOPSIS and weighted sum methodologies.

Despite some authors criticized additive aggregation methods, results show that no differences were found in the classification of farms regarding their sustainability (in terms of $C_{i\text{TOPSIS}}$ and $C_{i\text{WS}}$ values) between the two MCDM approaches followed in this research. Nonetheless, further comparative analysis using alternative MCDM techniques to calculate agricultural sustainability such as VIKOR, ELECTRE, PROMETHEE, MACBETH or MELCHIOR should be applied, to continue investigating the pros and cons of MCDM tools in analysing agricultural sustainability issues.

Our results indicate that any public policy intended to promote olive grove sustainability should be based on a structural policy that promotes the intensification of olive farms. Therefore, and taking into account the weights given to each indicator by the experts panel and the weights given to each of the sustainability dimension by the society, we can conclude that 'intensive olive farms' are the most sustainable in Andalusia. This result is mainly caused due to the high weight given to economic sustainability from the Andalusian citizens (58.6 %) and the high values achieved by economic indicators in this type of olive farms. In general, we can conclude that $C_{i\text{TOPSIS}}$ and $C_{i\text{WS}}$ values depend in a high degree on the particular weights given to the socio-economic and environmental indicators.

Despite the results show that an intensification of olive grove would result in higher agricultural sustainability, this intensification process is not always suitable for all plots. For instance, an olive farm located in high sloping lands or without irrigation water access could not be successfully intensified (i.e. higher erosion rates related to sloping agricultural land or lower yields of olive trees due to the lack of water).

Finally, it should be pointed out that further research is needed in order to validate the results obtained in this research. A sensibility analysis of the weights given to indicators and sustainability dimensions would be welcome. In addition, the implementation of this methodology in other olive grove systems would be valuable. Both analyses may contribute on the suitability of the methods proposed for olive grove sustainability calculation by distinguishing between the influence of the selection of indicators for each dimension of sustainability and the social preferences regarding olive grove sustainability.

Acknowledgments The authors gratefully acknowledge the financial aid received from the Spanish Ministry of Science and Innovation (AGL2009-12553-C02-02 and AGL2010-17560-C02-01) and the Regional Government of Andalusia (P09-SEJ-4992).

Disclaimer: The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the Water Regulatory Council of Mozambique.

References

Aguarón J, Moreno-Jiménez JM (2000) Local stability intervals in the analytic hierarchy process. Eur J Oper Res 125(1):114–133. doi:10.1016/S0377-2217(99)00204-0

Andersen E, Elbersen B, Godeschalk F, Verhoog D (2007) Farm management indicators and farm typologies as a basis for assessments in a changing policy environment. J Environ Manage 82 (3):353–362. doi:10.1016/j.jenvman.2006.04.021

Andreoli M, Tellarini V (2000) Farm sustainability evaluation: methodology and practice. Agric Ecosyst Environ 77(1):43–52. doi:10.1016/S0167-8809(99)00091-2

Arriaza M, Gómez-Limón JA (2011) Valoración social del carácter multifuncional de la agricultura andaluza. ITEA-Inf Tec Econ Ag 107(2):102–125

Arriaza M, Cañas Ortega JF, Cañas Madueño JA, Ruiz P (2004) Assessing the visual quality of rural landscapes. Landscape Urban Plan 69:115–125. doi:10.1016/j.landurbplan.2003.10.029

Awasthi A, Chauhan SS, Goyal SK (2011) A fuzzy multi criteria approach for evaluating environmental performance of suppliers. Int J Prod Econ 126:370–378. doi:10.1016/j.ijpe.2010.04.029

Beaufoy G, Cooper T (2009) The application of the high nature value impact indicator, European evaluation network for rural development. European Commission's Directorate-General for Agriculture and Rural Development, Brussels

Behzadian M, Otaghsara SK, Yazdani M, Ignatius J (2012) A state-of the-art survey of TOPSIS applications. Expert Syst Appl 39:13051–13069. doi:10.1016/j.eswa.2012.05.056

Bockstaller C, Girardin P, van der Werf HMG (1997) Use of agro-ecological indicators for the evaluation of farming systems. Eur J Agron 7(1–3):261–270. doi:10.1016/S1161-0301(97)00041-5

CAyP, Consejería de Agricultura y Pesca (2008) El sector del aceite de oliva y la aceituna de mesa en Andalucía. Consejería de Agricultura y Pesca—Junta de Andalucía—Regional Government of Andalusia, Seville

CHG, Confederación Hidrográfica del Guadalquivir (2008) Esquema de Temas Importantes. Confederación Hidrográfica del Guadalquivir, Seville

Deng H, Yeh CH, Willis RJ (2000) Inter-company comparison using modified TOPSIS with objective weights. Comput Oper Res 27:963–973. doi:10.1016/S0305-0548(99)00069-6

DG ENV (2010) LIFE among the olives—good practice in improving environmental performance in the olive oil sector. Directorate General for Environment, European Commission, Brussels, Belgium. http://ec.europa.eu/environment/life/publications/lifepublications/lifefocus/documents/oliveoil.pdf. Accessed 13 Oct 2012

Easley RF, Valacich JS, Venkataramanan MA (2000) Capturing group preferences in a multicriteria decision. Eur J Oper Res 125:73–83. doi:10.1016/S0377-2217(99)00196-4, http://dx.doi.org/10.1016/S0377-2217(99)00196-4

Ebert U, Welsch H (2004) Meaningful environmental indices: a social choice approach. J Environ Econ Manage 47(2):270–283. doi:10.1016/j.jeem.2003.09.001

Forman E, Peniwati K (1998) Aggregating individual judgments and priorities with the analytic hierarchy process. Eur J Oper Res 108(2):165–169. doi:10.1016/S0377-2217(97)00244-0

Gallego-Ayala J (2012) Selecting irrigation water pricing alternative using a multi-methodological approach. Math Comput Model 55(3–4):861–883. doi:10.1016/j.mcm.2011.09.014

Girardin P, Bockstaller C, van der Werf H (2000) Assessment of potential impacts of agricultural practices on the environment: the AGRO-ECO method. Environ Impact Assess 20:227–239. doi:10.1016/S0195-9255(99)00036-0

Gómez Calero JA, Giráldez JV (2009) Erosión y degradación de suelos. In: Gómez Calero JA (ed) Sostenibilidad de la producción de olivar en Andalucía. Consejería de Agricultura y Pesca-Junta de Andalucía—Regional Government of Andalusia, Seville

Gómez Calero JA, Sobrinho TA, Giráldez JV, Fereres E (2009) Soil management effects on runoff, erosion and soil properties in an olive grove of Southern Spain. Soil Till Res 102 (1):5–13. doi:10.1016/j.still.2008.05.005

Gómez-Limón JA, Riesgo L (2009) Alternative approaches to the construction of a composite indicator of agricultural sustainability. An application to irrigated agriculture in the Duero basin in Spain. J Environ Manage 90:3345–3362. doi:10.1016/j.jenvman.2009.05.023

Gómez-Limón JA, Riesgo L (2012) Sustainability assessment of olive groves in Andalusia: a methodological proposal. New Medit 12(2):39–49

Hansen JW (1996) Is agricultural sustainability a useful concept? Agr Syst 50:117–143. doi:10.1016/0308-521X(95)00011-S

Hwang CL, Yoon K (1981) Multiple attributes decision making methods and applications. Springer, New York

Kim G, Park CS, Yoon KP (1997) Identifying investment opportunities for advanced manufacturing systems with comparative-integrated performance measurement. Int J Prod Econ 50:23–33. doi:10.1016/S0925-5273(97)00014-5

Lai YJ, Liu TY, Hwang CL (1994) TOPSIS for MODM. Eur J Oper Res 76:486–500. doi:10.1016/0377-2217(94)90282-8

MAGRAMA, Spanish Ministry of Agriculture, Food and Environment (2010) Agricultural and statistics yearbooks. Spanish Ministry of Environment, and Rural and Marine Environments. http://www.magrama.gob.es/es/estadistica/temas/estad-publicaciones/anuario-de-estadistica/default.aspx. Accessed 13 Oct 2012

MAGRAMA, Spanish Ministry of Agriculture, Food and Environment (2012) El mercado del aceite de oliva. Spanish Ministry of Environment, and Rural and Marine Environments. http://www.magrama.gob.es/es/agricultura/temas/regulacion-de-los-mercados/organizaciones-comunes-de-mercado-y-regimenes-de-ayuda/ocm-aceite-de-oliva-y-aceitunas-de-mesa/. Accessed 13 Oct 2012

Morse S, McNamara N, Acholo M, Okwoli B (2001) Sustainability indicators: the problem of integration. Sustain Dev 9(1):1–15

Munda G (2008) Social multi-criteria evaluation for a sustainable economy. Springer, New York

Olson DL (2004) Comparison of weights in TOPSIS models. Math Comput Model 40:721–727. doi:10.1016/j.mcm.2004.10.003

Raman S (2006) Agricultural sustainability. Principles, processes and prospects. Haworth Press, New York

Rowley HV, Peters GM, Lundie S, Moore SJ (2012) Aggregating sustainability indicators: beyond the weighted sum. J Environ Manage 111:24–33. doi:10.1016/j.jenvman.2012.05.004

Russillo A, Pintér L (2009) Linking farm-level measurement systems to environmental sustainability outcomes: challenges and ways forward. International Institute for Sustainable Development, Winnipeg

Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York
Sauvenier X, Valckx J, van Cauwenbergh N, Wauters E, Bachev H, Biala K, et al (2006) Framework for assessing sustainability levels in Belgian agricultural systems—SAFE. Part 1: sustainable production and consumption patterns. Final Report—SPSD II CP 28. Belgian Science Policy, Brussels
Shih HS, Shyur HJ, Lee ES (2007) An extension of TOPSIS for group decision making. Math Comput Model 45:801–813. doi:10.1016/j.mcm.2006.03.023
Srdjevic B, Medeiros YDP, Faria AS (2004) An objective multi-criteria evaluation of water management scenarios. Water Resour Manage 18:35–54
van Cauwenbergh N, Biala K, Bielders C et al (2007) SAFE—a hierarchical framework for assessing the sustainability of agricultural systems. Agr Ecosyst Environ 120:22–242. doi:10.1016/j.agee.2006.09.006
van Passel S, Nevens F, Mathijs E, van Huylenbroeck G (2007) Measuring farm sustainability and explaining differences in sustainable efficiency. Ecol Econ 62:149–161. doi:10.1016/j.ecolecon.2006.06.008
Wu CS, Lin CT, Lee C (2010) Optimal marketing strategy: a decision-making with ANP and TOPSIS. Int J Prod Econ 127:190–196. doi:10.1016/j.ijpe.2010.05.013
Yan G, Ling Z, Dequn Z (2011) Performance evaluation of coal enterprises energy conservation and reduction of pollutant emissions base on GRD–TOPSIS. Energy Procedia 5:535–539. doi:10.1016/j.egypro.2011.03.093
Yunlong C, Smit B (1994) Sustainability in agriculture: a general review. Agr Ecosyst Environ 49:299–307. doi:10.1016/0167-8809(94)90059-0
Zahedi F (1987) A utility approach to the analytic hierarchy process. Math Model 9:387–395. doi:10.1016/0270-0255(87)90497-0
Zanakis S, Solomon A, Wishart N, Dublish S (1998) Multi-attribute decision making: a simulation comparison of select methods. Eur J Oper Res 107:507–529. doi:10.1016/S0377-2217(97)00147-1
Zhou P, Ang BW (2009) Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. Soc Indic Res 94:83–96

Chapter 16
On the Feasibility of Establishing a Northern-Western Australian Beef Abattoir as a Facility Location Problem

Rodolfo García-Flores and Andrew Higgins

16.1 Introduction

Beef production all over the world is currently undergoing pressure to become more efficient. While global demand continues to increase (Kearney 2012), thanks mostly to demand from developing countries, consumers everywhere are becoming more conscious of their choices in food quality and safety. This has imposed many requirements on beef producers. For example, international trade has made it necessary to introduce labelling and traceability regulations, whereas sustainability and impact on the environment have become of paramount concern to both society and industry. Unpredictable climate patterns have also forced the beef producers to plan for unforeseen changes in their physical landscapes, while at the same time the increasingly uncertain economic environment and changing regulatory policies have forced them to increase the efficiency of the whole supply chain. Optimisation studies in this area must now consider many of these factors simultaneously.

Despite their differences in environmental, economic, and productive conditions, the challenges that the beef supply chains face in many countries are similar. Beef supply chains are linear sequences of activities, from primary production through to the consumer and waste management. Figure 16.1 is a schematic of a generic agricultural supply chain, which also encompasses beef and cattle. In this diagram, industry drivers are highlighted in the top four rectangular boxes, and

R. García-Flores (✉)
CSIRO Mathematics, Informatics and Statistics, Gate 5 Normanby Rd., Clayton South, VIC 3168, Australia
e-mail: Rodolfo.Garcia-Flores@csiro.au

A. Higgins
CSIRO Ecosystem Sciences, Ecosciences Precinct, Dutton Park, QLD 4001, Australia
e-mail: Andrew.Higgins@csiro.au

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_16

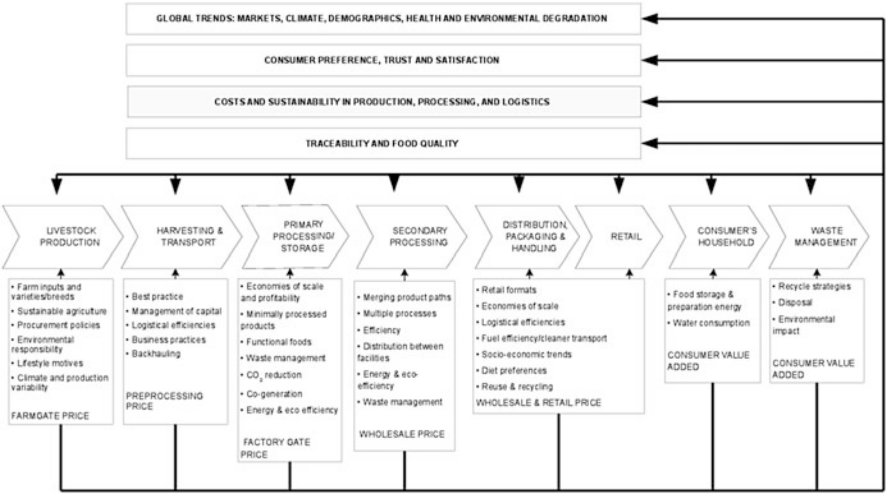


Fig. 16.1 Description of a general agricultural value chain (Higgins et al. 2010), highlighting some important features for achieving sustainability at each segment. Industry level drivers are listed at the *top*

some of the issues that can be modelled and improved using OR are listed as dot points in the lower boxes. Concrete problems to address include:

- Increasing the reliability of supply chain operations in the event of natural disasters and environmental changes.
- Adding value to the chain by, for example, finding methodologies that assure quality and quantity of beef regardless of the season.
- Exploiting potential synergies between different economic actors and between regions in the same country. A clear example is multi-modal transport in countries where agriculture and resources are important.
- Exploiting alternatives in the development of feeding, processing, and transportation operations to increase the value and efficiency of the industry.
- Exploring integrated approaches to production, for example, the development of multi-species abattoirs or integrated beef and milk production systems.
- Improving pasture and beef management to reduce the variability on the supply end of the chain.

Likewise, governments share responsibility for the efficient operation of the cattle producers in their countries by

- Directing rational exploitation of agricultural areas by mapping and zoning land resources.
- Fostering commercial and technological partnerships among all supply chain participants.
- Ensuring sustainability of the industry through environmental, safety, and quality legislation.

- Stimulating coordinated production programs and systems.
- Regulating licensing of abattoirs and other premises, staff training, export fees, and taxes.
- Investing in physical infrastructure.

In the particular case of northern Australia, the beef industry is a critical part of the economy: it is worth over one billion Australian dollars, covers 90 % of the land area, carries 30 % of the nation's cattle numbers, and produces 80 % of Australia's live cattle exports (Department of agriculture, Fisheries and Forestry 2012). The north is fundamentally different to the more intensive beef farming industry of the south because it takes place in an environment characterised by large scale enterprises on pastoral lease, low herd density, long distances to market, and significant annual interruptions of the production and distribution processes due to heat, drought, and tropical rainfall patterns.

In this chapter, we review common optimisation approaches that have been used to address some of the above issues and present the northern Australian beef supply chain as a case study. Section 16.2 introduces the beef supply chain, highlighting its agricultural and logistical aspects, and provides pointers to existing literature. We introduce our case study in Sect. 16.3, together with an optimisation tool devised to assist the stakeholders in deciding on the best choices for infrastructure investment. The tool solves a facility location and network design model. We discuss preliminary results in Sect. 16.4 and summarise and outline the path for future developments in Sect. 16.5.

16.2 Review of Relevant Features of the Beef Supply Chain

In this section we review relevant research on particular issues that complicate management of the beef and livestock supply chain and explain in detail some optimisation concepts widely used to address them. We organise the review in problems that stem from the food and agricultural nature of the supply chain and problems that are inherent to logistics and production. For an exhaustive review, the reader should refer to Ahumada and Villalobos (2009) and Lucas and Chhajed (2004).

16.2.1 The Food Supply Chain

The application of OR methods to agricultural chains expanded rapidly in the mid 1990s, as seen in the increase in publications in the OR and agricultural literature (Higgins et al. 2010). However, success has been limited, partly because published projects lack a supply chain-wide perspective, and partly because of the inherent complexity caused by the biological nature of agricultural supply chains.

This complexity is reflected in *high variability*, the difficulty to incorporate *sustainability* issues, and critical *timing* to process and market. These problems (and some of the solutions that have been proposed) are described in the following subsections.

16.2.1.1 Variability of Biological Processes

Perhaps the distinguishing feature of all agricultural supply chains is their variability, which is evident in all temporal and spatial scales. The causes of this variability are *biological*, e.g., the genetics of the cattle breeds and pastures of interest, *environmental*, as for example the changing weather patterns that affect the yields and the quality of meat and feed stocks, and *socio-economical*, as for instance the changing demand patterns and political decisions that often affect contracted production agreements.

The most used technique to deal with uncertainty is *stochastic programming*, as a survey of the literature shows. Stochastic programming and dynamic formulations are better suited to capture the temporal variation of uncertain variables. A stochastic program is an optimisation problem where some or all of the parameters are described by random variables. These formulations are not only harder to solve, but also require a deeper understanding from the decision maker. In contrast to deterministic solutions, where the solution "prescribes" unique values for the decision variables, the solution of a stochastic programming problem is the solution that truly optimises the *expected value* of the objective function. As a consequence, the stochastic solution is normally never optimal after the values of the variables are known, but at the same time, it is hardly ever really bad (Kall and Wallace 1994). When using this type of model, the decision maker must be aware that the value of the stochastic solution is in the *options* it provides in the sense that the solution suggests some investment decisions *in anticipation of* realisations of uncertain variables, e.g., high prices or demands.

Stochastic programming has been used extensively by many authors, for example, by Boyabatli et al. (2011), who developed a two-stage stochastic recourse problem for understanding the trade-offs facing a meat-processing company in the choice of alternative arrangements for sourcing cattle, when that company acts as a wholesaler into several final product markets. The results showed that higher variability increases the profits of the processing company, but decreases the reliance on the contract market relative to the spot market. Schütz and Tomasgard (2011) and Schutz et al. (2009) studied the relationship between uncertainty and flexibility in a meat supply chain by using a two-stage stochastic programming model. The authors compared the results of this model to the deterministic expected-value problem in order to discuss the impact of flexibility in the supply chain. Their results showed that, given sufficient flexibility in the supply chain, a deterministic approach to planning may result in better results than a stochastic model. Jiang et al. (2009) formulated a mixed-integer stochastic programming model of a meat supply chain and solved it using a Benders decomposition algorithm.

Other projects have made extensive use of *simulation* to deal with variability. McDermott et al. (2005) introduced a simulation model of the New Zealand beef value chain and analysed three simulation scenarios: changes in land price, wider use of beef semen in the dairy industry, and introduction of a gene to improve feed intake. The authors conclude that land price dominates this industry's ability to create value in the long run. While reviewing the distinguishing features of fresh-food supply chains, Bruzzone et al. (2009) presented a case study for the fresh-meat supply chain of a major retailer operating in northern Italy. They introduced a simulation system and proposed a heuristic to balance demand and manage variability. We also note that there are a number of simplified models for many inherently variable processes related to cattle production systems. An overview of these can be found in Hirooka (2010).

16.2.1.2 Sustainability

Many aspects of the management of food supply chains are central to ensuring sustainability, understood as satisfying the needs of the present human generation without compromising the ability of future generations to meet their needs. Ensuring sustainability requires the analysis of social, economic, and environmental dimensions (see, for example, Dake et al. 2005; White and Lee 2009).

Sustainability optimisation studies explore a variety of environmental protection strategies, in addition to the traditional focus on the economics of beef production only. These include the selection of processing sites (for example, disposal plants as in Caballero et al. 2007), improvement of farm efficiency (as in Dimara et al. 2005, who use data envelopment analysis), and disease outbreak and management (Zhao et al. 2006). Maintaining biodiversity and rationally exploiting natural resources are also objectives of operational research projects in this area. The study by Havlik et al. (2006) introduced a linear model which accounted for the joint production of beef and other agricultural goods in a region of the Czech Republic. This project focused in joint production of organic suckler cow farms, grassland, and crops. The model maximised the margin subject to the availability of land resources, production capability, and production agreements. The authors explained that European governments support farm multi-functionality on the grounds of maintaining biodiversity. Costa and Rehman (2005) analysed the reasons why farmers encourage overgrazing in Brazil and assessed the optimal policy using a bi-criteria optimisation model. The model maximised returns and the asset value of cattle subject to herd structure, forage and capital restrictions, pasture costs, and minimum herd requirements. The authors concluded that a certain level of over-grazing is economically rational. In Australia, Moloney and Hearne (2009) showed that the replacement of domestic livestock with native alternatives, or even mixed grazing, is economically and environmentally viable. The interested reader can find a review of optimisation and simulation models of livestock farming systems with emphasis in sustainability in Gouttenoire et al. (2011).

16.2.1.3 Time to Market

As in all agricultural supply chains, timing and seasonality are also of concern for beef producers. A good example of the kind of time restrictions imposed on this type of models is due to Nielsen et al. (2004), who devised a model to optimise grazing strategy, feed level in winter, and time of fattening and slaughter in organic steer production. Their dynamic programming model used a multi-level hierarchic Markov process. Crosson et al. (2006) introduced a model of Irish beef production systems and used it to investigate how farmers might optimally react to variations in beef and feed prices, alternative feed sources, and participation in an agro-environmental scheme that limits nitrogen usage. As is typical for this kind of system, monthly time intervals were chosen to describe seasonal variations.

Commer (1991) solved a spatio-temporal model of the south eastern United States to determine the optimal number, size, and regional locations of slaughter and processing facilities, and to determine optimal regional locations for backgrounding, feeding, and finishing, given spatially dispersed patterns of weaned calf supply and beef demand. Glen (1986) introduced a linear program to assess the performance of an integrated crop and intensive beef production enterprise. The model assumes that calf breeding and rearing activities form a separate part of the enterprise, and therefore they are not considered in the model.

16.2.1.4 Flat Payoff Functions

A consequence of the forgiving nature of biological processes is the commonly found flat payoff functions in economic studies of agricultural activities that strive to find economically optimal production levels. This has often been cited (e.g., Higgins et al. 2010) as a reason for the lack of widespread adoption of operational research techniques in the primary sector in particular, and of precision agriculture in general. Pannell (2006) states it very clearly and warns the OR practitioner that flat payoff curves have important consequences. “A modeller can usually be more helpful to the decision maker by identifying the shape of the payoff function, and specially the range over which it is relatively flat, rather than emphasising a single optimal solution”. Agricultural optimisation models are meant to be used to gain insight on the processes, not to prescribe “best” solutions.

16.2.2 Optimisation of Livestock Supply Chains

Although beef supply chains are inherently variable due to its biological nature, they have many features that make them akin to the supply chains of industries in the secondary sector. In particular, the location of processing facilities is affected by the closeness of consumption and distribution centres, and for this reason studies on cattle and livestock logistics have also focused on *infrastructure and investment* decisions, often taking the form of facility location problems. In many studies, they are considered as variations of the trans-shipment problem. These are reviewed next.

16.2.2.1 Infrastructure

The selection of roads and facilities such as bridges, abattoirs, and agistment farms is usually modelled as facility location and network design problems. Roads, bridges, and other facilities are expected to operate for decades, and planning their layouts and capacities is a non-trivial task that, more often than not, must consider multiple commodities, transportation modes, and economic interests in the region of concern. Multiple productive activities are often incorporated in agricultural supply chain models, as in Branco et al. (2010), who formulate a multi-commodity network flow problem that considers flows of sugar, alcohol, corn, soybean, soybean oil, and wheat, in order to identify potential hubs for multimodal transportation in central Brazil. In the same vein, the prototype mentioned in Sect. 16.5 considers cattle transportation in conjunction with other productive activities and incorporates a range of transportation modes and commodities. The cost of transportation infrastructure makes construction projects a long-term investment, which often is financed by governments.

This is a problem which can be modelled using static, deterministic models that are invariably limited to decisions taken in a single step. Examples concerning the location of abattoirs include Cassidy et al. (1970), who presented a study to find the optimum size and location of beef slaughter plants in the Eastern Central Queensland region. The authors showed that, for the year of the study, building abattoirs in the production-area would reduce the supply chain operating costs by 50 %. Domingues-Zucchi et al. (2011) introduced a model to determine the optimum location of new export-oriented slaughterhouses in the Brazilian state of Mato Grosso. The problem considered supply, demand, and production constraints. Their results showed that building abattoirs close to the export ports would minimise logistical costs. However, deterministic models fail to capture the uncertainty of important variables such as demands, distances, and travel times. Sensitivity analysis has been used to address this shortcoming, as in Broek et al. (2006), who tested various demand scenarios to assess the results of a linear programme intended to determine abattoir location. Sensitivity analysis assesses the effect of varying parameter values on the objective function value, but it does not address the variability in operation conditions directly. Deterministic models assume that the values of operating conditions are known, and the optimisation is made as one-time decision values.

Stochastic programming and dynamic formulations are better suited to capture the temporal variation of uncertain variables, as discussed in Sect. 16.2.1.1. Stochastic facility location has been applied by Schutz et al. (2008) to determine the best location of an abattoir as a two-stage stochastic programming model where both demand and short-run costs may be uncertain at investment time. Brown and

Drynan (1986) stressed the inability of deterministic models to produce results that adequately reflect variability in cattle supplies and produced a stochastic model to select potential sites for abattoir construction. The interested reader can find a review on strategic facility location in Owen and Daskin (1998) and an introduction to facility location problems with stochastic demands in Berman and Krass (2001).

16.2.2.2 Transportation and Logistics

Food distribution is nowadays a global business that enables most countries access to most foods all year round. Food supply chains have become more complex, leading to a dual emphasis on quality preservation and efficiency. The early paper by Cassidy et al. (1970) to determine the location of abattoirs in Queensland is formulated as a modified transshipment problem. Judge et al. (1973) analyse the inter-regional transport of meat in the United States in order to determine the optimal geographical flows and prices of livestock and meat.

Because the cattle and beef supply chain is a system where decision making is distributed, companies have also tried to learn from other participants' experience in the logistics of distribution. For example, Simons and Taylor (2007) report on a study where the participants of the red meat supply chain in the UK work together to cooperatively identify and implement improvements in chain performance that would not be available to companies working individually. Surprisingly, for many of the participants this was the first attempt to reach out and coordinate actions with their partners, other than immediate customers.

We would like to finish this section stressing that modelling beef supply chains is a complex task and modelling some of its aspects necessarily require simplifications. For example, a common simplification is to consider a spatial network by selecting representative locations, as is the case in the papers by Commer (1991), Branco et al. (2010), and Domingues-Zucchi et al. (2011). Other projects address complexity by focusing on individual stages of problems in the supply chain, as in Stott et al. (2003), who use linear programming to assess the relative contribution that disease prevention could make to farm income and to its variability, or Stygar and Makulska (2010), who provide a review of optimisation and simulation models used for herd management. Most studies neglect the loss of weight that the cattle experience when they are transported over long distances.

16.3 Case Study: The Northern Australia Beef Industry Strategy

Cattle and beef production is an important economic activity in Australia. The 2010–2011 Australian farm exports earned the country $32.5 billion (Australian dollars), of which beef and veal production contributed 17 %.

The increase on demand was of \$400 million from the 2008–2009 to 2009–2010 financial years, and of \$600 million from 2009–2010 to 2010–2011. Australia exported 937,301 tonnes of beef and veal in 2010–2011, worth \$4.5 billion. The major export markets for beef and veal are Japan (37 %), the United States (17 %), and Korea (15 %) (Department of Employment, Economic Development and Innovation 2010). Australian live cattle exports were worth \$660 million in 2010–2011, predominantly exported to Indonesia (57 %), Turkey (13 %), and Israel (7 %). Cattle processing is expected to rise 13 % to a capacity of 9.5 million heads, whereas production is projected to reach 2.4 million tonnes by 2015. Assuming continued export to foreign markets and average seasonal conditions, the Australian cattle herd is expected to increase to 29.7 million head by 2015, 11.9 % higher than in 2010.

During the 1980s and 1990s many abattoirs in Australia closed. One detailed case is presented in Piggot et al. (1987), who analysed the cost structure of a single abattoir through a simulation study and identified the problems that led to difficulties in covering its variable costs. The main of these is the fact that the abattoir did not adjust its fee for service in a timely manner, but it has been reported that other abattoirs did also close due to an increasing dominance of Indonesian buyers of live cattle (Strategic Design and Development, Meateng Pty Ltd 2010).

Recent evidence indicates that the economic environment is changing. According to the study in Meateng Pty Ltd (2012) and Kearney (2012), there is evidence of a shift in the global market for beef consumption, with demand stagnating in traditional northern hemisphere markets and growing rapidly in parts of Asia and the Middle East. In order to capitalise on this opportunity, the corresponding government's taskforce has identified potential to increase production from Australia's northern cattle herd through the intensification of production and greater diversification and flexibility in land use. To realise this potential, the governments of the northern states of Australia are working with industry to provide a comprehensive analysis of the livestock industry value chain. This analysis is part of a plan known as the *Northern Australia Beef Industry Strategy*, or NABIS.[1] The analysis of the livestock industry value chain presented in this section optimises the infrastructure and transport logistics of the beef supply chain while considering production, transport and handling of live cattle, processing, retail, and exports.

Cattle production in the north of Australia is fundamentally different to the more intensive beef farming industry of the south. Northern beef production takes place in an environment characterised by large scale enterprises on pastoral lease, low herd density, long distances to market, and significant annual interruptions to turn-off due to heat, drought, and tropical rainfall patterns. These are important challenges that must be addressed in a highly competitive global market (Strategic Design and Development, Meateng Pty Ltd 2010).

[1] www.regional.gov.au/regional/ona/nabis.aspx, retrieved 19 Sept 2012.

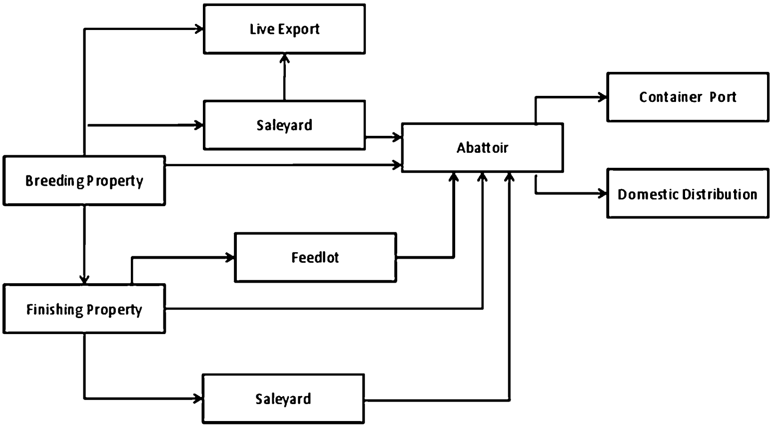


Fig. 16.2 Schematic of the northern Australian beef supply chain. Reproduced from the study in Meateng Pty Ltd 2012

Figure 16.2 shows a schematic of the northern Australian supply chain. *Breeding properties* typically produce weaner calves to about the age of about 8 months, when their weight is approximately 330 kg. These cattle can then be sold to *live export* for finishing in other countries. Many breeding properties do not have enough forage to produce cattle to slaughter weight. Such cattle are transported by road trains to *finishing properties* where they are grass-fed, or to a more intensive confined feeding system (or *feedlot*) where they are grain-fed. Cattle spend a minimum of 100 days in feedlots until they reach suitable weight categories for sale. In *sale yards*, cattle of multiple classes are sold by auction, which includes sales for abattoirs, breeding, and for further finishing. *Abattoirs* transform the finished cattle into frozen or chilled meat products. Abattoirs vary significantly in terms of throughput (up to 2,000 head per day) though Australia's largest 25 abattoirs account for 61 % of production. Once processed, meat is either transported in refrigerated containers to terminals or to domestic wholesale.

The preliminary model we introduce in the next subsection considers only two stages: the transportation from breeding to agistment farms (i.e., feedlots and finishing properties), and the transportation from agistment farms to abattoirs. The questions the case study addresses are:

- At what locations should meat processing take place so as to minimise transportation costs subject to infrastructure choices and budget?
- What transportation infrastructure is crucial to keep the beef value chain operating?
- What are the flows to be transported and processed among facilities during the time horizon?

16.3.1 The Model

The purpose of the preliminary model described in this subsection is:

> **Given** the locations of breeding and agistment farms, potential abattoir sites, unit transportation costs, road maintenance and construction costs, and cattle production, **determine** the transportation routes and location of the abattoir(s) that minimise the total costs **subject to** road network constraints, farm and abattoir capacity constraints, and budget.

The region of Western Australia under consideration is shown in Fig. 16.3. This model incorporates two key findings reported in the feasibility studies of Strategic Design and Development, Meateng Pty Ltd (2010) and Meateng Pty Ltd (2012). Firstly, that unpredictable supply and high labour costs caused by transporting livestock could be ameliorated if the agistment sector was developed. This is considered in the first stage of the solution methodology, which considers the transportation from properties to farms where agistment can take place. Secondly, that access to a processing stream would be of significant benefit to producers, who are exposed to tightening live export market constraints. However, this will not be viable in competition with a strong live export trade, and without tangible

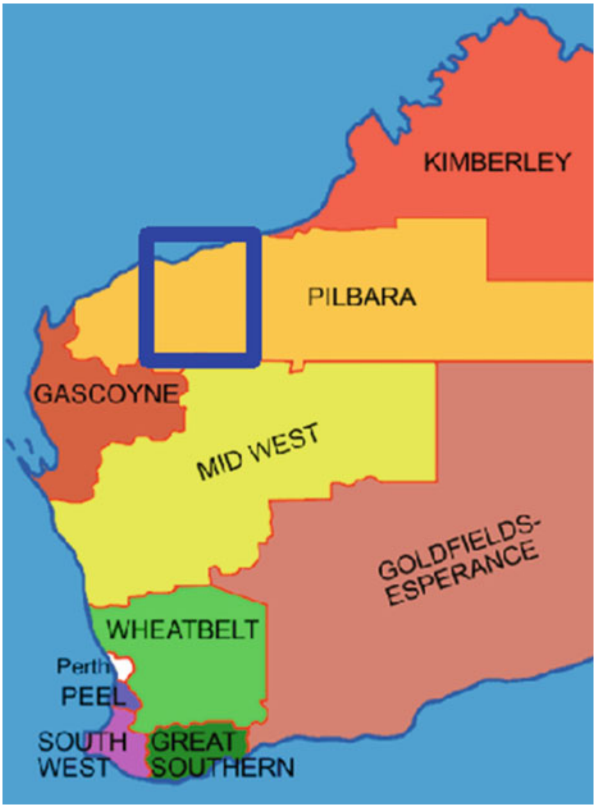


Fig. 16.3 The region under consideration for the optimisation model. The *highlighted area* covers an area of approximately 180,000 km^2

government support and producers' commitment. For this reason, the second stage of the methodology assumes that all cattle are processed in the abattoir, i.e., there are no live exports. An additional advantage of this assumption is that processed meat product can be transported more cheaply than live cattle.

The model we present considers only two stages: the transportation from breeding to agistment farms, and the transportation of agistment farms to abattoirs. As noted above, we disregard live exports, assuming all cattle are processed in the abattoir and reflecting the current restrictive constraints on the live export trade. We acknowledge that incorporating live exports is an important extension since, currently, most production goes to export. However, the model outlined here is a good starting point because it addresses the questions of the feasibility of an abattoir in the north and the design of alternative transportation routes in case of disruptions.

The problem considered is as follows: cattle are transported in truckloads from the properties to fattening farms, possibly through intermediate junctions, where they spend a number τ of periods until they are ready to be sent to the abattoir. The problem is solved in two stages. In the first stage, we solve a trans-shipment problem with inventory to determine the amount of cattle ready for slaughtering after the agistment period. In the second stage, we solve the facility location problem to find the locations of the abattoirs that minimise total costs. In contrast to Domingues-Zucchi et al. (2011), we do not calculate centroids for the demand centres, but consider the actual distances between facilities. The following subsections explain these sub-problems in greater detail.

16.3.1.1 Nomenclature

The parameters and variables used in the model are defined next.

Sets

D	The set of transhipment points.
F	The set of agistment farms.
A	The set of potential abattoir sites.
K_{FA}	The set of commodities to be transported from agistment farms to abattoirs.
L_{FA}	The set of valid links from farms to abattoirs. Locations of the abattoirs are not known a priori.
S	The set of properties.
T	The set of all periods.

Decision variables

DFA_{it}	The number of truckloads supplied to the abattoir(s) by farm $i \in \{F \cup A\}$. This decision variable is used for the facility location problem, which considers flows from farms F to abattoirs A.
p_{it}	Number of truckloads of cattle held at farm i at time t.

q_{ijut} Number of truckloads of livestock that flow on link $(i, j) \in L_{SF}$ that enter a time u and leave at time t, that is, after agistment (properties and farms).
w_{it}^{k} Fraction of demand of k served by node i at time t.
x_{ij} Indicator variable, takes value one if link $(i, j) \in L_{SF}$ is open (agistment farms and abattoirs) and zero otherwise.
y_{ijt}^{k} Fraction of demand of commodity $k \in K_{FA}$ that flows on link $(i, j) \in L_{FA}$ at time t (farms and abattoirs).
z_i Indicator variable, takes value one if abattoir is located at node $i \in \{F \cup A\}$ and zero otherwise.

Parameters

τ Number of periods required to fatten a truckload of livestock.
V_{ijt} Availability of link (i, j) between properties and farms at time t.
B Total budget allocated to facility construction.
c_{ij} Cost of constructing link (i, j).
DSF_{it} Demand (or supply) of site i at time t. This parameter is used for the trans-shipment problem, which considers flows from properties S to farms F.
f_i Cost of setting an abattoir in location i.
FK_j Total capacity (in truckloads) of farm j.
$\mathrm{TC}_{ijt}^{'}$ Transportation cost from nodes i to j at time t, $i,j \in \{F \cup A\}$ in dollars.
TC_{ijt} Transportation cost from nodes $i \in \{S \cup D\}$ to $j \in \{F \cup D\}$ at time t in dollars per truckload.

16.3.1.2 Cattle Flows from Properties to Farms

The first stage of the supply network is the transportation from properties to fattening farms, which is modelled as a trans-shipment problem with constraints:

$$\sum_{u \in T} \sum_{j \in \{S \cup F\}} V_{jit} q_{jiut} - \sum_{u \in T} \sum_{j \in \{S \cup F\}} V_{ijt} q_{ijut} = \begin{cases} \mathrm{DSF}_{it} & \text{if } \sum_{j \in \{S \cup F\}} V_{ijt} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

for all $i \in S$ and for all $t \in T$. For the farms,

$$\sum_{u \in T} \sum_{j \in \{S \cup F\}} V_{jit} q_{jiut} - \sum_{u \in T} \sum_{j \in \{S \cup F\}} V_{ijt} q_{ijut} = -\mathrm{DSF}_{it} + p_{it} \tag{2}$$

for all $i \in F$ and for all $t \in T$, where V_{ijt} is the availability of link (i, j) between properties and farms at time t and DSF_{it} is the number of truckloads supplied by the properties and farms. Note that the model considers that the agistment farms produce truckloads of cattle, just like any other property. Constraints (2) state that the sum of all inputs plus production must equal total outputs; if site i is a farm, accumulation (agistment) is permitted. The necessary number of periods to fatten each truckload of livestock, denoted by τ, is captured in

$$\sum_{j \in S} q_{jitt} = q_{ii,t+\tau} \quad \forall i \in F, \ \ \forall t \ \in T. \tag{3}$$

The model assumes that agistment farms can only accommodate a limited number of truckloads of cattle. This farm holding capacity constraints are expressed as

$$p_{it} \leq FK_i \quad \forall i \in F, \ \ \forall t \in T. \tag{4}$$

The *objective function* of this sub-problem minimises the total costs of transportation from properties to agistment farms, in order to determine the optimal design of the transportation network:

$$\text{Minimise} \quad \sum_{(i,j) \in L_{SF}} \sum_{t \in T} \sum_{u \in T} \mathrm{TC}_{ijtu} q_{ijtu} \ , \tag{5}$$

where q_{ijtu} is the number of truckloads transported between properties and farms and TC_{ijt} is the transportation cost of commodity k from nodes i to j at time t. The links correspond to the existing road network in the Pilbara region.

16.3.1.3 Cattle Flows from Farms to Abattoirs

For the second sub-problem it is necessary to define a commodity k, which represents the truckloads of live cattle sent from a given farm to an abattoir. Let x_{ij} be variables that indicate if link between nodes i and j are open between fattening farms and abattoirs, y^k_{ijt} the fraction of demand flows of commodity coming from node k on link (i, j) at time t between farms and abattoirs, z_i a variable that indicates if an abattoir is located at node i, and w^k_{it} the fraction of demand of k served by node i at time t.

This stage of the formulation is a combined network design and facility location problem that assumes that the demand of the nodes selected as facility is in fact served by a super-node; see Melkote and Daskin (2001a, b) for details. The flow conservation constraint for the nodes selected as abattoirs i is

$$z_i + \sum_{j \in \{F \cup A\}} x_{ij} = 1 \quad \forall i \in \{F \cup A\}, \tag{6}$$

which says that nodes selected as abattoirs fulfil the total demand, and that there are no outbound links transporting livestock from the sites chosen to be abattoirs. We assume that the sets $S, F \subset S$ are known, but the locations of A are not and need to be determined as from the set of farms and proposed abattoir sites as a facility location problem.

For the case where i is the destination of k, we have

$$z_k + \sum_{i \in \{F \cup A\},\, i \neq k} w_{it}^k = 1 \quad \forall k \in \{F \cup A\}, \ \forall t \in T, \tag{7}$$

which says that the demand of all other nodes that are not abattoirs is supplied by the abattoirs. Equations (6) and (7) specify zero demand for the nodes that are not selected as abattoirs. Conservation constraints for selected and non-selected links in the network design are

$$x_{ki} + \sum_{j \in \{F \cup A\}} y_{jit}^k = \sum_{j \in \{F \cup A\}} y_{ijt}^k + w_{it}^k \quad \forall i, k \in \{F \cup A\}, i \neq k, \ \forall (k, i) \in L_{FA}, \ \forall t \in T \tag{8}$$

$$\sum_{j \in \{F \cup A\}} y_{jit}^k = \sum_{j \in \{F \cup A\}} y_{ijt}^k + w_{it}^k \quad \forall i, k \in \{F \cup A\}, i \neq k, \ \forall (k, i) \notin L_{FA}, \ \forall t \in T \tag{9}$$

where w_{it}^k is the fraction of demand of k served by node i at time t. Constraints (8) and (9) are the conservation equations for links that remain open and those that do not, respectively. Also, flow is only permitted in the links that are part of the transportation design from farms to abattoirs,

$$y_{ijt}^k \leq x_{ij} \forall (i, j) \in L_{FA}, \quad \forall k \in \{F \cup A\}, i \neq k, \ \forall t \in T \tag{10}$$

and demand from an individual farm is served by a node only if this node is selected as an abattoir,

$$w_{it}^k \leq z_i \quad \forall i, k \in \{F \cup A\}, i \neq k, \ \forall t \in T. \tag{11}$$

Because we only care if links are open and not about their direction, we state that

$$x_{ij} + x_{ji} \leq 1 \quad \forall (i, j) \in L_{FA}, \ \forall t \in T. \tag{12}$$

To assess the trade-off between allocating resources to operations or facilities, we add the budget constraint

$$\sum_{i \in N} f_i z_i + \sum_{(i,j) \in L_{FA}} c_{ij} x_{ij} \leq B, \tag{13}$$

where B is the budget allocated to facility construction and c_{ij} is the cost of constructing link (i, j).

The objective function of the facility location sub-problem minimises the sum of the transportation costs.

$$\text{Minimise} \quad \sum_{i,j \in L_{FA}} \sum_{t \in T} \sum_{k \in K_{FA}, k \neq i} \left(\mathrm{TC}'_{ijt}{}_{ijt} y^k_{ijt} + \mathrm{TC}'_{jit} y^k_{jit} \right), \tag{14}$$

where TC'_{ijt} are the transportation costs in link (i, j) and time t.

Finally, the integrality and non-negativity constraints are

$$q_{ijtt'} \geq 0 \quad \forall (i,j) \in L_{SF}, k \neq i, \;\; \forall k \in K_{SF}, \tag{15}$$

$$y^k_{ijt} \geq 0, x_{ij} \in \{0,1\}, \quad \forall (i,j) \in L_{SF}, k \neq i, \;\; \forall k \in K_{SF}, \tag{16}$$

$$w^k_t \geq 0, z_i \in \{0,1\}, \quad \forall i,k \in \{F \cup A\}, k \neq i. \tag{17}$$

16.3.1.4 Solution Procedure

The network design and facility location problem is solved according to the following steps:

1. Solve the trans-shipment problem for the whole time horizon T to determine the number of truckloads sent from the properties q_{ijt} and the inventories in farms p_{it}.
2. Calculate the number of truckloads that must be transported from each farm to the abattoirs as

$$\mathrm{DFA}_{it} = q_{ii,t-\tau,t} \quad \forall i \in \{A \cup F\}.$$

3. Calculate the transportation cost from farms to abattoirs $TC'_{ij} = TC_{ijt} DFA_{it}$.
4. Solve the facility location problem using the transportation cost calculated in step 2 in objective function (14).

The rationale for separating the problem in two stages is as follows. The purpose of the facility location/network design problem in the second stage of the algorithm is to devise a longer-term, strategic plan to guide investment in road and abattoir construction. The funding of these is likely to come from a central source (e.g. State or Commonwealth governments), and it is of interest to policy makers. By contrast, the multi-period distribution of young animals for agistment in Step 1 of the above algorithm is funded mainly by the farmers. However, the total number of truckloads of cattle ready for processing after agistment must be calculated before agistment planning. Determining the optimal road network is critical to keep the beef supply chain operating, and it is of interest to policy makers only.

16.3.2 Data Set

The locations of a number of cattle stations in the Pilbara region are shown in Fig. 16.4. For the preliminary model presented in this paper, the amount of cattle produced by each property was simulated from a Uniform (500, 8,500) distribution and normalised so that the average is 280,000 head divided by the number of

Fig. 16.4 All properties, farms, and candidate abattoirs. Properties are marked in *green*, agistment farms in *orange*, and road junctions in *red*. Existing roads are shown in *gray*, proposed roads in *orange*. In the simulation, all the properties inside the area marked in *yellow* lose production and link with the rest of the network during the months of December, January, and February

properties. The number of heads is the current estimate of the total herd's size for this region. Most of the parameter values were taken from the studies in Strategic Design and Development, Meateng Pty Ltd (2010) and Meateng Pty Ltd (2012) for calculation.

16.4 Results and Discussion

All the results presented here were obtained using lpsolve 5.5.2.0[2] in a 64-bit Intel Xeon CPU with 2 processors of 8 cores (2.27 GHz) each and 48 GB of RAM.

To show changes in the actual designs as a function of budget, Fig. 16.6a–d show the structure of the supply network, including the selected abattoirs. In this figure,

[2] http://lpsolve.sourceforge.net, retrieved 12 May 2012.

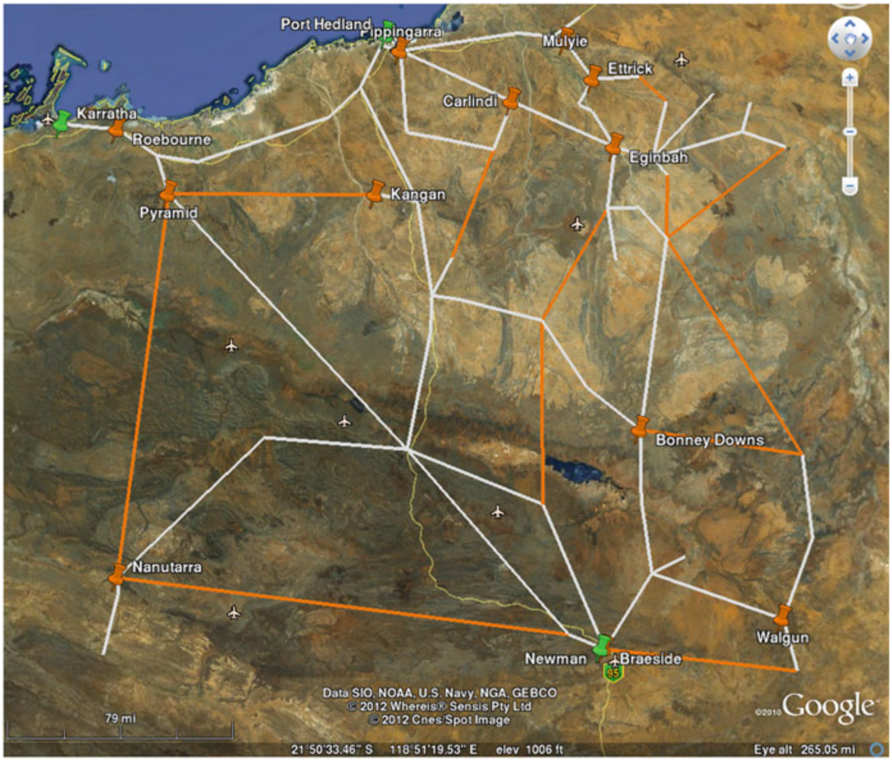


Fig. 16.5 All farms and candidate abattoirs to be used in the facility location/network design problem. Existing roads are shown in *gray*, proposed roads to be built are coloured in *orange*. All farms (marked in *orange*) are also potential abattoir sites, whereas sites marked in *green* are simply potential abattoir locations that do not produce cattle

sites selected as abattoirs are shown in red, existing roads are marked in green, and proposed roads are marked in blue. Designs a and b recommend the construction of one abattoir (Roebourne), design c recommends the construction of two abattoirs (Roebourne and Pippingarra), and design d recommends three abattoirs (Roebourne, Ettrick, and Braeside). It is interesting to note that the three sites considered by the feasibility study (Strategic Design and Development, Meateng Pty Ltd 2010) as more likely to host an abattoir in the region, namely Karratha, Newman, and Port Hedland, are very close to sites selected by the optimiser (Roebourne, Braeside, and Pippingarra). This can be appreciated in Fig. 16.6: Roebourne and Karratha are located in the northwest, Port Hedland and Pippingarra are located in the northern coast, and Braeside and Newman are in the south of the region under study. Although the closeness between the sites selected by the optimiser and those proposed by the feasibility study does not validate our model, it at least gives credibility to the magnitude of the parameters chosen.

There is one feature worth noting in these results. Figure 16.5 shows that there are not many agistment farms in the central part of the area under study (orange

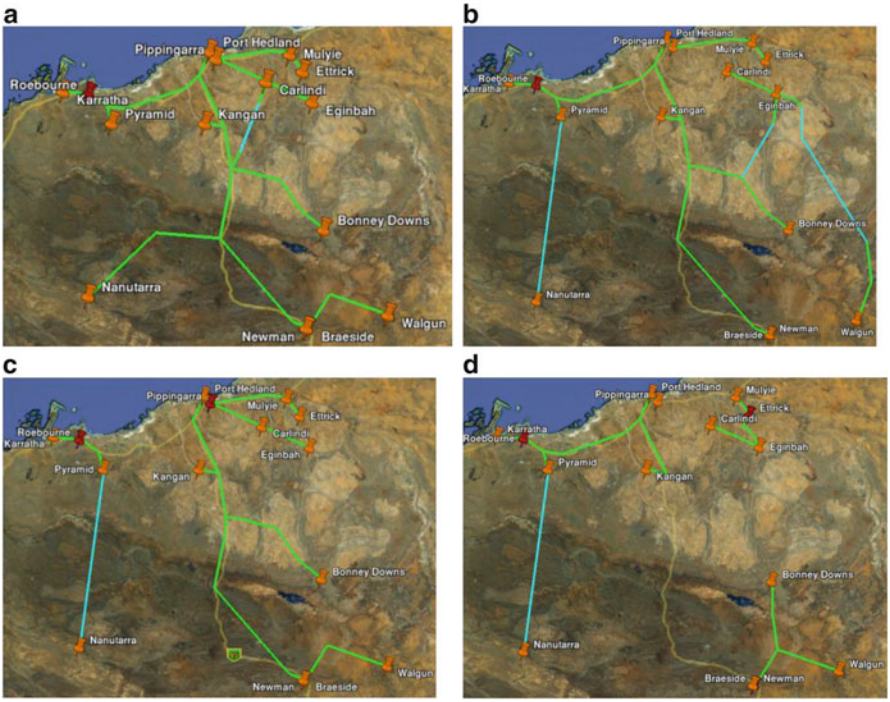


Fig. 16.6 Sites selected as abattoirs (marked in *red*) and network design. Existing roads are marked in *green* and proposed roads are marked in *blue*. The total budgets for each design are (**a**) $50M, (**b**) $93.3M, (**c**) $100M, and (**d**) 130M. Designs (**a**) and (**b**) recommend the construction of one abattoir (Roebourne), design (**c**) recommends the construction of two abattoirs (Roebourne and Pippingarra), and design (**d**) suggests building three abattoirs (Roebourne, Ettrick and Braeside)

sites). For this reason, the solution that corresponds to Figure 16.6a prescribes a road from Nanutarra to Carlindi, a path that seems counter-intuitive because the abattoir is located closer to Nanutarra. However, aggregating truckloads and transporting them through already-built roads is more economic, although at the expense of higher transportation costs (see Fig. 16.7). This result may be the product of not using accurate, realistic data. However, in an area as large as northern Australia, it is not unlikely to have very big extensions of land without sites of any type, so a more realistic model could easily produce similar results. This is something that deserves further investigation once we have access to more realistic data sets.

Figure 16.7 shows the expenses the policy maker would incur as a function of total budget. This figure provides insight into the question of resource allocation: as the total budget increases, both the transportation cost and road maintenance costs decrease, whereas more money is invested in new abattoirs and roads. According to the figure, an acceptable compromise would be attained with a budget of around $70M, although the parameter values used seem to produce a very flat curve of total costs in the region neighbouring this budget of $70M.

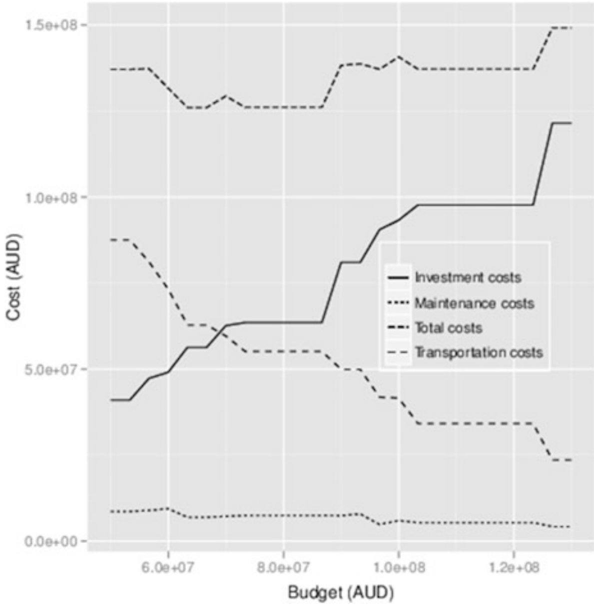


Fig. 16.7 Plot of all costs as a function of budget

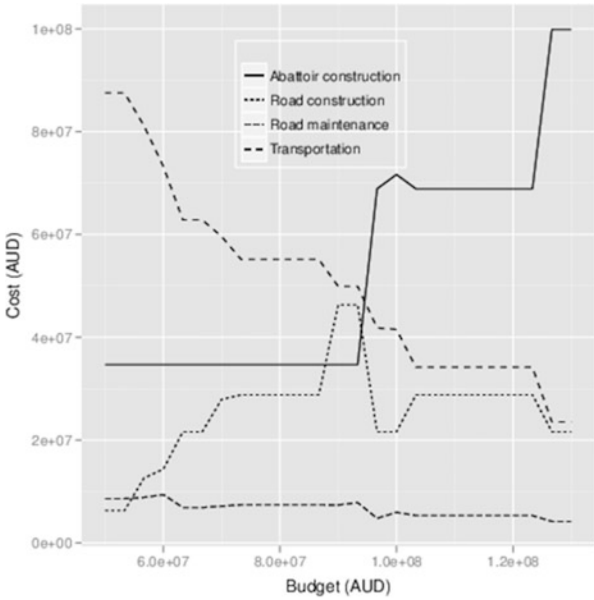


Fig. 16.8 Breakdown of all costs as a function of budget

Figure 16.8 is a breakdown of all costs. Total investment costs are made up of abattoir construction costs and road construction costs. The steps in the abattoir construction cost line show clearly that it is optimal to build a second abattoir when the budget is equal to or greater than $97M, and a third abattoir after the total budget

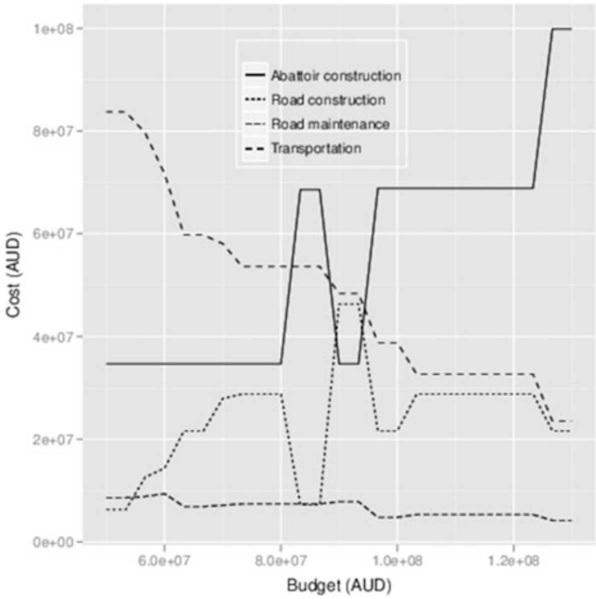


Fig. 16.9 Breakdown of all costs as a function of budget for affected coastal towns and roads during the austral summer

exceeds \$124M. However, the total investment costs in Fig. 16.8 when building a second abattoir with a budget of around \$97M are smoothed by a spike in road construction. Once there is an advantage on building a second abattoir, the construction costs suddenly decrease to increase again with the budget at around \$104M. It is also worth noting that there is a wide range where the cost components do not vary much, from around \$70M until \$87M, which explains the relatively flat payoff curve in Fig. 16.7.

Figure 16.9 shows the components of the total costs when the roads and properties in the coast of the Pilbara are affected by cyclones from December until February. In this case, the changes in the optimal investment decisions change substantially within a short budget range. The policy maker is advised:

- To build two abattoirs if the budget is of around \$84M and decrease the road construction costs,
- To build one abattoir only in case the budget is of around \$90M and invest more in-road construction,
- To build two abattoirs if the budget increases to approximately \$97M. The road construction costs decrease slightly.

Although the payoff curve for the supply chain affected by cyclones (not shown) is also flat and not too different to the payoff curve of the unaffected supply chain (Fig. 16.8), the cost breakdown is significantly different. This can be seen clearly in the budget interval spanning between \$80M and \$100M: the amounts of money are earmarked under different headings in Figs. 16.7 and 16.8 inside this interval. These are important changes that are not self-evident, and a study like the one presented here could unveil similar patterns that could potentially save the stakeholders hundreds of thousands of dollars.

16.5 Conclusions and Directions for Research

In this chapter, we have reviewed the food/agricultural and production/logistics aspects of the beef supply chain, and some commonly operational research techniques used to address them. We also presented a model of a trans-shipment and network design/facility location problem that selects segments of the road network to upgrade, and abattoirs from a set of potential sites in the Pilbara region of Western Australia. The model provides insight on the trade-off between resource allocation to facilities and links.

Cattle raising in the north of Australia is characterised by large scale enterprises on pastoral lease, low herd density, long distances to market, and significant annual interruptions to turn-off due to heat, drought, and tropical rainfall patterns. The trade of livestock from north Western Australia has recently become more vulnerable to policy and environmental changes. For instance, the dependence on live export as a market for cattle produced in the Rangelands of Western Australia has become a major source of risk to the viability of pastoral enterprises.

The model considers recommendations of two existing feasibility studies in that, first, unpredictability in supply and high labour costs caused by transporting livestock could be ameliorated if the agistment sector was developed, and second, that access to a processing facility would be of significant benefit to producers as long as this facility does not operate in competition with live exports. The former is considered in the first stage of our methodology (i.e., the trans-shipment problem), whereas the latter is a basic assumption of the model. The values of the parameters are close (for the most part) to what these feasibility studies recommend.

The results show a clear trade-off between investment and transportation costs. As observed by Pannell (2006), the pay-off curve of this primary industry's supply chain turns out to be quite flat. This is a result of aggregated investment costs; however, splitting these costs by heading provides valuable insight into long-term decisions that are not trivial. For example, there are narrow ranges of budget expenditure where it is optimal to build two abattoirs, and outside these ranges, it is optimal to build only one abattoir. Investment advice provided by this or similar models could potentially unveil savings of hundreds of thousands of dollars.

It is interesting to note that the three sites considered by the feasibility study as more likely to host an abattoir in the region, namely Karratha, Newman, and Port Hedland, are very close to sites selected by the optimiser (Roebourne, Braeside, and Pippingarra). The model presented here is preliminary and a number of issues remain to be investigated, including seasonality. For example, it has been reported that some sites, if commissioned as abattoirs, would most profitably operate as dual-species facility. This would balance out supply variations due to seasonality regarding cattle: goat supply is at its lowest in winter, when cattle supply is at its highest. The possibility of considering multiple species and seasonality will be incorporated in the model in the future.

We would like to stress that active assistance from the state and regional governments will be required to stimulate the development of the beef industry sector in the north. This assistance should balance societal factors, not necessarily relevant to the feasibility case study introduced in this chapter, which will influence the viability of new abattoirs in northern Australia. A clear example of these is the competition from the mining and resources industries to recruit suitably skilled labour in areas with relatively low population density. A regional development plan is a necessary pre-requisite for any abattoir development; the model presented is just an initial step assisting the stakeholders in devising such plan.

Finally, the abattoir selection model presented is part of a larger research effort in logistics. In addition to the problem described in this chapter, we are undertaking a number of projects in the agricultural, mining, and services industries that use network optimisation algorithms implemented in a newly developed software platform known as the Infrastructure Futures Analysis Platform (IFAP). IFAP (Fig. 16.10) makes extensive use of Geographic Information Systems, and in the near future its capabilities will be extended to answer a wide range of transport infrastructure planning questions. IFAP will be used to provide a commercial software implementation of the abattoir location problem.

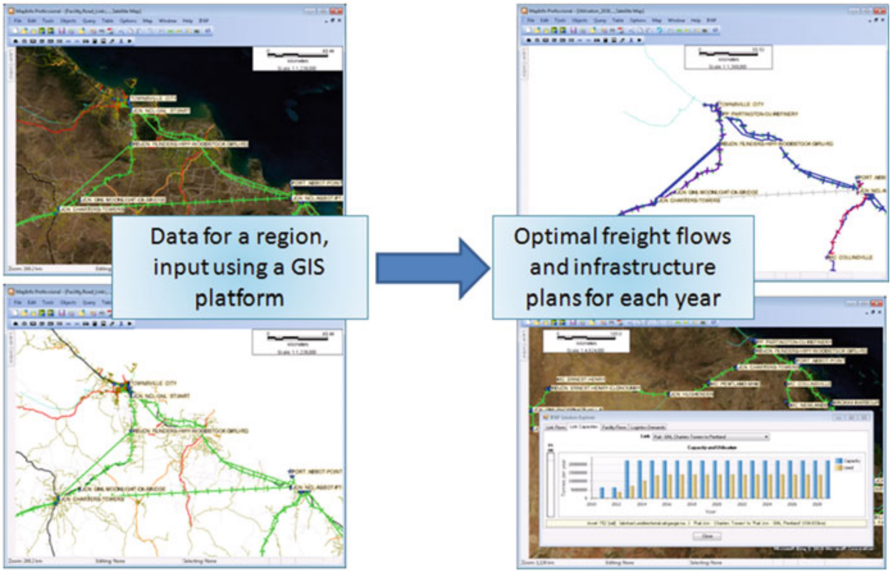


Fig. 16.10 The Infrastructures Futures Analysis Platform (IFAP) makes extensive use of Geographic Information Systems and implements network optimisation methods to determine the most cost-effective capacity, location, and maintenance sites of major transport links, interchanges, and ports in time horizons of up to 25 years

References

Ahumada O, Villalobos J (2009) Application of planning models in the agri-food supply chain: a review. Eur J Oper Res 195(1):1–20

Berman O, Krass D (2001) Facility location problems with stochastic demands and congestion. In: Hamacher H, Drezner Z (eds) Facility location: applications and theory. Springer, New York, pp 329–371

Boyabatli O, Kleindorfer P, Koontz S (2011) Integrating long-term and short-term contracting in beef supply chains. Manage Sci 57(10):1771–1787

Branco J, Caixeta-Filho J, Xavier C, Lopes R, Gameiro A (2010) Development of a mathematical model for the optimisation of multimodal transport logistic for agricultural products by Centre West Corridor. Informe GEPEC 14(1):84–100

Broek J, Schutz P, Stougie L, Tomasgard A (2006) Location of slaughterhouses under economies of scale. Eur J Oper Res 175(2):740–750

Brown C, Drynan R (1986) Plant location analysis using discrete stochastic programming. Austr J Agric Econ 30(1):1–22

Bruzzone A, Massei M, Bocca E (2009) Simulation-based case studies in logistics, Fresh-food supply chain. Springer, London, pp 127–146

Caballero R, González M, Guerrero F, Molina J, Paralera C (2007) Solving a multiobjective location routing problem with a metaheuristic based on tabu search. Application to a real case in Andalusia. Eur J Oper Res 177(3):1751–1763

Cassidy P, McCarthy O, Toft H (1970) An application of spatial analysis to beef slaughter plant location and size, Queensland. Austr J Agric Econ 14(1):1–20

Commer M (1991) A spatiotemporal analysis for fed beef in the southeastern United States. Agribusiness 7(1):71–89

Costa F, Rehman T (2005) Unravelling the rationale of 'overgrazing' and stocking rates in the beef production systems of Central Brazil using a bi-criteria compromise programming model. Agr Syst 83(3):277–295

Crosson P, O'Kiely P, O'Mara F, Wallace M (2006) The development of a mathematical model to investigate Irish beef production systems. Agr Syst 89(2–3):349–370

Dake C, Mackay A, Manderson A (2005) Optimal trade-offs between financial and environmental risks in pastoral farming. In: Zerger A, Argent RM (eds) Proceedings of the international conference in modelling and simulation (MODSIM) 2005, p 2421–2427

Department of Agriculture, Fisheries and Forestry (2012) National Food Plan Green Paper. Australian Government, Department of Agriculture, Fisheries and Forestry. Australian Government, Canberra

Department of Employment, Economic Development and Innovation (2010) The Queensland beef industry: current trends and future projections

Dimara E, Pantzios C, Skuras D, Tsekouras K (2005) The impacts of regulated notions of quality on farm efficiency: a DEA application. Eur J Oper Res 34(4–5):416–431

Domingues-Zucchi J, Zeng A, Caixeta-Filho J (2011) Optimum location for export-oriented slaughterhouses in Mato Grosso, Brazil: a dynamic mathematical model. Int J Logist Res Appl 14(3):135–148

Glen J (1986) A linear programming model for an integrated crop and intensive beef production enterprise. J Oper Res Soc 37(5):487–494

Gouttenoire L, Cournut S, Ingrand S (2011) Modelling as a tool to redesign livestock farming systems: a literature review. Animal 5(12):1957–1971

Havlik P, Jacqut F, Boisson J, Hejduk S, Vesely P (2006) Mathematical programming models for agri-environmental policy analysis: a case study from the White Carpatians. Agric Econ 52 (6):51–66

Higgins A, Miller C, Archer A, Ton T, Fletcher T, McAllister R (2010) Challenges of operations research practice in agricultural value chains. J Oper Res Soc 61(6):964–973

Hirooka H (2010) Systems approaches to beef cattle production systems using modelling and simulation. Anim Sci J 81(4):411–424

Jiang Y, Zhao L, Sun S (2009) A resilient strategy for meat-food supply chain network design. In: Proceedings of the IEEE international conference on industrial engineering and engineering management, p 1479–1483

Judge G, Hlavicek J, Rizek R (1973) A spatial analysis of the U.S. livestock economy. In: Studies in economic planning over space. American Elsevier Publishing Company, Inc., New York, p 261–273

Kall P, Wallace S (1994) Stochastic programming. Wiley, Chichester

Kearney J (2012) Food consumption trends and drivers. Philos Trans R Soc 365:2793–2807

Lucas M, Chhajed D (2004) Applications of location analysis in agriculture: a survey. J Oper Res Soc 55(6):561–578

McDermott A, Smeaton D, Sheath W, Dooley A (2005) A model of the New Zealand beef value chain: evaluating opportunities. Proc N Z Grass Assoc 67:81–86

Meateng Pty Ltd (2012) Evaluating the commercial viability of a northern outback Queensland meat processing facility. Meateng Pty Ltd., Hawthorn East

Melkote S, Daskin M (2001a) An integrated model of facility location and transportation network design. Transp Des Part A 35:515–538

Melkote S, Daskin M (2001b) Capacitated facility location/network design problems. Eur J Oper Res 129:481–495

Moloney P, Hearne J (2009) The population dynamics of converting properties from cattle to kangaroo production. In: Braddock R, Anderssen RS (eds) Proceedings of the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. University of Western Australia, Cairns, QLD, pp 561–566

Nielsen B, Kristensen A, Thamsborg S (2004) Optimal decisions in organic steer production. Livest Prod Sci 88(3):239–250

Owen S, Daskin M (1998) Strategic facility location: a review. Eur J Oper Res 111(3):423–447

Pannell D (2006) Flat earth economics: the far-reaching consequences of flat payoff functions in economic decision making. Rev Agric Econ 28(4):553–566

Piggot R, Small A, Dumsday R, Wright V (1987) Short-run costs and throughput variability for a NSW abattoir. Austr J Agric Econ 31(1):56–64

Schütz P, Tomasgard A (2011) The impact of flexibility on operational supply chain planning. Int J Prod Econ 134(2):300–311

Schutz P, Stougie L, Tomasgard A (2008) Stochastic facility location with general long-run costs and convex short-run costs. Comput Oper Res 35(9):2988–3000

Schutz P, Tomasgard A, Ahmed S (2009) Supply chain design under uncertainty using sample average approximation and dual decomposition. Eur J Oper Res 199(2):409–419

Simons D, Taylor D (2007) Lean thinking in the UK red meat industry: a systems and contingency approach. Int J Prod Econ 106(1):70–81

Stott A, Lloyd J, Humphry R, Gunn G (2003) A linear programming approach to estimate the economic impact of bovine oral diarrhoea (BVD) at the whole-farm level in Scotland. Prev Vet Med 59(1–2):51–66

Strategic Design and Development, Meateng Pty Ltd (2010) Feasibility of establishing a northern Australian beef abattoir. Australian Government, Rural Industries Research and Development Corporation. Australian Government, Canberra

Stygar A, Makulska J (2010) Application of mathematical modelling in beef herd management—a review. Ann Anim Sci 10(4):333–348

White L, Lee G (2009) Operational research and sustainable development: tackling the social dimension. Eur J Oper Res 193(3):683–692

Zhao Z, Wahl T, Marsh T (2006) Invasive species management: foot-and-mouth disease in the U.S. beef industry. Agric Resour Econ Rev 35(1):98–115

Chapter 17
Optimal Delivery of Pigs to the Abattoir

Lluís M. Plà-Aragonés and Sara V. Rodríguez-Sánchez

17.1 Introduction

During last years, the increment of competition between intensive pig producers has caused the marginal benefits per unit of product to reduce. Pig production has changed a lot during the last decade within the European Union (EU) as well as in the rest of the world. Statistics of the Food and Agriculture Organization of the United Nations (FAOSTAT 2013) has outlined the European Union (EU-28 members) as the second-largest producer in the world, after China. The evolution of the modern pork industry is a result of the global economy, advances in technology, scientific developments and changes in social and cultural attitudes (Taylor 2006; Trienekens et al. 2009). Economies of scale have continued to accelerate changes in the pork production industry (Perez et al. 2009; Ohlmann and Jones 2011). As general response, a concentration of production to maintain past profit levels is performed provoking a reduction in the number of farms although their sizes are increasing (Perez et al. 2010; Khamjan et al. 2013; Nadal-Roig and Plà 2014). Additionally, consumer concerns about environment, animal welfare, food safety and food quality are becoming new challenges for the pork industry (Backus and Dijkhuizen 2002; Rodríguez et al. 2014). As a result, the profile of the typical farm is changing from a family-based, small-scale, independent firm to one in which larger firms are more tightly aligned along the pig production and distribution processes (Perez et al. 2010; Rodriguez 2010).

Traditionally, judgement based on experience had been the basis for the production planning on individual farm units, but the increasing complexity of

L.M. Plà-Aragonés (✉)
Department of Mathematics, University of Lleida, Jaume II, 73 E-25001, Lleida, Spain

S.V. Rodríguez-Sánchez
Graduate Program in System Engineering, Universidad Autonoma de Nuevo Leon, 111-F, Ciudad Universitaria, San Nicolas de los Garza, NL 66450, Mexico

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_17

the pork industry made the development of more formal planning methods necessary (Boland et al. 1993; Plà 2007; Rodríguez et al. 2014). The usual planning problem in fattening pig units is related with the optimal delivery to the abattoir of fattened pigs. Visual inspection is used for selecting pigs to be delivered as the cheapest method, whilst novel information technology solutions remain expensive for this purpose (Plà-Aragonés et al. 2013). Hence, the objective of this paper is to formulate a mixed-integer linear programming model (MILP) describing a fattening pig unit operating under all-in-all-out (AIAO) strategy. The interest of the model is to maximise the revenue from deliveries of pigs according to a carcass classification. It is assumed that the farm is vertically integrated and all production is sold to an abattoir prescribed by the company into which the fattening farm is integrated.

The organisation of this chapter is as follows. We present in Sect. 17.2 a brief description of the problem highlighting the role of the growing process. In Sect. 17.3, we describe the proposed model whilst in Sect. 17.4, we present main computational results that are discussed in Sect. 17.5. Finally, we present the conclusions in the last section.

17.2 Fattening Pigs and the Delivery to the Abattoir

Fattening pig farms are the place where pigs are fattened before being sent to the abattoir. Therefore, these farms can be seen as growing places for pigs till they reach a marketable weight. According to Whittemore and Kyriazakis (2006), the live weight of slaughtering pigs in Europe is around 115 kg. Nevertheless, slaughtering weight can vary considerably according to the country, region and abattoir. The same authors value this variation at the 30–50 % of the mature size. For instance, in Spain, slaughtering weight is around 100 kg, but if pigs are intended to produce cured products (e.g. ham or chorizo) or dealing with the Iberian pig breed, then the slaughtering weight is higher reaching the 140 kg.

The most extended management of fattening farms is the AIAO management. AIAO management is practised by the majority of producers with large facilities or as a part of a vertically integrated company. As a result, the entire facility is emptied for cleaning before replacing the actual batch of pigs with a new one (Ohlmann and Jones 2011). The main advantage of AIAO management is the disease control and prevention. The easier cleaning and disinfection of facilities between batches makes difficult the spread of illness and disease. In practice, this means all young pigs are entered at the same time in the farm and sold also at the same time within a narrow time window. Once the farm is empty, a cleaning and drying period of a week as maximum follows before a new batch of young pigs arrive. The farmer must determine when to send pigs and deliver them to the abattoir because not all of them reach marketable weight at the same time even though they have the same age.

Spanish pig production is highly specialised and mostly controlled by big companies or cooperatives, namely, integrators (Ouden et al. 1996; Rodríguez et al. 2014). Then, optimal delivering policies of pigs for individual farms may be different when operating as an independent producer or as an integrated one. Integrated farmers has less control about the composition of the batch of pigs in terms of number of animals, suppliers, weight distribution at the beginning, feed-stuffs, maximum duration of fattening, etc. Differences in marketing strategies are mainly due to different weight distributions at the beginning of the marketable time window. Additionally, consumer concerns about environment, animal welfare, food safety, food quality and healthier diets have led to a grading system of payment prioritising leaner carcasses (Plà-Aragonés et al. 2013; Rodríguez et al. 2014). Hence, the determination of when to start to deliver pigs of a batch to the abattoir and which and how many pigs to sell is the central problem to the pig manager.

We model the decision-making problem of delivering pigs to the abattoir as an MILP that determines the marketing strategy that maximises expected profit. By discretising the population into appropriate growth clusters, meat quality and carcass weight categories, we formulate a mixed-integer programme. We solve the problem for one individual farm considering the full fattening process. Additional expected benefits of the model are the analysis of different marketing windows, transport costs and homogeneity in the growth of the batch of pigs.

17.3 Modelling a Fattening Pig Farm

17.3.1 Approaching the Life Weight of a Group of Pigs

The fattening process is basically a growing process of pigs. Many mathematical models are used in order to describe the growth of domestic animals in the attempt to predict optimal slaughter time and weight. As growing is not equal for each individual, a batch of pigs at the same age has a natural mean weight and variation. Our approach is based on a discretisation of a given regression growth curve (Castro 2001; Plà-Aragonés et al. 2013). Hence, the regression provides the normal weight distribution over time: mean and standard deviation of the live weight for a given herd of pigs at the same age. The growth curve is based on experimental data representing a particular breed and fitted for male pigs. If we assume a normal distribution of live weight, $W \sim N(\mu, \sigma^2)$, then the truncated normal distribution of W between A and B, $TW \sim N\left(\mu_{A-B}, \sigma^2_{A-B}, A, B\right)$, has as expectation:

$$\mu_{A-B} = E(TW) = \mu + \frac{\sigma\left(\varphi\left(\frac{A-\mu}{\sigma}\right) - \varphi\left(\frac{B-\mu}{\sigma}\right)\right)}{\Phi\left(\frac{B-\mu}{\sigma}\right) - \Phi\left(\frac{A-\mu}{\sigma}\right)} \tag{17.1}$$

where φ is the standard normal probability density function and Φ the cumulative distribution function of a normal standard distribution. Hence, the truncated normal allows us to calculate the expected live weight of a group of pigs falling into a specific weight range within a herd.

17.3.2 *The Model for Optimal Delivery*

The decision variables involved in the problem are related either to the number of pigs present on the farm (i.e. the inventory over time) or those sold to the abattoir forcing to update the inventory on the farm. Complementary to that, the number of trucks to deliver pigs can be considered. The nature of these decision variables is integer: pigs or trucks. Other logical operation rules of pig deliveries to the abattoir have to be taken into account. For instance, weight distribution and visual inspection of pigs lead to consider the selection of heavier pigs to be delivered to the abattoir. There would not have sense the delivery of lighter pigs to the abattoir being some heavier ones present. Modelling these rules may lead to consider additional variables. The formulation of the complete model follows.

17.3.2.1 Set and Indexes

$P = \{i\}$ set of partitions applied to the batch of pigs, P. Each individual belonging to a partition is assumed that will grow according to similar parameters. Thus, this subscript represents a growth category preserved along the fattening period.

$T = \{t\}$ set of periods of time in which the fattening phase is divided, i.e. number of weeks.

17.3.2.2 Parameters

N: number of pigs in a batch.

$|P|$: number of groups in which the herd is divided.

$W_{\mathrm{t}} \sim N(\mu_{\mathrm{t}}, \sigma_t^2)$: live weight distribution of the batch at week t.

$TW_{\mathrm{ti}} \sim N(\overline{w}_{ti}, \sigma_{ti}^2, i-, i+)$: truncated live weight distribution of the partition i at week t, being $i-$ and $i+$ the extreme weights bounding the partition.

Cct: load capacity of trucks in number of fattened pigs covering the path to the abattoir.

Lw: averaged maximum live weight of loaded pigs to calculated the maximum load capacity of trucks in kg of weight.

Pv: is the base selling price in € per kg of live pig.

Bonus (gender, % lean$_{it}$, wcarcass$_{it}$): is a bonus if positive, or a penalisation if negative, on the base selling price that the abattoir computes depending on carcass weight and the lean percentage.

Gender: can be male, female or castrated.

% Lean$_i$: lean composition of a pig estimated from the live weight of pigs.

wcarcass$_i$: carcass weight estimated from the live weight of pigs.

p_0: represents the unitary cost of purchasing a growing pig to be fattened.

Ct: is the unitary cost for one trip to the abattoir.

$\overline{Cf}_t$: is the accumulated cost of feedstuff or concentrates consumed on average by a pig till the t-week.

k_{it}: represents the unitary cost associated to other costs per sold pig.

K: fixed costs associated to a batch like cleaning and disinfection.

17.3.2.3 Decision Variables

X_{it}: number of pigs at growth stage i, sent to the abattoir at week t.

N_{it}: inventory of pigs at growth stage i and week t.

h_{it}: binary variable, $\{0, 1\}$, representing two consecutive growth stages ($i-1$ and i) sending pigs to the abattoir at week t when it takes the value 1 or 0 otherwise.

d_{it}: binary variable, $\{0, 1\}$, representing whether pigs at growth stages i and time t are sent to the abattoir when it takes the value 1 or 0 otherwise.

Y_t: integer variable representing the number of trucks needed to transfer pigs from the farm to the abattoir at week t.

The proposed deterministic model maximises the total profit of the production and delivery of a batch of fattened pigs to the abattoir, and the feasible solutions must satisfy a set of constraints that mainly concern the population dynamic behaviour, given by the following optimisation problem:

$$
\begin{aligned}
\text{Maximise R} &= I(X_{it}) - C(X_{it}) \\
&= \sum_{it} X_{it} \cdot \overline{w}_{it} \cdot (\text{Pv} + \text{Bonus}(\text{gender},\ \%\ \text{lean}_{it},\ \text{wcarcass}_{it})) \\
&\quad - \left(\text{p}_0 \cdot \text{N} + \sum_t \text{Ct} \cdot Y_t + \sum_{it} \left(\overline{Cf}_t \cdot X_{it} \right) + \sum_{it} (k_{it} \cdot X_{it}) + K \right)
\end{aligned}
\tag{17.2}
$$

s.t.

$$N_{i1} = \left(N / |P| \right) \qquad \forall i \in P \tag{17.3}$$

$$N_{it+1} = (N_{it} - X_{it}) \qquad \forall i \in P,\ \ t \in \text{T} \backslash \{|\text{T}|\} \tag{17.4}$$

$$X_{it} \leq N_{it} + N(1 - h_{it}) \qquad \forall i \in P,\ \ t \in \text{T} \tag{17.5}$$

$$X_{it} \geq N_{it} - N(1 - h_{it}) \qquad \forall i \in P,\ t \in \mathrm{T} \tag{17.6}$$

$$N_{i|\mathrm{T}|} - \mathrm{X}_{i|\mathrm{T}|} \leq 0 \qquad \forall i \in P \tag{17.7}$$

$$\sum_{i=1}^{|P|} X_{it} \leq Cct \cdot Y_t \qquad \forall t \in \mathrm{T} \tag{17.8}$$

$$\sum_{i=1}^{|P|} X_{it} \cdot \overline{w}_{it} \leq Cct \cdot Lw \cdot Y_t \qquad \forall t \in \mathrm{T} \tag{17.9}$$

$$X_{it} \leq N \ \cdot d_{it} \qquad \forall i \in P,\ t \in \mathrm{T} \tag{17.10}$$

$$h_{it} \leq d_{it} \qquad \forall i \in P,\ t \in \mathrm{T} \tag{17.11}$$

$$d_{it} + d_{i+1t} \leq 1 + h_{i+1t} \qquad \forall i \in P \backslash \{|P|\},\ t \in \mathrm{T} \tag{17.12}$$

The objective function (17.2) accounts for income and cost during a fattening period, involving one batch of fattening pigs. The income, I, is generated by selling the batch of fattened pigs to the abattoir. The cost, C, is calculated from the summation of four concepts: purchase of growing pigs, transport to the abattoir, feed consumption over the fattening period and other costs including the cleaning or disinfection of facilities. The relevant decision variables are X_{it} (how many pigs in the category growth i at week t have to be sent to the abattoir) and Y_t representing the number of trucks needed the week t to perform the transport.

The feasible solutions of the model have to satisfy a set of constraints (17.3–17.12). It is assumed that a batch of N growing pigs is introduced in the fattening farm and kept fattening till T weeks as maximum. Animals at the beginning of the growing process, $t = 1$, have a weight following a normal distribution, $W_1 \sim N(\mu_1, \sigma_1)$. This distribution is partitioned into $|P|$ percentiles representing each one different growing categories ($i \in P$). By default, at the beginning of the process, an equal number of pigs in each category are considered (17.3). The truncated normal (17.1) allows us to derive conveniently the expected live weight of each category week by week, $\overline{w}_{ti}$. The model represents the growth and flow of pigs over time. The model assumes that pigs belonging to specific growing categories (initial partition of the population) do not change over time. In principle, once deliveries to the abattoir start, pigs growing to the next week (i.e. N_{it+1}) are those present in the current week minus sales to the abattoir (i.e. N_{it}–X_{it}). The equality constraint (17.4) can be relaxed if casualties are considered from week to week. The model does not impose specific constraints about when pigs have to be sent to the abattoir (17.4). It is the grid of bonus and penalties paid by the abattoir that regulates implicitly the opening of the marketing window for fattened pigs. This is so because pigs with a low weight are penalised and even not accepted by abattoirs. The number of pigs available to be sent to the abattoir, X_{it}, is bounded by N_{it}, i.e. the inventory of pigs in i-growth category at week t. The binary variable h_{it} is introduced (17.5 and 17.6) to detect when pigs in i and ($i-1$)-growth category at week t are sent to the abattoir, then $h_{it}=1$, otherwise $h_{it}=0$. This way, we can represent the rational behaviour of the farmer: all pigs remaining in the i-growth category at week t has to be sold before pigs of the lighter growth category ($i-1$) can be also selected. Hence a complementary constraint (17.6) is necessary. Note that N could be replaced by any arbitrary bigger number without affecting the functionality of (17.5 and 17.6). As AIAO strategy is considered an additional constraint, (17.7) for the last week of the process has to be set assuring the emptying of the facility. Note that for a good modelling of the system (17.7) makes necessary to consider a fattening period, i.e. set T of weeks, big enough to represent the duration of the whole fattening process, unless T is fixed for management reasons.

The number of pigs in i-growth category at time t sold to the abattoir, X_{it}, must be loaded and transported to the abattoir in trucks, Y_t, with a limited capacity (17.8). It is usual to consider capacities between 200 and 240 pigs depending on individual live weight of the load. A complementary constraint (17.9) can be considered regarding the load weight capacity of trucks. Then, the total weight of the load, i.e. the live weight of pigs, has to be considered. Let us note that constraint (17.9) could be redundant in regular fattening unit when only fattened pigs are delivered to the abattoir because load capacity of trucks is far enough. A different situation is met when trucks have to transport adult animals like culled sows or boars from breeding farms. They are much heavier and this constraint can be relevant for sow farms, but this consideration is out of the scope of the present study.

The delivery of pigs has to follow several rational rules imposed by the fact that at present no individual measures of weight are available in most farms. In that case, the heaviest pigs are the first to be selected for loading a truck and being transported to the abattoir. Remaining pigs have the chance to keep growing further and being selected later for the next delivery and so on till the farm is emptied. This aspect is modelled assisted by binary variables and a set of constraints. Hence, the binary variable h_{it} has been introduced to detect if two consecutive groups of growth categories are sending pigs to the abattoir. Another auxiliary binary variable, d_{it}, was also introduced. It was intended to detect if a group of pigs in i-growth category at time t is sent or not to the abattoir (17.10). Note that N could be replaced by any arbitrary bigger number without affecting the functionality of (17.10). We complement this constraint (17.10) with two more giving full sense to the binary variable h_{it}. As result, the introduction of constraints (17.11) and (17.12) provokes that deliveries of intermediate categories are not feasible, unless all the heavier ones are sent.

17.4 Computational Results

The algebraic modelling language IBM ILOG OPL Studio was used with CPLEX 12.4 as the linear optimisation solver for implementing and solving the different instances developed for this case study in a laptop computer (Dual-Core i5 CPU at 2.5 GHz and 4Gb RAM). Microsoft Excel has been used for storing data, both input parameters and outputs. All the cases were solved in few seconds and results were reported in a spreadsheet for easy inspection. The case presented here is based on a typical Spanish fattening farm integrated into a private pig company or integrator who provides piglets, feedstuff, veterinary care and medicines, technical advice during the fattening process and regular control over the growth of animals. The company owns the abattoir where pigs are slaughtered and determines the pricing grid to reward integrated farmers. Furthermore, the integrator fixes also the maximum duration of fattening according to the production plans of the company, including the procurement to the abattoir and the supply of piglets to conform a new batch to be fattened. Biological parameters related basically with growth and consumption are referred to a crossbreed of (Large White × Landrace) × Pietrain.

17.4.1 Parameters of a Case Study

Let us consider a batch of pigs being fattened on a farm during a fattening period under AIAO management. Fattening comprises from the arrival of the first young pigs to the farm (with pigs weighing around 20 kg) till the week the last pig of the batch is delivered to the abattoir. Thereafter, the farm can be prepared for another incoming batch of pigs. The purchase value of a young pig was of 40.55 €/pig in average for 2013 (DARP 2014). The duration of the fattening phase typically ranges from 14 to 20 weeks resulting in approximately 2 or 3 cycles per year, being $T = 17$ the maximum range selected for this instance. The duration of the fattening period has to be also fixed according to the breed selected, the achievable market weight and the specific growing traits. The objective of our model is to determine the delivering policy, i.e. marketing strategy, that maximises the profit of a batch of fattened pigs. In our case, we consider a fattening farm with a housing capacity of $N = 1{,}000$ pigs. Five groups (two deciles) of 200 pigs ($N/|P|$) are considered to split the total population ($|P| = 5$).

Pig producers are paid for their pigs based on carcass weight and predictions of fat-free lean. Each abattoir uses a unique grid to establish discounts or premiums, according to their percent of lean estimates. Pig carcass classification is regulated across the EU. The SEUROP classification was applied by abattoirs since 1984 (Regulation 3220/84). Later on, the present legislation was updated incorporating more details to previous regulation on carcass classification and commercial grading (Commission Regulation (EC) No.1249/2008).

In view of producing uniformly sized lean products, abattoirs specify a bonus or penalisation of €/kg based on the percentage of lean as quality indicator (Table 17.1). Other common penalisation applied depends on carcass weight when it is out of range, the gender of animals (male, female or castrated), live weight or other traits that the abattoir may consider to fit better pig meat demand. In the end, bonus or penalisations are applied in order to motivate producers to deliver pigs in homogeneous groups and offer a homogeneous product to end customers. The appropriate delivering strategy and the corresponding revenue will depend on the price grids of the abattoirs (see Table 17.2). Base prices are

Table 17.1 Carcass classification according to lean percentage and an example of bonus applied by an Spanish abattoir

% lean	Classification	Bonus (€)
>60	S	+0.12
55–60	E	+0.07
50–55	U	0.00
45–50	R	−0.07
40–45	O	−0.12
<40	P	−0.30

Table 17.2 Bonus and penalisation of €/kg of carcass is a function of weight and percentage carcass lean. The base considers a 81 % of carcass weight over live weight

Carcass	Bonus	S	E	U	R	O	P
Weight (kg)	(€)	0.12	0.07	0.00	−0.07	−0.12	−0.3
50	−0.60	−0.48	−0.53	−0.6	−0.67	−0.72	−0.90
60	−0.40	−0.28	−0.33	−0.4	−0.47	−0.52	−0.70
65	−0.23	−0.11	−0.16	−0.23	−0.30	−0.35	−0.53
67.5	−0.13	−0.01	−0.06	−0.13	−0.20	−0.25	−0.43
70	−0.05	0.07	0.02	−0.05	−0.12	−0.17	−0.35
72.5	−0.04	0.08	0.03	−0.04	−0.11	−0.16	−0.34
75	0.00	0.12	0.07	0.00	−0.07	−0.12	−0.30
80	−0.05	0.07	0.02	−0.05	−0.12	−0.17	−0.35
85	−0.10	0.02	−0.03	−0.10	−0.17	−0.22	−0.40
97.5	−0.02	0.10	0.05	−0.02	−0.09	−0.14	−0.32
102.5	−0.05	0.07	0.02	−0.05	−0.12	−0.17	−0.35
107.5	−0.11	0.01	−0.04	−0.11	−0.18	−0.23	−0.41

agreed in auctions markets weekly and used by abattoirs being Mercolleida (http://www.mercolleida.com) the common reference for Spain. The annual mean in 2013 used for this instance is $Pv = 1.377$ €/kg of live pig. Figure 17.1 shows the evolution of the base price during 2013 in Spain that can be considered as good (i.e. over the mean of recent past historical prices). These figures have kept the sector out of the general crisis other Spanish economic sectors are experimenting at present.

A delay in the optimum delivering time of pigs implies an increment in the feeding cost whilst growth rate is getting worse, with the risk of getting fatter carcasses. Whether the weight of pigs overtakes specific ranges, they depose too much fat, i.e. the lean percentage is reduced, and penalisations are applied. Furthermore, a long duration of the fattening phase is neither interesting as the annual number of batches the farm produces decreases. And so, the annual profit per kg of meat produced will decrease. Otherwise, an anticipatory delivery generates penalisations mainly in terms of carcass weight out of range and a waste of better growing rates and efficient conversion of feed into meat.

A shortcoming of the process is that no objective measure of weight is available from individual animals and the pig weight is assessed by eye before making a

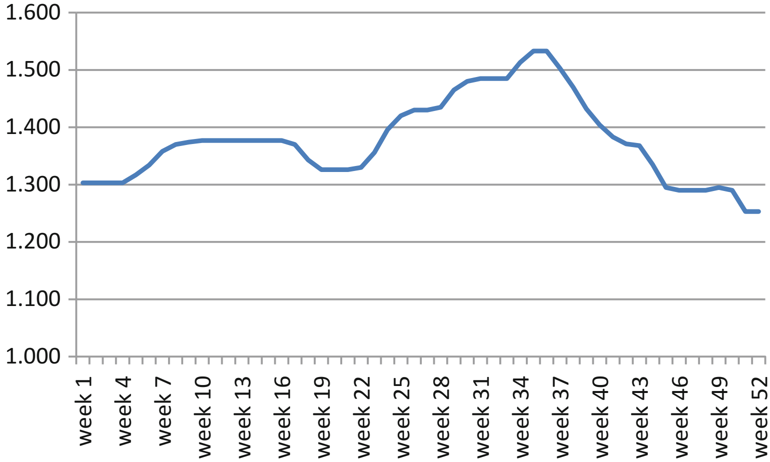


Fig. 17.1 Price (€/kg) for pigs in 2013 published in Mercolleida, the main pig auction market in Spain (DARP 2014)

decision (Plà-Aragonés et al. 2013). This fact makes impossible to select animals with an intermediate weight because the risk of error is very big. Only after slaughtering, the abattoir provides the individual measures per pig of relevant traits (e.g. live weight, carcass weight, lean and fat composition); however, the correlation between relevant variables as live weight, carcass weight and lean percent is quite high and useful to estimate the percentage of lean of pigs (Castro 2001). In general, pigs are delivered over a period of few weeks (marketing time window) using trucks with a maximum capacity of 240 (Cct) pigs and having a cost per trip of 475 € (Ct). It is of interest to determine the timing of these deliveries near the end of the fattening phase to maximise expected profit given the natural variability of growth of the batch. In this context, the growth model of pigs is essential for this problem. To this respect, we follow the proposed model by Castro (2001) and derived from experimental data (Fernández et al. 2011) with individual controls for hybrid pigs selected for meat production. Note that other breeds should require the calibration of the corresponding growth curve to be used in the proposed model.

Besides considering specific growth curves, the total feed intake of pigs is also relevant to calculate the feeding cost and the consumption over time, in particular for remaining animals in the farm when deliveries to the abattoir start. A unitary feeding cost of 0.28 €/kg of feedstuff is taken into account to calculate the total feeding cost of a pig sold the t-week $\left(\overline{cf_t}\right)$. The distributions over time for live weight and for the accumulated feed intake are shown in Fig. 17.2.

Load capacity of trucks has to respect European regulation imposing minimum room per animal depending on live weight and distance. The common practice is delivering pigs when there are enough in number and within an optimal weight range to fill a truck. Normally, it ranges from 200 to 240 fattened pigs of about 100 kg.

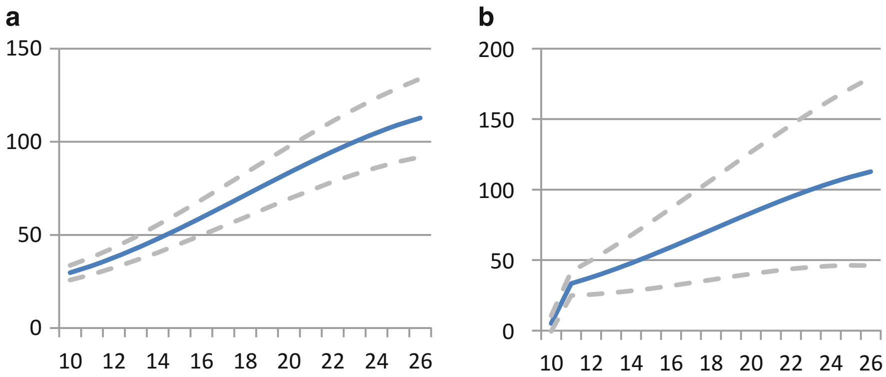


Fig. 17.2 Mean live weight and accumulated feed intake over time, valid for the range of 10–26 weeks of pig lifespan, according to Castro (2001). (**a**) Mean live weight and ±SD. (**b**) Mean feed intake and ±SD

17.5 Results

The proposed model with the parameters presented in the previous section was solved. Reasonable results were obtained under practical point of view as shown in Table 17.3. The maximum revenue was of 23.744 € or 23.74 €/pig or 0.224 €/kg of live weight sold. This result was a 54 % better compared with the delivery of all the pigs on the last week of the marketing window (15.47 €/pig). According to what was modelled, pigs sent to the abattoir are always the heaviest present on the farm. Trips to cover the path to the abattoir were 5 in total from which the number of trucks needed was easily derived. On the other hand, the marketing time window opened at week 22 and closed at week 25. So, the optimal duration of the fattening stage was of 16 weeks as maximum (from week 9 to week 25) plus an additional week for cleaning and drying.

Other complementary results were obtained for extreme situations. For example, the optimal solution of the problem when the cost of transport was extremely high was the same as that shown in Table 17.3, and this situation means that the model already provides the maximum income from the fattened pigs even though when total cost arises. However, a change was observed when the unitary cost of transport was reduced enlarging from week 22 to 25 the optimal marketing time window. In all the cases computed, the mean live weight of pigs sent to the abattoir is around 100 kg, which is a regular practice in Spain.

The abattoir prefers uniformly sized fattened pigs and establishes corresponding marketing system that penalises carcasses with weights and lean composition outside a desired range in order to motivate farmers to deliver pigs in homogeneous batches (Boland et al. 1993). Thus, an additional analysis was performed to observe the impact of reductions in the live weight variability. A hard reduction was applied to the standard deviation of live weight at each growth category. In this case, the delivery of pigs to the abattoir was sending all of them the last week to the abattoir.

Table 17.3 Delivery of pigs (X_{it}) when binary variables h and d are included in the model

$(i, t+9)$	22	24	25
1	0	0	200
2	0	0	200
3	0	0	200
4	0	100	100
5	100	100	0

Table 17.4 Maximum revenue per batch and per day, depending on the fattening period length (T) plus 1 extra week for cleaning

	Revenue (€)	
T (weeks)	Per batch	Per day
17	23.744	188.44
16	23.744	199.53
15	22.558	201.41
14	20.756	197.68

In that case, the consideration of different categories of pigs would not have sense and all the batches are treated as one unity. In this sense, the optimal profit and the average weight of animals sent to the abattoir were higher as weight distribution was more concentrated around the mean. A similar result was obtained if no bonus was considered on pig sales. The optimal policy in that case was to sell all pigs at week 26 with a similar averaged weight (112 kg) and a profit per batch of 29.53 € per pig. Furthermore, if no price grid is considered, the optimal solution of the model indicates the trend to deliver fatter and weightier pigs.

The results presented previously correspond to a fattening period of 17 weeks. As no pigs are sent the last week, it is expected to have the same result even if that period is limited to 16. However, it could be questioned if a shorter period would be better. To answer this question, different instances varying the value of T were calculated as shown in Table 17.4. In addition, as the model calculated only the revenue per batch, the resulting revenue per day is computed. The farmer intends to keep producing pigs, not just one batch, so the maximum revenue per day is a more appropriate criterion to optimise delivering policies. Thus, the maximum revenue per batch is observed when $T \geq 16$, but if we add the time needed for cleaning and disinfection (on week more), then the maximum revenue per day corresponds to a $T = 15$, representing a 1 % of improvement. However, there is a 6 % of improvement when going from a $T = 17$ to the one $T = 16$.

17.5.1 General Discussion

The solution proposed by the model can be performed in practice. The selection of fattened pigs by eye is feasible when the heaviest pigs are the first to be sent to the abattoir (Plà-Aragonés et al. 2013). On the contrary, selecting pigs from one category not adjacent to another as other authors propose (Khamjan et al. 2013) would be difficult without individual measures of weight.

The optimal marketing time window of 4 weeks is in agreement with practical recommendations of most companies in Spain. However, compared with other results in literature, the length of the marketing window differs likely because of different objective function, growth model and grid of payment (Plà-Aragonés et al. 2013). For instance, Ohlmann and Jones (2011) used transition probabilities to represent the growth of animals whilst in this paper a specific growth model and its discretisation is proposed to better estimate herd distribution of live weight over time and make decisions accordingly. However, the differences are not very big.

The optimal solution when no bonus was considered on pig sales indicated the trend to deliver fatter and weightier pigs (result not shown). This result is in agreement with Boland et al. (1993) and Ohlmann and Jones (2011) who stated that price grids provoke changes in marketing policies. A similar situation is observed when variability in live weight distribution is reduced (Plà-Aragonés et al. 2013). More variability in live weight provokes a wider marketing window (Boland et al. 1993), whilst no variability would allow the farmer to deliver all the animals at a time (the smallest time window).

Different results were reported by Pla et al. (2013) using a similar approach. These differences are explained in part by the set of parameters and price grid reflecting the better situation for the pig sector in Spain than before. For instance, at present, the sector allows the farmer to sell all the pigs and slower growing pigs to be marketed did not represent a problem. It is known and assumed that the model proposed incurs errors because we do not have individual measures to allow the application of the model individually. However, if this was possible, the optimal delivery of individual pigs to the abattoir could be determined with additional gains. This has been confirmed by several papers in literature considering individual measures on live weight and consumption (Kure 1997; Kristensen et al. 2012). Individual measures would allow the farmer to value more precisely the feed cost and the expected profit of pigs individually making eventually a more informed decision. However, the feasibility of these solutions will depend on the practical means available to select and keep apart these pigs at the moment of loading.

17.6 Conclusions

Modern pig farming is changing and pig farms are becoming more and more specialised. In addition, whilst the size of pig facilities is increasing, the number of farms is decreasing. Pig farms have tended to integrate and coordinate their operations into vertically integrated companies. This integration and coordination affects the decision-making process at each pig production unit, in particular on fattening farms. So, this paper presents a mixed-integer linear programming model description of a fattening pig unit operating under AIAO strategy. The discrete growth model approach presented assumes flexible growing categories based on a variable partition of the herd. The results presented confirm preliminary outcomes found in the literature advocating for the benefit of implementing a price grid to get different qualities of carcass. The reduction on variability at the entry of the process permits to reduce the marketing window of pigs and rises the efficiency of the process. It is also shown how a time window of 5 weeks delivering animals to the abattoir suffices to empty the farm and prepare it for the next batch of animals. The optimal result per batch does not correspond with the optimal result per day. The latter would imply saving 1 week in the time window and increments of 5 % in the daily or annual revenue. Summarising, our contribution confirms the findings of past studies and envisions the importance of future trends relying on individual measures to avoid inefficiencies related to managing grouped animals.

Acknowledgement Lluís M. Plà wishes to acknowledge the financial support of the Spanish research program (MTM2009-14087-C04-01).

References

Backus G, Dijkhuizen AA (2002) Kernkamp lecture: the future of the European pork chain. Allen D. Leman Swine Conference, Minnesota, pp 8–11

Boland MA, Preckel PV, Schinckel AP (1993) Optimal hog slaughter weights under alternative pricing systems. J Agric Appl Econ 25(2):148–163

Castro R (2001) Optimització del process d'engreixament porcí. Un enfocament operatiu. Ph.D. Dissertation. Department of Industrial Engineering, University of Girona, Spain. http://www.tdx.cat/handle/10803/7710. Accessed 20 Apr 2013

DARP (2014). Statistics and surveys of agricultural production in Catalonia. Livestock prices downloadable from the webpage http://www20.Gencat.Cat/Docs/Dar/De_Departament/De02_Estadistiques_Observatoris/20_Observatori_Agroalimentari_De_Preus/03_Preus/01_Preus_En_Origen/01_Preus_En_Llotges_I_Mercats/Fitxers_Estatics_Nou/Pllotja_Ramaders/Catalunya/Po_Cat_Ramaders_2013.Pdf. Accessed 23 Jul 2014

FAOSTAT (2013). Statistics of livestock primary in the section of production. Downloadable from the webpage http://faostat.fao.org/site/573/default.aspx#ancor. Accessed 23 Feb 2013

Fernández J, Fàbrega E, Soler J, Tibau J, Ruiz JL, Puigvert X, Manteca X (2011) Feeding strategy in group-housed growing pigs of four different breeds. Appl Anim Behav Sci 134 (3–4):109–120

Khamjan S, Piewthongngam K, Pathumnakul S (2013) Pig procurement plan considering pig growth and size distribution. Comput Ind Eng 64(4):886–889

Kristensen AR, Nielsen L, Hielsen MS (2012) Optimal slaughter pig marketing with emphasis on information from on-line live weight assessment. Livest Sci 145:95–108

Kure H (1997) Marketing management support in slaughter pig production. Ph.D. Dissertation. Department of Animal Science and Animal Health, The Royal Veterinary and Agricultural University, Copenhagen

Nadal-Roig E, Plà LM (2014) Multiperiod planning tool for multisite pig production systems. J Anim Sci 92(9):4154–4160

Ohlmann JW, Jones PC (2011) An integer programming model for optimal pork marketing. Ann Oper Res 190(1):271–287

Ouden M, Dijkhuizen AA, Huirne RBM, Zuurbier PJP (1996) Vertical cooperation in agricultural production marketing chains, with special reference to product differentiation in pork. Agribusiness 12:277–290

Perez C, de Castro R, Font i Furnols M (2009) The pork industry: a supply chain perspective. Br Food J 111(3):257–274

Perez C, de Castro R, Simons D, Gimenez G (2010) Development of lean supply chains: a case study of the Catalan pork sector. Int J Supply Chain Manag 11:271–280
Plà LM (2007) Review of mathematical models for sow herd management. Livest Sci 106:107–119
Plà-Aragonés LM, Rodriguez-Sanchez S, Rebillas-Loredo V (2013) A mixed integer linear programming model for optimal delivery of fattened pigs to the abattoir. J Appl Oper Res 5:164–175
Rodriguez SV (2010) Models under uncertainty to support sow herd management in the context of the pork supply chain supply. Ph.D. Dissertation. Department of Mathematics. University of Lleida, Spain. http://www.tesisenred.net/handle/10803/8253. Accessed 20 Apr 2013
Rodríguez SV, Faulin J, Plà LM (2014) New opportunities of operations research to improve pork supply chain efficiency. Ann Oper Res 219:5–23
Taylor DH (2006) Strategic consideration in the development of lean agri-food supply chains: a case study of the UK pork sector. Int J Supply Chain Manag 11(3):271–280
Trienekens J, Petersen B, Wognum N, Brinkmann D (2009) European pork chains. Diversity and quality challenges in consumer-oriented production and distribution. Wageningen Academic Publishers, Wageningen. ISBN 978-90-8686-103-3
Whittemore CT, Kyriazakis I (2006) Whittemore's science and practice of pig production. Blackwell, London

Chapter 18
Diet Problems

E. Joannopoulos, F. Dubeau, J.-P. Dussault, and C. Pomar

18.1 Introduction

In its usual form, the diet problem is formulated as a linear program. It has been introduce for the first time in (Stigler, 1945). It was enlarged with the development of the simplex algorithm in 1963 and latter in 1990 (Dantzig 1998, 1990). It was also revisited in (Garille and Gass, 2001). All the models developed in these studies aim to optimize the unit cost of a diet. In this case, the energy density of the diet is considered as fixed. Since the animals' appetite is considered proportional to their energy requirements, feed consumption is fixed and the modeling is in proportions. In our study, the proposed models consider the energy density in diets as a decision variable. We no longer optimize the unitary cost of feeds but the total cost of the feeds that will be consumed during the entire growing period. We will present evidence that this approach is more advantageous for farmers. All the models presented here will be applied to the pig problem.

E. Joannopoulos • F. Dubeau (✉)
Département de mathématiques, Université de Sherbrooke, 2500 Boul. Université, Sherbrooke, QC, Canada, J1K 2R1
e-mail: emilie.joannopoulos@usherbrooke.ca; francois.dubeau@usherbrooke.ca

J.-P. Dussault
Département d'informatique, Université de Sherbrooke, 2500 Boul. Université, Sherbrooke, QC, Canada, J1K 2R1
e-mail: jean-pierre.dussault@usherbrooke.ca

C. Pomar
Agriculture and Agri-Food Canada, 2000 College Street, Sherbrooke, QC, Canada, J1M 0C8
e-mail: Candido.Pomar@agr.gc.ca

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_18

18.2 The Feeding Problem: Diet Formulation

A diet is a mixture of available ingredients which provide the required nutrients for subsistence and growth. The optimal dietary nutrient concentration changes over time following the animal's natural patterns of appetite and daily required nutrients. Because some nutrients (e.g., amino acids, phosphorous, etc.) are expensive, optimal diets are those minimizing feed cost while providing the amount of nutrients that are required by animals.

We will consider in this study the problem of formulating optimal diets (minimum cost) for growing-finishing pig operations. We will quantify the animals' appetite and requirements over time, give the conditions to be met by a diet to be eligible, and introduce the formulation of several models.

The daily requirements described here and the diet characteristics to make it eligible had been given using recognized standards. We will base our models on these standards which can be found in Subcommittee on Swine Nutrition, Committee on Animal Nutrition, National Research Council (1998).

18.2.1 Elements of the General Problem

The growth period of pigs begins at weaning time and ends at the slaughter time. It last T days and will be noted $D = \{1, \ldots, T\}$.

Let us assume that we have n available ingredients to formulate a diet, and let $j \in J = \{1, \ldots, n\}$ be the index set of the ingredients. Let $x_j(t)$ be the amount (in kg) of the jth ingredient of the diet at day $t \in T$. Thus, $\mathbf{x}(\mathbf{t}) \in \Re^n$ is the diet vector for day t, and the weight of the diet on day t is

$$W(t) = \sum_{j=1}^{n} x_j(t). \tag{18.1}$$

Let c_j be the unit cost of the jth ingredient which we assume constant all along the growth period. The cost of the diet for the day t is given by

$$z(t) = \sum_{j=1}^{n} c_j x_j(t) \tag{18.2}$$

and the total cost over the period D is then

$$Z = \sum_{t=1}^{T} z(t) = \sum_{j=1}^{n} c_j \left(\sum_{t=1}^{T} x_j(t) \right). \tag{18.3}$$

For its growth, it is assumed here that pigs have daily digestible energy requirements and that these requirements are mean driver for its daily feed intake. Under this assumption, pigs will eat the available feeds until the energy requirements are met. In this context, let $E(t)$ be the required digestible energy, in kilocalories (kcal), to be supplied at day t by feeds to ensure the optimal growth of the pig. If e_j is the energy content (kcal/kg) of the jth ingredient, the diet $\mathbf{x(t)}$ must satisfy the following equation

$$\sum_{j=1}^{n} e_j x_j(t) = E(t). \tag{18.4}$$

We say that a diet of weight $W(t)$ with an energy $E(t)$ has an energy density $e(t)$, in kilocalories per kilogram, given by

$$e(t) = \frac{E(t)}{W(t)}. \tag{18.5}$$

The diet should also supply minimum quantities of amino acids. Let I_{AA} be the set of amino acids and $AA_i(t)$ be the daily minimal amount of the ith amino acid required at day t. Let $aa_{i,j}$ be the amount of the ith digestible amino acid per kilogram of the jth ingredient. Thus, the diet must verify:

$$\sum_{j=1}^{n} aa_{i,j} x_j(t) \geq AA_i(t). \tag{18.6}$$

Moreover, some nutrients must appear in the diet in minimal amounts without exceeding maximum quantities. Let I_B be the set of all these nutrients and $B_i^{\min}(t)$ and $B_i^{\max}(t)$ be the minimum and maximum intakes of the ith nutrient of type at day t. If $a_{i,j}$ is the ith nutrient intake in 1 kg of the jth ingredient, we must have

$$B_i^{\min}(t) \leq \sum_{j=1}^{n} a_{i,j} x_j(t) \leq B_i^{\max}(t). \tag{18.7}$$

The animal's intake capacity is restricted each day, so we need to add a constraint on the weight of diet to swallow at day t. This maximal limit is given by $W^{\max}(t)$, and we must have

$$W(t) \leq W^{\max}(t). \tag{18.8}$$

Other types of constraints on the ingredients and nutrients are imposed to obtain a diet of good quality. We consider two of them. The first type of constraints concerns the importance, in proportion, of some ingredients in the diet. Some proportions of ingredients must not exceed minimum or maximum thresholds set in advance at day t or even be in fixed proportion in the diet. Let J_p be the set of these ingredients and $p_j^{\min}(t)$ and $p_j^{\max}(t)$ be the minimum and maximum proportions of the jth ingredient in the diet. Thus, we have the following constraint:

$$p_j^{\min}(t) \leq \frac{x_j(t)}{W(t)} \leq p_j^{\max}(t). \tag{18.9}$$

The second type of constraints concerns pairs of nutrients for which the ratio of their intake in the diet is between a minimum and a maximum value at day t. Let I_R^2 be this set of pairs of nutrients and $r_{i_1,i_2}^{\min}(t)$ and $r_{i_1,i_2}^{\max}(t)$ be the minimum and maximum ratios of nutrient inputs i_1 and i_2. So we have

$$r_{i_1,i_2}^{\min}(t) \leq \frac{\sum_{j=1}^{n} a_{i_1,j}x_j(t)}{\sum_{j=1}^{n} a_{i_2,j}x_j(t)} \leq r_{i_1,i_2}^{\max}(t). \tag{18.10}$$

Considering the needs and diet characteristics, let us define by $S^V(t)$ the set of **feasible diets** at day t. A diet $\mathbf{x(t)} \in \mathfrak{R}^n$ is feasible at day t, i.e., $\mathbf{x(t)} \in S^V(t)$, if and only if it satisfies the following conditions:

$$S^V(t) \quad \begin{cases} x_j(t) \geq 0 & (j \in J = \{1,\dots,n\}) \\ \sum\limits_{j=1}^{n} e_j x_j(t) = E(t) & \\ \sum\limits_{j=1}^{n} aa_{i,j}x_j(t) >= AA_i(t) & (i \in I_{AA}) \\ B_i^{\min}(t) \leq \sum\limits_{j=1}^{n} a_{i,j}x_j(t) \leq B_i^{\max}(t) & (i \in I_B) \\ p_j^{\min}(t) \leq \dfrac{x_j(t)}{\sum\limits_{j-1}^{n} x_j(t)} \leq p_j^{\max}(t) & \left(j \in J_p\right) \\ r_{i_1,i_2}^{\min}(t) \leq \dfrac{\sum\limits_{j=1}^{n} a_{i_1,j}x_j(t)}{\sum\limits_{j=1}^{n} a_{i_2,j}x_j(t)} \leq r_{i_1,i_2}^{\max}(t) & \left((i_1,i_2) \in I_R^2\right) \\ \sum\limits_{j=1}^{n} x_j(t) \leq W^{\max}(t) & \end{cases} \tag{18.11}$$

A diet $\mathbf{x(t)} \in S^V(t)$ of weight $W(t) = \sum_{j=1}^{n} x_j(t)$ has its own energy density $e(t) = \frac{E(t)}{W(t)}$. This energy density might change for different diets in $S^V(t)$; it is not a predetermined quantity. For a diet in this set $S^V(t)$, we say that we have a **variable energy density diet**.

If we a priori fix the energy density $e(t)$ of a kilogram of diet, we will give the quantity (in kg) $W(t) = \frac{E(t)}{e(t)}$ of diet to animals. Considering the weight constraint (18.8), this fixed energy density must satisfy the following inequality:

$$e(t) \geq \frac{E(t)}{W^{\max}(t)} = e^{\min}(t). \tag{18.12}$$

In this case, let us write $\mathbf{y(t)} = (y_1(t), \ldots, y_n(t)) \in \mathfrak{R}^n$ where $y_j(t)$ represent the proportion of the jth ingredient in the diet at day t. We write $\mathbf{y(t)} \in \Delta_n$ where $\Delta_n = \left\{\mathbf{y} \in R^n \middle| y_j \geq 0 \quad (j \in J) \quad \text{and} \quad \sum_{j=1}^{n} y_j = 1\right\}$. It is necessary that the selected proportion vector $\mathbf{y(t)}$ gives a feasible diet $\mathbf{x(t)} = W(t)\mathbf{y(t)} = \frac{E(t)}{e(t)}\mathbf{y(t)}$, in other words $\mathbf{x(t)} \in S^V(t)$, or equivalently $\mathbf{y(t)} \in \frac{e(t)}{E(t)} S^V(t)$. The set of feasible proportions with a fixed energy density $e(t)$ is defined by

$S^F(t, e(t)) = \Delta^n \cap \frac{e(t)}{E(t)} S^V(t)$.

Thus, $\mathbf{y(t)} \in S^F(t, e(t))$ if and only if $\mathbf{y(t)}$ satisfies the following conditions:

$$S^F(t, e(t)) \begin{cases} \Delta_{\text{n}} \begin{cases} y_j(t) \geq 0 \\ \displaystyle\sum_{j=1}^{n} y_j(t) = 1 \end{cases} & (j \in J = \{1, \ldots, n\}) \\ \displaystyle\sum_{j=1}^{n} e_j y_j(t) = e(t) & \\ \displaystyle\sum_{j=1}^{n} aa_{i,j} y_j(t) >= \frac{AA_i(t)}{E(t)} e(t) & (i \in I_{AA}) \\ \displaystyle\frac{B_i^{\min}(t)}{E(t)} e(t) \leq \sum_{j=1}^{n} a_{i,j} y_j(t) \leq \frac{B_i^{\max}(t)}{E(t)} e(t) & (i \in I_B) \\ p_j^{\min}(t) \leq y_j(t) \leq p_j^{\max}(t) & (j \in J_p) \\ r_{i_1,i_2}^{\min}(t) \leq \dfrac{\sum_{j=1}^{n} a_{i_1,j} y_j(t)}{\sum_{j=1}^{n} a_{i_2,j} y_j(t)} \leq r_{i_1,i_2}^{\max}(t) & \left((i_1, i_2) \in I_R^2\right) \end{cases} \tag{18.13}$$

In this case we say that the diet $\mathbf{x(t)}$, with an a priori fixed energy content $e(t)$, is a **fixed energy density diet**.

18.2.2 Minimal Cost Diet

We will consider the two types of diets described above: the variable and the fixed energy density diets and their consequences on the formulation of minimal cost diets.

First, consider the variable energy density diet. We look for a feasible variable energy density diet at minimal cost. This is an optimal solution of the following mathematical program:

$$P^V(t) \quad \begin{cases} z^V(t) = \min_{x(t)} \sum_{j=1}^{n} c_j x_j(t) \\ \text{subject to } \mathbf{x}(\mathbf{t}) \in S^V(t) \end{cases}. \tag{18.14}$$

The problem for all the growing period will be noted P^V, and the optimal total cost for the whole period is

$$Z^V = \sum_{t \in D} z^V(t). \tag{18.15}$$

Next we consider the fixed energy density diet. On day t we look for the minimal unit cost of a diet that has a given and fixed energy density $e(t)$ satisfying the inequality (18.12). This is an optimal solution of the following mathematical program:

$$P^F(t;e(t)) \quad \begin{cases} z_u^F(t;e(t)) = \min_{y(t)} \sum_{j=1}^{n} c_j y_j(t) \\ \text{subject to } \mathbf{y}(\mathbf{t}) \in S^F(t;e(t)) \end{cases}. \tag{18.16}$$

The daily cost is given by

$$z^F(t;e(t)) = W(t) z_u^F(t;e(t)). \tag{18.17}$$

As for the variable energy density problem, the problem for the whole growing period will be denoted $P^F(e)$, and the optimal total cost is

$$Z^F = \sum_{t \in D} z^F(t;e(t)). \tag{18.18}$$

18.2.3 The Multiphase Formulation

Although it is mathematically possible to solve the optimal diet problem for each day during the growing period, we obtain a diet plan not useful in practice because of physical limitations of feed transportation and storage capacity. A technique used by the swine industry to overcome these limitations is then to feed pigs in phases. A **feeding phase** is a time interval during which pigs are fed with the same diet. In our formulation models, this diet should be feasible for each day of the phase to ensure that pigs will always eat all the required nutrients to express their full potential for growth. We partition the duration of growth by introducing K phases ($1 \leq K \leq T$).

We take an increasing sequence of time $0 = t_0 < t_1 < \cdots < t_{K-1} < t_K = T$ and define the kth phase by the set $D_k = \{t_{k-1} + 1, \ldots, t_k\}$ for $k = 1, \ldots, K$. The length of a phase is given by $\Delta D_k = t_k - t_{k-1}$. If $\Delta D_k = 1$ for all k, it is a daily feeding phase program and $D_k = \{t_k\}$. For each phase k, the diet chosen should provide the total energy required during the phase. This energy is given by

$$\overline{E}(k) = \sum_{t=t_{k-1}+1}^{t_k} E(t). \tag{18.19}$$

To distinguish phase quantities from daily ones, we use bar symbol for phases. So $\bar{E}(k)$ represent the energy requirement for the phase k and $E(t)$ represent the energy requirement for the day t.

For the variable energy density problem, we look for a diet $\overline{x}(k) = (x_1(k), \ldots, x_n(k)) \in \mathfrak{R}^n$ which provides the total energy of the phase $\bar{E}(k)$. The total weight of this diet is $\overline{W}(k) = \sum_{j=1}^{n} \overline{x}_j(k)$. For each day $t \in D_k$, we give to the pigs the quantity $x(t) = \frac{E(t)}{\overline{E}(k)}\overline{x}(k)$ which must be feasible on day t and especially must satisfy energy needs. Hence, for all $t \in D_k$, $x(t) = \frac{E(t)}{\overline{E}(k)}\overline{x}(k) \in S^V(t)$ or equivalently $\overline{x}(k) \in \frac{\overline{E}(k)}{E(t)} S^V(t)$. So, we define the set of feasible diets for the phase k by $\overline{S}^V(k) = \bigcap_{t \in D_k} \dfrac{\overline{E}(k)}{E(t)} S^V(t)$.

Therefore, $\overline{x}(k) \in \overline{S}^V(k)$ if and only if $\overline{x}(k)$ satisfy the following conditions:

$$\overline{S}^V(k) \begin{cases} \overline{x}_j(k) \geq 0 & (j \in J) \\ \displaystyle\sum_{j=1}^{n} e_j \overline{x}_j(k) = \overline{E}(k) & \\ \displaystyle\sum_{j=1}^{n} aa_{i,j} \overline{x}_j(k) \geq \widetilde{aa}_i(k) \overline{E}(k) & (i \in I_{AA}) \\ \widetilde{b}_i^{\min}(k) \overline{E}(k) \leq \displaystyle\sum_{j=1}^{n} a_{i,j} \overline{x}_j(k) \leq \widetilde{b}_i^{\max}(k) \overline{E}(k) & (i \in I_B) \\ \widetilde{p}_j^{\min}(k) \leq \dfrac{\overline{x}_j(k)}{\sum_{j=1}^{n} \overline{x}_j(k)} \leq \widetilde{p}_j^{\max}(t) & (j \in J_P) \\ \widetilde{r}_{i_1,i_2}^{\min}(k) \leq \dfrac{\sum_{j=1}^{n} a_{i_1,j} \overline{x}_j(k)}{\sum_{j=1}^{n} a_{i_2,j} \overline{x}_j(k)} \leq \widetilde{r}_{i_1,i_2}^{\max}(k) & \left((i_1, i_2) \in I_R^2\right) \\ \displaystyle\sum_{j=1}^{n} \overline{x}_j(k) \leq \widetilde{W}^{\max}(k) \overline{E}(k) & \end{cases} \tag{18.20}$$

where we have used the definitions and notations given in Table 18.1.

Table 18.1 Definitions and notations

Definitions	Notations	Definitions	Notations
$\max_{t\in D_k} \frac{AA_i(t)}{E(t)} =$	$\tilde{aa}_i(k)$	$\max_{t\in D_k} p_j^{\min}(t) =$	$\widetilde{p}_j^{\min}(k)$
$\min_{t\in D_k} \frac{W^{\max}(t)}{E(t)} =$	$\widetilde{W}^{\max}(k)$	$\min_{t\in D_k} p_j^{\max}(t) =$	$\widetilde{p}_j^{\max}(k)$
$\max_{t\in D_k} \frac{B_i^{\min}(t)}{E(t)} =$	$\widetilde{b}_i^{\min}(k)$	$\max_{t\in D_k} r_{i_1,i_2}^{\min}(t) =$	$\widetilde{r}_{i_1,i_2}^{\min}(k)$
$\min_{t\in D_k} \frac{B_i^{\max}(t)}{E(t)} =$	$\widetilde{b}_i^{\max}(k)$	$\min_{t\in D_k} r_{i_1,i_2}^{\max}(t) =$	$\widetilde{r}_{i_1,i_2}^{\max}(k)$

For each phase, the problem is to find a variable energy density diet $\overline{x}(k)$ by solving the following mathematical program:

$$\overline{P}^V(k) \begin{cases} \overline{z}^V(k) = \min_{\overline{x}(k)} \sum_{j=1}^{n} c_j \overline{x}_j(k) \\ \text{subject to } \overline{x}(k) \in \overline{S}^V(k) \end{cases}. \tag{18.21}$$

As $\overline{z}^V(k)$ is the feeding cost for the phase k, the total feeding cost for the overall growing period using K phases is given by

$$\overline{Z}^V(K) = \sum_{k=1}^{K} \overline{z}^V(k). \tag{18.22}$$

This problem using K feeding phases for the overall growing period will be noted $\overline{P}^V(K)$. Let us remark that $\overline{P}^V(T)$ reduces to P^V a daily-phase program.

When we choose to fix the energy density of the diet at $\bar{e}(k)$ during the phase k, it must satisfy $W(t) = \frac{E(t)}{\bar{e}(k)} \leq W^{\max}(t)$ for all $t \in D_k$, and we have

$$\overline{e}(k) \geq \overline{e}^{\min}(k) = \max_{t\in D_k} e^{\min}(t) = \max_{t\in D_k} \frac{E(t)}{W^{\max}(t)}. \tag{18.23}$$

In this case, we are looking for proportions $\overline{y}(k) = (\overline{y}_1(k), \ldots, \overline{y}_n(k)) \in \Delta_n$ with energy density $\overline{e}(k) = \sum_{j=1}^{n} e_j \overline{y}_j(k)$. For each day t, we provide the quantity $W(t) = \frac{E(t)}{\overline{e}(k)}$ of diet to the pig, the daily diet is determined by $x(t) = W(t)\overline{y}(k) = \frac{E(t)}{\overline{e}(k)}\overline{y}(k)$. Therefore, we must have $x(t) \in S^V(t)$ for all $t \in D_k$, i.e.,

$$\overline{y}(k) \in \frac{\overline{e}(k)}{E(t)} S^V(t) \quad (\forall t \in D_k). \tag{18.24}$$

Then let us define set of feasible proportions by

$$\overline{S}^F(k; \overline{e}(k)) = \Delta_n \cap \bigcap_{t\in D_k} \frac{\overline{e}(k)}{E(t)} S^V(t). \tag{18.25}$$

Therefore, $\overline{y}(k) \in \overline{S}^F(k;\overline{e}(k))$ if and only if $\overline{y}(k)$ satisfy the following conditions:

$$\overline{S}^F(k;\overline{e}(k)) \begin{cases} \overline{y}_j(k) \geq 0 & (j \in J) \\ \sum_{j=1}^{n} \overline{y}_j(k) = 1 & \\ \sum_{j=1}^{n} e_j \overline{y}_j(k) = \overline{e}(k) & \\ \sum_{j=1}^{n} aa_{i,j} \overline{y}_j(k) \geq \widetilde{a}\widetilde{a}_i(k)\overline{E}(k) & (i \in I_{AA}) \\ \widetilde{b}_i^{\min}(k)\overline{E}(k) \leq \sum_{j=1}^{n} a_{i,j}\overline{y}_j(k) \leq \widetilde{b}_i^{\max}(k)\overline{E}(k) & (i \in I_B) \\ \widetilde{p}_j^{\min}(k) \leq \dfrac{\overline{y}_j(k)}{\sum_{j=1}^{n} \overline{y}_j(k)} \leq \widetilde{p}_j^{\max}(t) & (j \in J_P) \\ \widetilde{r}_{i_1,i_2}^{\min}(k) \leq \dfrac{\sum_{j=1}^{n} a_{i_1,j}\overline{y}_j(k)}{\sum_{j=1}^{n} a_{i_2,j}\overline{y}_j(k)} \leq \widetilde{r}_{i_1,i_2}^{\max}(k) & \left((i_1,i_2) \in I_R^2\right) \end{cases} . \tag{18.26}$$

Then the problem is to solve

$$\overline{P}^F(k,\overline{e}(k)) \begin{cases} \overline{z}_u^F(k;\overline{e}(k)) = \min_{\overline{y}(k)} \sum_{j=1}^{n} c_j \overline{y}_j(k) \\ \text{subject to } \overline{y}(k) \in \overline{S}^F(k;\overline{e}(k)) \end{cases} . \tag{18.27}$$

The feeding cost during the phase k is given by

$$\overline{z}^F(k;\overline{e}(k)) = W(k)\overline{z}_u^F(k;\overline{e}(k)) \tag{18.28}$$

where $W(k) = \frac{\overline{E}(k)}{\overline{e}(k)}$. The feeding cost for the overall growth period is

$$\overline{Z}^F(K) = \sum_{j=1}^{n} \overline{z}^F(k;\overline{e}(k)). \tag{18.29}$$

This problem using K phases during the growing period will be noted $\overline{P}^F(K;\overline{e})$. Let us note that $\overline{P}^F(T\overline{e})$ coincide with $P^F(e)$ for $\overline{e} = e$.

18.2.4 Modeling with Premixes

Pig production with 2- and 3-phase feeding periods are today frequently used by the swine industry because of the transportation and storage cost of using more feeds during the growing period. The development of feeding systems that allow blend feeding and the automatic distribution of two premixes that, combined in variable ratios, could meet the requirements of pigs throughout their growing period allow for significant reductions of feeding costs and nutrients' excretion. We call in this study **premix** a proportion vector, known or unknown, of ingredients used to make a unit of mix. In what follow, we consider two premixes (A and B) such that

$$\begin{cases} A = (A_1, \ldots, A_n) \geq 0 \\ \sum_{j=1}^{n} A_j = 1 \end{cases} \quad \text{and} \quad \begin{cases} B = (B_1, \ldots, B_n) \geq 0 \\ \sum_{j=1}^{n} B_j = 1 \end{cases}, \tag{18.30}$$

i.e., $A, B \in \Delta_n$. The unit costs of these premixes are given by

$$c_A = \sum_{j=1}^{n} c_j A_j \quad \text{and} \quad c_B = \sum_{j=1}^{n} c_j B_j. \tag{18.31}$$

Similarly, the energy densities are

$$e_A = \sum_{j=1}^{n} e_j A_j \quad \text{and} \quad e_B = \sum_{j=1}^{n} e_j B_j. \tag{18.32}$$

In the variable energy density diet problem, we are looking for a premix combination $\overline{x}(k) = \alpha_k A + \beta_k B$ where $\alpha_k \geq 0$ and $\beta_k \geq 0$ and such that $\overline{x}(k)$ is a feasible diet for the phase k. For the phase k, the weight of the diet is $\overline{W}(k) = \alpha_k + \beta_k$, its total energy is $\overline{E}(k) = \alpha_k e_A + \beta_k e_B$, and its energy density is given by

$$\overline{e}(k) = \frac{\overline{E}(k)}{\overline{W}(k)} = \frac{\alpha_k e_A + \beta_k e_B}{\alpha_k + \beta_k} \in [\min\{e_A; e_B\};\ \max\{e_A; e_B\}]. \tag{18.33}$$

On day $t \in D_k$, we provide the quantity $W(t) = \frac{E(t)}{\overline{e}(k)} = \frac{E(t)}{\overline{E}(t)} \overline{W}(k)$ of diet. This quantity is determined to satisfy the energy needs as well as amino acids and minerals needs. Thus, the daily diet is

$$x(t) = \frac{W(t)}{\overline{W}(k)} \overline{x}(k) = \frac{E(t)}{\overline{E}(k)} \overline{x}(k) = \frac{E(t)}{\overline{E}(k)} (\alpha_k A + \beta_k B). \tag{18.34}$$

The cost of this diet is

$$\begin{aligned}\bar{z}^V(k;A,B) &= \sum_{j=1}^{n} c_j \bar{x}_j(k) \sum_{j=1}^{n} c_j\big(\alpha_k A_j + \beta_k B_j\big) \\ &= \alpha_k \sum_{j=1}^{n} c_j A_j + \beta_k \sum_{j=1}^{n} c_j B_j,\end{aligned} \tag{18.35}$$

and the total feeding cost during the overall growth period is given by

$$\overline{Z}^V(K;\ A,B) = \sum_{k=1}^{K} \bar{z}^V(k;\ A,B) = \left(\sum_{k=1}^{K} \alpha_k\right) c_A + \left(\sum_{k=1}^{K} \beta_k\right) c_B. \tag{18.36}$$

We consider two types of premixes: known or so-called fixed premixes and unknown or so-called variable premixes.

If premixes are known and fixed, we have an optimization problem for each phase k given by

$$\overline{P}^V\left(k;\ A^f, B^f\right) \begin{cases} \bar{z}^V\left(k; A^f, B^f\right) = \min\limits_{\alpha_k,\beta_k} \alpha_k c_{A^f} + \beta_k c_{B^f} \\ \text{subject to} \\ \begin{cases} \alpha_k \geq 0, \beta_k \geq 0 \\ \bar{x}(k) = \alpha_k A^f + \beta_k B^f \in \overline{S}^V(k) \end{cases} \end{cases} \tag{18.37}$$

For this problem, we must use two known premixes to ensure that the problem has a solution for each phase. The problem during the overall growth period will be noted $\overline{P}^V\left(K;\ A^f, B^f\right)$.

In the case of variable (unknown) premixes, as the premixes are part of the optimization problem and they appear in each phase, we cannot decompose the problem. Thus, we have a global problem that considers all phases at once. Our problem is then to find $A^v, B^v \in \Delta_n$ and scalars $\alpha_k \geq 0$ and $\beta_k \geq 0$ for $k = 1, \ldots, K$ which give feasible diets for each phase and minimize the total feeding cost during the overall growth period. Thus, we have the following problem:

$$\overline{P}_G^V(K; A^v, B^v) \begin{cases} \overline{Z}_G^V(K; A^v, B^v) = \min\limits_{\alpha,\beta,A^v,B^v} \left(\sum_{k=1}^{K} \alpha_k\right) c_{A^v} + \left(\sum_{k=1}^{K} \beta_k\right) c_{B^v} \\ \text{subject to} \begin{cases} A^v \in \Delta_n, B^v \in \Delta_n \\ \text{and for } k = 1, \ldots, K \\ \alpha_k \geq 0,\ \beta_k \geq 0 \\ \bar{x}(k) = \alpha_k A^v + \beta_k B^v \in \overline{S}^V(k) \end{cases} \end{cases}. \tag{18.38}$$

The decision variables of this problem are α, β, A^v and B^v, and a product of two of them appears in the objective function. Thus, this is not a linear problem but a bilinear problem.

If we are using the fixed energy density diet problem, we are looking for a combination of two premixes $\overline{y}(k) = \alpha_k A + \beta_k B$ in which $\alpha_k + \beta_k = 1$ $\alpha_k \geq 0$, and $\beta_k \geq 0$ such that $\overline{y}(k)$ is a feasible proportion in phase k. In this case, considering the feasibility constraint for the k combinations $\overline{y}(k)$, we set the energy density $e_A = e_B = e$, and we have $\overline{e}(k) = e$ for each $k = 1, \ldots, K$.

The unit cost of this diet is $\overline{z}_u^F(k; A, B) = \alpha_k c_A + \beta_k c_B$, and the cost during the phase k is $\overline{z}^F(k; A, B) = \overline{W}(k)\overline{z}_u^F(k; A, B) = \overline{W}(k)(\alpha_k c_A + \beta_k c_B)$ where $\overline{W}(k) = \frac{\overline{E}(k)}{\overline{e}(k)}$. The total cost during the overall growth period is given by

$$\overline{Z}^F(K; A, B) = \sum_{k=1}^{K} \overline{z}^F(k; A, B) = \left(\sum_{k=1}^{K} \overline{W}(k)\alpha_k\right) c_A + \left(\sum_{k=1}^{K} \overline{W}(k)\beta_k\right) c_B. \quad (18.39)$$

If premixes are known, we have the following problem to solve

$$\overline{P}^F\left(k; \overline{e}(k); A^f, B^f\right) \quad \begin{cases} \overline{z}_u^F\left(k; A^f, B^f\right) = \min\limits_{\alpha_k, \beta_k} \alpha_k c_{A^f} + \beta_k c_{B^f} \\ \text{subject to} \\ \begin{cases} \alpha_\mathrm{k} \geq 0, \beta_k \geq 0, \alpha_k + \beta_k = 1 \\ \overline{y}(k) = \alpha_k A^f + \beta_k B^f \in \overline{S}^F(k; \overline{e}(k)) \end{cases} \end{cases} . \quad (18.40)$$

The problem during the overall growth period will be noted $\overline{P}^F\left(K; \overline{e}; A^f, B^f\right)$.

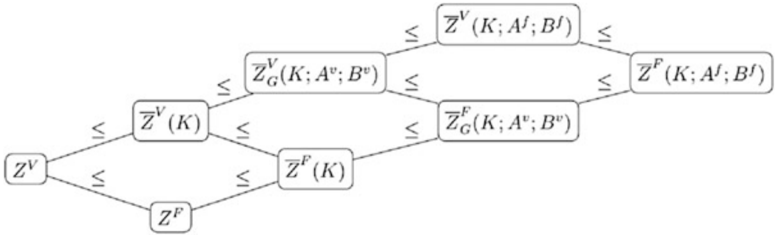
In the case of variable premixes, we consider the following problem:

$$\overline{P}_G^F(K; \overline{e}(k), A^v, B^v) \begin{cases} \overline{Z}_G^F(K; A^v, B^v) = \min\limits_{\alpha, \beta, A^v, B^v} \sum_{k=1}^{K} \overline{W}(k)(\alpha_k c_{A^v} + \beta_k c_{B^v}) \\ \text{Subject to} \begin{cases} A^v \in \Delta_n, B^v \in \Delta_n \\ \text{and for } k = 1, \ldots, K \\ \alpha_k \geq 0, \beta_k \geq 0, \alpha_k + \beta_k = 1 \\ \overline{y}(k) = \alpha_k A^v + \beta_k B^v \in \overline{S}^F(k; \overline{e}(k)) \end{cases} \end{cases} . \quad (18.41)$$

Remark 1 All the previous problems were linear problems. This is not the case for the two problems, $\overline{P}_G^V(K; A^v, B^v)$ and $\overline{P}_G^F(K; \overline{e}(k), A^v, B^v)$, which are bilinear and non-convex problems due to the simultaneous optimization of the α's, β's, A and B.

18.2.5 Global Comparison

Some relationships between the costs of the several problems can be highlighted. The following partially ordered graph summarizes these relationships.



We remark that the feasible set of these problems are described by a similar graph, replacing "$\leq$" by "$\supseteq$." For example, the optimal solution to the fixed energy density diet problem using fixed premixes $\overline{P}^F(K; A^f, B^f)$ is a feasible solution to the fixed energy density diet problem using variable premixes $\overline{P}^F(K; A^v, B^v)$ and also to the variable energy density diet problem using fixed premixes $\overline{P}^V(K; A^f, B^f)$.

Another example, the optimal solution to the fixed energy density problem using phases $\overline{P}^F(K; \overline{e}(k))$, is only a feasible solution to the variable energy density problem using the same number of phases $\overline{P}^V(K)$.

18.3 An Example: Data from the Québec Context

To illustrate the formulations presented in the preceding section, we use data from the Canadian context. We assumed that the growing and finishing period in pigs takes $T = 111$ days, and body weight goes from 20 to 130 kg. The set D described above is $D = \{1, \ldots, 111\}$. The main need of animals is energy. An animal eat to satisfy his energy requirements (Whittemore & Fawcett 1976; van Milgen et al. 2008), and so it will eat more of a diet with a low energy density and less of a diet with a high energy density. In this study, the feed is formulated to satisfy or exceed the needs estimated by nutritionists. Thus, the animal's growth will not be affected by the feeding.

18.3.1 Ingredient Cost and Technical Coefficients

In order to compare the feeding costs in commercial conditions, the formulations described above used 16 ingredients listed in the Table 18.2 with their corresponding variables used in the models. The prices are the average of those recorded by a food manufacturer (Breton Food, Inc, Saint-Bernard, QC, Canada) from November 2011 to October 2012. All costs in this study are expressed in \$CAD. Ingredients' composition values used in this study are those from NRC (Subcommittee on Swine Nutrition, Committee on Animal Nutrition, National

Table 18.2 Variables associated with ingredients

Associated variable	Ingredient
X1	DL-methionine
X2	Hard wheat
X3	Calcium carbonate
X4	Sodium chloride
X5	Meat meal
X6	Animal fat
X7	L-lysine HLC
X8	L-threonine
X9	L-tryptophan
X10	Corn
X11	Barley
X12	Dicalcium phosphate
X13	Premix
X14	Wheat shorts
X15	Canola meal
X16	Soybean meal

Research Council 1998). All these data and the minimal and maximal proportion of each ingredient, constant throughout the period, are collected in Table 18.3.

Two kinds of phosphorus appear in Table 18.3, total phosphorus and available phosphorus. The last one is part of the other and will be used in the nutrient constraint. Total phosphorus will be used in the calcium/phosphorus ratio constraint.

18.3.2 Growth Model

Daily requirements are estimated based on NRC (Subcommittee on Swine Nutrition, Committee on Animal Nutrition, National Research Council 1998). The energy requirement increases during the growth period (Fig. 18.1). The maximal amount of feed intake is given by the formula $W^{\max}(t) = \max\left\{\frac{E(t)}{3,400}; \quad 0.111 \times p(t)^{0.803}\right\}$ which is the maximum between the required quantity with energy density equals to 3,400 kcal/kg and the formula $0.111 \times p(t)^{0.803}$ given by Black et al. (1986), in which $p(t)$ is the weight of the animal on day t. This maximal quantity of feed intake increases over all the growth period as well as the energy requirement (Fig. 18.1).

The appetite that limits feed intake grows faster than the energy requirement, so that the minimal energy density of a diet is a decreasing function of time (Fig. 18.2).

The requirements of amino acids are increasing during the first half and decreasing during the second half of the growing period. The needs of other nutrients are

Table 18.3 Ingredients' composition, costs, minimal and maximal proportions

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
e_j (kcal/kg)																
Energy	5,640	3,310	0	0	2,695	8,281	2,820	4,780	4,120	6,570	3,390	3,070	0	0	2,760	3,520
$aa_{i,j}^{dig}$ (g/kg)																
Lysine	0	2.5	0	0	25.5	0	7.0	798	0	0	1.9	2.8	0	0	13.6	24.9
Threonine	0	2.7	0	0	16.2	0	5.7	0	990	0	2.5	2.6	0	0	10.8	15.3
Methionine	990	1.5	0	0	7.0	0	2.5	0	0	0	1.5	1.4	0	0	6.0	5.9
Methionine + Cystine	990	3.7	0	0	10.4	0	5.3	0	0	0	3.3	3.4	0	0	12.6	11.6
Tryptophan	0	1.1	0	0	2.8	0	2.2	0	0	985	0.4	1.0	0	0	3.3	5.2
Isoleucine	0	3.4	0	0	13.4	0	5.8	0	0	0	2.7	2.9	0	0	10.6	18.7
Valine	0	4.0	0	0	21.3	0	8.7	0	0	0	3.6	4.1	0	0	13.1	19.3
Leucine	0	6.4	0	0	31.9	0	10.2	0	0	0	9.5	5.7	0	0	18.6	29.8
Phenylalanine	0	4.5	0	0	18.4	0	7.0	0	0	0	3.7	4.1	0	0	10.9	20.6
Phenylalanine + Tyrosine	0	7.1	0	0	29.6	0	12.1	0	0	0	6.8	6.5	0	0	18.8	34.5
Histidine	0	2.2	0	0	9.5	0	4.3	0	0	0	2.1	1.8	0	0	7.4	10.9
Arginine	0	4.7	0	0	31.0	0	10.7	0	0	0	3.5	4.0	0	0	17.6	31.6
NEAA	0	57.4	0	0	254.4	0	88.0	0	0	0	36.3	46.6	0	0	144.9	206.1

(continued)

Table 18.3 (continued)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
$a_{i,j}$ (g/kg)																
Sodium	0	0.1	0.8	395	8	0	0	0	0	0.04	0.1	1.8	0	0.2	0.4	0.3
Calcium	0	0.8	385	3	76	0	0	0	0	0.4	0.7	220	0	0.9	8.3	3.4
Total phosphorus	0	3.4	0.2	0	38.8	0	0	0	0	2.6	3.4	185	0	8.4	11.4	6.2
Available phosphorus	0	1.7	0.2	0	31.4	0	0	0	0	0.364	1.02	180.4	0	3.44	2.394	1.922
Total nitrogen	158.4	23.2	0	0	86.4	0	127.96	116.96	136.48	12.96	16.16	0	0	25.6	53.92	72.48
c_j ($/kg)																
Cost	5.750	0.321	0.080	0.198	0.532	1.236	2.808	3.750	57.000	0.320	0.288	0.906	5.269	0.242	0.317	0.455
$p_j^{\min}$ and $p_j^{\max}$ (kg/kg)																
Minimal proportion	0	0	0	0	0	0	0	0	0	0	0	0	0.005	0	0	0
Maximal proportion	1	0.40	1	1	0.03	0.05	1	1	1	0.6	0.6	1	0.005	0.25	0.05	1

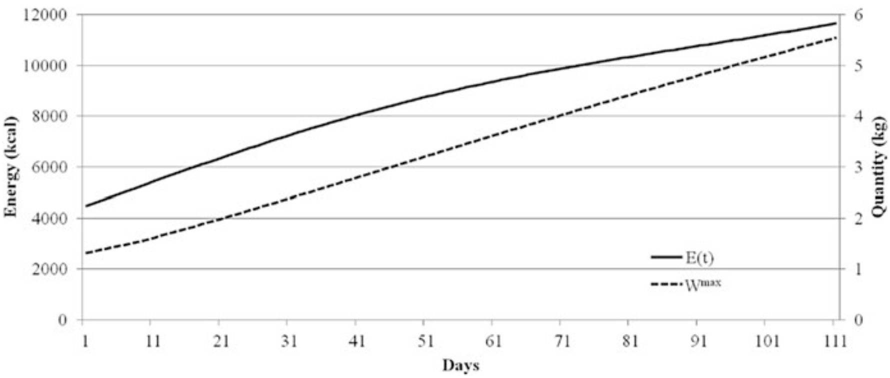


Fig. 18.1 Evolution of the energy requirement $E(t)$ and the intake capacity $W^{\max}(t)$

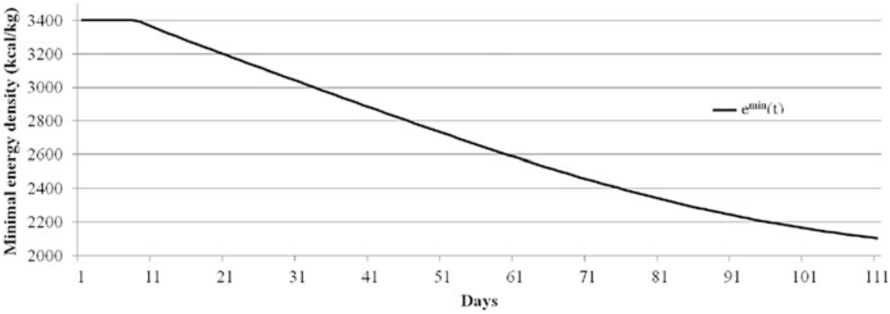


Fig. 18.2 Evolution of the minimal energy density in the diet $e^{\min}(t)$

increasing during the growth period. The behavior of all the amino acids belonging to the set $I_{AA} = \{$lysine, threonine, methionine, methionine + cystine, tryptophan, isoleucine, valine, leucine, phenylalanine, phenylalanine + tyrosine, histidine, arginine, NEAA$\}$ is similar to the lysine. Also, the behavior of the minerals, which belong to the set $\mathrm{I_B} = \{$sodium, calcium, available phosphorus, nitrogen$\}$, is similar to the calcium (Fig. 18.3).

When they are expressed in g/kcal, all the requirements are decreasing (Fig.18.4). This is the form in which they appear in the problems with phases or premixes.

Three nutrients are lower and/or upper bounded. Two of the three nutrients are not upper bounded. The last one must not exceed 2.5 g in 1 kg of the diet. Thus, we can summarize in the three equations:

$$B_{Na}^{\max}(t) = 2.5W(t) \qquad B_{Ca}^{\max}(t) = +\infty \qquad B_{Pav}^{\max}(t) = +\infty$$

In our case, the set J_P of the ingredients which are restricted in the diet corresponds to the set J of all the ingredients. If an ingredient has no lower

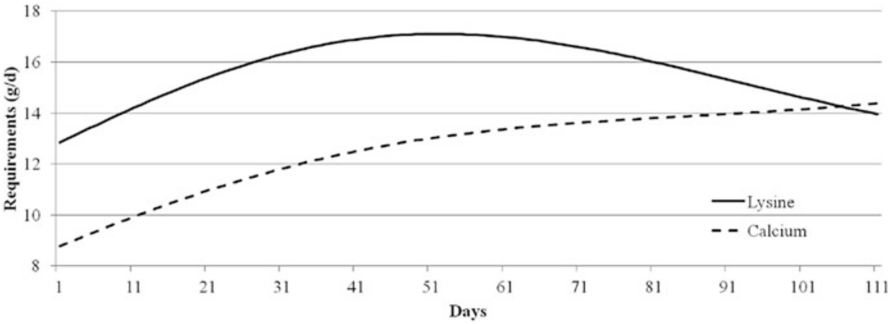


Fig. 18.3 Behavior of daily amino acid requirements $AA_i(t)$ and daily mineral needs $B_i^{\min}(t)$ during the overall growth period

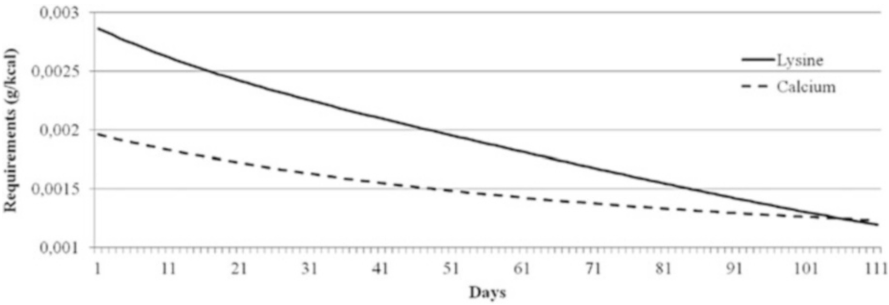


Fig. 18.4 Behavior of amino acid requirements $\frac{AA_i(t)}{E(t)}$ and nutrient needs $\frac{B_i^{\min}(t)}{E(t)}$ expressed in g/kcal during the overall growth period

bound, we impose its lower bound to be equal to 0. As well, if an ingredient has no upper bound, the maximal proportion of this ingredient will be 1.

The ratio constraints only concern, in our case, the calcium/total phosphorus ratio, so we have $J_R^2 = \{(\text{calcium, total phosphorus})\}$. The lower and upper bounds for this ratio are constant along the whole period, and we have

$$r_{Ca,Ptot}^{\min}(t) = 1 \quad \text{and} \quad r_{Ca,Ptot}^{\max}(t) = 1.5$$

18.3.3 Modeling with Phases

Concerning the feeding with phases, we will impose that all phases have the same length. If this is not possible, the longest phases will be placed at the end. For example, to feed with 4 phases, we will have 1 phase of 27 days and next 3 phases of 28 days.

When we feed animals with phases, the most restrictive value for each nutrient during the phase as defined in Table 18.1 must be considered. In the case of the pigs, the most restrictive value corresponds to the first day of the phase for each nutrient.

18.3.4 Comparison and Discussion

In this part, we compare the models described in Sect. 18.2. The optimization was realized to minimize the feeding cost. However, we also look for the environmental impact, particularly phosphorus and nitrogen excretions. The results are reported in Table 18.4.

Each problem will be compared to the current situation in the industry which corresponds to the problem $\overline{P}^F(3;\ 3400)$, the asterisk marked problem in Table 18.4.

Thus, using a variable energy density diet without premixes, i.e., a different diet in each phase, and feeding animals in 3 phases, corresponding to the problem $\overline{P}^V(3)$, decrease cost by 5 %, while nitrogen and phosphorus excretions increase, respectively, by 7.5 % and 11 % (Table 18.4, Fig. 18.5). The use of a daily-phase feeding

Table 18.4 Feeding costs and phosphorus and nitrogen intake and excretion of the several models

Number of phases K	Cost ($/pig)	P intake (kg/pig)	P excreted (kg/pig)	N intake (kg/pig)	N excreted (kg/pig)
$\overline{P}^V(K)$					
3	103.07	1.579	1.371	6.787	4.691
111	100.62	1.526	1.318	6.313	4.217
$\overline{P}^F(K;3400)$					
3[a]	108.68	1.445	1.273	6.462	4.366
111	107.02	1.342	1.134	5.763	3.667
$\overline{P}^V(K;A^f,B^f)$					
3	104.39	1.563	1.355	6.821	4.725
111	103.51	1.542	1.334	6.685	4.589
$\overline{P}_G^V(K;A^v,B^v)$					
3	104.29	1.615	1.407	7.007	4.911
111	102.38	1.530	1.322	6.591	4.495
$\overline{P}^F(K;3400;A^f,B^f)$					
3	108.82	1.456	1.248	6.509	4.413
111	107.55	1.391	1.183	6.063	3.967
$\overline{P}_G^F(K;3400;A^v,B^v)$					
3	108.82	1.456	1.248	6.503	4.407
111	107.29	1.349	1.141	5.827	3.731

[a]Current situation in the swine industry

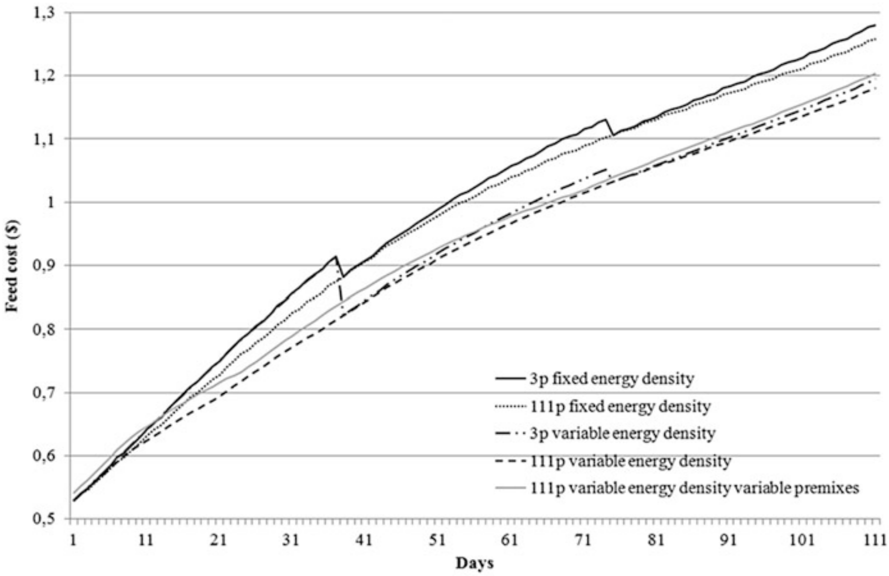


Fig. 18.5 Evolution of the feeding cost per day

program, corresponding to the optimal solution of the problem $\overline{P}^{V}(111)$, contributes even more to the cost decreasing. Indeed, in this case, feeding costs decrease by 7.5 % and nitrogen excretion by 3.5 %. By contrast, phosphorus excretion increases by 11 % (Table 18.4, Fig. 18.5). The cost decreasing is mainly due to the energy density of the diet. When the energy density of the diet decreases, the unit cost is also reduced. Despite the increase of the consumption, the daily cost is always lower than the currently feeding system. We expected the feeding costs decrease with the relationships that we determined above in the comparison graph.

As also expected, feeding animals with a daily phases but with fixed energy density ($\overline{P}^{F}(111;\ 3400)$) contributes to reduce feeding costs by 1.5 %. The excretions also reduced by 8 % concerning the phosphorus and by 16 % regarding nitrogen excretions (Table 18.4, Fig. 18.5).

These feeding plans are optimal but impossible to implement in practice. Indeed, if animals are fed with different diet every day, the storage costs will increase significantly. That is why we used two premixes. This method allows to feed animals with a different mix every day without increasing the storage cost. When we are talking about fixed premixes, premix A correspond to the optimal mixture of the first day of the feeding, and premix B correspond to the optimal mixture of the first day of the last phase. From now on, we consider A^f and B^f to be determined in this way.

So, the use of a variable energy density feeding system with two fixed premixes, and three ($\overline{P}^{V}(3; A^f, B^f)$) or daily phases ($\overline{P}^{V}(111; A^f, B^f)$), is interesting compared to the current situation because it reduce the feeding cost by 4 % and 5 %, respectively. However, these feeding plans are less interesting than the optimal solution of the problem $\overline{P}^{V}(3)$. These two solutions increase the feeding cost by about 1 %.

The use of a 3-phase feeding program with premixes ($\overline{P}^{V}\left(3;\ A^{f}, B^{f}\right)$, $\overline{P}^{V}_{G}(3;\ A^{v}, B^{v})$, $\overline{P}^{F}\left(3; 3400; A^{f}, B^{f}\right)$, and $\overline{P}^{F}_{G}(3;\ 3400;\ A^{v}, B^{v})$) is not interesting. Whatever the problem (fixed or variable energy density diet problem) and the premixes used (fixed or variable premixes), the use of premixes increases the feeding cost compared to the same problem without using premixes ($P^{V}(3)$ or $P^{F}(3)$).

On the other side, the use of variable premixes allows us to reduce the feeding cost by 6 % using daily phases ($\overline{P}^{V}_{G}(111;\ A^{v}, B^{v})$), compared to the current situation (Table 18.4, Fig. 18.5). Nonetheless, this solution has a problem; phosphorus and nitrogen excretions are increased (phosphorus, 7 %; nitrogen, 3 %).

Whether we use fixed or variable energy density formulations, the use of variable premixes further reduces the cost when compared to A^{f} and B^{f}. Thus, the use of variable premixes in the fixed energy density model with daily phases (P^{F}_{G} (111; 3400; A^{v}, B^{v})) reduces feeding cost by 1 %. In this case, nitrogen and phosphorus excretions are also reduced by 10 % concerning the phosphorus and 14.5 % concerning the nitrogen.

To conclude, the most compatible with production practices feeding plan concerning costs is to feed animals according to the variable energy density model using daily phase and variable premixes. The drawback of this formulation is the environmental impact. Indeed, phosphorus and nitrogen excretions are significantly increased. The problem should be solved using a multi-criteria model, which optimizes cost and excretions of the diet proposed in (Dubeau et al. 2011).

References

Stigler GJ (1945) The cost of subsistence. Am J Agric Econ 27:303–314. doi:10.2307/1231810

Dantzig GB (1998) Linear programming and extensions. Princeton University Press, Princeton

Dantzig GB (1990) The diet problem. Interfaces 20:43–47. doi:10.1287/inte.20.4.43

Garille SG, Gass SI (2001) Stigler's diet problem revisited. Oper Res 49:1–13. doi:10.1287/opre.49.1.1.11187

Subcommittee on Swine Nutrition, Committee on Animal Nutrition, National Research Council (1998) Nutrient Requirements of Swine. 10th Revised edn. National Academic Press, Washington, DC

Whittemore C, Fawcett RH (1976) Theoretical aspects of a flexible model to stimulate protein and lipid growth in pigs. Anim Prod 22:87–96. doi:10.1017/S0003356100035455

van Milgen J, Valancogne A, Dubois S, Dourmad JY, Sève B, Noblet J (2008) InraPorc: a model and decision support tool for the nutrition of growing pigs. Anim Feed Sci Technol 143:387–405. doi:10.1016/j.anifeedsci.2007.05.020

Black JL, Campbell RG, Williams IH, James KJ, Davies GT (1986) Simulation of energy and amino acid utilisation in the pig. Res Dev Agr 3:121–145

Dubeau F, Julien PO, Pomar C (2011) Formulating diets for growing pigs: economic and environmental considerations. Ann Oper Res 190:239–269. doi:10.1007/s10479-009-0633-1

Chapter 19
Markov Decision Processes to Model Livestock Systems

Lars Relund Nielsen and Anders Ringgaard Kristensen

19.1 Introduction

Mathematical models for livestock farming systems have been used since the 1950s. Examples of techniques used include deterministic optimization such as linear programming (for an early example, see Fisher and Schruben 1953) and dynamic programming (with White 1959, as one of the first applications to livestock farming) as well as stochastic models based on Monte Carlo simulation (e.g. Sørensen et al. 1992) and Markov decision processes (*MDPs*).

The nature of livestock systems differ from other industrial systems. Compared to, e.g., modeling the state of a machine, modeling the state of, e.g., a cow is more complex. First, the traits of an animal is harder to estimate and animals like humans differ, i.e., the variance between animals is much higher and it is harder to determine which state the animal is in. Second, livestock systems have a cyclic nature. In most cases an animal is inserted into the herd and after some cyclic periods (lactations, parity, feeding cycle) replaced with a new animal. Decisions regarding which cycle and when to replace the animal within the cycle have to be taken. Finally, often the supply of animals is not unlimited, e.g., a cow cannot be replaced if we do not have a heifer available. These three characteristics have also been referred to as the uniformity, reproductive cycle, and availability features of livestock systems (Ben-Ari et al. 1983).

Livestock farming is often sequential in nature. For instance at a specific time instance the decision on whether or not to replace an animal is based on observed

---

L.R. Nielsen (✉)
CORAL, Department of Economics and Business, Aarhus University, Aarhus, Denmark
e-mail: lars@relund.dk

A.R. Kristensen
Department of Large Animal Sciences, University of Copenhagen, Copenhagen, Denmark
e-mail: ark@life.ku.dk

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_19

information and expectation about the future. At the next decision epoch updated information is available and the decision choice is re-evaluated. Since random variation is a core property of a livestock system, MDPs have often been used to model livestock decision problems over the last decades (see Kristensen 1994, for an overview). At a specified point in time, the decision maker observes the state of a system and makes a decision. The decision and the state of the process produce two results: the decision maker receives an immediate reward (or incurs an immediate cost), and the system evolves probabilistically to a new state at a subsequent discrete point in time. At this subsequent point in time, the decision maker faces a similar problem. However, the observed state may be different from the previously observed state. The goal is to find a policy of decisions (dependent on the observation of the state) that maximizes, for example, the expected discounted reward.

In the MDP the state of the animal is defined by a set of state variables, each representing a trait relevant for the livestock system under consideration, e.g. for a dairy cow state variables could be milk yield level, lactation number, days in milk, reproductive status etc. It is assumed that the value of the state variable belongs to a finite set of levels/classes that represent the value of the trait. Often a trait is continuous and must be discretized into a set of levels. If we consider a realistic number of levels we may face the problem known as the "curse of dimensionality": the number of possible states grows exponentially with the number of state variables (the state space is often formed as the cartesian product of the number of levels of each of the state variables). This is one of the major drawbacks of using a MDP to model a livestock system.

Hierarchical MDPs (*HMDPs*) are an attempt to decompose the state space and to reduce the number of states in the MDP. The model is a series of finite time MDPs built together into one MDP called the founder process. As a result, the age of the animal can be omitted in the state space compared with an ordinary MDP model. Moreover, it takes into account that the production is cyclic. When a replacement occurs, not just a regular state transition takes place but rather the process (life cycle of the replacement animal) is restarted. HMDPs were first considered by Kristensen (1988) assuming two levels in the HMDP. Later, Kristensen and Jørgensen (2000) extended the methodology to multi-level HMDPs such that MDPs can be built together at multiple levels. Note that an HMDP is an infinite-stage MDP with parameters defined in a special way, but nevertheless in accordance with all usual rules and conditions relating to such processes. The basic idea of the hierarchic structure is that stages of the process can be expanded to the so-called child processes, which again may expand stages further to new child processes leading to multiple levels. Even though that HMDPs may help to reduce the number of state variables, the curse of dimensionality is still a problem.

In most papers an MDP is used to model a single animal and its successors (single-component). Hence herd constraints (heifers, feed, milk-quota, etc.) are not taken into account. To represent the whole herd a multi-component MDP has to be considered as discussed in Ben-Ari and Gal (1986) and Kristensen (1992). The multi-component model is based on single-component MDPs representing a single animal and its future successors. However, the model is far too large for optimization in practice. Therefore, the need for an approximate method emerged, and a method called *parameter iteration* was introduced by Ben-Ari and Gal (1986) and later modified by Kristensen (1992) to whom reference is made for details. To the authors' knowledge the parameter iteration method has only been applied under a constraint of a limited supply of heifers (Kristensen 1992).

The state of an MDP must be directly observable. Since the state in the model represents the present traits of the animal in question, this means that the traits are assumed to be well defined and directly observable. This is not always the case. Traits of an animal vary no matter whether we are considering the milk yield of a dairy cow or the litter size of a sow. Moreover, it is not obvious to what extent the observed trait is a result of a permanent property of the animal or a temporary random fluctuation. Most often the observed value is the result of several permanent and random effects. This problem can be solved by modeling the trait as a stochastic process and embedding the parameters of the process into the MDP instead of the observed value of the trait. The technique is referred to as *Bayesian updating*. As observations are done, the Bayesian approach is used to increase the knowledge on the true value of the trait. The technique was first used in practise by Kennedy and Stott (1993) for milk yield and has been described in detail by Kristensen (1993) and generalized in Nielsen et al. (2011).

For an MDP to be valid the *Markov property* must be fulfilled. It implies that the state space at any decision epoch (or *stage*) must contain sufficient information for determination of the probability distribution of the state to be observed at next decision epoch. In a straight forward formulation of a decision problem this is rarely the case, and various tricks must be used in order to make the process Markovian. The most common trick is to include *memory variables* in the state space (for instance the milk yield of previous lactation(s) in dairy cow models). This approach has been used in numerous models in practice. A more elaborate approach is to use Bayesian updating to estimate latent traits (for instance an abstract milk yield capacity of a dairy cow) as observations are done over time.

The objective of this chapter is to review the increasing amount of papers using MDPs to model livestock farming systems and provide an overview over the recent advances within this branch of research. Moreover, theory and algorithms for solving both ordinary and hierarchical MDPs are given and possible software for solving MDPs are considered. The chapter provides and updated overview compared to the latest survey (Kristensen 1994) which is almost 20 years old. The authors have tried to include all peer-review articles using MDPs to model livestock systems which resulted in more than 80 papers in total. Some very old applications (mainly from the 1960s and 1970s) have been omitted in the overview. Most of those early applications were deterministic, and some of them were published in research reports which are not available online. Readers who are interested in those papers are referred to Kennedy (1986) who gives an overview of applications until the early 1980s.

The chapter is organized as follows. In Sect. 19.2 a short introduction to ordinary MDPs and hierarchical MDPs is given and algorithms for optimizing the process are described. Next, a survey over papers using MDPs applied to cattle farming problems is given in Sect. 19.3. Dairy production is the most successful area on which MDPs have been applied. The chapter is continued in Sect. 19.4 with a survey over papers within the area of pig production. Finally, a few papers which lies outside these two areas are considered in Sect. 19.5. Software for solving both ordinary and hierarchical MDPs are discussed in Sect. 19.6. At last conclusions and directions for future research are discussed in Sect. 19.7.

19.2 Methodology

We briefly introduce the methodology of MDPs and describe the different algorithms which can be used to find an optimal policy under different criteria. Many papers using MDPs to solve livestock problems consider a stochastic process where the length of a stage is not constant. This is actually an extension of the MDP methodology (where a constant stage length is assumed), referred to as a semi MDP (Tijms 2003). However, due to the use of the term MDP instead of a semi MDP in the past we will stick to this. Indeed, throughout the rest of the paper we will use the term MDP for both ordinary and hierarchical (semi) MDPs and explicit write ordinary or hierarchical if needed.

19.2.1 Finite-Horizon Markov Decision Processes

We consider an ordinary finite-horizon MDP with N *stages*. At stage n the system occupies a *state* belonging to the finite set of system states S_n. Given that the decision maker observes state $s \in S_n$ at stage n, he must choose an *action* a from the set of finite allowable actions $A_{s,n}$ generating an immediate *reward* $r_s^a(n)$. Let $t_s^a(n)$ denote the expected length of stage n, i.e., the time until the system evolves probabilistically to a new state (decision epoch) and $\beta_s^a(n)$ the corresponding *discount rate* of the stage. Note that if α denotes the interest rate per time unit, and the stage length is L, then the discount factor is $\exp(-\alpha L)$ if we assume continuous compounding or $1/(1+\alpha)^L$ if we assume periodic compounding. Let $p_{s\hat{s}}^a(n)$ denote the *transition probabilities* of observing state $\hat{s} \in S_{n+1}$ at stage $n + 1$ given state s and action a.

A *policy* δ is a function that assigns to each state s a fixed action $a = \delta(s)$, i.e., a policy provides the decision maker with a plan of which action to take given stage and state. Under a given policy we write $r_s^a(n)$, $t_s^a(n)$ and $p_{s\hat{s}}^a(n)$ as $r_s^\delta(n)$, $t_s^\delta(n)$ and $p_{s\hat{s}}^\delta(n)$, respectively.

Let X_n denote the state of the system at the n'th decision epoch. Under a finite time-horizon the *total expected discounted reward* criterion may be relevant when consider livestock problems:

$$h(\delta) = \mathbb{E}\left(\sum_{n=1}^{N} r^{\delta}_{X_n}(n) \prod_{i=1}^{n-1} \beta^{\delta}_{X_i}(i)\right), \tag{19.1}$$

where the product is the total discount factor need to discount the reward at stage n back to stage 1. Moreover, if no discounting is used ($\alpha = 0$), then (19.1) calculates the *total expected reward*. It is assumed that no decision is taken at decision epoch N, i.e., a deterministic dummy action $a_N = \delta(X_N)$ is taken. The reward $r^{a_N}_{X_N}(N)$ is often referred to as the *terminal or salvage reward*.

Having introduced the notation for an MDP, we are also able to give a formal definition of the Markov property mentioned in the introduction. The Markov property is satisfied in an MDP if and only if

$$\mathrm{P}_a(X_{n+1}|X_n) = \mathrm{P}_a(X_{n+1}|X_n, \ldots, X_1) = p^a_{X_n X_{n+1}}, \forall n < N, X_n \in S_n, a \in A_{X_n,n}, \tag{19.2}$$

where P_a denotes the probability function under the decision a. In words it means that the state at next stage is only allowed to depend on the present state and action. Any other historical information is of no relevance. It is essential for the correctness of the results from an MDP that this property is satisfied.

An optimal policy maximizing (19.1) can be found using the following *Bellman equations*, Bellman (1957):

$$v_n(s) = \begin{cases} \max\limits_{a \in A_{s,n}} \left\{ r^a_s(n) + \beta^a_s(n) \sum\limits_{\hat{s} \in S_{n+1}} p^a_{s\hat{s}}(n) v_{n+1}(\hat{s}) \right\} & n < N \\ r^{a_N}_s(N) & n = N \end{cases}, \tag{19.3}$$

where $v_n(s)$ is the total expected discounted reward in state s at stage n under the optimal policy until the process terminates. Equation (19.3) shows that the optimal policy can be found by analyzing a sequence of simpler inductively defined single-stage problems. This is often referred to as *value iteration*.

19.2.2 Infinite-Horizon Markov Decision Processes

A situation where the stage of termination is unknown (or at least far ahead) is usually modeled using an infinite planning horizon ($N = \infty$). Given that the process is time homogeneous, i.e., the states and actions are independent of stage number and the policy stationary (constant over stages), we can drop the index n from the notation given in Sect. 19.2.1. Criterion (19.1) can still be considered (now an infinite sum) and will converge toward a fixed value when increasing N if discount rates are less than one.

Let $Z(t)$ denote the total reward incurred until time t and assume that the MDP is unichain (see Tijms 2003 for a formal definition). As an alternative criterion we may consider the *average reward per time unit*:

$$g(\delta) = \lim_{t\to\infty} \frac{Z(t)}{t} = \frac{\sum_{s\in S} \pi_s^\delta r_s^\delta}{\sum_{s\in S} \pi_s^\delta t_s^\delta} \tag{19.4}$$

where π_s^δ are the limiting state probabilities or *equilibrium distribution probabilities* given policy δ. Other criteria such as the *average reward per physical output* can also be considered and are defined as in (19.4) by redefining t_s^a as the physical output instead. For instance, Nielsen et al. (2004) maximize the average reward per steer. Furthermore, if all stages have equal length the denominator of (19.4) equals one and (19.4) reduces to the well-known formula for an ordinary MDP.

Various optimization techniques can be used to find the optimal policy such as value iteration, policy iteration, and linear programming. We will restrict ourselves to the first two here since linear programming has only been used in two of the papers reviewed.

Value iteration can be used to approximate the optimal policy. It has been used in the majority of papers since it is relatively straightforward to implement the algorithm. Moreover, the algorithm is good for solving large-scale MDP problems since there is no need for solving a large set of equations simultaneously. However, the number of iterations is problem dependent and typically increases in the number of states of the problem under consideration. The value iteration algorithm is given in Fig. 19.1. The algorithm is initialized in Step 0 where a pre-specified small accuracy number ε is chosen. Next, we use the recursive equations to update $v_s(n)$, which under criterion (19.1) denotes the total expected discounted reward in state s with n periods left and a terminal cost of $v_s(0)$. Under criterion (19.4) the recursive equation is based on a data transformation method (see Tijms 2003). This is repeated until the stopping condition is met (Step 3).

Note that if ε is sufficiently small and the same policy is found during several iterations, we may be rather sure that the optimal policy has been found. However, there is no guarantee but for practical purposes the deviation will have no

```
Step 0: Set v_s(0) such that 0 ≤ v_s(0) ≤ min_{a∈A_s}{r_s^a/t_s^a}, ∀s ∈ S. Choose a number ε > 0, set
  n←0 and τ = min_{s∈S,a∈A_s}{t_s^a} (under criterion (19.4)).
Step 1: For each s ∈ S compute v_s(n) using the recursive equation in Table 19.1 and
  let δ be the policy whose actions maximize v_s(n).
Step 2: Compute the bounds m_n = min_{s∈S}{v_s(n) − v_s(n−1)} and M_n = max_{s∈S}{v_s(n) − v_s(n−1)}
Step 3: If the condition in Table 19.1 is statisfied then stop; otherwise set n←
  n+1 and go to Step 1.
```

Fig. 19.1 Value iteration algorithm for an infinite-horizon ordinary MDP

Table 19.1 Equations and expressions to be used in the value iteration algorithm

Criterion	Step 1—recursive equation	Step 3—condition
(19.1)	$v_s(n) = \max_{a \in A_s} \left\{ r_s^a + \sum_{\hat{s} \in S} \beta_s^a p_{s\hat{s}}^a v_{\hat{s}}^{\delta}(n-1) \right\}$	$M_n \leq \varepsilon$
(19.4)	$v_s(n) = \max_{a \in A_s} \left\{ \frac{r_s^a}{t_s^a} + (1 - \frac{\tau}{t_s^a}) v_s(n-1) + \frac{\tau}{t_s^a} \sum_{\hat{s} \in S} p_{s\hat{s}}^a v_{\hat{s}}^{\delta}(n-1) \right\}$	$0 \leq M_n - m_n \leq \varepsilon m_n$

Step 0: Choose a policy δ.
Step 1: Solve the set of linear equations in Table 19.2
Step 2: For each state s, find the action a that maximizes the expression given in Table 19.2, and set $\delta'(s) = a$.
Step 3: If $\delta' = \delta$ then stop; otherwise go to Step 1.

Fig. 19.2 Policy iteration algorithm for an infinite-horizon ordinary MDP

significance. Under criterion (19.4) the stopping criterion ensures that $0 \leq (g^* - g(\delta))/g^* \leq \varepsilon$, where g^* denotes the optimal value to (19.4), i.e., the average reward per time unit $g(\delta) \in [m_n, M_n]$ is at most $100\varepsilon\%$ away from the optimal average reward per time unit. Finally observe that if the time between each decision epoch is constant ($t_s^a = 1$ and $\beta_s^a = \beta$), then the recursive formulas in Table 19.1 reduces to the well-known formulas for an ordinary MDP. During the years more advanced variants of value iteration algorithms have been developed which provide faster convergence and better stopping conditions. The interested reader is referred to Tijms (2003) and Puterman (1994) for details.

Policy iteration unlike value iteration finds an optimal policy in a finite number of steps. The algorithm is robust in the sense that in general it converges very fast, the number of iterations are independent of the number of states and varies typically between 3 and 15 (Tijms 2003). However, to use the algorithm $|S|$ linear equations must be solved simultaneously which may be computational costly for large state spaces. The policy iteration algorithm is given in Fig. 19.2. In Step 0 an arbitrary policy is chosen and in Step 1 the set of equations is solved. Under criterion (19.1) v_s denotes the total expected discounted reward of a process starting in state s and running over an infinite number of stages. Under criterion (19.4) v_s is the relative value compared to state $\hat{s}$. The difference between the relative value of two states denotes the amount we are willing to pay for stating in the state with the highest relative value. In Step 2 we update the current policy. This is repeated until a better policy can not be found (Step 3). Finally, observe that if the time between each decision epoch is constant ($t_s^a = 1$ and $\beta_s^a = \beta$), then the recursive formulas in Table 19.2 reduce to the well-known formulas for an ordinary MDP. For more advanced variants of the policy iteration algorithm see Puterman (1994).

Table 19.2 Equations and expressions to be used in the policy iteration algorithm

	Step 1		Step 2
Criterion	Equations	Unknowns	Expression
(19.1)	$v_s = r_s^\delta + \sum_{\hat{s}\in S} \beta_s^\delta p_{s\hat{s}}^\delta v_{\hat{s}}, \forall s \in S$	$v_1, \ldots, v_{\lvert S\rvert}$	$r_s^\delta + \sum_{\hat{s}\in S} \beta_s^\delta p_{s\hat{s}}^\delta v_{\hat{s}}^\delta$
(19.4)	$v_s = r_s^\delta - g t_s^\delta + \sum_{\hat{s}\in S} p_{s\hat{s}}^\delta v_{\hat{s}}, \forall s \in S, v_{\hat{s}} = 0$	$v_1, \ldots, v_{\lvert S\rvert}, g$	$r_s^\delta - t_s^\delta g(\delta) + \sum_{\hat{s}\in S} p_{s\hat{s}}^\delta v_{\hat{s}}^\delta$

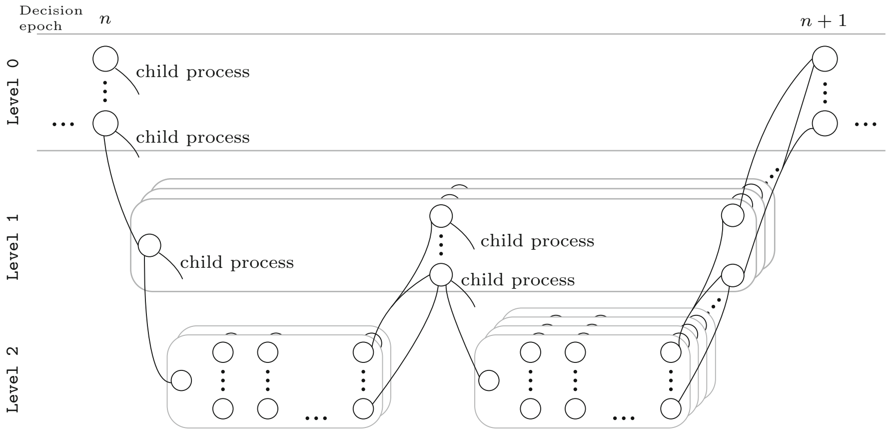


Fig. 19.3 Illustration of a stage in a hierarchial MDP. Level 0 indicates the founder level, and the nodes indicate states at the different levels and stages. A child process (*oval box*) represents a finite horizon MDP and is uniquely defined by a given state and action of its parent process (the specific link/edge from the parent to the child). Links at the last stage of a process illustrate the possible transitions back to the parent process when the child process ends

19.2.3 Hierarchical MDPs

Hierarchical MDPs are an attempt to decompose the state space and reduce the number of states in the MDP. The approach also provide a more intuitively way of modeling the stochastic process. Moreover, it reduces the number of equations which must be solved simultaneously under policy iteration. We consider hierarchical MDPs with multiple levels also referred to as multi-level hierarchic Markov processes. A hierarchical MDP is an infinite stage MDP with parameters defined in a special way, but nevertheless in accordance with all usual rules and conditions relating to such processes. The basic idea of the hierarchic structure is that stages of the process can be expanded to a so-called child processes which again may expand stages further to new child processes leading to multiple levels.

A stage in a process with three levels is illustrated in Fig. 19.3. The infinite horizon process at level 0 is named the *founder process* and is the only process in the structure which is not the child of a parent process. Each node corresponds to a state at different levels and stages. A child process (oval box) is a finite horizon

MDP and is uniquely defined by a given stage, state, and action of its parent process (the specific link/edge from the parent to the child). For each finite horizon process an initial probability distribution of the states at stage 1 is assumed, i.e., a fictitious stage 0 with only one state and one action is added to the model. As a result given a state and action at the parent level a transition to the child process can be represented deterministically (edges in Fig. 19.3). Moreover, a set of terminal probabilities are given representing the transition probabilities back to the parent process when the last stage ends (the links from the last stage in the child in Fig. 19.3).

Note that a finite horizon process at level $l > 0$ is uniquely defined by a sequence of stages, states, and actions $\rho = (s_0, a_0, n_1, s_1, a_1, \ldots, n_{l-1}, s_{l-1}, a_{l-1})$ and at level 0 we only have the infinite horizon founder process which we will denote ρ_0. We will use the notation in Sects. 19.2.1 and 19.2.2 given a specific process ρ; however, an action a is not necessarily identical to an action as it is usually defined in an MDP. In addition to the selection of a specific process we also have to choose which policy to follow during its child processes.

Let δ_ρ denote an *expanded policy* of process ρ, i.e., a function that assigns to each state s a fixed action $a = \delta_\rho(s)$, i.e., an expanded policy provides the decision maker with a plan of which action to take given stage and state in the parent process and all its child processes. Then the reward $r_s^{\delta_\rho}(n)$, expected length $t_s^{\delta_\rho}(n)$, discount factor $\beta_s^{\delta_\rho}(n)$, and transition probabilities $p_{s\hat{s}}^{\delta_\rho}(n)$ can be calculated recursively by processing the child processes from the lowest levels and upward toward the parent process ρ. Hence an *expanded value iteration* can be applied. Under the *total expected discounted reward* criterion (19.1) and given a set of terminal rewards, the optimal policy δ_ρ of a finite horizon process can be found by recursively applying value iteration (19.3) from the lowest levels and upward toward the parent process ρ. The same holds when considering the *average reward per time unit* criterion (19.4) where we must solve the following recursive equations:

$$v_n(s) = \begin{cases} \max\limits_{a \in A_{s,n}} \left\{ r_s^a(n) - g t_s^a + \sum\limits_{\hat{s} \in S_{n+1}} p_{s\hat{s}}^a(n) v_{n+1}(\hat{s}) \right\} & n < N \\ r_s^{a_N}(N) & n = N \end{cases}, \tag{19.5}$$

Note that an additional average reward g must be chosen together with the terminal values. For further details see Kristensen and Jørgensen (2000).

We can also apply a single iteration of expanded value iteration to the founder process to determine all the parameters needed to solve the set of equations when considering policy iteration. A hierarchical policy iteration algorithm can now be formulated in Fig. 19.4. It combines policy iteration at the founder level and value iteration at the other levels. First some initial values are chosen in Step 0 and the expanded policy and the parameters of the founder process are calculated. Next the linear equations at the founder level are solved in Step 1 and used as terminal values in the expanded value iteration in Step 2. If no new policy is found the algorithm stops in Step 3.

```
Step 0: Set $v_s(0)=0, \forall s \in S_{\rho_0}$ and $g=0$ (under criterion (19.4)). Perform an expanded
  value iteration to find the expanded policy $\delta_{\rho_0}$ and parameters $r_s^{\delta_{\rho_0}}$, $t_s^{\delta_{\rho_0}}$, $\beta_s^{\delta_{\rho_0}}$
  and $p_{s\hat{s}}^{\delta_{\rho_0}}$.
Step 1: Solve the set of linear equations in Table 19.2 using the parameters of
  the founder process.
Step 2: Perform expanded value iteration to find the expanded policy $\delta'_{\rho_0}$ and
  parameters $r_s^{\delta'_{\rho_0}}$, $t_s^{\delta'_{\rho_0}}$, $\beta_s^{\delta'_{\rho_0}}$ and $p_{s\hat{s}}^{\delta'_{\rho_0}}$.
Step 3: If $\delta'_{\rho_0} = \delta_{\rho_0}$ then stop; otherwise redefine $\delta_{\rho_0}$ to the new policy and go
  to Step 1.
```

Fig. 19.4 Hierarchical policy iteration algorithm for an hierarchical MDP

19.3 MDP Models Applied to Cattle Farming

This section gives an overview of MDPs applied to cattle farming problems. Around 60 papers describing more than 40 different models were found in this area. Table 19.3 summarizes the models by listing their structure in terms of the number of levels (the value 1 indicates an ordinary MDP), the criterion of optimality, the state variables with number of levels/classes, stage lengths with maximum number of stages, decisions being optimized, application area, and supplementary information. Each row in the table corresponds to a model and reference to the paper(s) describing it is given in the first column. It should be noticed that it is not always clear whether a paper should be classified as describing a new model (by further developing an existing model) or it should be classified as just an application of an existing model.

Only *decision* models are included in the survey. Simple Markov chain models are not mentioned even though they are, of course, closely related to MDPs since an MDP with a predefined policy is a Markov chain. Examples of such, not included, Markov chain models are Giordano et al. (2012), Cabrera (2012), Allore et al. (1998), Noordegraaf et al. (1998), as well as Jalvingh et al. (1993a,b, 1994).

Many of the models mentioned in the survey are by the authors themselves presented as *dynamic programming* models and the term Markov decision process is seldom mentioned. Dynamic programming exists in a deterministic version and a stochastic version, and particularly the stochastic version is identical to the MDP concept described in this chapter. Very often, however, the use of the term dynamic programming implies that the optimization method is value iteration. The deterministic version is also compatible with an MDP, but such models are degenerate in the sense that for any stage n, state s, and action a there exists a state s' at stage $n+1$ where $p_{ss'}^{a} = 1$. Accordingly, we have for any state $\hat{s} \neq s'$ that $p_{s\hat{s}}^{a} = 0$.

In a book Kennedy (1986) reviewed dynamic programming applications to agriculture until the early 1980s. As a main rule, models mentioned in that book are omitted, but for the most important application area, which is dairy cow replacement, also models mentioned by Kennedy (1986) are included. The main reason is that the study by Giaever (1966) is so important that it would be preposterous to omit it.

Table 19.3 Overview over literature using MDPs for modeling within cattle farming

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Kalantari and Cabrera (2012)	1	*DR* (VI)	Lactation (9), days in pregnancy (282), DIM (750), milk yield (5)	Day (∞)	K, R	Dairy (US)	Study the effect of reproductive performance
Heikkila et al. (2012)	1	*DR* (PI)	Month (78), culling reason (3), mastitis cases (5)	Month (∞)	K, R	Dairy (FIN)	Focus on clinical mastitis
Langford and Stott (2012)	1	*DR* (VI)	Parity (12), milk yield level (15)	Parity (20)	K, R	Dairy (UK)	Extension of Stott (1994) which study the effect on welfare
Cha et al. (2011)	3	*DR* (HPI)	Permanent milk yield level (5); dummy (1); temporary milk yield level (5), pregnancy state (9), clinical mastitis state (13)	Cow life (∞); parity (8); month (20)	I, K, R	Dairy (US)	Lactation number and stage of lactation known from stage number. Extension of the work by Bar et al. (2008b) and Cha et al. (2010)
Demeter et al. (2011)	4	*DR* (HPI)	Permanent milk yield potential (PMYP) estimated at first calving (13); PMYP estimated at the beginning of lactation (13), months open previous lactation (8); PMYP estimated this month (13), temporary milk yield capacity (13), pregnancy state (2); PMYP estimated this month (13), temporary milk yield capacity (13)	Cow life (∞); parity (12); month/gestation period (18); month (9)	I, K, R	Dairy (NL)	Used to assess herd level implication of genetic selection strategies. Lactation number, stage of lactation, and month of pregnancy known from stage numbers

(continued)

Table 19.3 (continued)

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Cabrera (2010)	1	R/T (LP)	Parity (15), month in lactation (24), pregnancy status (10)	Month (∞)	K, R	Dairy (US)	Consider different diets and nitrogen excretion
Cha et al. (2010)	3	*DR* (HPI)	Permanent milk yield level (5); dummy (1); temporary milk yield level (5), pregnancy state (9), lameness state (13)	Cow life (∞); parity (8); month (20)	I, K, R	Dairy (US)	Lactation number and stage of lactation known from stage number. Extension of the work by Bar et al. (2008b) with focus on lameness
Kalantari et al. (2010)	1	DR (VI)	Lactation (12), month after calving (24), milk production class (15), pregnancy status (10)	Lactation (180)	K, R	Dairy (IR)	A modification of van Arendonk and Dijkhuizen (1985) applied to Iran conditions
Nielsen et al. (2010)	3	*DR* (HPI)	Dummy (1); milk yield potential (MYP) esti-mated at the beginning of lactation (13); combi-nation of MYP estimated until present day and temporary milk yield level (45 combinations), drying off week (32)	Cow life (∞); parity (10); day (483)	K, R	Dairy (DK)	Lactation number and stage of lactation known from stage number. Focus on management. Bayesian updating used
Bar et al. (2008a,b)	3	*DR* (HPI)	Permanent milk yield level (5); mastitis in previous lactation (2); temporary milk yield level (5), pregnancy state (9), mastitis state in present lactation (13)	Cow life (∞); parity (8); month (20)	I, K, R	Dairy (US)	Lactation number and stage of lactation known from stage number. Focus on cost of clinical mastitis

Heikkila et al. (2008)	1	*DR* (PI)	Lactation (10), milk yield (3), health status (3)	Lactation (∞)	K, R	Dairy (FIN)	Focus on diseases and milk yield
Nielsen and Kristensen (2007); Nielsen et al. (2004)	4	*R/T* (HPI), *R/Q* (HPI)	Birth month (12); live weight (up to 26) previous winter feeding level (2), weigh gain (5); weight gain at fattening (3)	Steer life (∞); seasons (summer/winter) (6); month (up to 6); month (4)	G, Fe, Fa, R	Steer (DK)	Nielsen et al. (2004) consider average reward per steer while in Nielsen and Kristensen (2007) the average reward per time unit is maximized
de Vries (2006)	1	*DR* (VI)	Lactation (12), days open (10), month of lactation (24), milk yield (15)	Month (∞)	K, R	Dairy (US)	Extension of model by de Vries (2004)
Stott et al. (2005)	1	*DR* (VI)	Lactation (12), milk yield (15)	Lactation (20)	K, R	Dairy (UK)	Studies financial incentive to control paratuberculosis. Extension of model by Stott (1994)
de Vries (2004)	1	*DR* (VI)	Lactation (12), days open (10), month of lactation (24), milk yield (15), month of calving (12)	Month (∞)	K, R	Dairy (US)	Studies the effect of delayed replacement with seasonal cow performance
Grohn et al. (2003)	1	*DR* (VI)	Lactation (12), days open (10), month of lactation (20), milk yield (5), month of calving (12), disease state (240)	Month (60)	I, K, R	Dairy (US)	Extension of models by Delorenzo et al. (1992) and Mccullough and Delorenzo (1996b)
Stott et al. (2002)	1	*DR* (VI)	Lactation (12), milk yield (15), somatic cell count (11)	Lactation (20)	K, R	Dairy (UK)	Extension of model by Stott (1994)

(continued)

Table 19.3 (continued)

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Pihamaa and Pietola (2002)	1	*DR* (VI)	Live weight (507)	Week (326)	Fe, K, R	Beef (FIN)	Study the effect of agricultural policy reforms in Finland
Rajala-Schultz and Grohn (2001)	1	*DR* (VI)	Lactation (12), production level (5), month of calving (12), month of lactation (19), days open (10)	Month (60)	I, K, R	Dairy (FIN)	Compares optimal decisions with farmer decisions. Use of model by Mccullough and Delorenzo (1996b)
Vargas et al. (2001)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15)	Month (180)	I, K, R	Dairy (CR)	Based on model by van Arendonk and Dijkhuizen (1985)
Rajala-Schultz et al. (2000a,b)	1	*DR* (VI)	Parity (12), days open (10), stage of lactation (19), production level (3, 5, 7), month of calving (12)	Month (48-120)	I, K, R	Dairy (FIN)	Use of model by Mccullough and Delorenzo (1996b)
Yalcin and Stott (2000)	1	*DR* (VI)	Lactation (12), milk yield (15), somatic cell count (11)	Lactation (20)	K, R	Dairy (UK)	Extension of work by Stott (1994)
Cardoso et al. (1999b,a)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15)	Month (240)	I, K, R	Dairy (BR)	Use of model by van Arendonk and Dijkhuizen (1985)
Mourits et al. (1999a,b)	2	*DR* (HPI)	Month of birth (12); body weight (173), reproductive state (32), prepubertal growth rate (3)	Rearing period (∞); month (30)	Fe, I, K, R	Heifers (NL)	Age of heifer known from stage number. The keep and inseminate decisions can be done under different growth strategies

Yates and Rehman (1998)	1	*DR* (LP)	Lactation (12), genetic level (4)	Year (10)	K, R	Dairy (UK)	The keep decision has two options: produce calf for replacement or for beef
Dekkers et al. (1998)	1	*DR* (VI)	Lactation l (12), month in lactation (16), milk yield l (15), calving intervals (6)	Month (180)	I, K, R	Dairy (CDN)	Quantify the impact of persistency of lactation. Adaptation of the work in van Arendonk and Dijkhuizen (1985)
Haran (1997)	2	*DR* (HPI)	Month of first calving (12); current month (12), milk production level (15), time of conception (5)	Cow life (∞); lactation stage (72)	I, K, R	Dairy (IRL)	Lactation number and stage of lactation known from stage number
Mccullough and Delorenzo (1996b,a)	1	*DR* (VI)	Lactation (12), production level (15), month of calving (12), month of lactation (19), days open (10)	Month (60)	I, K, R	Dairy (US)	Focus: levels of state variables, milk price and management inputs. Model based on Delorenzo et al. (1992)
Houben et al. (1994)	2	*R/T* (HPI)	Dummy (1); milk production l (15), $l-1$ (15), calving interval (18), mastitis current month (2), mastitis cases l (4), $l+1$ (4)	Life span of a cow (∞); month (204)	I, K, R	Dairy (NL)	Focus on mastitis
Stott (1994)	1	*DR* (VI)	Lactation (12), yield class (15)	Lactation (∞)	K, R	Dairy (UK)	Uses Bayesian updating for milk yield
Kennedy and Stott (1993)	1	*DR* (VI)	Lactation l (12), yield class (5), mastitis status $l-1$ (2)	Lactation (∞)	K, R	Dairy (UK)	Focus: model and Bayesian updating

(continued)

Table 19.3 (continued)

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Stott and Kennedy (1993)	1	*DR* (VI)	Lactation number (12), mastitis state (2)	Lactation (∞)	K, R	Dairy (UK)	Focus on clinical mastitis
Delorenzo et al. (1992)	1	*DR* (VI)	Lactation (12), production level (15), month of calving (12), month of lactation (16), days open (7)	Month (240)	I, K, R	Dairy (US)	Model based on van Arendonk (1986)
Dekkers (1991)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15), time of conception (6)	Month (180)	I, K, R	Dairy (CDN)	Studies economic values for breeding goals. Adaptation of the work in van Arendonk and Dijkhuizen (1985)
Boichard (1990)	1	*DR* (VI)	Lactation l (6), lactation stage (22), stage of conception (7), calving date (18), milk yield in l (9), $l-1$ (9)	20 days (200)	I, K, R	Dairy (F)	Focus: economic value of conception
Harris (1990)	1	*DR* (VI)	Lactation (10), best linear prediction of future milkfat production, milk volume production, milk protein production, breed, calving date (6)	Year (20)	K, R	Dairy (NZ)	It is not clear from the description whether an optimization is performed or the model is only used for simulation

Kristensen (1989); Kristensen and Thysen (1991a,b)	2	*R/Q* (HPI)	Estimated genetic class at first calving (5); milk yield of present lactation (15), milk yield of previous lactation (15), length of calving interval (8)	Cow life (∞); 4 week period (108)	K, R	Dairy (DK)	Lactation number and stage of lactation known from stage number. Average reward per kg milk is maximized. Extension of work by Kristensen (1987). The model is later applied by Kristensen and Thysen (1991b,a)
Rogers et al. (1988a,b)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15), time of conception (6)	Month (180)	I, K, R	Dairy (US)	Adaptation of the work in van Arendonk and Dijkhuizen (1985)
Kristensen (1987)	2	*DR* (HPI)	Estimated genetic class at first calving (5); milk yield of present lactation (15), milk yield of previous lactation (15), length of calving interval (8)	Cow life (∞); lactation stage (24)	K, R	Dairy (DK)	Lactation number and stage of lactation known from stage number
van Arendonk (1986)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), time of conception (6), month of calving (12)	Month (180)	I, K, R	Dairy (NL)	Extension of the work in van Arendonk (1985b)
van Arendonk and Dijkhuizen (1985); van Arendonk (1988)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15), time of conception (6)	Month (180)	I, K, R	Dairy (NL)	Extension of the work in van Arendonk (1985b)

(continued)

Table 19.3 (continued)

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
van Arendonk (1985b,a)	1	*DR* (VI)	Lactation l (12), stage in lactation (11), milk yield l (15), $l-1$ (15)	Month (240)	I, K, R	Dairy (NL)	The model has had a huge impact on later models
Ben-Ari et al. (1983); Ben-Ari and Gal (1986)	1	DR (VI)	Lactation, milk yield, body weight	Lactation (∞)	K, R	Dairy (IL)	Ben-Ari and Gal (1986) consider how to solve the multi-component system
Killen and Kearney (1978)	1	*R* (VI)	Lactation number (9)	Lactation (20)	R, K	Dairy (IRL)	Very small model
Stewart et al. (1977, 1978)	1	*DR* (VI)	Lactation (7), body weight (5), 305d milk yield (11), milk fat pct (7)	Lactation (10)	R, K	Dairy (CDN)	Stewart et al. (1977) describe the model and Stewart et al. (1978) consider different breeds. Culling decisions were assumed to occur at 60 days postcalving
McArthur (1973)	1	*R* (VI)	Lactation number (7), milk production level (80)	Lactation (15)	K, R	Dairy (NZ)	Milk yield represented as average over lactations
Smith (1973, 1971)	1	*DR* (VI)	Lactation l (6), yield in l (29), $l-1$ (29), calving interval (3)	Lactation (15)	R, K	Dairy (US)	Far more detailed model than the one by Giaever (1966)

Giaever (1966)	1	*DR* (VI)	Lactation number (5), calving interval (3), milk yield (7)			Dairy (US)	Alternative optimization methods described. Important considerations about Markov property

[a]Papers have been ordered in reverse order of year

[b]Number of levels in the MDP. If 1 then the MDP is an ordinary MDP

[c]DR = expected discounted reward, R = expected reward, R/T = average reward per time unit, R/Q average reward per quantity unit. Algorithm used is given in parentheses (VI = value iteration, PI = policy iteration, HPI = hierarchical policy iteration, LP = Linear programming)

[d]State variables for each level in the process (separated with semicolon). The number of levels/classes of each state variable is given in parentheses

[e]Stage length at each level in the process (separated with semicolon). Maximum number of stages given in parentheses

[f]R = replace, K = keep, I = Inseminate, G = Grazing, Fe = Feeding intensity, Fa = Fattening

[g]Animal group applied to. The country from which the parameters has been estimated is given in parentheses

The vast majority of papers and models address problems related to dairy cows. A few models consider growing cattle (the review by Kennedy 1986, contains several very early applications to growing cattle). Nielsen et al. (2004) and Nielsen and Kristensen (2007) consider the raising of steers and Pihamaa and Pietola (2002) study the effect of beef cattle management under agricultural policy reforms in Finland. Also management of heifers (Mourits et al. 1999a,b) has been studied. All models are defined at the individual animal level and since all of them also basically consider the replacement problem, they reflect a chain of animals successively replacing each others over a finite or infinite time horizon. They therefore all have the action "Replace" as an option. The alternative to replacement is, of course, to keep the animal, and many models only have "Keep" as an alternative to "Replace" Many models describing cows and heifers also have an "Inseminate" action, and the models optimizing raising of steers and heifers have actions defining the feeding level in some sense.

The first models published until the mid-1980s were ordinary MDPs solved by value iteration over a number of stages typically aiming at approximating an infinite horizon. The criterion of optimality was typically maximization of expected discounted reward, which is still today the most commonly used criterion. The concept of hierarchical MDPs was described by Kristensen (1988), and over the following years it has been increasingly used in cattle models. In total, 11 of the models mentioned in Table 19.3 are hierarchical. Most of the recent hierarchical models have been implemented in the MLHMP software system developed by Kristensen (2003). The technique has made it possible to handle even very large models with millions of states like Demeter et al. (2011), Nielsen et al. (2010), and Houben et al. (1994). The introduction of hierarchical models also implies that policy iteration has become a common optimization technique (for the founder process).

When it comes to state variables, the models include age of the animal as a state variable. For dairy cows it is typically measured by lactation number and often also stage of lactation. Also the reproductive state (typically measured by month of conception or length of calving interval) and the milk yield level are usually included in the dairy cow models. In the beginning the health status was not included in the models, but starting with Stott and Kennedy (1993), Kennedy and Stott (1993), and Houben et al. (1994) mastitis has often been included in the state space. In recent years (Bar et al. 2008a,b; Cha et al. 2011; Heikkila et al. 2012) mastitis has been studied intensively. Also other diseases have occasionally been included (Cha et al. 2010; Grohn et al. 2003; Heikkila et al. 2008).

When comparing state variables across models it is important to remember that in hierarchical models some of the state variables are typically implicitly given by stage number. This is typically the case for properties like age (lactation number and lactation stage for dairy cows) and/or season. Thus, in hierarchical models it is most often not necessary to include state variables for such properties because they are given by the model structure. Hence, the same problem formulated as a hierarchical model will typically have fewer state variables than if it had been formulated as an ordinary MDP.

Stage lengths (for hierarchical models at the most detailed level) vary from one day as in Kalantari and Cabrera (2012), Nielsen et al. (2010) to typically a lactation period in many early models. Geographically, the largest number of models (12) describe US conditions, but also models for UK conditions (8), Dutch conditions (6), Danish (4), and Finish conditions (4) are common. Two models describe MDPs developed for New Zealand, two for Ireland, two for Canada, and for each of the countries Iran, Costa Rica, France, and Israel one model has been developed.

Very few papers actively discuss how to satisfy the Markov property, but in many papers it is obvious that the problem is considered (in other papers it is ignored). The preferred method for (approximate) fulfilment of the Markov property has been by use of memory variables where milk yield of previous lactation is remembered. This tradition goes back to van Arendonk (1985b) and has been continued in many subsequent models using that model as a basis (see the "Misc" column of Table 19.3). The same approach was used by Kristensen (1987, 1989). The main drawback of memory variables is that they contribute considerably to the curse of dimensionality. This was realized already by Giaever (1966) who instead defined milk yield as a weighted index of all lactations until now. He showed how it was possible to define the weight coefficients of the index in such a way that the Markov property was not violated. Also McArthur (1973) defined an index which in his case was a simple average of lactation yields. Thus, the state space was reduced, but the Markov property was not satisfied.

Another approach used in several models is to express the milk yield as partly resulting from a permanent property of the cow. This approach was used by Kristensen (1987, 1989) (as a supplement to the memory variable also included). In the models developed at Cornell University (Bar et al. 2008a,b; Cha et al. 2010, 2011) the permanent property was the only approach used to satisfy the Markov property. All the models mentioned are hierarchical MDPs which are particularly well suited for handling permanent traits. Nevertheless, Harris (1990) seems to have used a similar principle in an ordinary MDP.

When the principles of Bayesian updating was described by Kristensen (1993) and (independently) applied by Kennedy and Stott (1993) a new tool became available for model builders. Instead of memory variables, the Bayesian updating focuses on estimating an abstract latent milk yield capacity of a cow based on *all* observed milk yield records. It was, however, not until the models by Nielsen et al. (2010) and Demeter et al. (2011) that it was implemented as a main feature. In other application areas (Kristensen and Søllested 2004a,b; Lien et al. 2003; Verstegen et al. 1998) it was used earlier.

19.4 MDP Models Applied to Pig Farming

Table 19.4 summarizes MDPs applied to pig farming along the same guidelines as for the cattle applications in Table 19.3. A total of 17 papers describing 12 different models were identified. As with the cattle models only decision models are included implying that simple Markov chain models are excluded. Examples of such not included Markov chain models are Jalvingh et al. (1992a,b) and Pla et al. (2003).

Table 19.4 Overview over literature using MDPs for modeling pig farming

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Kristensen et al. (2012)	2	*R*/*T* (HPI)	Dummy (1); number of pigs remaining (21), estimated permanent growth potential (7), estimated temporary growth potential (7), estimated within pen standard deviation (9)	Prod. cycle in pen (∞); week (17)	D_δ	Finishers (DK)	Embeds a dynamic linear model linking automatically recorded live weights to state variables. Group level: models a pen
Rodriguez et al. (2011)	3	*R*/*T* (HPI)	Dummy (1); exp. serially correlated effect (21), exp. permanent litter size potential (21)[h]; health status (2), gestation status (3), litter size (21), weak sow index of previous parity (5), weak sow index of present parity (5)	sow life (∞); parity (12); parity phases (3)	NM, AI, R, K, M_i	Sows (DK)	Extension of work by Kristensen and Søllested (2004a,b). The weak sow index is based on clinical observations
Toft et al. (2005)	2	? (HPI)	Disease transition (5); configurations of susceptible and infectious pigs (?), fraction of pigs still present (5)	Prod. cycle in pen (∞); day and week (88)	V, T, D_π, K	Finishers (DK)	Group level: models a batch

Niemi (2006)	1	*DR* (VI)	Lean tissue weight (37), fat tissue weight (52)	Day (1800)	R, P, E	Finishers (FIN)	Deterministic model. Very detailed control options
Kristensen and Søllested (2004a,b)	3	*R*/*T*	Dummy (1); exp. serially correlated effect (21), exp. permanent litter size potential (21)[h]; health status (2), gestation status (3), litter size (21)	Sow life (∞); parity (12); parity phases (3)	NI, AI, R, K, M_i	Sows (DK)	Uses Bayesian updating to estimate litter size
Pla et al. (2004)	1	*R*/*T* (PI)	Reproductive state (9), parity (11)	Variable (from event to event)	R, K	Sows (E)	Uses herd data for estimation of transition probabilities
Kure (1997b,a,c)	2	*DR* (HPI)	Observed live weigh, observed carcass leanness	Prod. cycle in pen (∞); weeks of delivery (4)	D_δ, E	Finishers (DK)	Uses recursive dynamic programming in child process. Group level: models a batch
Jørgensen (1993)	2	*DR* (VI)	Dummy (1); weeks since start (5), pigs in pen (32) [161]	Prod. cycle in pen (∞); week (5)	D_δ, E	Finishers (DK)	The first period at second level is actually 10 weeks (minimum feeding time)
Jørgensen (1992)	2	*L*/*T* (VI)	Dummy (1); parity (20), exp. random effect and influence on litter size (100) [2001]	Sow life (∞); parity (20)	R, K	Sows (DK)	Litter size based on Bayesian updating. Hierarchical structure imply reduced state space compared to (Huirne et al. 1993)

(continued)

Table 19.4 (continued)

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Huirne and Hardaker (1998); Huirne et al. (1993, 1991)	1	*DR* (VI)	Parity p (11), litter size in $p-1, p-2$ (12), unsuccessful breedings in p (4) [5633]	Parity (70)	R, K	Sows (NL)	Huirne and Hardaker (1998) uses the MDP as a sub-model.
Huirne et al. (1988)	1	*DR* (VI)	Parity p (15), litter size in $p-1, p-2, p-3$ (20)	Parity (50)	R, K	Sows (NL)	Litter size based on a dynamic formula
Glenn (1983)	1	*R* (VI)	Live weight (80), carcass composition ()	5 days (17)	G, P	Finishers (UK)	Deterministic model

[a]Papers have been ordered in reverse order of year
[b]Number of levels in the MDP. If 1 then the MDP is an ordinary MDP
[c]DR = discounted reward, L/T = avg. litter size per time unit, R/T = avg. reward per time unit. Algorithm used is given in parentheses (VI = value iteration, PI = policy iteration, HPI = hierarchical policy iteration)
[d]State variables for each level in the process (separated with semicolon). The number of levels/classes of each state variable is given in parentheses
[e]Stage length at each level (separated with semicolon). Maximum number of stages given in parentheses
[f]V = vaccinate, T = treat, D_π = deliver π pigs, R = replace, K = keep, D_δ = deliver pigs with weight above δ, E = empty the pen, NM = natural mating, AI = Artificial Insemination, M_i = allow i matings, P = protein level, E = energy level, G = gain
[g]Animal group applied to. The country from which the parameters has been estimated is given in parentheses
[h]A dummy state representing the pig has been culled is also included in the model

Analogously to the many dairy cow replacement models in the previous section a total of six sow replacement models were found. The remaining papers (6) address problems related to production of finishers. Also the pig models are in some sense replacement models, but unlike the cattle models there are also examples of MDPs defined at group level. Thus, Kristensen et al. (2012) model a pen, and Toft et al. (2005) as well as Kure (1997a,b,c) model a batch of finishers. There are, however, also examples of finisher models (Jørgensen 1993; Glenn 1983; Niemi 2006) defined at individual animal level. The sow models are all defined at individual animal level.

Decisions considered in the sow models are in addition to "Keep" and "Replace" also insemination method and number of inseminations to accept before culling for infertility. In finisher models decisions are the marketing policy and, some times, the feeding level. As concerns the optimization method the first models published were ordinary MDPs based on value iteration optimizing expected reward or expected discounted reward. Later hierarchical models became the norm with the deterministic model by Niemi (2006) as an exception. Also for the hierarchical pig models the preferred software tool has been the MLHMP system described by Kristensen (2003).

In all models the age of the animal(s) is included either as a state variable or indirectly through the stage number in hierarchical models. In the sow models litter size is often included either directly or through Bayesian updating of a latent litter size potential as in Kristensen and Søllested (2004a,b) and Rodriguez et al. (2011). Also, the number of unsuccessful inseminations is sometimes directly or indirectly (through the model structure) taken into account. One model by Rodriguez et al. (2011) included a weak sow index defined by clinical observations in the state space.

Stage lengths vary from one day as in Niemi (2006) to a reproduction period (parity) in several models. Geographically, the largest number of models (7) describe Danish conditions, but also models for Dutch, UK, Spanish, and Finish conditions are found.

As concerns the Markov property, the approach has been the same as with dairy models. Dutch models (Huirne and Hardaker 1998; Huirne et al. 1988, 1991, 1993) used memory variables (2 or 3 previous litter sizes). Later models (Jørgensen 1992; Kristensen and Søllested 2004a,b; Kristensen et al. 2012; Rodriguez et al. 2011) used Bayesian updating.

19.5 MDP Models Applied to Other Areas

Even though most models have been developed for applications within cattle and pig production, a few papers within other applications exist in the literature.

Table 19.5 summarizes MDPs applied to other areas within livestock farming along the same guidelines as in the tables for cattle and pig applications (Tables 19.3 and 19.4). A total of five papers were identified. In addition to those

Table 19.5 Overview over literature using MDPs within other areas than cattle and pig farming

Paper[a]	Levels[b]	Criterion[c]	State variables[d]	Stage length[e]	Decisions[f]	Application[g]	Misc
Viet et al. (2012)	1	*DR* (VI)	Current month (12), S = susceptible ($N+1$), I = infected ($N+1$), V = vaccinated ($N+1$); only state combinations where $S+I+V=N$	1 month (∞)	V, no V	Disease control (F)	Total number of herds N varied from 50 to 400
Ge et al. (2010b)	2	*DR* (HPI)	Epidemic situation (3), export ban (2), infection index (5), estimated growth potential of epidemic (5), uncertainty of growth potential (5)	Duration of epidemic (∞); 10 day periods (10)	BP, V, PC	FMD control (NL)	Modification of work by Ge et al. (2010a)
Ge et al. (2010a)	3	(HPI)	Epidemic situation (2), infection index (5), estimated growth potential of epidemic (5), uncertainty of growth potential (5)	Duration of epidemic (∞); 10 day periods (10); 1 day (10)	SP, BP, V, PC, STOP	FMD control (NL)	See also Ge et al. (2010b)
van Asseldonk et al. (1999)	1	*DR* (VI)	IT investment status (11^5): automatic concentrate feeder (11), activity measurement (11), milk production measurement (11), milk temperature measurement (11), conductivity measurement (11)	Year (20)	Invest	IT investment (NL)	Studies investments in IT equipment at farm level. Deterministic model.

Verstegen et al. (1998)	2	? (HPI)	Number of production weeks, yield per production week	Project life (∞); year (10)	R, K	MIS[i] evaluation (NL)	Project age known from stage number. Model with Bayesian updating. Use the MDP as a tool for comparing against farmers choice

[a]Papers have been ordered in reverse order of year

[b]Number of levels in the MDP. If 1 then the MDP is an ordinary MDP

[c]*DR* = discounted reward, *L/T* = litter size per time unit, *R/T* = reward per time unit. Algorithm used is given in parentheses (*VI* = value iteration, *PI* = policy iteration, *HPI* = hierarchical policy iteration)

[d]State variables for each level in the process (separated with semicolon). The number of levels/classes of each state variable is given in parentheses

[e]Stage length at each level (separated with semicolon). Maximum number of stages given in parentheses

[f]*V* = vaccinate, *BP* = Basic control (FMD), *PC* = Preemptive culling (FMD), *SP* = Stop program (FMD), *R* = replace, *K* = keep

[g]Animal group applied to. Country parameters have been estimated from given in parentheses

[i]Management information system

listed in the table, Kennedy (1986) reviews a number of very early applications to laying hens, broilers, and sheep.

Verstegen et al. (1998) used an MDP as a tool for comparing different management information systems performance against the optimal decisions found by the MDP and van Asseldonk et al. (1999) used an MDP to optimize which IT solutions to implement on farm. The remaining papers focus on food and mouth disease (FMD) (Ge et al. 2010a,b) and how to compute an adaptive control strategy of an animal disease among a set of farms (Viet et al. 2012). Decisions considered in the models are “Keep”, “Replace”, if the farm should investment in a certain IT solution, vaccination strategy, and different FMD control options.

Due to the various applications state variables differ much. Examples are IT investment status, epidemic situation, infected and month, etc. Stage lengths vary from one day as in Ge et al. (2010a) to a year (van Asseldonk et al. 1999). Two papers use ordinary MDPs based on value iteration optimizing expected discounted reward and three papers use hierarchical models, with two implemented using the MLHMP software (Kristensen 2003).

The models by Ge et al. (2010a,b) use Bayesian updating to estimate the disease spread properties of the FMD virus causing the FMD outbreak, and Verstegen et al. (1998) use Bayesian updating to estimate the properties of hypothetical projects.

19.6 Software for Solving MDP Models

The value iteration algorithm for ordinary MDPs is relatively easy to implement and most papers have implemented the algorithm using various programming languages. The policy iteration is harder to implement since we have to invert a matrix when solving the set of linear equations. That is probably the reason that most studies reported in literature have used the more straightforward value iteration algorithm. In a few cases software packages in MATLAB[1] have been used to perform policy iteration (Heikkila et al. 2008, 2012). Linear programming can also be used to find optimal policies but have only been used in two papers (Cabrera 2010; Yates and Rehman 1998).

When considering hierarchical MDPs implementation becomes harder due to the nested structure of the processes. Fortunately a general software system MLHMP for construction, editing, and optimization of Markov decision processes ranging from finite time ordinary MDPs to hierarchical MDPs has been developed by Kristensen (2003). MLHMP is implemented in Java[2] with the possibility of building models as plug-ins. Moreover, it can handle all the criteria mentioned in this paper. MLHMP has been used to solve almost all hierarchical MDPs in the

[1] MathWorks Inc. http://www.matlab.com.

[2] Oracle http://www.java.com/.

literature. Recently, a package "Markov decision processes (MDPs) in R" (Nielsen 2011) has been developed for model building in R.[3] It is based on a C++ implementation for fast execution of policy and value iteration and can be used to solve both ordinary and hierarchical MDPs under all criteria.

19.7 Conclusions and Directions for Further Research

In this chapter MDPs have been considered to model livestock systems. Livestock farming problems are often sequential in nature and hence MDPs are suitable as a modeling tool.

A total of approximately 80 papers using a MDP for modeling the livestock system have been reviewed with the first paper dating back to 1966 and the last paper in 2012. Only decision models are included in the survey, i.e., simple Markov chain models are not mentioned even though they are, of course, closely related to MDPs. Most papers have been considered within dairy and some within pig production; however, MDPs have also been applied to other areas.

The papers may be divided into two categories, namely, papers using MDPs as a tool for evaluating different herd effects, e.g., different reproductive programs (Kalantari and Cabrera 2012) and papers formulating MDP models which may be embedded into a management decision support system (*DSS*), e.g., a model for slaughter pig marketing (Kristensen et al. 2012).

The first category is mainly used by researchers as an evaluation tool and giving advise to the industry. Several of the most advanced recent models are in this category. Thus, the models by Bar et al. (2008a,b) and Cha et al. (2011) use the models to estimate the costs of clinical mastitis in dairy cows and evaluate the treatment and prevention options, and Demeter et al. (2011) use their model to estimate the long-term consequences of different breeding strategies in dairy cows. It is expected that many models developed in the future will belong to this category.

The aim of models in the second category is that they ultimately should be used within the DSS on farm. However, the actual use of such models on farm has been limited. Reasons for this may be that MDPs require access to good data for estimating the many parameters needed in the model. Moreover the estimation process may be cumbersome and error-prone. As a result there have been a growing focus on using on-farm biosensors for retrieving data and algorithms for data filtration and parameter estimation based on Bayesian updating as in Nielsen et al. (2010) for a dairy cow replacement model. An example from pig production is the work by Bono et al. (2012) where important litter size parameters to be used in a sow replacement models are automatically and dynamically estimated from herd registrations and fed into the replacement model. Furthermore, the states of the individual sows are automatically identified so that the optimal decision can be

[3] R Development Core Team http://www.R-project.org/.

returned by the optimization model. Providing direct links from data is crucial if MDP models should be applied within farms since the parameter settings may be quite different among farms.

Another issue is violated herd constraints. MDP models often are applied at animal level and given replacement it is assumed that a new animal is available. As a result MDP models have to be coordinated with other information streams and other models used in the farm DSS. This calls for further research.

Due to the large number of state variables there is a trend in using hierarchical MDPs, since here state variables such as lactation number and lactation stage are implicitly given by the model structure. Hence, the same problem formulated as a hierarchical model will typically have fewer state variables than if it had been formulated as an ordinary MDP. Moreover, finding the optimal policy using policy iteration is often faster.

Finally, the number of state variables may be so large that models may face the curse of dimensionality. This calls for research in models which finds an approximate good policy using techniques such as approximate dynamic programming (Powell 2011).

Acknowledgements The authors are grateful to Dr. Anna-Maija Heikkilä, MTT Economic Research, for her valuable information about several of the Finnish models referenced.

This chapter has been compiled with support from The Danish Council for Strategic Research (The PigIT project, Grant number 11-116191).

References

Allore H, Schruben L, Erb H, Oltenacu P (1998) Design and validation of a dynamic discrete event stochastic simulation model of mastitis control in dairy herds. J Dairy Sci 81(3):703–717. DOI 10.3168/jds.S0022-0302(98) 75626-7

van Arendonk J (1985a) A model to estimate the performance, revenues and costs of dairy cows under different production and price situations. Agr Syst 16(3):157–189. DOI 10.1016/0308-521X(85)90010-1

van Arendonk J (1985b) Studies on the replacement policies in dairy cattle. ii. optimum policy and influence of changes in production and prices. Livest Prod Sci 13(2):101–121. DOI 10.1016/0301-6226(85)90014-4

van Arendonk J (1986) Studies on the replacement policies in dairy cattle. iv. influence of seasonal variation in performance and prices. Livest Prod Sci 14(1):15–28. DOI 10.1016/0301-6226(86)90093-X

van Arendonk J (1988) Management guides for insemination and replacement decisions. J Dairy Sci 71(4):1050–1057. DOI 10.3168/jds.S0022-0302(88) 79651-4

van Arendonk J, Dijkhuizen A (1985) Studies on the replacement policies in dairy cattle. iii. influence of variation in reproduction and production. Livest Prod Sci 13(4):333–349. DOI 10.1016/0301-6226(85)90025-9

van Asseldonk M, Huirne R, Dijkhuizen AA, Beulens A (1999) Dynamic programming to determine optimum investments in information technology on dairy farms. Agr Syst 62 (1):17–28. DOI 10.1016/S0308-521X(99) 00051-7

Bar D, Tauer L, Bennett G, Gonzalez R, Hertl J, Schukken Y, Schulte H, Welcome F, Groehn Y (2008a) The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J Dairy Sci 91(6):2205–2214. DOI 10.3168/jds.2007-0573

Bar D, Tauer L, Bennett G, Gonzalez R, Hertl J, Schulte H, Schukken Y, Welcome F, Grohn Y (2008b) Use of a dynamic programming model to estimate the value of clinical mastitis treatment and prevention options utilized by dairy producers. Agr Syst 99(1):6–12. DOI 10.1016/j.agsy.2008. 09.001

Bellman R (1957) Dynamic programming. Princeton University Press, Princeton

Ben-Ari Y, Gal S (1986) Optimal replacement policy for multicomponent systems—an application to a dairy-herd. Eur J Oper Res 23(2):213–221. DOI 10.1016/0377-2217(86)90240-7

Ben-Ari Y, Amir I, Sharar S (1983) Operational replacement decision model for dairy herds. J Dairy Sci 66:1747–1759. DOI 10.3168/jds. S0022-0302(83)82002-5

Boichard D (1990) Estimation of the economic value of conception rate in dairy cattle. Livest Prod Sci 24(3):187–204. DOI 10.1016/0301-6226(90) 90001-M

Bono C, Cornou C, Kristensen A (2012) Dynamic production monitoring in pig herds I: modeling and monitoring litter size at herd and sow level. Livest Sci 149(3):289–300. DOI 10.1016/j. livsci.2012.07.023

Cabrera V (2010) A large markovian linear program to optimize replacement policies and dairy herd net income for diets and nitrogen excretion. J Dairy Sci 93(1):394–406. DOI 10.3168/ jds.2009-2352

Cabrera V (2012) A simple formulation and solution to the replacement problem: a practical tool to assess the economic cow value, the value of a new pregnancy, and the cost of a pregnancy loss. J Dairy Sci 95(8):4683–4698. DOI 10.3168/jds.2011-5214

Cardoso V, Nogueira J, van Arendonk J (1999a) Optimum replacement and insemination policies for crossbred cattle (holstein friesian x zebu) in the south-east region of brazil. Livest Prod Sci 58(2):95–105. DOI 10.1016/ S0301-6226(98)00205-X

Cardoso V, Nogueira J, Van Arendonk J (1999b) Optimal replacement and insemination policies for holstein cattle in the southeastern region of brazil: the effect of selling animals for production. J Dairy Sci 82(7):1449–1458. DOI 10.3168/jds.S0022-0302(99)75372-5

Cha E, Hertl J, Bar D, Groehn Y (2010) The cost of different types of lameness in dairy cows calculated by dynamic programming. Prev Vet Med 97(1):1–8. DOI 10.1016/j. prevetmed.2010.07.011

Cha E, Bar D, Hertl J, Tauer L, Bennett G, Gonzalez R, Schukken Y, Welcome F, Groehn Y (2011) The cost and management of different types of clinical mastitis in dairy cows estimated by dynamic programming. J Dairy Sci 94(9):4476–4487. DOI 10.3168/jds.2010-4123

Dekkers J (1991) Estimation of economic values for dairy cattle breeding goals: bias due to suboptimal management policies. Livest Prod Sci 29(2–3):131–149. DOI 10.1016/0301-6226(91) 90062-U

Dekkers JCM, Ten Hag JH, Weersink A (1998) Economic aspects of persistency of lactation in dairy cattle. Livest Prod Sci 53(3):237–252. DOI 10.1016/S0301-6226(97)00124-3

Delorenzo M, Spreen T, Bryan G, Beede D, van Arendonk J (1992) Optimizing model: insemination, replacement, seasonal production, and cash flow. J Dairy Sci 75(3):885–896. DOI 10.3168/jds.S0022-0302(92) 77829-1

Demeter R, Kristensen A, Dijkstra J, Lansink AO, Meuwissen M, van Arendonk J (2011) A multilevel hierarchic markov process with bayesian updating for herd optimization and simulation in dairy cattle. J Dairy Sci 94(12):5938–5962. DOI 10.3168/jds.2011-4258

Fisher W, Schruben L (1953) Linear programming applied to feed-mixing under different price conditions. J Farm Econ 35(4):471–483. URL http://www.jstor.org/stable/1233362

Ge L, Kristensen A, Mourits M, Huirne R (2010a) A new decision support framework for managing foot-and-mouth disease epidemics. Ann Oper Res. DOI 10.1007/s10479-010-0774-2

Ge L, Mourits M, Kristensen A, Huirne R (2010b) A modelling approach to support dynamic decision-making in the control of fmd epidemics. Prev Vet Med 95(3–4):167–174. DOI 10.1016/j.prevetmed.2010.04.003

Giaever H (1966) Optimal dairy cow replacement policies. Ph.D. thesis, University of California, Berkeley, University Microfilms, Ann Arbor

Giordano J, Kalantari A, Fricke P, Wiltbank M, Cabrera V (2012) A daily herd markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrus detection. J Dairy Sci 95(9):5442–5460. DOI 10.3168/jds.2011-4972

Glenn J (1983) A dynamic programming model for pig production. J Oper Res Soc 34:511–519. DOI 10.1057/jors.1983.118

Grohn Y, Rajala-Schultz P, Allore H, DeLorenzo M, Hertl J, Galligan D (2003) Optimizing replacement of dairy cows: modeling the effects of diseases. Prev Vet Med 61(1):27–43. DOI 10.1016/S0167-5877(03)00158-2

Haran P (1997) Markov decision processes in the optimisation of culling decisions for irish dairy herds. Master's thesis, School of Computer Applications, Dublin City University

Harris B (1990) Recursive stochastic programming applied to dairy cow replacement. Agr Syst 34 (1):53–64. DOI 10.1016/0308-521X(90)90093-6

Heikkila A, Nousiainen J, Jauhiainen L (2008) Optimal replacement policy and economic value of dairy cows with diverse health status and production capacity. J Dairy Sci 91(6):2342–2352. DOI 10.3168/jds.2007-0736

Heikkila A, Nousiainen J, Pyorala S (2012) Costs of clinical mastitis with special reference to premature culling. J Dairy Sci 95(1):139–150. DOI 10. 3168/jds.2011-4321

Houben E, Huirne R, Dijkhuizen A, Kristensen A (1994) Optimal replacement of mastitis cows determined by a hierarchic Markov process. J Dairy Sci 77:2975–2993. DOI 10.3168/jds.S0022-0302(94)77239-8

Huirne R, Hardaker J (1998) A multi-attribute utility model to optimise sow replacement decisions. Eur Rev Agric Econ 25(4):488–505. DOI 10.1093/ erae/25.4.488

Huirne R, Hendriks T, Dijkhuizen A, Giesen G (1988) The economic optimisation of sow replacement decisions by stochastic dynamic programming. J Agric Econ 39:426–438. DOI 10.1111/j.1477-9552.1988. tb00602.x

Huirne R, Dijkhuizen A, Renkema JA (1991) Economic optimization of sow replacement decisions on the personal computer by method of stochastic dynamic programming. Livest Prod Sci 28:331–347. DOI 10. 1016/0301-6226(91)90014-H

Huirne R, van Beek P, Hendriks T, Dijkhuizen A (1993) An application of stochastic dynamic programming to support sow replacement decisions. Eur J Oper Res 67:161–171. DOI 10.1016/0377-2217(93)90059-V

Jalvingh A, Dijkhuizen A, van Arendonk J, Brascamp E (1992a) Dynamic probabilistic modelling of reproductive and replacement in sow herds. general aspects and model description. Agr Syst 39:133–152. DOI 10.1016/ 0308-521X(92)90105-W

Jalvingh A, Dijkhuizen A, van Arendonk J, Brascamp E (1992b) An economic comparison of management strategies on reproduction and replacement in sow herds using a dynamic probabilistic model. Livest Prod Sci. 32:331–350. DOI 10.1016/0301-6226(92)90004-N

Jalvingh A, van Arendonk J, Dijkhuizen A (1993a) Dynamic probabilistic simulation of dairy-herd management-practices.i. model description and outcome of different seasonal calving patterns. Livest Prod Sci 37(1–2):107–131. DOI 10.1016/0301-6226(93)90067-R

Jalvingh A, van Arendonk J, Dijkhuizen A, Renkema J (1993b) Dynamic probabilistic simulation of dairy-herd management-practices. ii. comparison of strategies in order to change a herds calving pattern. Livest Prod Sci 37(1–2):133–152. DOI 10.1016/0301-6226(93)90068-S

Jalvingh A, Dijkhuizen A, van Arendonk J (1994) Optimizing the herd calving pattern with linear-programming and dynamic probabilistic simulation. J Dairy Sci 77(6):1719–1730. DOI 10.3168/jds.S0022-0302(94) 77113-7

Jørgensen E (1992) Sow replacement: reduction of state space in dynamic programming model and evaluation of benefit from using the model. Dina Research Report 6, National Institute of Animal Science

Jørgensen E (1993) The influence of weighing precision on delivery decisions in slaughter pig production. Acta Agric Scand 43(3):181–189. DOI 10. 1080/09064709309410163

Kalantari A, Cabrera V (2012) The effect of reproductive performance on the dairy cattle herd value assessed by integrating a daily dynamic programming model with a daily markov chain model. J Dairy Sci 95(10):6160–6170. DOI 10.3168/jds.2012-5587

Kalantari A, Mehrabani-Yeganeh H, Moradi M, Sanders A, de Vries A (2010) Determining the optimum replacement policy for holstein dairy herds in iran. J Dairy Sci 93(5):2262–2270. DOI 10.3168/jds.2009-2765

Kennedy J (1986) Dynamic programming. Applications to agriculture and natural resources. Elsevier Applied Science Publishers, London/New York

Kennedy J, Stott A (1993) An adaptive decision-making aid for dairy cow replacement. Agr Syst 42(1–2):25–39. DOI 10.1016/0308-521X(93) 90066-B

Killen L, Kearney B (1978) Optimal dairy cow replacement policy. Ir J Agric Econ Rural Sociol 7:33–40. URL http://www.jstor.org/stable/25556435

Kristensen A (1987) Optimal replacement and ranking of dairy cows determined by a hierarchical markov process. Livest Prod Sci 16(2):131–144. DOI 10.1016/0301-6226(87)90015-7

Kristensen A (1988) Hierarchic Markov processes and their applications in replacement models. Eur J Oper Res 35(2):207–215. DOI 10.1016/ 0377-2217(88)90031-8

Kristensen A (1989) Optimal replacement and ranking of dairy-cows under milk quotas. Acta Agric Scand 39(3):311–318. DOI 10.1080/ 00015128909438523

Kristensen A (1992) Optimal replacement in the dairy-herd: a multicomponent system. Agr Syst 39(1):1–24. DOI 10.1016/0308-521X(92)90002-6

Kristensen A (1993) Bayesian updating in hierarchical Markov-processes applied to the animal replacement-problem. Eur Rev Agric Econ 20(2):223–239

Kristensen A (1994) A survey of Markov decision programming techniques applied to the animal replacement-problem. Eur Rev Agric Econ 21(1):73–93. DOI 10.1093/erae/21.1.73

Kristensen A (2003) A general software system for Markov decision processes in herd management applications. Comput Electron Agric 38(3):199–215. DOI 10.1016/S0168-1699(02) 00183-7

Kristensen A, Jørgensen E (2000) Multi-level hierarchic Markov processes as a framework for herd management support. Ann Oper Res 94:69–89. DOI 10.1023/A:1018921201113

Kristensen A, Søllested T (2004a) A sow replacement model using Bayesian updating in a three-level hierarchic Markov process I. Biological model. Livest Prod Sci 87(1):13–24. DOI 10.1016/j.livprodsci.2003.07.004

Kristensen A, Søllested T (2004b) A sow replacement model using Bayesian updating in a three-level hierarchic Markov process, II. optimization model. Livest Prod Sci 87(1):25–36. DOI 10.1016/j.livprodsci.2003.07.005

Kristensen A, Thysen I (1991a) Economic value of culling information in the presence and absence of a milk quota. Acta Agric Scand 41(2):129–135. DOI 10.1080/00015129109438594

Kristensen A, Thysen I (1991b) Ranking of dairy-cows for replacement—alternative methods tested by stochastic simulation. Acta Agric Scand 41(3):295–303. DOI 10.1080/ 00015129109439912

Kristensen A, Nielsen L, Nielsen M (2012) Optimal slaughter pig marketing with emphasis on information from on-line live weight assessment. Livest Sci 145(1–3):95–108. DOI 10.1016/j. livsci.2012.01.003

Kure H (1997a) Marketing management support in slaughter pig production. Ph.D. thesis, The Royal Veterinary and Agricultural University. URL http://www.prodstyr.ihh.kvl.dk/pub/phd/ kure_thesis.pdf

Kure H (1997b) Optimal slaughter pig marketing. In: Proceedings of the dutch/danish symposium on animal health and management economics, Copenhagen, January 23–24 1997. Dina Notat No. 56, 39-47. URL http://www.prodstyr.ihh.kvl.dk/pub/symp/hk/Kure97a.pdf

Kure H (1997c) Slaughter pig marketing management: utilization of highly biased herd-specific data. In: Kure et al. (eds.) Proceedings of the first european conference for information technology in agriculture, Copenhagen, June 15–18

Langford F, Stott A (2012) Culled early or culled late: economic decisions and risks to welfare in dairy cows. Anim Welf 21(1):41–55. DOI 10.7120/ 096272812X13345905673647

Lien G, Kristensen A, Hegrenes A, Hardaker J (2003) Optimal length of leys in an area with winter damage problems. Grass Forage Sci 58(2):168–177. DOI 10.1046/j.1365-2494.2003.00367.x

McArthur A (1973) Application of dynamic programming to the culling decision in dairy cattle. Proc N Z Soc Anim Prod 33:141–147

Mccullough D, Delorenzo M (1996a) Effects of price and management level on optimal replacement and insemination decisions. J Dairy Sci 79(2):242–253. DOI 10.3168/jds.S0022-0302(96)76357-9

Mccullough D, Delorenzo M (1996b) Evaluation of a stochastic dynamic replacement and insemination model for dairy cattle. J Dairy Sci 79(1):50–61. DOI 10.3168/jds.S0022-0302(96)76333-6

Mourits M, Huirne R, Dijkhuizen A, Galligan D (1999a) Optimal heifer management decisions and the influence of price and production variables. Livest Prod Sci 60(1):45–58. DOI 10.1016/S0301-6226(99)00037-8

Mourits M, Huirne R, Dijkhuizen A, Kristensen A, Galligan D (1999b) Economic optimization of dairy heifer management decisions. Agr Syst 61(1):17–31. DOI 10.1016/S0308-521X(99)00029-3

Nielsen B, Kristensen A (2007) Optimal decisions in organic beef production from steers effects of criterion of optimality and price changes. Livest Sci 110(1–2):25–32. DOI 10.1016/j.livsci.2006.09.024

Nielsen B, Kristensen A, Thamsborg S (2004) Optimal decisions in organic steer production—a model including winter feed level, grazing strategy and slaughtering policy. Livest Prod Sci 88 (3):239–250. DOI 10.1016/j. livprodsci.2003.11.010

Nielsen L (2011) Markov decision processes (MDPs) in R. URL https://r-forge.r-project.org/projects/mdp/

Nielsen L, Jørgensen E, Kristensen A, Østergaard S (2010) Optimal replacement policies for dairy cows based on daily yield measurements. J Dairy Sci 93(1):77–92. DOI 10.3168/jds.2009-2209

Nielsen L, Jørgensen E, Højsgaard S (2011) Embedding a state space model into a markov decision process. Ann Oper Res 190(1):289–309. DOI 10.1007/s10479-010-0688-z

Niemi J (2006) A dynamic programming model for optimising feeding and slaughter decisions regarding fattening pigs. Ph.D. thesis, MTT Agrifood Research Finland

Noordegraaf A, Buijtels J, Dijkhuizen A, Franken P, Stegeman JA, Verhoeff J (1998) An epidemiological and economic simulation model to evaluate the spread and control of infectious bovine rhinotracheitis in the netherlands. Prev Vet Med 36(3):219–238. DOI 10.1016/S0167-5877(98)00081-6

Pihamaa P, Pietola K (2002) Optimal beef cattle management under agricultural policy reforms in finland. Agric Food Sci Finl 11(1):3–12. URL http://jukuri.mtt.fi/handle/10024/452290

Pla L, Pomar C, Pomar J (2003) A Markov decision sow model representing the productive lifespan of herd sows. Agr Syst 76(1):253–272. DOI 10. 1016/S0308-521X(02)00102-6

Pla L, Pomar C, Pomar J (2004) A sow herd decision support system based on an embedded markov model. Comput Electron Agric 45(1):51–69. DOI 10.1016/j.compag.2004.06.005

Powell W (2011) Approximate dynamic programming solving the curses of dimensionality. Wiley, Hoboken

Puterman M (1994) Markov decision processes. Wiley series in probability and mathematical statistics. Wiley, New York

Rajala-Schultz P, Grohn Y (2001) Comparison of economically optimized culling recommendations and actual culling decisions of finnish ayrshire cows. Prev Vet Med 49(1–2):29–39. DOI 10.1016/S0167-5877(01)00180-5

Rajala-Schultz P, Grohn Y, Allore H (2000a) Optimizing breeding decisions for finnish dairy herds. Acta Vet Scand 41(2):199–212

Rajala-Schultz P, Grohn Y, Allore H (2000b) Optimizing replacement decisions for finnish dairy herds. Acta Vet Scand 41(2):185–198

Rodriguez S, Jensen T, Pla L, Kristensen A (2011) Optimal replacement policies and economic value of clinical observations in sow herds. Livest Sci 138(1–3):207–219. DOI 10.1016/j.livsci.2010.12.026

Rogers G, van Arendonk J, McDaniel B (1988a) Influence of involuntary culling on optimum culling rates and annualized net revenue. J Dairy Sci 71(12):3463–3469. DOI 10.3168/jds.S0022-0302(88)79952-X

Rogers G, van Arendonk J, McDaniel B (1988b) Influence of production and prices on optimum culling rates and annualized net revenue. J Dairy Sci 71(12):3453–3462. DOI 10.3168/jds.S0022-0302(88)79951-8

Smith B (1971) The dairy cow replacement problem. An application of dynamic programming, Bulletin 745. Florida Agricultural Experiment Station, Gainesville

Smith B (1973) Dynamic programming of the dairy cow replacement problem. Am J Agric Econ 55(1):100–104. DOI 10.2307/1238671

Sørensen J, Kristensen E, Thysen I (1992) A stochastic model simulating the dairy herd on a pc. Agric Syst 39(2):177–200. DOI 10.1016/0308-521X(92) 90107-Y

Stewart H, Burnside E, Wilton J, Pfeiffer W (1977) Dynamic-programming approach to culling decisions in commercial dairy herds. J Dairy Sci 60(4):602–617. DOI 10.3168/jds.S0022-0302(77)83908-8

Stewart H, Burnside E, Pfeiffer W (1978) Optimal culling strategies for dairy cows of different breeds. J Dairy Sci 61:1605–1615. DOI 10.3168/jds. S0022-0302(78)83772-2

Stott A (1994) The economic advantage of longevity in the dairy cow. J Agric Econ 45:113–122. DOI 10.1111/j.1477-9552.1994.tb00382.x

Stott A, Kennedy J (1993) The economics of culling dairy-cows with clinical mastitis. Vet Rec 133 (20):494–499. DOI 10.1136/vr.133.20.494

Stott A, Jones G, Gunn G, Chase-Topping M, Humphry RW, Richardson H, Logue DN (2002) Optimum replacement policies for the control of subclinical mastitis due to s.aureus in dairy cows. J Agric Econ 53(3):627–644. DOI 10.1111/j.1477-9552.2002.tb00041.x

Stott A, Jones G, Humphry R, Gunn G (2005) Financial incentive to control paratuberculosis (Johne's disease) on dairy farms in the United Kingdom. Vet Rec 156(26):825–831. DOI 10.1136/vr.156.26.825

Tijms H (2003) A first course in stochastic models. Wiley, West Sussex

Toft N, Kristensen A, Jorgensen E (2005) A framework for decision support related to infectious diseases in slaughter pig fattening units. Agr Syst 85(2):120–37. DOI 10.1016/j.agsy.2004.07.017

Vargas B, Herrero M, van Arendonk J (2001) Interactions between optimal replacement policies and feeding strategies in dairy herds. Livest Prod Sci 69(1):17–31. DOI 10.1016/S0301-6226(00)00250-5

Verstegen J, Sonnemans J, Huirne R, Dijkhuizen A, Cox J (1998) Quantifying the effects of sow-herd management information systems on farmers' decision making using experimental economics. Am J Agric Econ 80(4):821–829. DOI 10.2307/1244066

Viet A, Jeanpierre L, Bouzid M, Mouaddib A (2012) Using markov decision processes to define an adaptive strategy to control the spread of an animal disease. Comput Electron Agric 80:71–79. DOI 10.1016/j.compag.2011.10. 015

de Vries A (2004) Economics of delayed replacement when cow performance is seasonal. J Dairy Sci 87(9):2947–2958. DOI 10.3168/jds.S0022-0302(04) 73426-8 de Vries A (2006) Economic value of pregnancy in dairy cattle. J Dairy Sci 89(10):3876–3885. DOI 10.3168/jds.S0022-0302(06)72430-4

White W (1959) The determination of an optimal replacement policy for a continually operating egg production enterprise. J Farm Econ 41(5):1535–1542. URL http://www.jstor.org/stable/1235311

Yalcin C, Stott A (2000) Dynamic programming to investigate financial impacts of mastitis control decisions in milk production systems. J Dairy Res 67(4):515–528. DOI 10.1017/S0022029900004453

Yates C, Rehman T (1998) A linear programming formulation of the Markovian decision process approach to modelling the dairy replacement problem. Agr Syst 58(2):185–201. DOI 10.1016/S0308-521X(98)00054-7

ERRATUM

A Hierarchical Planning Scheme Based on Precision Agriculture

Víctor M. Albornoz, Néstor M. Cid-García, Rodrigo Ortega, and Yasmín A. Ríos-Solís

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_6,

DOI 10.1007/978-1-4939-2483-7_6

In chapter 6 titled "A Hierarchical Planning Scheme Based on Precision Agriculture," the departments in the affiliations for two of the authors are updated as follows:

Victor M. Albornoz: Departamento de Industrias

Rodrigo Ortega: Departamento de Ingeniería Comercial

V.M. Albornoz (✉)
Departamento de Industrias, Universidad Técnica Federico Santa María, Vitacura, Santiago, Chile
e-mail: victor.albornoz@usm.cl

R. Ortega
Departamento de Ingeniería Comercial, Universidad Técnica Federico Santa María, Vitacura, Santiago, Chile
e-mail: rodrigo.ortega@usm.cl

The online version of the original chapter can be found at
http://dx.doi.org/10.1007/978-1-4939-2483-7_6

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_20,

N.M. Cid-García • Y.A. Ríos-Solís
Graduate Program in Systems Engineering,
Universidad Autónoma de Nuevo León (UANL),
San Nicolás de Los Garza, Mexico
e-mail: nestor.cidgrc@uanl.edu.mx; yasmin.riossls@uanl.edu.mx

Index

L.M. Plà-Aragonés (ed.), *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7

Q

R