UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author LEVIS, A.

COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOAN

Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The University Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

- [x] This copy has been deposited in the Library of UCL
- [ ] This copy has been deposited in the University of London Library, Senate House, Malet Street, London WC1E 7HU.

Multi-Site Capacity Planning and Business Optimisation for Process Industries

by

Aaron Levis

A thesis submitted for the degree of

Doctor of Philosophy of University of London

Department of Chemical Engineering

University College London

London WC1E 7JE

September 2004

ABSTRACT

Changing market conditions, volatile customer demand, intense competition and tightness of capital are some of the primary characteristics of the global economy that affect process industries nowadays. The main objective of the thesis is to facilitate business decision-making in today's increasingly complex and highly uncertain market environment by applying mathematical programming techniques for multi-site capacity planning and business optimisation in process industries.

In the first part of the thesis, the problem of multi-site capacity planning under uncertainty in the pharmaceutical industry is addressed. A comprehensive two-stage, multi-scenario mixed-integer linear programming (MILP) model is proposed able to determine an optimal product portfolio and multi-site investment plan in the face of clinical trials uncertainty. A hierarchical algorithm is also developed in order to reduce the computational effort needed for the solution of the resulting large-scale MILP model. The applicability of the proposed solution methodology is demonstrated by a number of illustrative examples.

The second part addresses the problem of business optimisation for customer demand management in process industries. A customer demand forecasting approach is developed based on support vector regression analysis. The proposed three-step algorithm is able to extract the underlying customer demand patterns from historical sales data and derive an accurate forecast as demonstrated through a number of illustrative examples. An active demand management approach for close substitute products is also developed based on price optimisation. The proposed methodology is able to determine optimal pricing policies as well as output levels, while taking into consideration manufacturing costs, resource availability, customer demand elasticity, outsourcing and market competition. An iterative algorithm is developed able to determine Nash equilibrium in prices for competing companies as demonstrated by the illustrative examples.

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr Lazaros Papageorgiou for introducing me to the fascinating world of optimisation theory and mathematical programming. His excellent supervision, contribution and guidance throughout my PhD studies are more than appreciated. Financial support from the European Union VIP-NET project is also gratefully acknowledged, while special thanks belong to Dr Michael Georgiadis. Many thanks to all the past and present members of the Centre for Process Systems Engineering, UCL Department of Chemical Engineering and particularly to my friends Dimitris and Vangelis.

My friends from Stepney Green, Kilburn and King's Cross Hillel Houses, especially Sorana, Mendel, Anthony and Regis, deserve all my gratitude for being a constant source of encouragement, sharing most moments of my non-academic life in London and nourishing my hopes and ambitions with their great enthusiasm and passion for life.

Finally, I would like to thank my parents Makis and Suzi for providing me with their unconditional love and constant support all these years as well as my sister Nina for always being there for me.

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Process industries in the 21st century

Process industries cover all industries involving a chemical or physical change in the manufacturing process such as chemicals, pharmaceuticals, food, steel and mineral processing. They are recognised as being the most successful manufacturing industries in the UK where total sales in 1997 were £33 billion, with profits £4.8 billion and they are fast becoming the largest single export group (Munir *et al.*, 1999).

Volatile customer demand, intense competition, tightness of capital and global operations are the primary characteristics of the 21st century economy that affect every stage of business decision-making in process industries. In particular, process industries nowadays operate in a market environment that is highly uncertain and increasingly complex. Modern economy favours the production of high-value added and low-volume products. The pressure of global competition is frequently cited as a primary driver for greater customer demands and improved quality products. Furthermore, globalisation of business operations introduces the need for manufacturing a wide variety of products, sold in multiple geographically distributed locations around the world.

Given the competitive business landscape both at local and global level and the ever-changing market conditions across different time and space scales, process industries are faced with the question of how to ensure high profitability as well as guarantee business sustainability. Overall, the main concern for process industries is how to make the most out of their resources so as to lead a sustainable business life in the 21st century. The decisions involved have to be considered simultaneously and they are usually taken in the face of significant uncertainty and risk that further magnify the importance of optimal business design and operation.

During the last decades, the industrial sector has experienced a major shift from product-oriented companies to customer-centric supply chains and from mass production modes to tailor-made products where product specifications are literally dictated by each individual customer. The Internet explosion and the subsequent rise of e-commerce enabled process industries to adopt a more responsive way of manufacturing and delivering products to end-customers. Towards the same direction, information technology provided process industries with novel applications such as vendor-managed inventories, radio frequency identification (RFID) product tags and electronic shelf labels resulting in real-time responsiveness, exceptional visibility across the entire supply chain and increased customer satisfaction.

However, it is still very often the case where process engineers find themselves data rich but information poor. Despite recent technological advances, the abundance of data in process industries is not always translated into extra money or time. As it is typical with any emerging technology, although new applications provide a wide range of solutions, they also create a new array of technology-specific problems that necessitate the full attention of process industries wishing to stay ahead of market competition and outrun their slower competitors.

As firms increase their participation in global economy, developing a deep understanding of business optimisation issues and realising the potential benefits and opportunities becomes an increasingly important matter for process industries. There exists a clearly identified need for advanced decision-making tools able to facilitate business management and potentially lead to significant economic benefits for process industries.

1.2 Mathematical programming for process industries

It has long been recognised that mathematical programming techniques provide such a fertile environment that can sustain a wide range of business applications in process industries. More than 40 years ago, Forrester (1958) identified a major management breakthrough in understanding how industrial company success depends on the interaction between the flows of information, materials, money, manpower and capital equipment. Discussing the shape of the future, he proposed that "there will come general recognition of the advantage enjoyed by the pioneering managements who have been the first to improve their understanding of the interrelationships between separate business functions and the company and its markets, its industry and its national economy".

Since then, numerous practitioners in the academia and industry have made vital contributions in the emerging area of process systems engineering with the most celebrated one being the pioneering work of Prof. Roger WH Sargent (Sargent, 1977). However, prohibitively expensive and unavailable computational resources hindered the wide acceptance of mathematical programming techniques in process industries. It was only until recently when computer power became fairly cheap and widely available that process industries began to realise the true potential of mathematical programming by harnessing the number-crunching capabilities of super-computers coupled with major advances in the field of process systems engineering.

Recently, Kallrath and Wilson (1997) summarised a number of interesting business optimisation problems and proposed mathematical programming techniques for their solution. The most important among those problems include transportation and assignment problems, production planning and scheduling, distribution planning, yield management, project planning and facility location problems. An indicative example of the scope of mathematical programming in assisting process industries is described by Arntzen *et al.* (1995) who reported an optimisation model used by Digital Equipment Corporation that achieved savings in the order of $100 million US dollars.

According to Williams (1999), apart from the apparent economic benefits, the motivation for model building in mathematical programming is three-fold:

- To gain insight into the problem. The actual exercise of building a mathematical model often reveals relationships that were not apparent previously. As a result greater understanding of the problem is achieved.

- To identify non-obvious solutions to the problem. Having built a model it is then possible to analyse it mathematically and help suggest course of actions that might not otherwise be obvious.

- To investigate extreme aspects of the problem. Computational experiments can be conducted when it is not possible or desirable to conduct an experiment in real-life (e.g. accident simulation models) and provide us with useful information concerning the problem under investigation.

1.3 Aims and objectives

Motivated by the promise of better understanding and enhanced problem-solving capabilities offered by mathematical programming, the aim of our work is *to facilitate business decision-making by applying mathematical programming techniques for multi-site capacity planning and business optimisation in process industries*. More specifically our goal is to develop a number of mathematical modelling frameworks able to accommodate different business functions and serve as an advanced toolbox for strategic business decision-making that complement human expertise in process industries.

In order to achieve our goal, the following areas will be addressed:

- *Multi-site capacity planning*: this area is concerned with determining a multi-site investment strategy in the pharmaceutical industry as a key representative of the R&D-intensive, risk-prone and highly-regulated process industries including among others the agrochemical, biotechnology, food and drinks industries.

- *Efficient solution methodologies*: this area is focused on the development of alternative solution approaches able to accommodate the combinatorial nature of the proposed multi-site capacity planning model and derive near-optimal solutions within reasonable computational time without compromising the quality of the obtained solution.

- *Customer demand forecasting*: this area deals with a systematic approach for customer demand forecasting via the employment of support vector regression analysis.
- *Active demand management*: this area is concerned with a novel, non-conventional approach for customer demand management in competitive marketplaces through price optimisation.

1.4 Thesis outline

The rest of the thesis is structured in two parts as follows. Part I addresses the problem of multi-site capacity planning under uncertainty in the pharmaceutical industry and consists of Chapters 2 and 3. Part II tackles the problem of business optimisation for customer demand management in process industries and consists of Chapters 4 and 5.

Chapter 2 presents a simultaneous approach for multi-site capacity planning under uncertainty. The overall problem is formulated as a two-stage, multi-scenario mixed-integer linear programming (MILP) mathematical model incorporating issues related to product management, clinical trials uncertainty, capacity management and trading structure of the company. The applicability of the proposed mathematical model is illustrated by a number of illustrative examples.

A hierarchical approach for multi-site capacity planning under uncertainty is proposed in Chapter 3. Based on an aggregate formulation of the original MILP model, a hierarchical solution methodology is developed able to accommodate the inherent complexity of the multi-site capacity problem by decoupling the strategic and operational decision-making levels. The illustrative examples are revisited so as to perform a valid comparison between the hierarchical and the simultaneous approach.

Chapter 4 presents a customer demand forecasting approach via support vector regression analysis. Based on the recently developed statistical learning theory, a three-step algorithm is proposed comprising both nonlinear (NLP) and linear (LP) mathematical model formulations to determine the regression function while the final step employs a recursive methodology to perform customer demand forecasting. Based on historical sales data, the proposed algorithm identifies underlying customer demand patterns and captures customer behaviour before deriving an accurate forecast as demonstrated by three illustrative examples.

An active demand management approach for close substitute products through price optimisation is presented in Chapter 5. The proposed methodology identifies price as the ultimate driver behind customer purchasing decisions and maximises company profits while taking into consideration manufacturing costs, resource availability, customer demand elasticity, outsourcing and market competition. An iterative algorithm is proposed able to simulate the decision-making process by solving a series of non-linear programming (NLP) mathematical models before determining Nash equilibrium in prices for competing multi-product companies. The applicability of the proposed methodology is demonstrated by a number of illustrative examples.

Finally, Chapter 6 summarises the main contributions of the thesis and also provides recommendations for further research work.

PART I

MULTI-SITE CAPACITY PLANNING UNDER UNCERTAINTY

Chapter 2

A simultaneous approach for multi-site capacity planning

2.1 Introduction and literature survey

Every year the pharmaceutical industry spends a large amount of research funds developing new chemical entities (NCEs). This process involves producing and screening vast libraries of chemical compounds before a limited number of promising NCEs enters the clinical trials phase. Despite the costly research and development (R&D) effort, only a few of the initial chemical compounds actually become marketed drugs, depending on the clinical trials outcomes and the estimate is that a successful drug might take ten years from the laboratory to the pharmacy shelf (EFPIA, 2003).

Pharmaceutical companies are constantly faced with the question of best use of limited resources available to obtain the highest possible profit and the decisions involved are usually taken in the presence of significant uncertainty. It would be fatal for a pharmaceutical company to wait for well-informed investment decisions in the end of the clinical trials phase. Late investment decisions could lead to a prohibitively long time before the new product reaches the market and therefore would jeopardise both the profitability and the sustainability of the entire company.

In order for a pharmaceutical company to stay competitive and lead a sustainable business life, it is absolutely vital to recoup its investment in developing new products as soon as possible. Since the effective patent life is diminishing and generic drugs will inevitably enter the marketplace, posing a valid threat to the profitability of the new products, it is of crucial importance that the company decides early enough on its capacity investment plan so as to guarantee the availability of the necessary manufacturing capacity resources to produce the new products as soon as the marketing authorisation is granted. This enables the company to take a proactive course of action in the face of uncertain clinical trials outcomes and market launch its new products on time, making the most out of the patent lifetime and outrun its slower competitors while ensuring a high product profitability and a steady financial flow back to the R&D product pipeline.

Schmidt and Grossmann (1996) were the first to address the problem of task scheduling for new product development. Later on, Jain and Grossmann (1999) proposed two MILP models to account for the resource-constraint scheduling of testing tasks. Rotstein *et al.* (1999) presented a stochastic capacity planning model incorporating clinical trials uncertainty, using a scenario-based approach. Although they managed to integrate the problem of capacity planning and new product development under uncertainty in the pharmaceutical industry, their proposed model was limited to the case of single-site capacity planning. Samuelsson (1999) developed both a deterministic and a stochastic single-site capacity planning model while he proposed various heuristic solution approaches. Blau *et al.* (2000) considered the decision-making process involved in the development of the product portfolio under uncertainty of a pharmaceutical company while they also calculated the ratio of reward to risk. In order to address the combinatorial nature of the R&D product pipeline problem, Subramanian *et al.* (2000) developed a computing architecture based on mathematical programming and discrete-event system simulation so as to facilitate decision-making for new product development. Gatica *et al.* (2001) described a stochastic single-site capacity planning model considering multiple clinical trials outcomes per candidate product. Maravelias and Grossmann (2001) proposed a multi-period model able to accommodate simultaneously new product development and capacity planning of manufacturing facilities. Papageorgiou *et al.* (2001) developed a multi-site, multi-period capacity planning model incorporating the internal trading structure of the company. However, their model assumes a deterministic demand profile for the potential products with no consideration of the uncertainty of clinical trials outcomes. Rogers *et al.* (2002) proposed a novel approach for optimal pharmaceutical R&D portfolios based on real-options analysis. They developed a stochastic mathematical model using a real-option decision tree for making optimal project selection decisions in the face of market uncertainty. More recently, Subramanian *et al.* (2003) extended their previous work for the stochastic optimisation of the R&D pipeline by employing a three-step heuristic procedure. Finally, Cheng *et al.* (2003) formulated the problem of design and planning under uncertainty as a multi-objective Markov decision process while they recognised the intense computational requirements arising from the "curse of dimensionality" when dealing with multiple scenario realisations.

Different solution approaches have been proposed in the literature in order to tackle efficiently large-scale mathematical models involving uncertainty. Pistikopoulos (1995) introduced a theoretically unified way for characterization and quantification of uncertainty in process design and operations. Mulvey *et al.* (1995) developed a scenario decomposition method for large-scale stochastic optimisation problems. Subrahmanyam *et al.* (1996) developed decomposition approaches for batch plant design and planning. Recent approaches include two-stage stochastic mathematical models accompanied by their respective solution strategy: Ierapetritou and Pistikopoulos (1994) (decomposition-based solution approach), Liu and Sahinidis (1996) (decomposition algorithm) and Gupta and Maranas (2000) (two-step algorithm). Gupta and Maranas (1999) also developed a hierarchical lagrangean relaxation for solving mid-term planning problems. Rotstein *et al.* (1999) used a hierarchical solution methodology based on a reduced scenario space of the original problem. In their work on capacity planning, Maravelias and Grossmann (2001) proposed a heuristic algorithm based on lagrangean decomposition that exploits the special structure of the problem and is able to yield near-optimal solutions within reasonable computational time. Finally, a scenario aggregation-disaggregation solution approach was introduced by Gatica *et al.* (2003). Wolsey (1998) discusses several heuristic algorithms for the solution of integer programming problems (dive-and-fix, relax-and-fix, cut-and-fix), while Reeves (1995) presents a number of modern heuristic techniques for large-scale combinatorial problems (simulated annealing, tabu search, genetic algorithms, lagrangean relaxation and decomposition).

Clearly, the problem of product portfolio and capacity planning has previously been addressed in the process systems engineering literature. However, the existing approaches mainly focus their attention on individual aspects of the general problem assuming for example the specific case of single-site capacity planning under uncertainty (Gatica *et al.*, 2001) or dealing with multi-site capacity planning that postulates a deterministic customer demand profile without considering the issue of uncertainty explicitly (Papageorgiou *et al.*, 2001). Another common drawback of nearly all the existing approaches is that they ignore the issue of the internal trading structure of the company thus failing to address the financial flows that take place among the different business centres of the company. However, the trading structure plays a dominant role in the after-tax profitability of the company. Especially in the pharmaceutical industry, it is very often the case when multi-national companies operate in many geographically distributed manufacturing facilities while dealing with an international clientele located in different customer zones. Therefore, the issues related to the trading structure of the company have to be taken into account when deciding on the optimal multi-site investment strategy of the company.

Furthermore, the inherent complexity of the problem of product portfolio and capacity planning under uncertainty usually gives rise to large-scale mathematical models that are computationally expensive to solve via direct application of commercially available solvers. The resulting mathematical models cannot be solved efficiently and in many cases are intractable through the employment of traditional branch-and-bound methods. The solution of such large-scale instances of the problem necessitates the implementation of alternative solution approaches.

In this chapter, we present a systematic simultaneous approach multi-site capacity planning in the face of clinical trials uncertainty while considering the trading structure of the company. The rest of the chapter is structured as follows. In the next section, the main characteristics of the problem are discussed, while a formal problem statement is given as well. The proposed mathematical model along with the key assumptions are described in section 2.3, while its applicability is demonstrated by five illustrative examples in section 2.4. Finally, some concluding remarks are drawn in section 2.5.

2.2 Problem description

The problem of multi-site capacity planning in the pharmaceutical industry is very challenging since it aims to integrate traditionally isolated areas such as product development, manufacturing, accounting and marketing. The complexity of the problem is mainly attributed to the great variety of parameters and decision-making levels involved. A strategic investment plan should simultaneously address and evaluate in a proper manner the following four main issues: product management, clinical trials uncertainty, capacity management and trading structure. The different main areas will be further analysed in full detail in the following sections.

2.2.1 Product management

Pharmaceutical products are typically manufactured in two main stages, namely the primary and secondary manufacturing stage. The primary stage is responsible for the production of the active ingredient (AI) of the drug. The second stage is responsible for converting the AI to a final product for direct use (e.g. vials, tablets, etc). The primary manufacturing step is the highest value-added step of the overall process and is considered to be the most critical one for product portfolio planning. Therefore, this is the main stage that will be further analysed in this chapter.

Product management is concerned with the main features of each product considered as a suitable candidate for manufacturing and commercialisation. Such features include R&D costs associated with the development of each new product and commercial characteristics of each product such as manufacturing costs, selling price and marketing expenses (Papageorgiou *et al.*, 2001).

Overall, product management addresses the problem of selecting the optimal product portfolio based on the individual characteristics of each candidate product. It is very important to understand that the decisions involved in product management cannot be viewed in isolation from the overall problem of multi-site capacity planning, since they are strongly dependant on the uncertain clinical trials outcomes, while they are also directly linked to both the capacity investment decisions and the trading structure of the company.

2.2.2 Clinical trials uncertainty

No pharmaceutical product can be placed in the market without receiving prior authorisation from the regulatory authorities, upon successful completion of a lengthy procedure for evaluating the quality, safety and efficacy of the product. New products in the process of development are legally required to undergo strict, extensive and stringent tests before they are approved and allowed in the market.

Once a new compound has been identified in the laboratory, medicines are developed as follows. During rigorous pre-clinical testing, pharmaceutical companies conduct laboratory and animal studies to assess the chemical, biological and toxicological properties of the compound against the targeted disease.

It is only when these tests show favourable results that a company can proceed with clinical trials, which are experiments conducted within very strict ethical and technical rules on human beings. The clinical trials testing comprises three distinct phases:

- In phase I, the medicine is tested on a few (about 20 to 100) healthy volunteers under strict hospital control.
- In phase II, controlled trials are carried out on volunteer patients (approximately 100 to 500) to gather information on the compound's efficacy (relation between dose and effect) and safety (identification of possible adverse side effects).
- In phase III, more comprehensive studies are carried out (usually involving 1000 to 5000 voluntary patients in clinics and hospitals), especially on any long-term effects. The proposed new treatment is also compared with other treatments already in use.

Upon successful completion of the clinical trials phases, a marketing authorisation application is submitted to the regulatory authorities for approval. Once marketing authorisation has been granted, the use of a medicine is still carefully monitored in accordance with approved current medical practices (EFPIA, 2003). In our mathematical model, we focus on the late stages of the clinical trials testing procedure (Phase III).

In order for a pharmaceutical company to make up for the patent time spent on the lengthy clinical trials phases and market launch its new products as soon as possible, it should decide early enough on its capacity investment plan. The decisions involved are taken in the face of uncertain clinical trials outcomes, usually relying on the pre-clinical results of each candidate product. They are considered as valid indicators that can produce rough estimates of the probability of success in the clinical trials procedure for each candidate product.

2.2.3 Capacity management

The manufacturing of the selected products can take place either in one production site or in many different geographically distributed production sites. Different production sites usually offer different tax rates, capital and operating costs that have to be taken into account by the optimisation algorithm. Additional data such as scale-up and qualification runs costs, production rates per product is also required for an accurate analysis of the capacity needs at each candidate production site.

The manufacturing equipment at each site is organised into blocks and each block involves a number of manufacturing suites. A single manufacturing suite comprises a production line coupled with a purification line. These suites are assumed to be available in identical capacities and known fixed cost. Each suite makes use of services such as resource utilities, administration and analytical/laboratory facilities. A maximum number of suites can share one service centre, creating a manufacturing block as illustrated in Figure 2.1.

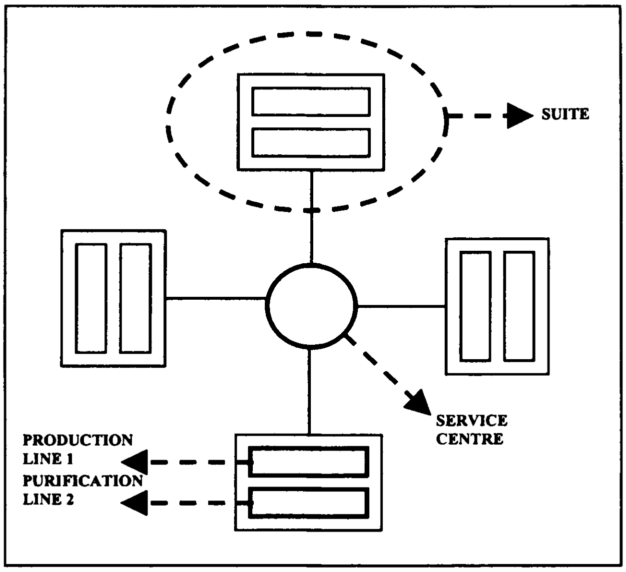

Figure 2.1: *Manufacturing block with one service centre and four suites*

The first suite to be attached to the block is denoted as the header suite, while the rest of the suites that complete the block are referred to as the non-header suites. Each block does not need to have all of its suites constructed at the same time, while the investment strategy must take into account the construction lead-time before an invested suite becomes available for production. Usually, the construction time and cost for the header suite of each block is relatively larger than that of the rest of suites belonging to the same block, reflecting the additional cost and time needed for the construction of the service centre.

Initially, there might be a number of existing blocks. Capacity management faces the problem of allocating the existing manufacturing resources to the selected product portfolio and also considers the option of investing into additional manufacturing capacity in order to satisfy the anticipating customer demand.

The first time a product is manufactured in a new production site, it must undergo a scale-up activity before actual production starts. This reflects the time period needed in order to learn how to manufacture the product in a satisfactorily repeatable fashion. Upon completion of the scale-up process, the first few batches of the product ever produced at a site (qualification amount), must be sent to the relevant regulatory authorities for approval.

Before a campaign for a particular product is started, the suite must be cleaned thoroughly and that takes a long time (e.g. one month). Additionally, the product-to-suite allocation should consider the case of manufacturing many different products in the same facility by taking into account the associated changeover time. During the campaign products are produced at a nominal production rate. However, spoiled batches may occur during production and therefore production loss factors need to be included to account for the actual production levels that tend to be lower than the nominal ones (Papageorgiou *et al.*, 2001).

2.2.4 Trading structure

Pharmaceutical companies nowadays are multi-product, multi-purpose and multi-site facilities operating in different countries and dealing with a global-wide international clientele. In such enterprise networks, the internal trading structure of the company plays a key role in business performance and necessitates the appropriate attention.

The internal trading structure of the company is mainly concerned with the financial flows between the various manufacturing and commercialisation business centres of the company. The supply chain network of a pharmaceutical company usually involves the following types of business centres:

- *Production sites*: these locations are responsible for product manufacturing.
- *Intellectual property owner (IP-Centre)*: this location is responsible for the funding and development of new products.
- *Sales regions*: these locations are responsible for product marketing and sales.

Each one of the aforementioned business centres that is present in the trading structure of the company, can operate either as a cost or a profit centre. The different operating modes are explained below:

- *Cost Centre:* this centre usually covers only its own costs plus a small profit, while products are sold according to a "cost-plus" formula (cost plus a predetermined percentage).

- *Profit Centre:* significant profits are realised in this centre where products are sold according to a "resale-minus" formula (sales revenue minus a predetermined percentage).

The adopted trading structure often has a significant impact on the after-tax profitability of the company by affecting the profits made at each business centre and consequently the taxes paid at every corresponding location. Furthermore, it may also affect decisions concerning the location of a new manufacturing plant, when candidate locations with different tax regimes are considered.

A typical trading structure for a pharmaceutical company is shown in Figure 2.2. In this case, all production sites operate as cost centres, while the *IP-Centre* and the sales regions operate as profit centres retaining most of the overall profits (Papageorgiou *et al.*, 2001).

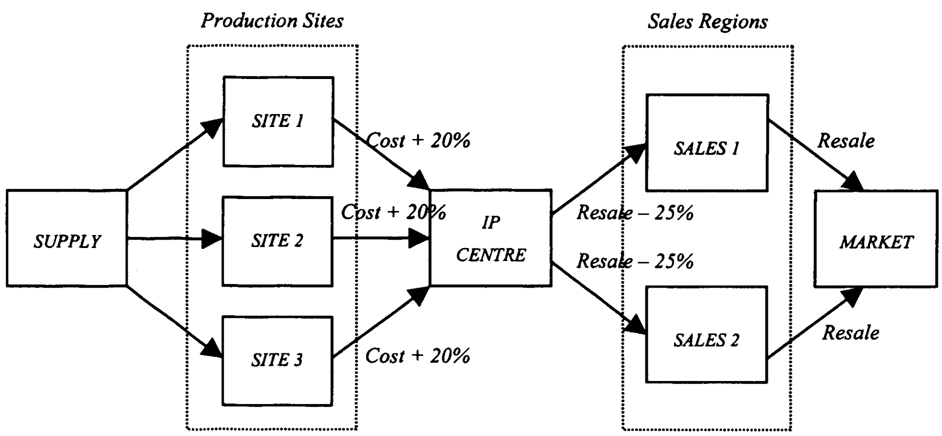

Figure 2.2: *Typical internal trading structure*

2.2.5 Problem statement

A holistic approach is needed in order to consider simultaneously product management, clinical trials uncertainty, capacity management and trading structure and resolve the dominant trade-offs in an optimal manner, so as to facilitate business decision-making in the pharmaceutical industry. Overall, the problem of multi-site capacity planning under uncertainty can be formally stated as follows:

Given:

- A set of potential products.
- Probability of success in clinical trials for each product.
- Production rates, expected production losses and shelf-life for each product.
- Manufacturing costs for each product.
- Commercialisation costs for each product.
- Forecasted nominal demand and selling price for each product.
- A set of potential production sites and products involved.
- Fixed and variable operating costs for every production site.
- Construction lead-times and capital investment costs for each production site.
- Depreciation rates for capital investment at each production site.
- Taxation, interest and inflation rate for each location.
- Trading structure of the company.

Determine:

- The product portfolio (which products from the candidate portfolio to manufacture).
- The manufacturing network (where to manufacture the selected products).
- The multi-site investment strategy (what capacity and when to invest at each production site).
- Detailed production plans (how much product to manufacture in each suite at each production site per year).
- Sales and inventory planning profiles (how much product to sell and how much inventory to maintain).

So as to maximise the expected nett present value (eNPV) of the company.

2.3 Detailed mathematical formulation

In formulating the detailed mathematical model, we follow the notation of Papageorgiou *et al.* (2001) while adding a stochastic dimension to the problem in order to account for the uncertain clinical trials outcomes and the associated customer demand. We employ a discrete time formulation, where the time horizon is discretised into time intervals of equal duration, one year each.

The following assumptions are made:

- Start-up and shutdown periods are considered to be negligible compared to the duration of each time interval.
- Two different clinical trials outcomes are considered for each candidate product (*Success, Failure*). The product probability of success is assumed to be given from the pre-clinical tests.
- In case a product fails in the clinical trials, the customer demand is consequently zero over all remaining time periods. Otherwise, the customer demand equals a forecasted nominal baseline demand.
- Once a product has been selected for manufacturing, it undergoes the tasks of scale-up and qualification runs before actual production starts.

The decision variables involved in our problem can be partitioned into two different sets, namely the strategic and the operational decisions. The strategic decisions reflect the decisions that must be made immediately (here-and-now) in the face of uncertainty and they include:

- Product selection (binary variables).
- Allocation of products to production sites (binary variables).
- Capacity investment decisions for the selected production sites (binary variables).

Generating all possible scenarios for p potential products, each one with two outcomes, results in 2^p scenarios. Each individual scenario is a fairly small deterministic problem. The demand and its associated probability for the different outcomes of each product are assumed to be known. The probability of occurrence for scenario k (π_k) is a function of the individual probabilities of success ($prob_p$) and failure ($1 - prob_p$) of each product p and is given by the following formula:

$$\pi_k = \prod_{p \in \Delta_k} prob_p \cdot \prod_{p \notin \Delta_k} (1 - prob_p) \tag{2.1}$$

where Δ_k is the set of products that are successful in scenario k.

The multi-site investment strategy is common to all possible scenarios present in the second stage. However, due to the different product demand patterns, every scenario has its own characteristic production, inventory and sales profile. The operational decisions reflect the scenario-dependant decisions made upon completion of the clinical trials and resolution of the uncertainty (wait-and-see) and they include:

- Timings of scale-up and qualifications runs (binary variables).
- Allocation of products to manufacturing suites (binary variables).
- Detailed production plans at each production site (continuous variables).
- Inventory profiles (continuous variables).
- Sales profiles at each sales region (continuous variables).

Based on the given probabilities of success for each potential product, the problem is then to find the optimal product portfolio and investment decisions together with detailed production and sales plans so as to maximise the expected nett present value (eNPV). The eNPV is simply the summation of all scenario NPVs, weighted by their associated probabilities of occurrence.

Overall, the problem of multi-site capacity planning under uncertainty is formulated as a two-stage, multi-scenario mixed integer linear programming (MILP) mathematical model. The proposed model constitutes an extension of the mathematical model presented by Papageorgiou *et al.*, (2001). The following nomenclature is used in our mathematical model formulation:

Indices

i, j manufacturing suites

k scenarios

l production sites

p products

s sales regions

t, θ time periods

Sets

Γ^l set of products that can be manufactured at production site l

Z^s set of products that can be sold at sales region s

Δ_k set of products that are successful in scenario k

Π_k^l set of products that can be manufactured at production site l and are successful in scenario k ($\Pi_k^l = \Gamma^l \cap \Delta_k$)

Λ_k^s set of products that can be sold at sales region s and are successful in scenario k ($\Lambda_k^s = Z^s \cap \Delta_k$)

Ψ_{pk} set of production sites that can manufacture product p in scenario k

Θ_{pk} set of sales regions that can sell product p in scenario k

I^l set of manufacturing suites at production site l

F^l set of header suites at production site l

N_i^l set of non-header suites that belong to the same block as header suite i at production site l

T set of time periods

Parameters

D_{ptk}^s forecasted demand of product p at sales region s at time t in scenario k

f inflation rate

g interest rate

H_t available suite production time at time period t

L_p production loss factor of product p

M maximum number of suites in each block

r_p production rate of product p in suite i at production site l

α_p qualification amount of product p

β^l_{pt} scale up cost for product p at time t at production site l

γ^l_{pt} qualification runs cost for product p at time t at production site l

δ_i construction lead time for suite i

ε_t discount factor

ζ_p shelf-life of product p

η^l_i fixed operating cost of suite i at production site l

λ_p royalties costs for product p

μ_p marketing costs for product p

ν_p forecasted sales price of product p

ξ^l_p variable operating cost of product p at production site l

π_k probability of occurrence for scenario k

ρ^l tariff price for production site l

ρ^s tariff price for sales region s

σ_{pt} research and development cost of product p at time t

$\hat{\tau}_p$ scaleup time of product p

$\bar{\tau}$ product changeover time

φ^l tax life period for production site l

ψ_t^l tax rate for production site *l* at time *t*

ψ_t^s tax rate for sales region *s* at time *t*

ψ_t^{IP} tax rate for the intellectual property centre *IP* at time *t*

ω_i^l capital investment cost of suite *i* at production site *l*

Next, the following key variables are introduced:

Binary Variables

A_{it}^l 1 if suite *i* is available for production at time *t* at production site *l*, 0 otherwise

E_{it}^l 1 if investment decision is taken for suite *i* at time t at production site *l*, 0 otherwise

U_p 1 if product *p* is selected for development and manufacturing, 0 otherwise

V_p^l 1 if product *p* is selected at production site *l*, 0 otherwise

X_{ptk}^l 1 if qualification runs of product *p* take place at time *t* in scenario *k* at production site *l*, 0 otherwise

Y_{iptk}^l 1 if product p is manufactured in suite *i* at time *t* in scenario *k* at production site *l*, 0 otherwise

Z_{iptk}^l 1 if scaleup of product *p* takes place in suite *i* at time *t* in scenario *k* at production site *l*, 0 otherwise

$\hat{Z}_{ptk}^l$ 1 if scaleup of product p takes place at time *t* in scenario *k* at production site *l*, 0 otherwise

Continuous Variables

C_{itk}^l number of changeovers in suite *i* at time *t* in scenario *k* at production site *l*

CI_t^l cost of capital investment at time *t* at production site *l*

$CSQR_{tk}^{l}$ scale-up and qualification runs cost at time t at production site l in scenario k

DC_t^l depreciation at time t at production site l

I_{ptk} inventory of product p held at the end of time t in scenario k

MC_{tk}^{s} marketing costs at time t at sales region s in scenario k

OC_{tk}^{l} operating cost at time t at production site l in scenario k

RDC_t^{IP} R&D costs at time t for the *IP-Centre*

RC_{tk}^{IP} royalty costs at time t for the *IP-Centre* in scenario k

S_{ptk}^{s} amount of product p sold at time t in scenario k at sales region s

SR_{tk}^{s} revenues of sales at time t in scenario k at sales region s

T_{iptk}^{l} production time of product p in suite i at time t in scenario k at production site l

Tax_{tk}^{IP} taxes paid at the *IP-Centre* location at time period t at scenario k

T^{IP} taxes paid at the *IP-Centre* location

T^{S} taxes paid at all sales regions

T^{L} taxes paid at all production sites

W_{ptk} wasted amount of product p at time t in scenario k

Φ^{A} expected nett present value after taxes

Φ^{B} expected nett present value before taxes

The derivation of the detailed mathematical model formulation is described next.

2.3.1 Detailed constraints

Product Existence Constraints

If a product is not selected for development and manufacturing (i.e. $U_p=0$), then this product is not allowed for manufacturing in any candidate production site.

$$V_p^l \leq U_p \qquad \forall\, l, p \in \Gamma^l \tag{2.2}$$

Suite Availability and Investment Constraints

At every production site *l*, manufacturing suite *i* is available at time period *t*, only if it is available at the previous time period (*t-1*), or an investment decision is taken for that suite δ_i periods before, reflecting the construction time required.

$$A_{it}^l = A_{i,t-1}^l + E_{i,t-\delta_i}^l \qquad \forall\, l, i \in I^l, t \tag{2.3}$$

Constraints that relate the construction times of the header suite with the subsequent non-header suites belonging to the same block are necessary, in order to ensure that a non-header suite *j* can be invested at time *t* only if the header suite *i* has already been invested in at least δ_i-δ_j time periods before period *t* or it is initially available (i.e. $A_{i0}^l = 1$):

$$\sum_{\theta=1}^{t-(\delta_i-\delta_j)} E_{i\theta}^l + A_{i0}^l \geq E_{jt}^l \qquad \forall\, l, i \in F^l, j \in N_i^l, t \tag{2.4}$$

In order to suppress possible solution degeneracy and tighten the problem, a constraint is necessary that allows suite *i* to be invested at time *t* only if suite *i-1* has been invested at a time period up to a time period *t* inclusive or it is initially available (i.e. $A_{i0}^l = 1$). In order to take into account that different suites may have different construction lead times, the constraints take the following form:

$$\sum_{\theta=1}^{t-(\delta_i-\delta_{i+1})} E_{i\theta}^l + A_{i0}^l \geq E_{i+1,t}^l \qquad \forall\, l, i = 1, \ldots, I^l - 1, t \tag{2.5}$$

Scale-up Constraints

If a product is selected for manufacturing at production site *l*, then the scale-up process will occur only once.

$$\sum_t \hat{Z}_{ptk}^l \leq V_p^l \qquad \forall\; l, k, p \in \Pi_k^l \tag{2.6}$$

If scale-up occurs (i.e. $\hat{Z}^{l}_{ptk} = 1$), then a single suite should be allocated:

$$\sum_{i \in I^l} Z^{l}_{iptk} = \hat{Z}^{l}_{ptk} \qquad \forall\, t, l, k, p \in \Pi^l_k \tag{2.7}$$

A suite i at production site l can be used for scale-up only if it is available:

$$\sum_{p \in \Pi^l_k} Z^{l}_{iptk} \leq \left|\Pi^l_k\right| \cdot A^l_{it} \qquad \forall\, l,\ i \in I^l,\ t,\ k \tag{2.8a}$$ *

Alternatively, constraints (2.8a) can be disaggregated to result in a tighter form at the expense of a larger problem size:

$$Z^{l}_{iptk} \leq A^l_{it} \qquad \forall\, l, i \in I^l, t, k, p \in \Pi^l_k \tag{2.8b}$$

Finally, a product is allowed to be produced during a certain period t, only if scale-up has taken place up to that period. Mathematically, we have:

$$\sum_{\theta=1}^{t} \hat{Z}^{l}_{p\theta k} \geq Y^{l}_{iptk} \qquad \forall\ l, i \in I^l, t, k, p \in \Pi^l_k \tag{2.9}$$

Qualifications Constraints

After the scale-up is completed, qualification runs should be performed to ensure that the plant is capable of producing that product in full compliance with the regulatory authorities. The first batches produced are called qualification batches and they should coincide with the first time that production of that product occurs at the specific production site. The qualification constraints take the following form:

$$X^{l}_{ptk} \geq Y^{l}_{iptk} - \sum_{j \in I^l} \sum_{\theta=1}^{t-1} Y^{l}_{jp\theta k} \qquad \forall\ l, i \in I^l, k, p \in \Pi^l_k, t \tag{2.10}$$

In order to ensure that every product is qualified only once, the following constraints should be included:

$$\sum_{t} X^{l}_{ptk} \leq V^l_p \qquad \forall\ l, k, p \in \Pi^l_k \tag{2.11}$$

---

* Symbol $|X|$ represents the cardinality of set X

Production Constraints

Each product p can be produced at production site l only if it is selected for manufacturing in that specific site:

$$\sum_{i\in I^l}\sum_{t} Y^l_{iptk} \leq \left|I^l\right|\cdot\left|T\right|\cdot V^l_p \quad \forall \;\; l,k,p\in \Pi^l_k \tag{2.12a}$$

Constraints (2.12a) can be disaggregated to provide a tighter alternative in the following form:

$$\sum_{i\in I^l} Y^l_{iptk} \leq \left|I^l\right|\cdot V^l_p \quad \forall \; l,k,p\in \Pi^l_k,t \tag{2.12b}$$

Constraints (2.12b) can further be disaggregated to result in an even tighter model formulation:

$$Y^l_{iptk} \leq V^l_p \quad \forall \;\; l,\, i\in I^l,k,p\in \Pi^l_k,t \tag{2.12c}$$

Furthermore, each product p can be produced in a specific suite i at production site l only if that suite is available:

$$\sum_{p\in \Pi^l_k} Y^l_{iptk} \leq \left|\Pi^l_k\right|\cdot A^l_{it} \quad \forall \;\; l,\, i\in I^l,\, t,k \tag{2.13a}$$

Constraints (2.13a) can be disaggregated to result in a tighter form at the expense of a larger problem size:

$$Y^l_{iptk} \leq A^l_{it} \quad \forall \;\; l,i\in I^l,t,k,p\in \Pi^l_k \tag{2.13b}$$

When a product is qualified ($X^l_{ptk} = 1$), a minimum required amount must be produced:

$$\sum_{i\in I^l} r_p \cdot T^l_{iptk} \geq \alpha_p \cdot X^l_{ptk} \quad \forall \; l,k,p\in \Pi^l_k,t \tag{2.14}$$

If product p is manufactured in suite i at production site l under scenario k ($Y^{l}_{iptk} = 1$), then the production time is bounded by an upper bound of available production time. Otherwise, the production time variable is forced to zero:

$$T^{l}_{iptk} \leq H_t \cdot Y^{l}_{iptk} \qquad \forall\, l, i \in I^{l}, k, p \in \Pi^{l}_{k}, t \tag{2.15}$$

When multiple products are produced in the same suite, a minimum number of changeovers between the different product campaigns should taken into account:

$$C^{l}_{itk} \geq \sum_{p \in \Pi^{l}_{k}} Y^{l}_{iptk} - A^{l}_{it} \qquad \forall\, l, i \in I^{l}, t, k \tag{2.16}$$

The total suite production time for all manufactured products should not exceed the total available production time compensated for the time required for any necessary changeover and/or scale-up processes:

$$\sum_{p \in \Pi^{l}_{k}} T^{l}_{iptk} \leq H_t - \overline{\tau} \cdot C^{l}_{itk} - \sum_{p \in \Pi^{l}_{k}} \hat{\tau}_p Z^{l}_{iptk} \qquad \forall\, l, i \in I^{l}, t, k \tag{2.17}$$

Inventory Constraints

For every scenario k, the amount of product p stored globally at the end of period t will be equal to the amount produced at the previous period t-1 plus the nett amount produced in all suites at all production sites during period t (considering production losses), minus the amount sold to the sales regions minus the amount wasted due to the limited product lifetime.

$$I_{ptk} = I_{p,t-1,k} + (1 - L_p) \sum_{l \in \Psi_{pk}} \sum_{i \in I^{l}} r_p \cdot T^{l}_{iptk} - \sum_{s \in \Theta_{pk}} S^{s}_{ptk} - W_{ptk} \qquad \forall\, k, p \in \Delta_k, t \tag{2.18}$$

Lifetime Constraints

The amount of product p stored in each period t cannot be sold after the next ζ_p periods. The product lifetime constraints take the following form:

$$I_{ptk} \leq \sum_{s \in \Theta_{pk}} \sum_{\theta = t+1}^{t+\zeta_p} S^{s}_{p\theta k} \qquad \forall\, k, p \in \Delta_k, t \tag{2.19}$$

Sales Constraints

The amount of product p sold at each time period should be less or equal to the forecasted customer demand of that product:

$$S^{s}_{ptk} \leq D^{s}_{ptk} \qquad \forall\ k, s, p \in \Lambda^{s}_{k}, t \tag{2.20}$$

Furthermore, different sales strategies could be adopted to reflect alternative marketing policies of the company (see for example, Papageorgiou *et al.*, 2001).

2.3.2 Detailed objective function

The objective function employed in our model is the expected nett present value (eNPV) after taxes (Φ^A), which constitutes an appropriate performance criterion of the company. First, we present the formulation of the objective function before taxes (Φ^B), that includes revenues from sales, marketing costs, royalties costs, R&D costs, costs of scale-up and qualification runs, operating costs, and capital investment costs similarly to Papageorgiou *et al.* (2001). Next, we consider these terms in sequence.

Sales revenue at sales regions

For every scenario k, revenue due to product sales at sales region s over time period t, can be calculated as follows:

$$SR^{s}_{tk} = \sum_{p \in \Lambda^{s}_{k}} v_p S^{s}_{ptk} \qquad \forall\ t, k, s \tag{2.21}$$

Marketing costs at sales regions

In order to capture the marketing costs incurred at sales region s, the following cost term should be included in the objective function. The marketing costs are assumed to be proportional to the product amounts sold at sales region s over time period t in every scenario k.

$$MC^{s}_{tk} = \sum_{p \in \Lambda^{s}_{k}} \mu_p v_p S^{s}_{ptk} \qquad \forall\ t, k, s \tag{2.22}$$

Royalty costs at the IP-Centre

The royalty costs for the *IP-Centre* are also assumed to be proportional to the product amounts sold at all potential sales regions over time period t in every scenario k.

$$RC_{tk}^{IP} = \sum_{s \in \Theta_{pk}} \sum_{p \in \Lambda_k^s} \lambda_p v_p S_{ptk}^s \qquad \forall\, t,k \tag{2.23}$$

R&D costs at the IP-Centre

The cost spent at the *IP-Centre* associated with the R&D over time period t is calculated using the following constraint:

$$RDC_t^{IP} = \sum_p \sigma_{pt} \cdot U_p \qquad \forall\, t \tag{2.24}$$

Scale-up and qualification costs at production sites

The scale-up and qualification runs costs at production site l over time period t in scenario k are calculated as follows:

$$CSQR_{tk}^l = \sum_{p \in \Pi_k^l} (\beta_{pt}^l \hat{Z}_{ptk}^l + \gamma_{pt}^l X_{ptk}^l) \quad \forall\, t,k,l \tag{2.25}$$

Operating costs at production sites

The operating cost at production site l over time period t in scenario k is a function of two terms. The first one is associated with the fixed operating cost incurred if suite i is available for production (i.e. $A_{it}^l = 1$). The second term represents the variable operating costs depending on the amount of product produced in suite i at production site l during time period t at scenario k. The total operating cost is given by the following equation:

$$OC_{tk}^l = \sum_{i \in I^l} \eta_i^l A_{it}^l + \sum_{i \in I^l} \sum_{p \in \Pi_k^l} \xi_p^l \cdot r_p \cdot T_{iptk}^l \quad \forall\, t,k,l \tag{2.26}$$

Capital investment costs at production sites

The capital investment cost depends on the cost of each suite i invested at production site l in time period t.

$$CI_t^l = \sum_{i \in I^l} \omega_i^l E_{it}^l \quad \forall\ t, l \tag{2.27}$$

For the eNPV calculation, we must introduce a discount factor, ε_t, associated with the inflation rate, f, and the interest rate, g, according to the following formula:

$$\varepsilon_t = \left(\frac{1+f}{1+g} \right)^{t-1} \quad \forall\ t \tag{2.28}$$

Overall, the objective function before taxes, Φ^B, includes sales revenues minus marketing costs, capital investment costs, operating costs, scale-up and qualification costs, royalties and R&D costs. The aforementioned terms are discounted in a summation over all time periods and over all possible scenarios multiplied by the associated scenario probability of occurrence:

$$\Phi^B = \sum_k \pi_k \sum_t \varepsilon_t \left(\sum_s (SR_{tk}^s - MC_{tk}^s) - \sum_l (CI_t^l + OC_{tk}^l + CSQR_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP} \right) \tag{2.29}$$

In order to calculate the taxes paid from the company, we adopt the trading structure along with the internal transfer pricing policies as presented in Figure 2.2. The production sites are classified as cost centres operating with a cost-plus pricing formula. Each production site incurs the operating costs, cost of scale-up and qualification runs and capital investment costs. The profit made at each production site equals $(1+\rho^l)Cost\text{-}Cost$. Overall, the taxes paid by all production sites are given by the following equation, where the corresponding profit $\rho^l Cost$ is multiplied by the associated tax rate of each production site in a summation over all time periods and possible scenarios:

$$T^L = \sum_k \pi_k \sum_t \varepsilon_t \sum_l \psi_t^l \rho^l (DC_t^l + OC_{tk}^l + CSQR_{tk}^l) \tag{2.30}$$

Notice that in order to calculate the taxes in constraint (2.30), we employ a linear depreciation term that is defined as follows:

$$DC_t^l = \sum_{i \in I^l} \sum_{\theta = t-\delta_i - \phi^l + 1}^{t-\delta_i} \frac{\omega_i^l}{\phi^l} E_{i\theta}^l \qquad \forall\, t,l \tag{2.31}$$

The sales regions are classified as profit centres operating with a resale-minus pricing formula. Each sales region has associated sales revenues and marketing costs. Overall, the taxes paid by all sales regions are given by the following equation:

$$T^S = \sum_k \pi_k \sum_t \varepsilon_t \sum_s \psi_t^s (\rho^s SR_{tk}^s - MC_{tk}^s) \tag{2.32}$$

The intellectual property owner (*IP-Centre*) operates as a profit centre as well. The product selling from the production sites to the *IP-Centre* and from the *IP-Centre* to the different sales regions generates the profit retained in the *IP-Centre*. However, in calculating the taxes paid by the *IP-Centre*, the associated royalties and R&D costs should also be included. Overall, the corresponding tax paid by the *IP-Centre* is given by the following equation:

$$T^{IP} = \sum_k \pi_k \sum_t \varepsilon_t \cdot Tax_{tk}^{IP} \tag{2.33}$$

In order to ensure that the taxes paid by the *IP-Centre* at every time period *t* and scenario *k* have a positive sign, the following inequalities are necessary:

$$Tax_{tk}^{IP} \geq \psi_t^{IP} \left(\sum_s (1-\rho^s) SR_{tk}^s - \sum_l (1+\rho^l)(DC_t^l + OC_{tk}^l + CSQR_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP} \right) \quad \forall\, t,k \tag{2.34}$$

$$Tax_{tk}^{IP} \geq 0 \quad \forall\, t,k \tag{2.35}$$

2.3.3 Summary of detailed model

In conclusion, the detailed mathematical model can be summarised as follows:

[Problem D]

$$\max \quad \Phi^A = \Phi^B - T^L - T^S - T^{IP}$$

Subject to:

Constraints (2.2)-(2.7), (2.8a), (2.9)-(2.11), (2.12a), (2.13a), (2.14)-(2.27), (2.29)-(2.35).

The resulting optimisation problem (Problem D) corresponds to a mixed-integer linear programming (MILP) mathematical model. The applicability of the proposed mathematical model is demonstrated by the illustrative examples presented in the next section.

2.4 Illustrative examples

Five instances of a stochastic, multi-site, multi-period capacity planning problem are solved in order to validate the applicability of the proposed mathematical model. Consider four alternative locations (*A-D*), where *A* and *B* are the sales regions, *A* is the intellectual property owner (*IP-Centre*), while *B*, *C* and *D* are the candidate production sites. Five examples, namely 3PROD, 4PROD, 5PROD, 6PROD and 7PROD consider the manufacturing of three, four, five, six and seven potential products (*P1-P7*), respectively.

The entire time horizon of interest is thirteen years. In the first three years, no production takes place and the outcomes of the clinical trials are not yet known. Initially, there are two suites already in place at production site *B*. Further decisions for investing in new manufacturing suites are to be determined by the optimisation algorithm. We assume that the trading superstructure is given together with the internal pricing policies as shown in Figure 2.3.

The tax rate profiles for each location are shown in Table 2.1 All product-related data is presented in Table 2.2 while the forecasted customer demand is given in Table 2.3. All additional parameters are shown in Table 2.4. It should be added that the associated costs for production sites other than *B* can be found by multiplying the relevant costs of *B* with the relative capital and operating costs given in Table 2.1.

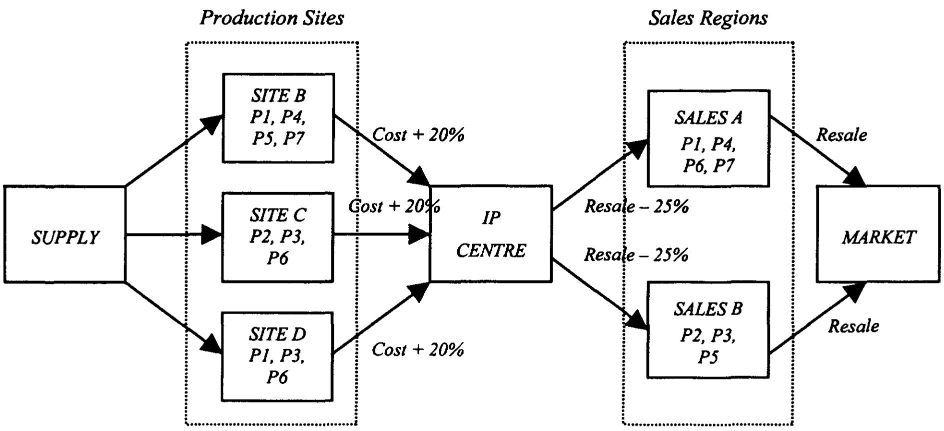

Figure 2.3: *Trading superstructure of the company*

Table 2.1: *Tax regions*

Location	Tax Rate Profile	Relative Operating Cost	Relative Capital Cost
A	0.28	-	-
B	0.25	1	1
C	0 (for *t1-t6*) 0.20 (for *t7-t13*)	1.1	1.1
D	0.30	0.27	1.3

Table 2.2: *Product data*

Product	Success Probab. $prob_p$	R&D Cost σ_{pl}	Scaleup Cost β^l_{pt}	Qual. Cost γ^l_{pt}	Qual. Amount a_p	Prod. Rate r_p	Var. Oper. Cost in *B* ξ^B_p	Royalty Cost λ_p	Selling Price v_p
P1	0.36	40	10	15	6.10	2.2	0.15	0.10	5
P2	0.33	45	20	20	2.55	0.9	0.20	0.10	4
P3	0.61	25	15	25	2.55	1.9	0.30	0.20	6
P4	0.32	50	10	20	5.20	0.9	0.15	0.15	7
P5	0.30	30	20	30	4.30	1.2	0.20	0.25	3
P6	0.31	60	10	15	7.10	1.4	0.25	0.15	3
P7	0.38	35	15	10	5.60	1.6	0.20	0.20	5

Table 2.3: *Customer demand in case of successful clinical trials outcomes*

Product\Year	*t4*	*t5*	*t6*	*t7*	*t8*	*t9*	*t10*	*t11*	*t12*	*t13*
P1	*14.7*	*31.5*	*78.4*	*75.6*	*77.7*	*77.7*	*77.7*	*77.7*	*77.7*	*77.7*
P2	*0.0*	*3.5*	*15.4*	*30.1*	*31.5*	*38.5*	*38.5*	*42.0*	*42.5*	*49.0*
P3	*7.7*	*22.4*	*36.4*	*47.6*	*56.0*	*56.7*	*56.7*	*57.4*	*57.4*	*57.4*
P4	*51.1*	*52.5*	*58.8*	*65.1*	*72.8*	*73.5*	*73.5*	*73.5*	*73.5*	*73.5*
P5	*0.0*	*26.6*	*50.4*	*64.4*	*74.2*	*79.8*	*79.8*	*79.8*	*79.8*	*79.8*
P6	*0.0*	*0.0*	*38.5*	*49.0*	*65.8*	*72.8*	*71.4*	*71.4*	*75.6*	*74.2*
P7	*4.2*	*6.3*	*8.4*	*23.8*	*30.8*	*39.2*	*64.4*	*65.8*	*65.8*	*65.8*

Table 2.4: *Additional parameters*

Definition and Symbol of Parameter	Value of Parameter
Inflation rate (f)	3 %
Interest rate (g)	15 %
Cost-plus percentage (ρ^l)	20 %
Resale-minus percentage (ρ^s)	25 %
Available production time (H_t)	11 months/year
Scale-up time ($\hat{\tau}$)	2 months
Changeover time ($\bar{\tau}$)	1 month
Shelf lifetime (ζ_p)	4 years
Tax life period (φ^l)	5 years
Production losses (L_p)	10 %
Maximum number of suites per site (M_{tot})	4
Maximum of suites per block (M)	4
Construction lead time (δ_i)	3 years for header suite 2 years for non-header suite
Capital investment cost for suite in location B (ω_i^B)	100 rmu* for header suite 50 rmu for non-header suite
Fixed operating costs for suite in location B (η_i^B)	10 rmu for header suite 5 rmu for non header suite

All five example problems were implemented in GAMS (Brooke *et al.*, 1998) using the XPRESS-MP (Dash Associates, 1999) MILP solver with a 5% margin of optimality. All runs were performed on an IBM RS/6000 workstation with a maximum computational time limit of 3 hours (10800 seconds).

Concerning the product portfolio, the obtained results for the 3PROD and 4PROD examples are summarised in the following table.

Table 2.5: *Product portfolio results*

Problem	Selected Products	No of Selected Products / Total Candidate Products
3PROD	P1, P3	2 / 3
4PROD	P1, P3, P4	3 / 4

The enterprise-wide pharmaceutical supply chain determined by the optimisation algorithm is common to both problem instances and results in the business network illustrated in Figure 2.4.

Despite the attractive zero tax rate profile for the first six years offered by location *C*, the solution determined by the optimisation algorithm suggests that it is more

* rmu = relative monetary units

profitable to invest at production sites *B* and *D* that provide lower operating costs, while no suite investment decisions are taken at production site *C*.

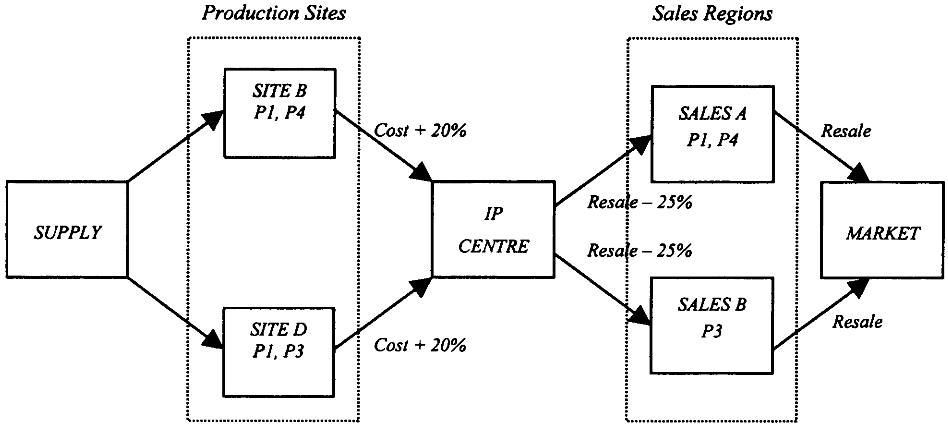

Figure 2.4: *Optimal business network*

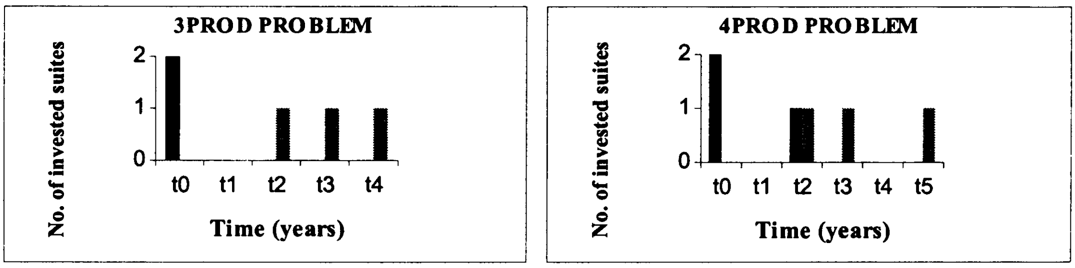

Figure 2.5: *Investment decision calendar (Site B: black, Site D: grey)*

Investment decision calendars for problems 3PROD and 4PROD are illustrated in Figure 2.5. Note in both examples the investment decisions for the very first additional manufacturing suites are taken in the early time periods while the clinical trials are still on going. The proposed investment plans take into account the construction lead-time (2 and 3 years for non-header and header suites respectively) and safeguard the availability of the newly invested equipment right after the end of the clinical trials phase. Such an investment strategy favours production in the second stage, making the most out of the products' patent life. Note also that the number of invested suites increases with the number of final products included in the portfolio, in order to satisfy the additional customer demand (three versus four invested suites for the 3PROD and 4PROD examples respectively).

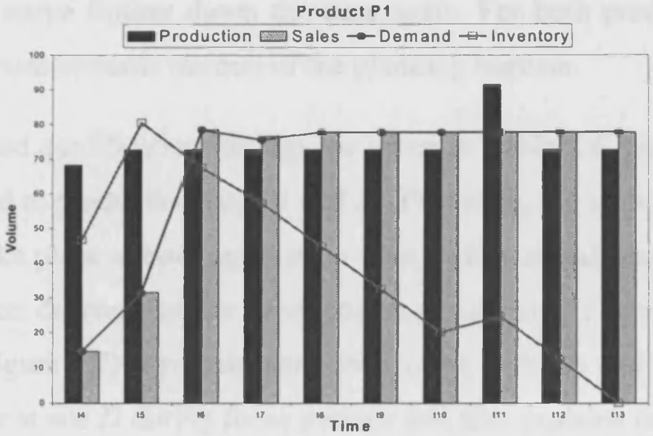

Figure 2.6: *Characteristic profiles for product P1*

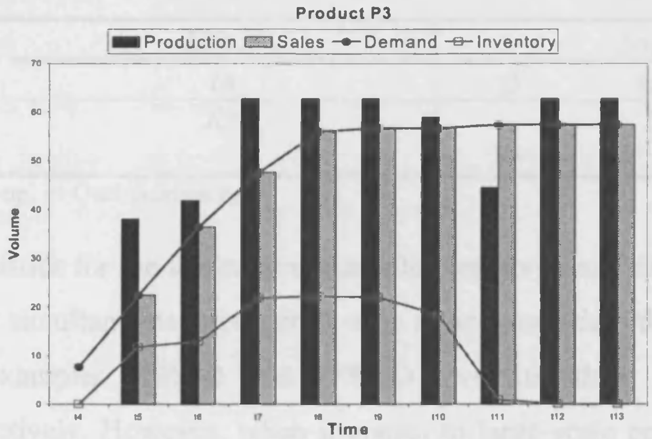

Figure 2.7: *Characteristic profiles for product P3*

According to the simultaneous approach, operational decision variables such as detailed production plans, inventory and sales profiles are also determined by the solution of the MILP model. Such characteristic profiles for the successful products *P1* and *P3* in scenario 3 of problem 4PROD are shown in Figures 2.5 and 2.7. Note that the black bars (showing the produced amounts at each year) are the total produced amounts, while the nett produced amounts are 10% lower than the total ones, due to production losses considered by the loss factor L_p in our model. It can be seen that customer demand is fully satisfied in nearly all time periods for both products *P1* and *P3*. Furthermore, the proposed production policy implies a similar inventory profile pattern for both products. In both cases we notice an inventory build-up in the beginning of the planning horizon while inventory levels gradually decline as we move further down the time scale. For both products inventories are kept at a minimum towards the end of the planning horizon.

The scale-up and qualification timings are given in Table 2.6. Notice that product *P1* is manufactured at production sites *B* and *D*. Therefore, the scale-up and qualification runs have to take place at both production sites, before actual production starts at each site. The sudden decrease in the production amounts of *P3* during time periods *t10* and *t11* (see Figure 2.7) is mainly attributed to the scale-up and qualification runs of *P1* taking place at site *D* during those periods and also explains the sudden increase in the overall production amounts of *P1* at time period *t11* (see Figure 2.6).

Table 2.6: *Scale-up and qualification timings*

Product	*Site B*	*Site D*		
	t4	*t5*	*t10*	*t11*
P1	X+		X	+
P3		X+		

X: Scale-up, +: Qualification runs

The model statistics for the illustrative examples are given in Table 2.7. According to the results, the simultaneous approach is able to accommodate the case of small and modest size examples 3PROD and 4PROD involving three and four candidate products respectively. However, when it comes to large-scale problems 5PROD and 6PROD the curse of dimensionality does not allow Problem D to converge within the 5% margin of optimality with reasonable computational effort. Moreover, in the case of the 7PROD example, Problem D fails to return an integer solution after 3 hours of computational time.

Table 2.7: *Model statistics for the illustrative examples*

Example	3PROD	4PROD	5PROD	6PROD	7PROD
Constraints	8524	18033	38434	86580	183829
Continuous Var.	3189	6933	14981	33477	71429
Discrete Var.	2320	5122	11524	29127	64329
Obj. Fun. (rmu)	223	276	281*	129**	No Int. Sol.♦
CPU (sec)	308	4526	10800	10800	10800

* Integer solution found within 12% margin of optimality.
** Integer solution found within 71% margin of optimality.
♦ LP relaxation not finished.

2.5 Concluding remarks

Chapter 2 presented a simultaneous solution approach for the problem of multi-site capacity planning under uncertainty in the pharmaceutical industry. The overall problem was formulated as a two-stage, multi-scenario mixed-integer mathematical programming MILP model incorporating issues related to product management, clinical trials uncertainty, capacity management and trading structure of the company. Five illustrative examples were then solved in order to validate the applicability of the simultaneous approach. According to the results, small and modest size example problems were solved successfully by employing the proposed methodology.

However, the solution of large-scale instances of the problem proved to be a very demanding task in terms of computational effort needed due to the combinatorial nature of the problem. The resulting large-scale MILP model could not be solved efficiently and in many cases was intractable through the employment of traditional branch-and-bound methods via the commercially available XPRESS-MP MILP solver.

Multi-site capacity problems involving an increased number of candidate products are inevitably accompanied by an increased number of possible scenarios that further magnify the inherent complexity of the problem. The solution of the resulting large-scale MILP models necessitates the implementation of an alternative solution approach that can alleviate the computational burden and yield near-optimal solutions within reasonable computational time. Such an efficient solution approach is presented in the next chapter.

Chapter 3

A hierarchical approach for multi-site capacity planning

3.1 Introduction

In this chapter, we propose a hierarchical solution approach able to accommodate the combinatorial nature of the multi-site capacity planning problem and reduce the computational effort needed for its solution. The proposed methodology is based on the decoupling of the strategic and operational decision-making levels identified in our problem. In particular, the hierarchical solution approach employs a suite-aggregate mathematical formulation to determine the strategic decision variables which are then fed into the original detailed model in order to derive the operational decision variables.

The suite-aggregate mathematical model formulation is an approximation of the detailed model formulation (Problem D) presented in Chapter 2. The basic idea behind its formulation is to exploit the fact that the manufacturing suites to be invested are identical pieces of equipment. Therefore, in case we are interested in determining only a capacity investment plan without considering product-to-suites allocation variables, there is no particular need to discriminate between the manufacturing suites and consider each suite individually since there exists both equipment and task equivalence between them. Alternatively, the manufacturing capacity to be invested in every production site can be modelled as an overall aggregate capacity resource, representing the summation of all individual manufacturing suites.

The main advantage of the aggregate model formulation is that it does not compromise neither on the scenario or the time dimension while it leads to a much smaller problem size, both in terms of constraints and variables. On the other hand, the reduced problem size of the resulting model comes at the expense of less detailed production plans. However, for the purposes of determining the strategic here-and-now decisions, aggregate production plans can still capture the various trade-offs among the candidate products and production sites. The suite-aggregate model is a coarse model, yet an accurate one that sufficiently approximates the detailed model by focusing on the strategic decisions, while it adopts a myopic behaviour towards the second stage operational decision variables, thus providing a valid upper bound by overestimating the objective function of the original problem. Next, the aggregate model formulation is presented.

3.2 Aggregate mathematical formulation

First, some new notation related to the aggregate model formulation is given in addition to the nomenclature presented in Chapter 2.

Parameters

δ_h construction lead time for header suite

δ_n construction lead time for non-header suite

η_h^l fixed operating cost of header suite at production site l

η_n^l fixed operating cost of non-header suite at production site l

M_{tot}^l maximum total number of allowed suites at production site l

ω_h^l capital investment cost of header suite at production site l

ω_n^l capital investment cost of non-header suite at production site l

Integer Variables

$h\tilde{A}_t^l$ number of header suites available for production at time t at production site l

$n\tilde{A}_t^l$ number of non-header suites available for production at time t at production site l

$h\tilde{E}_t^l$ number of header suites invested at time t at production site l

$n\tilde{E}_t^l$ number of non-header suites invested at time t at production site l

3.2.1 Aggregate constraints

Some of the constraints included in the aggregate model formulation are exactly the same as the ones used in the detailed model formulation. These include: product existence constraints (2.2), product lifetime constraints (2.19), sales constraints (2.20), sales revenue constraints (2.21), marketing costs constraints (2.22), royalty costs constraints (2.23), R&D costs constraints (2.24), sales regions taxes constraints (2.32) and *IP-Centre* taxes constraints (2.33), (2.35). Next, we describe the aggregate model constraints that take a different form from the corresponding constraints in the detailed model.

Suite Availability and Investment Constraints

The suite availability and investment constraints derive from the detailed model by simply dropping the manufacturing suite index. Consequently, the capacity investment decisions are now treated as integer variables instead of binary ones. During each time period, the optimisation algorithm has to determine an integer number of manufacturing suites to be invested at each candidate production site. In order though to capture the difference in capital and operating costs between header and non-header suites, it is necessary to introduce two distinct types of integer variables to account for the two different types of suites.

$$h\widetilde{A}_t^l = h\widetilde{A}_{t-1}^l + h\widetilde{E}_{t-\delta_h}^l \quad \forall\ l,t \tag{3.1}$$

$$n\widetilde{A}_t^l = n\widetilde{A}_{t-1}^l + n\widetilde{E}_{t-\delta_n}^l \quad \forall\ l,t \tag{3.2}$$

For the same reason, the suite investment upper bound constraints have to be considered separately for each type of suite. The parameter M_{tot}^l equals the cardinality of the manufacturing suite set I^l as described in the detailed model while $h\widetilde{A}_0^l$ and $n\widetilde{A}_0^l$ are the initially available header and non-header suites.

$$\sum_t h\widetilde{E}_t^l \le \frac{M_{tot}^l}{M} - h\widetilde{A}_0^l \quad \forall\ l \tag{3.3}$$

$$\sum_t n\widetilde{E}_t^l \le M_{tot}^l - \frac{M_{tot}^l}{M} - n\widetilde{A}_0^l \quad \forall\ l \tag{3.4}$$

The solution degeneracy constraints take the following form in the aggregate model, allowing for minimum and maximum number of non-header suites to be invested at production site l according to the number of header suites already invested in previous time periods.

$$n\widetilde{A}_0^l + \sum_{\theta=1}^{t} n\widetilde{E}_\theta^l \le (M-1)\left(h\widetilde{A}_0^l + \sum_{\theta=1}^{t-(\delta_h-\delta_n)} h\widetilde{E}_\theta^l \right) \quad \forall\ l,t \tag{3.5}$$

$$n\widetilde{A}_0^l + \sum_{\theta=1}^{t} n\widetilde{E}_\theta^l \ge (M-1)\left(h\widetilde{A}_0^l + \sum_{\theta=1}^{t-(\delta_h-\delta_n)} h\widetilde{E}_\theta^l - 1 \right) \quad \forall\ l,t \tag{3.6}$$

For example, if one header suite is initially available ($h\widetilde{A}_0^l = 1$) at production site l and no additional investment decisions are taken, then according to constraints (3.5), the maximum number of non-header suites equals the maximum number of suites per block (M) minus one, to account for the already existing header suite. Furthermore, constraints (3.6) safeguard that additional investment decisions for non-header suites in a new manufacturing block are taken only if all the previously invested manufacturing blocks are fully completed with one header suite and M-1 non-header suites for each one.

Production Constraints

The suite-aggregate model does not account for the scale-up and qualification runs and therefore no variables or constraints related to scale-up and qualification are considered. Every product included in the selected product portfolio goes on for manufacturing according to the following production constraints. In case a product is not selected for manufacturing at site l, its corresponding production time is forced to zero. The total time for manufacturing the selected product portfolio is less than or equal to the suite production time multiplied by the number of available header and non-header suites at site l in time period t. Note that the aggregate model does not consider product-to-suite allocation binary variables and therefore no changeover production constraints are included in the aggregate formulation.

$$\sum_t T_{ptk}^l \leq |T| \cdot M_{tot}^l \cdot H_t \cdot V_p^l \quad \forall\, l, k, p \in \Pi_k^l \tag{3.7}$$

$$\sum_p T_{ptk}^l \leq H_t \cdot (h\widetilde{A}_t^l + n\widetilde{A}_t^l) \quad \forall\, l, k, t \tag{3.8}$$

Inventory Constraints

The inventory constraints resemble the ones in the detailed model with the only difference being the second term on the right hand side of the equation. In the aggregate model, the overall amount of every product manufactured at each production site is calculated without considering every suite individually.

$$I_{ptk} = I_{p,t-1,k} + (1 - L_p) \sum_{l \in \Psi_{pk}} r_p \cdot T_{ptk}^l - \sum_{s \in \Theta_{pk}} S_{ptk}^s - W_{ptk} \quad \forall\, k, p \in \Delta_k, t \tag{3.9}$$

3.2.2 Aggregate objective function

The derivation of the objective function is the same as in the detailed mathematical model formulation with the main difference being that now the operational cost term, capital cost term and the depreciation term are mathematically expressed using integer suite availability and investment variables instead of binary ones. Furthermore, no scale-up and qualifications runs cost terms are included in the aggregate objective function. Finally, the tax constraints are reformulated accordingly.

The operating costs calculation is now based on the integer number of manufacturing suites and the time available for production:

$$OC_{tk}^{l} = \eta_h^l \cdot h\widetilde{A}_t^l + \eta_n^l \cdot n\widetilde{A}_t^l + \sum_{p \in \Pi_k^l} \xi_p^l \cdot r_p \cdot T_{ptk}^l \quad \forall\ t,k,l \tag{3.10}$$

The depreciation term is now a function of the integer number of invested suites:

$$DC_t^l = \sum_{\theta = t-\delta_h-\phi^l+1}^{t-\delta_h} \frac{\omega_h^l}{\phi^l} \cdot h\widetilde{E}_\theta^l + \sum_{\theta = t-\delta_n-\phi^l+1}^{t-\delta_n} \frac{\omega_n^l}{\phi^l} \cdot n\widetilde{E}_\theta^l \quad \forall\ t,l \tag{3.11}$$

The capital investment cost is also a function of the integer number of invested suites:

$$CI_t^l = \omega_h^l \cdot h\widetilde{E}_t^l + \omega_n^l \cdot n\widetilde{E}_t^l \quad \forall\ t,l \tag{3.12}$$

The objective function before taxes does not include the scale-up and qualifications runs cost terms:

$$\Phi^B = \sum_k \pi_k \sum_t \varepsilon_t \left(\sum_s (SR_{tk}^s - MC_{tk}^s) - \sum_l (CI_t^l + OC_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP} \right) \tag{3.13}$$

Furthermore, the taxes paid by the production sites and the *IP-Centre* do not include the scale-up and qualifications runs cost terms:

$$T^L = \sum_k \pi_k \sum_t \varepsilon_t \sum_l \psi_t^l \rho^l (DC_t^l + OC_{tk}^l) \tag{3.14}$$

$$Tax_{tk}^{IP} \geq \psi_t^{IP} \left(\sum_s (1-\rho^s) SR_{tk}^s - \sum_l (1+\rho^l)(DC_t^l + OC_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP} \right) \quad \forall\ t,k \tag{3.15}$$

3.2.3 Summary of aggregate model

In conclusion, the entire aggregate model formulation described in this section can be summarised as follows.

[Problem A]

max $\Phi^A = \Phi^B - T^L - T^S - T^{IP}$

Subject to:

Product Existence Constraints

$$V_p^l \leq U_p \qquad \forall\, l, p \in \Gamma^l$$

Suite Availability and Investment Constraints

$$h\widetilde{A}_t^l = h\widetilde{A}_{t-1}^l + h\widetilde{E}_{t-\delta_h}^l \qquad \forall\ l, t$$

$$n\widetilde{A}_t^l = n\widetilde{A}_{t-1}^l + n\widetilde{E}_{t-\delta_n}^l \qquad \forall\ l, t$$

$$\sum_t h\widetilde{E}_t^l \leq \frac{M_{tot}^l}{M} - h\widetilde{A}_0^l \qquad \forall\ l$$

$$\sum_t n\widetilde{E}_t^l \leq M_{tot}^l - \frac{M_{tot}^l}{M} - n\widetilde{A}_0^l \qquad \forall\ l$$

$$n\widetilde{A}_0^l + \sum_{\theta=1}^{t} n\widetilde{E}_\theta^l \leq (M-1)\left(h\widetilde{A}_0^l + \sum_{\theta=1}^{t-(\delta_h-\delta_n)} h\widetilde{E}_\theta^l \right) \qquad \forall\ l, t$$

$$n\widetilde{A}_0^l + \sum_{\theta=1}^{t} n\widetilde{E}_\theta^l \geq (M-1)\left(h\widetilde{A}_0^l + \sum_{\theta=1}^{t-(\delta_h-\delta_n)} h\widetilde{E}_\theta^l - 1 \right) \qquad \forall\ l, t$$

Production Constraints

$$\sum_t T_{ptk}^l \leq |T| \cdot M_{tot}^l \cdot H_t \cdot V_p^l \qquad \forall\ l, k, p \in \Pi_k^l$$

$$\sum_p T_{ptk}^l \leq H_t \cdot (h\widetilde{A}_t^l + n\widetilde{A}_t^l) \quad \forall\, l, k, t$$

Inventory Constraints

$$I_{ptk} = I_{p,t-1,k} + (1 - L_p) \sum_{l \in \Psi_{pk}} r_p \cdot T_{ptk}^l - \sum_{s \in \Theta_{pk}} S_{ptk}^s - W_{ptk} \qquad \forall\, k, p \in \Delta_k, t$$

Product Lifetime Constraints

$$I_{ptk} \leq \sum_{s \in \Theta_{pk}} \sum_{\theta=t+1}^{t+\zeta_p} S_{p\theta k}^s \qquad \forall\, k, p \in \Delta_k, t$$

Sales Constraints

$$S_{ptk}^s \leq D_{ptk}^s \qquad \forall\ k, p, s \in \Theta_{pk}, t$$

Sales Revenue

$$SR_{tk}^{s} = \sum_{p \in \Lambda_k^s} v_p S_{ptk}^{s} \qquad \forall\, t,k,s$$

Marketing Costs

$$MC_{tk}^{s} = \sum_{p \in \Lambda_k^s} \mu_p v_p S_{ptk}^{s} \qquad \forall\, t,k,s$$

Royalty Costs

$$RC_{tk}^{IP} = \sum_{s \in \Theta_{pk}} \sum_{p \in \Lambda_k^s} \lambda_p v_p S_{ptk}^{s} \qquad \forall\, t,k$$

R&D Costs

$$RDC_t^{IP} = \sum_p \sigma_{pt} \cdot U_p \qquad \forall\, t$$

Operating Costs

$$OC_{tk}^{l} = \eta_h^l \cdot h\widetilde{A}_t^l + \eta_n^l \cdot n\widetilde{A}_t^l + \sum_{p \in \Pi_k^l} \xi_p^l \cdot r_p \cdot T_{ptk}^{l} \qquad \forall\; t,k,l$$

Depreciation

$$DC_t^l = \sum_{\theta = t-\delta_h-\phi^l+1}^{t-\delta_h} \frac{\omega_h^l}{\phi^l} \cdot h\widetilde{E}_\theta^l + \sum_{\theta = t-\delta_n-\phi^l+1}^{t-\delta_n} \frac{\omega_n^l}{\phi^l} \cdot n\widetilde{E}_\theta^l \qquad \forall\; t,l$$

Capital Investment Costs

$$CI_t^l = \omega_h^l \cdot h\widetilde{E}_t^l + \omega_n^l \cdot n\widetilde{E}_t^l \quad \forall\; t,l$$

Objective function before taxes

$$\Phi^B = \sum_k \pi_k \sum_t \varepsilon_t \left(\sum_s (SR_{tk}^s - MC_{tk}^s) - \sum_l (CI_t^l + OC_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP} \right)$$

Production sites taxes

$$T^L = \sum_k \pi_k \sum_t \varepsilon_t \sum_l \psi_t^l \rho^l (DC_t^l + OC_{tk}^l)$$

Sales regions taxes

$$T^S = \sum_k \pi_k \sum_t \varepsilon_t \sum_s \psi_t^s (\rho^s SR_{tk}^s - MC_{tk}^s)$$

IP-Centre taxes

$$T^{IP} = \sum_k \pi_k \sum_t \varepsilon_t \cdot Tax_{tk}^{IP}$$

$$Tax_{tk}^{IP} \geq \psi_t^{IP}\left(\sum_s (1-\rho^s)SR_{tk}^s - \sum_l (1+\rho^l)(DC_t^l + OC_{tk}^l) - RC_{tk}^{IP} - RDC_t^{IP}\right) \quad \forall\, t,k$$

$$Tax_{tk}^{IP} \geq 0 \quad \forall\, t,k$$

3.3 Hierarchical solution algorithms

3.3.1 Algorithm H1

The proposed hierarchical algorithm H1 comprises the following steps:

[Algorithm H1]

```
Step 1.    Solve aggregate MILP model (Problem A) and fix
           strategic decision variables.
Step 2.    Solve original detailed MILP model (Problem D) in
           the reduced variable space to determine operational
           decision variables.
```

In the first step, the aforementioned suite-aggregate model (Problem A) is solved in order to determine the strategic here-and-now decisions: U_p, V_p^l, hE_t^l, nE_t^l, hA_t^l, nA_t^l. The derived strategic decision variables are then fixed. The levels of the binary variables U_p and V_p^l are fixed to their current values as determined from the aggregate model. On the other hand, the levels of the integer variables hE_t^l, nE_t^l, hA_t^l, nA_t^l can not be fixed in a straightforward way, since the suite investment and availability decision variables in the detailed model are modelled as binary variables (E_{it}^l and A_{it}^l) representing each suite individually. Therefore, the integer variables should first undergo an intermediate translation-fixing step* before their original values are fed into the detailed model.

---

* See Appendix A

In the second step of Algorithm H1, the original detailed model is solved in the reduced variable space in order to determine the optimal levels for the operational wait-and-see decisions variables. Furthermore, there is no need to consider any constraints related to product existence, suite availability and investment decisions since the levels of the variables U_p, V_p^l, E_{it}^l, A_{it}^l involved in constraints (2.2)-(2.5) are now treated as given parameters. Therefore, the goal for the reduced detailed MILP model is to maximise the eNPV subject to constraints (2.6), (2.7), (2.8a), (2.9)-(2.11), (2.12a), (2.13a), (2.14)-(2.27), (2.29)-(2.35). Additionally, the sets involving products and manufacturing suites (Δ_k, Π_k^l, Λ_k^s, Ψ_{pk}, Θ_{pk}, I^l) are now updated based on the aggregate solution information.

Summarising all the previously described manipulations, the detailed model is now reduced down to a model involving only a portion of the variables defined in the original problem. The discrete variables domain is limited to the scale-up, qualification and product-to-suite allocation binary variables ($\hat{Z}_{ptk}^l$, Z_{iptk}^l, X_{ptk}^l, Y_{iptk}^l) while the rest of the variables lie in the continuous space (production time, T_{iptk}^l, inventory levels, I_{ptk}, wasted amounts, W_{ptk}, and sales levels, S_{ptk}^s). Overall, due to its reduced size, the detailed mathematical model is now much easier to tackle computationally than the original one.

3.3.2 Algorithm H2

Instead of solving the reduced model as a single MILP, the final step of the proposed Algorithm H1 can further be decoupled by solving a series of single-scenario MILPs. Each scenario present in the reduced model is a fairly small deterministic problem with its own characteristic customer demand pattern. In the face of this unique customer demand, every scenario can be solved as a separate MILP model with the objective function being the maximisation of the scenario NPV subject to constraints (2.6), (2.7), (2.8a), (2.9)-(2.11), (2.12a), (2.13a), (2.14)-(2.27), (2.29)-(2.35). Recall that the product portfolio and multi-site investment strategy are both scenario-independent decision variables already determined and fixed from the previous step of the algorithm. The revised hierarchical algorithm comprises the following steps:

[Algorithm H2]

Step 1. Solve aggregate MILP model (Problem A) and fix strategic decision variables.

Step 2. For every scenario k:

i. Update dynamic sets (Δ_k, Π_k^l, Λ_k^s, Ψ_{pk}, Θ_{pk}, I^l).

ii. Solve original detailed MILP model (Problem D) in the reduced variable space of scenario k.

It is worth mentioning that the scenario-decoupling structure of the proposed hierarchical Algorithm H2 can further be exploited by employing a parallel computing solution strategy leading to a potentially unmatched computational efficiency.

3.3.3 Algorithm C

Rotstein *et al.* (1999) presented the cut-off hierarchical algorithm (Algorithm C). Their proposed algorithm was originally developed to address the problem of single-site capacity planning under uncertainty formulated as a two-stage, multi-scenario MILP model. For the sake of completeness, Algorithm C is briefly described below so as to facilitate its comparison with Algorithms H1 and H2.

Algorithm C comprises the following steps:

[Algorithm C]

Step 1. Solve detailed MILP model (Problem D) in the reduced scenario space and fix strategic decision variables.

Step 2. For every scenario k:

i. Update dynamic sets (Δ_k, Π_k^l, Λ_k^s, Ψ_{pk}, Θ_{pk}, I^l).

ii. Solve original detailed MILP model (Problem D) in the reduced variable space of scenario k.

According to Algorithm C, the scenarios are first prioritised based on their probability of occurrence (the scenario with highest probability comes first). Then, starting from the top-probability scenario, scenarios are included until their cumulative probability exceeds a predetermined value (cut-off probability). In the first step (Step 1), the algorithm solves the detailed mathematical model in the reduced scenario space of the included scenarios, to determine the strategic decisions. Then in the second step (Step 2), each original scenario is solved separately with the strategic decision variables fixed from the previous step, similarly to Algorithm H2.

The philosophy of Algorithm C is based on the assumption that since the scenarios with higher probabilities have a greater contribution in the objective function, we can therefore solely rely on them when trying to determine the strategic decision variables. The more scenarios included in the first step (higher cut-off probability), the more accurate the final solution will be. However, the problem size increases with the number of scenarios and that in turn results in an increased computational effort to solve the problem. Rotstein *et al.* (1999), proposed an iterative version of Algorithm C by employing a slightly higher cut-off probability in every iteration. The algorithm then terminates when a convergence criterion is satisfied. According to the authors' experience (Rotstein *et al.*, 1999), a value of 0.5 for the cut-off probability gives satisfactory results.

Overall, the proposed hierarchical algorithms presented in this section are simple yet powerful solution approaches, able to accommodate the combinatorial nature of the multi-site capacity planning problem, alleviate the associated computational burden and yield near-optimal solutions within reasonable computational time, as it is demonstrated by the illustrative examples revisited in the following section.

3.4 Illustrative examples revisited

The five illustrative examples presented in Chapter 2 are now revisited and solved using the proposed hierarchical algorithm H2. All five example problems were implemented in GAMS (Brooke *et al.*, 1998) using the XPRESS-MP (Dash Associates, 1999) MILP solver with a 5% margin of optimality while all runs were performed on an IBM RS/6000 workstation.

Concerning the product portfolio, the obtained results are summarised in Table 3.1. The enterprise-wide pharmaceutical supply chain determined by the optimisation algorithm is common to all five problem instances and results in the business network illustrated in Figure 3.1, same as the one obtained using the simultaneous approach.

Table 3.1: *Product portfolio results*

Problem	Selected Products	No of Selected Products / Total Candidate Products
3PROD	P1, P3	2 / 3
4PROD	P1, P3, P4	3 / 4
5PROD	P1, P3, P4, P5	4 / 5
6PROD	P1, P3, P4, P5	4 / 6
7PROD	P1, P3, P4, P5, P7	5 / 7

The investment decision calendars for all five examples are illustrated in Figure 3.2. Similar to the simultaneous approach, investment decisions for additional manufacturing suites are taken in the early time periods while the clinical trials are still on going. Again the number of invested suites increases with the number of final products included in the portfolio, in order to satisfy the additional customer demand, resulting in a total of five invested suites for the 7PROD example as opposed to only three invested suites for the 3PROD example (see Figure 3.2).

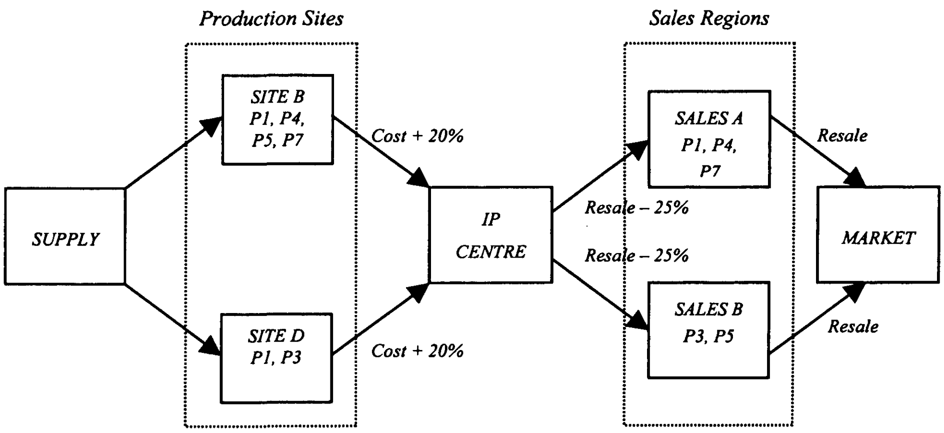

Figure 3.1: *Optimal business network*

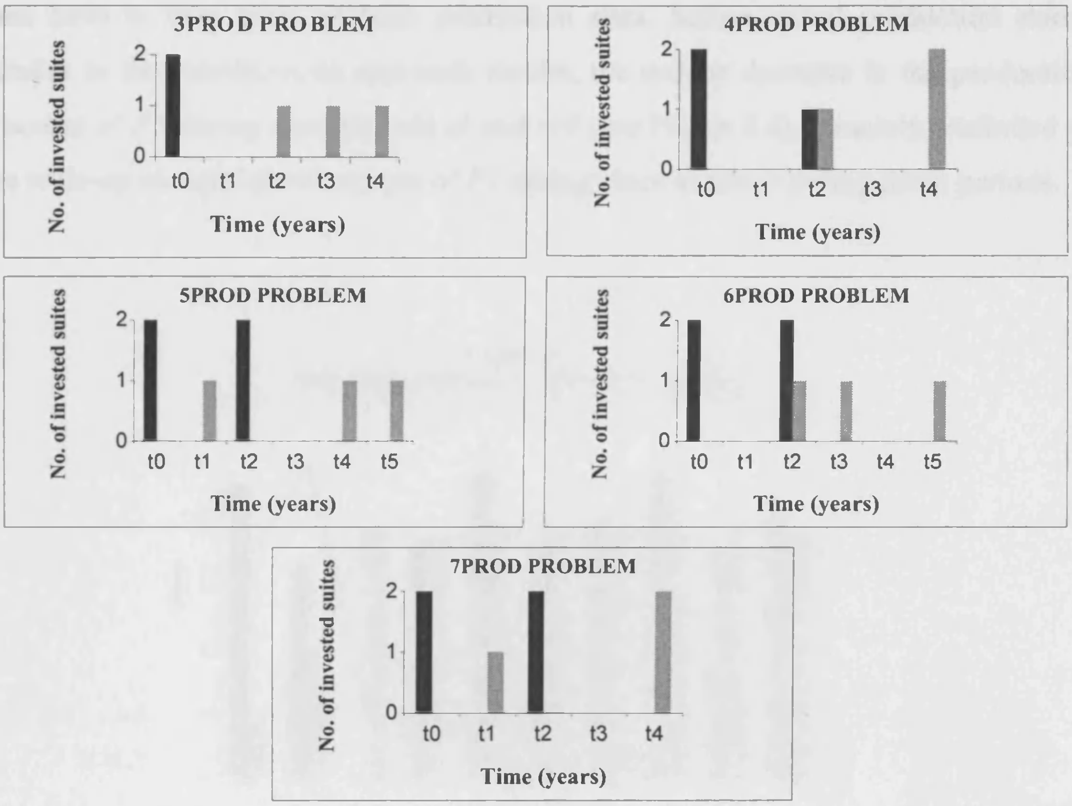

Figure 3.2: *Investment decisions calendars (Site B: black, Site D: grey)*

The aforementioned strategic decisions concerning product portfolio and multi-site investment plans are determined and fixed from Step 1 of hierarchical Algorithms H2 while Step 2 determines all the operational variables (detailed production plans, inventory and sales profiles). Such characteristic profiles for the successful products *P1, P3* and *P7* in scenario 11 of problem 7PROD are shown in Figures 3.3-3.5. It can be seen that for products *P1* and *P3*, customer demand is fully satisfied in nearly all time periods, while for product *P7*, the amounts produced can not meet the customer demand in the late four periods. It is very interesting to notice how the manufacturing capacity resources are shared between products *P1* and *P7* at site *B*. During time periods *t4* and *t5*, large amounts of product *P1* are being produced so as to satisfy the anticipating customer demand in later periods. However, in time period *t6* no production is taking place for product *P1*, while all the existing and invested manufacturing suites are used entirely for the production of *P7* that is going to be sold in the subsequent time periods.

The scale-up and qualification timings are given in Table 3.2. Notice that product *P1* is manufactured at production sites *B* and *D*. Therefore, the scale-up and qualification runs have to take place at both production sites, before actual production starts. Similar to the simultaneous approach results, the sudden decrease in the production amounts of *P3* during time periods *t8* and *t10* (see Figure 3.4) is mainly attributed to the scale-up and qualification runs of *P1* taking place at site *D* during those periods.

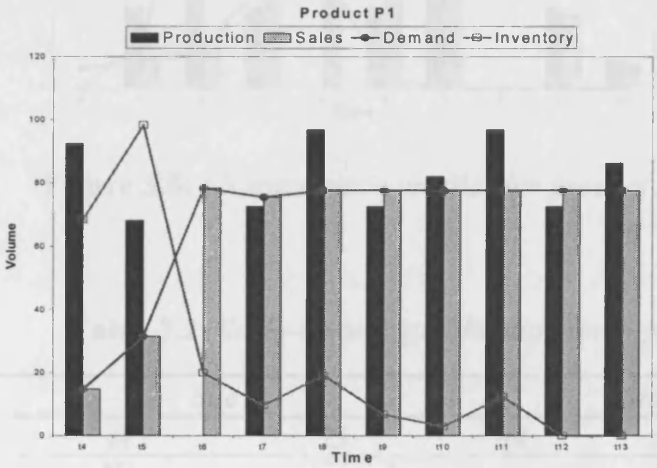

Figure 3.3: *Characteristic profiles for product P1*

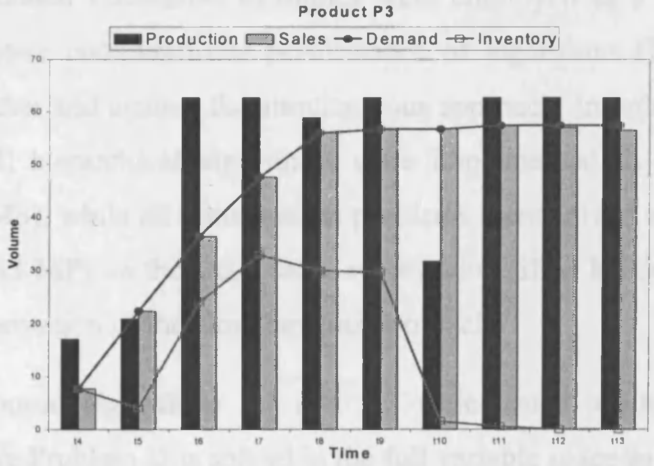

Figure 3.4: *Characteristic profiles for product P3*

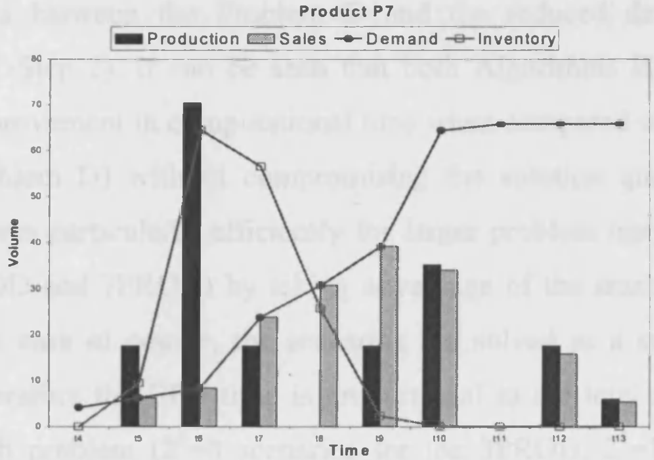

Figure 3.5: *Characteristic profiles for product P7*

Table 3.2: *Scale-up and qualification timings*

Product	*Site B*		*Site D*		
	t4	*t5*	*t4*	*t8*	*t10*
P1	X+			X	+
P3			X+		
P7		X+			

X: Scale-up, +: Qualification runs

3.4.1 Comparison of solution approaches

The aforementioned illustrative examples were employed as a test bed in order to assess the relative computational performance of algorithms H1, H2 and C tested against each other and against the simultaneous approach. In order to perform a valid comparison, all hierarchical algorithms were implemented in the same modelling platform (GAMS), while all optimisation problems were solved using the same MILP solver (XPRESS-MP) on the exact same workstation (IBM RS/6000) as the one used for the implementation of the simultaneous approach.

First, the proposed algorithms H1 and H2 were tested against the simultaneous approach, where Problem D is solved in the full variable space as a single-level MILP problem. The model statistics for the five problems can be found in Tables 3.3-3.7.

As expected, the obtained objective function value for Problem A is always higher than the one obtained from Problem D, since it serves as an upper bound to the optimal solution. It is worth mentioning the difference in both the number of variables and constraints between the Problem D and the reduced detailed MILP model (Algorithm H1-Step 2). It can be seen that both Algorithms H1 and H2 provide a significant improvement in computational time when compared with the simultaneous approach (Problem D) without compromising the solution quality. Algorithm H2 seems to perform particularly efficiently for larger problem instances (see examples 5PROD, 6PROD and 7PROD) by taking advantage of the smaller MILPs solved in Step 2. In that case of course, the scenarios are solved as a series of independent MILPs and therefore the CPU time is proportional to the total number of scenarios present in each problem (2^3=8 scenarios for the 3PROD, 2^4=16 scenarios for the 4PROD, 2^5=32 scenarios for the 5PROD, 2^6=64 scenarios for the 6PROD, 2^7=128 scenarios for the 7PROD). The CPU time in parentheses denotes the time spent for solving the most computationally demanding scenario in Step 2 of Algorithm H2.

Table 3.3: *Model statistics for the 3PROD example*

	Problem D	Problem A♦	Algorithm H1 (Step 2)	Algorithm H2 (Step 2)
Constraints	8524	890	3585	-
Continuous Var.	3189	760	1285	-
Discrete Var.	2320	73	880	-
Obj. Fun. (rmu)	223	250	221	221
CPU (sec)	308	4	1	2 (0.5)

Table 3.4: *Model statistics for the 4PROD example*

	Problem D	Problem A	Algorithm H1 (Step 2)	Algorithm H2 (Step 2)
Constraints	18033	1863	9349	-
Continuous Var.	6933	1744	3655	-
Discrete Var.	5122	75	2560	-
Obj. Fun. (rmu)	276	305	268	269
CPU (sec)	4526	9	11	12 (2)

Table 3.5: *Model statistics for the 5PROD example*

	Problem D	Problem A	Algorithm H1 (Step 2)	Algorithm H2 (Step 2)
Constraints	38434	4056	24021	-
Continuous Var.	14981	4032	9569	-
Discrete Var.	11524	77	7360	-
Obj. Fun. (rmu)	281*	326	281	282
CPU (sec)	10800	10	82	47 (4)

*Integer solution found within 12% margin of optimality.

♦ Step 1 of both Algorithms H1 and H2.

Table 3.6: *Model statistics for the 6PROD example*

	Problem D	Problem A	Algorithm H1 (Step 2)	Algorithm H2 (Step 2)
Constraints	86580	8970	48037	-
Continuous Var.	33477	9568	19263	-
Discrete Var.	29127	80	14720	-
Obj. Fun. (rmu)	129*	324	283	284
CPU (sec)	10800	67	281	94 (4)

*Integer solution found within 71% margin of optimality.

Table 3.7: *Model statistics for the 7PROD example*

	Problem D	Problem A	Algorithm H1 (Step 2)	Algorithm H2 (Step 2)
Constraints	183829	19755	107141	-
Continuous Var.	71429	21600	43135	-
Discrete Var.	64329	82	35840	-
Obj. Fun. (rmu)	No Int. Sol.*	369	309**	317
CPU (sec)	10800	619	6645	291 (10)

*LP relaxation not finished.

** Integer solution found within 6% margin of optimality.

The proposed algorithm H2 was also tested against the cut-off hierarchical algorithm (Algorithm C). In order to perform a valid comparison between them, the five problem instances previously described were solved using two different values for the cut-off probability, 0.5 and 0.7 (Algorithms C_0.5 and C_0.7) respectively. The comparative results between Algorithm H2, Algorithms C_0.5 and C_0.7 and Problem D are shown in Tables 3.8 and 3.9.

Table 3.8: *Objective function values (in rmu)*

Problem	Solution Approach			
	Algorithm H2	Algorithm C_0.5	Algorithm C_0.7	Problem D
3PROD	221	222	221	223
4PROD	269	272	273	276
5PROD	282	222	277	281*
6PROD	284	219	278	129**
7PROD	317	313♦	316♦♦	No Int. Sol.

* Integer solution found within 12% margin of optimality.

** Integer solution found within 71% margin of optimality.

♦ Integer solution in Step 1 found within 13% margin of optimality.

♦♦ Integer solution in Step 1 found within 32% margin of optimality.

Note that the reported CPU time for Algorithm H2 corresponds to the combined CPU time of the aggregate MILP model plus the summation of the CPU time over all scenario MILPs, while the CPU time for both Algorithms C_0.5 and C_0.7 corresponds to the combined CPU time for Step 1 and 2. A computational limit of 3 hours (10800 seconds) was used for all runs.

Table 3.9: *CPU times (in seconds)*

Problem	Solution Approach			
	Algorithm H2	Algorithm C_0.5	Algorithm C_0.7	Problem D
3PROD	4 + 2 (0.5)	29 + 2 (1)	49 + 2 (0.5)	308
4PROD	9 + 12 (2)	91 + 10 (2)	110 + 9 (2)	4526
5PROD	10 + 47 (4)	99 + 7 (0.5)	1740 + 18 (2)	10800
6PROD	67 + 94 (4)	358 + 13 (0.5)	8462 + 37 (2)	10800
7PROD	619 + 291 (10)	10800 + 129 (3)	10800 + 226 (6)	10800

According to the results, the performance of Algorithm C is dramatically influenced by the value of the employed cut-off probability. In case where the problems are solved in half of the original scenario space (Algorithm C_0.5), the overall CPU time needed is reduced when compared with the CPU time needed for solving Problem D. However, the reduced CPU time comes at the expense of a relatively poor solution quality. By employing an increased value for the cut-off probability (Algorithm C_0.7), we manage to obtain an improved solution quality. However, in this case the computational effort is dramatically increased as it is clearly shown by the CPU time for all problems.

For example in the 6PROD problem, using Algorithm C_0.5 results in a reduced MILP problem in Step 1 involving only 13 out of the 64 original scenarios, which explains the 8-fold decrease in terms of number of discrete variables when compared with the original Problem D (3427 vs. 29127). However, in this case the derived values for the strategic decision variables result in a poor solution quality for the overall problem when compared with the one obtained using the Algorithm H2 (219 rmu vs. 284 rmu). On the other hand, when the cut-off probability is increased to the value of 0.7, more scenarios are included in Step 1 of the Algorithm C_0.7 (24 out of 64) and therefore we end up with an improved, near-optimal solution (278 rmu).

Nevertheless, the increased number of scenarios is inevitably accompanied by an increased number of discrete variables (8327) that further magnifies the combinatorial nature of the problem resulting in a prohibitively long CPU time for Step 1 of the Algorithm C_0.7 (8462 seconds).

3.5 Concluding remarks

In this chapter, a hierarchical solution approach for multi-site capacity planning under uncertainty was presented. Two hierarchical algorithms (Algorithm H1 and H2) were proposed for the solution of the resulting large-scale MILP problem based on the decoupling of the decision-making levels (strategic and operational) identified in our problem. Without compromising the solution quality, significant savings in computational effort were achieved by employing the proposed algorithms in five illustrative examples.

In terms of CPU time, the hierarchical approach outperformed the simultaneous approach by approximately two orders of magnitude. Furthermore, Algorithm H2 performed significantly better than the previously developed cut-off algorithm C (Rotstein *et al.*, 1999). Finally, we should emphasize that the proposed hierarchical methodology features a highly parallel solution structure that can be further exploited for increased computational efficiency.

PART II

BUSINESS OPTIMISATION FOR CUSTOMER DEMAND MANAGEMENT

Chapter 4

Customer demand forecasting via support vector regression analysis

4.1 Introduction and literature survey

Recent studies have clearly identified customer demand as the ultimate driver of business management in process industries ranging from the traditional oil and gas industry (Lasschuit and Thijssen, 2004) and the paper-converting industry (Roslof *et al.*, 2002) to the high-risk agrochemical and pharmaceutical industries (Maravelias and Grossmann, 2001), (Papageorgiou *et al.*, 2001). Given the importance of customer demand, one can easily realise the potential benefits of an accurate customer demand forecasting tool in process industries.

Forecasting has gained widespread acceptance as an integral part of business planning and decision-making in areas such as sales planning, marketing research, pricing, production planning and scheduling (Makridakis and Wheelwright, 1978). From a historical perspective, exponential smoothing methods and decomposition methods were the first forecasting approaches developed back in the mid-1950s. During 1960s, as computer power became more available and cheaper, more sophisticated forecasting methods appeared. Box-Jenckins (Box and Jenckins, 1969) methodology gave rise to the autoregressive integrated moving average (ARIMA) models. Later on during 1970s and 1980s, ever more sophisticated forecasting approaches were developed including econometric methods and Bayesian methods (Makridakis and Wheelwright, 1982). The consolidation and improvement of the aforementioned approaches provided forecasting tools of ever increasing complexity, before artificial neural networks (ANN) emerged as a novel and promising forecasting approach in the 1990s taking full advantage of the number-crunching capabilities of super-computers (Foster *et al.*, 1992).

However, it is very interesting to notice that the increasing complexity of forecasting approaches is not always accompanied by the desired increased predictive accuracy as pointed out by Makridakis and Hibon (Makridakis and Hibon, 1979). Their remark is consistent with recent criticism towards the excessive use of parameters in unnecessarily complex artificial neural networks applications in chemical engineering problems (Bhat and McAvoy, 1992). There exists a clearly identified need for a new generation of forecasting tools that share all the benefits of artificial neural networks while at the same time maintain an underlying formulation as simple as possible.

Support vector machines (SVM) constitute such a novel learning paradigm that provides an inherently simple formulation and yet offers the promise of increased flexibility. The growing popularity of the SVM is mainly attributed to the solid theoretical foundations and the practical applications in a broad range of the scientific spectrum. Based on the statistical learning theory recently developed by Vapnik (Vapnik, 1998), support vector machines applications have been proposed for a number of classification and regression problems ranging from discrete manufacturing (Prakasvudhisarn *et al.*, 2003) to bioinformatics (Myasnikova *et al.*, 2002). Agrawal *et al.* (2003) portrays support vector machines as a useful tool for process engineering applications while Kulkarni *et al.* (2004) and Chiang *et al.* (2004) provide support vector classification applications in process engineering problems.

However, in order to present a balanced perspective, we must also mention that so far the applicability of support vector regression (SVR) is hindered by the notorious problem of parameter selection. Although, the number of parameters to be tuned is not prohibitively large, parameter values affect significantly the predictive capabilities. Exhaustive grid-search (Chang and Lin, 2001) and heuristic-based rules for parameter selection (Cherkassky and Ma, 2004) are currently used for SVR while further research in this area is in progress. As a typical case of any emerging forecasting research field, those heuristic rules can be regarded an initial step towards the identification of a more formal way for parameter selection in the near future.

Despite the parameter selection problem, SVR still enjoys numerous advantages when compared with other forecasting methodologies. Similar to ANN, SVR employs an adaptive basis regression function without postulating any pre-determined family of basis functions (e.g. high-order polynomial parametric regression). Support vectors provide a completely new way of parameterisation of the regression function (Cherkassky and Mulier, 1998) leading to increased flexibility while avoiding the trap of overcomplexity. Unlike ANN, SVR employs only a handful of parameters thus justifying "Occam's razor"* principle in the most illustrative way. Furthermore, its unique mathematical formulation guarantees a computationally tractable global optimal solution. This is a very attractive feature for applications in the process industries where repeatability and consistency are of paramount importance. Moreover, SVR requires no a priori fundamental understanding of the process being studied since it is a training data-driven methodology and therefore is very well-suited for process industries forecasting applications where historical data is abundant.

Overall, support vector regression is identified as a novel emerging forecasting technique and the aim of this chapter is to validate the applicability of SVR analysis for forecasting customer demand in process industries. The rest of the chapter is organised as follows. In the next section, the main characteristics of the customer demand forecasting problem are described. Section 4.3 provides a detailed mathematical description of support vector regression. A three-step algorithm is then proposed in section 4.4 and is validated through a number of illustrative examples presented in section 4.5. Finally, some conclusions are drawn in section 4.6.

* "One should not increase, beyond what is necessary, the number of entities required to explain anything", Occam's razor principle (also known as principle of parsimony) attributed to mediaeval English philosopher William of Occam (c. 1280-1349) who stressed the Aristotelian principle that entities must not be multiplied beyond what is necessary.

4.2 Problem description

We assume that customer demand at time period t (y_t) depends on a number of Z independent variables (x_{1t}, x_{2t}, ..., x_{Zt}) that are called attributes and form the associated input vector $\mathbf{x}_t$*. Therefore, the dependant variable y_t is a function of the input vector $\mathbf{x}_t$ which in turns contains the multiple independent variables as shown in the following equations.

$$\mathbf{x}_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ . \\ . \\ . \\ . \\ . \\ x_{zt} \end{pmatrix} \tag{4.1}$$

$$y_t = F(\mathbf{x}_t) \tag{4.2}$$

The problem of customer demand forecasting via support vector regression analysis can be formally stated as follows. *Given* a set of training data (time series) in the form of N training points ($\mathbf{x}_1$, y_1), ($\mathbf{x}_2$, y_2), ...,($\mathbf{x}_N$, y_N) where $\mathbf{x}_t$ is the input vector and y_t is the associated customer demand for every $\mathbf{x}_t$, as well as a forecasting horizon of size M, we would like to *determine* the output values $\hat{y}_{N+1}, \hat{y}_{N+2}, ..., \hat{y}_{N+M}$. The mean absolute percent error (MAPE) is a commonly used forecasting error metric for quantifying and assessing the accuracy of the predicted output values. Mathematically, it is given by the following formula (Makridakis and Wheelwright, 1978):

$$MAPE = 100 \cdot \frac{\sum_{t=N+1}^{N+M} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{M} \tag{4.3}$$

* Symbols in bold fonts represent vectors

Where y_t is the actual customer demand and $\hat{y}_i$ is the predicted demand at time period t. Clearly, accurate predictions would result in low MAPE values, which implies small absolute deviations between the actual and predicted output values.

Based on the available training points (x_t, y_t), the ultimate goal of support vector regression analysis is to extract as much information as possible from the historical data so as comprehend the complicating relationships between customer demand and all the different attributes before identifying an appropriate regression function F able to accurately predict future unknown output values from a given input vector of attributes.

In our time series forecasting problem, customer demand attributes can be classified into a number of different main categories such as:

- *Past Demand Attributes:* those attributes represent customer demand for a predetermined number of previous time periods. Employing past demand attributes can be extremely helpful to relate present customer demand with historical customer demand values. According to our experience, past demand attributes prove to be very efficient when dealing with periodical customer demand patterns.
- *Calendar attributes:* those attributes illustrate a specific characteristic of the time period under investigation and are usually treated as binary parameters representing true or false statements with one and zero values respectively. For example, calendar attributes could be employed to represent the day of the week, the month or week of the year etc. Moreover, calendar attributes could also be used to identify customer demand patterns on national holidays or weekends. Therefore, calendar attributes prove to be a very critical source of information when trying to predict time-sensitive output values such as electricity load demand or seasonality-dominated customer demand patterns such as swimware sales.

Although there may exist more categories of attributes other than the two mentioned above, we restrict ourselves to only those two main categories since any other attribute is viewed as problem specific. For example, in an ice-cream demand forecasting case, it would be very beneficial to incorporate a temperature attribute or any other weather attribute that can reflect the dependancy between ice-cream consumption and environmental conditions.

Based on our ability to know their future values or not, attributes can further be classified into deterministic and stochastic attributes. Deterministic attributes are those whose future values are known (or can be predicted with very high-accuracy). Calendar attributes fall under the deterministic category since anyone can accurately predict the date of next weekend or next Monday. However, there exist a number of attributes that affect output values considerably, whose values unfortunately can not be predicted or accurately estimated. Such attributes are called stochastic and include for example future temperature profiles, oil prices and dollar-to-pound exchange rates.

In our proposed methodology, customer demand forecasting is based entirely on past demand attributes. Past demand attributes belong to a very special case of attributes that can be regarded as semi-deterministic attributes explained in detail in section 4.4. According to equations (4.1) and (4.2), knowing the attributes of the input vectors is only the first step towards a valid prediction. What is foremost needed is to establish a solid relationship between the input vector attributes and the target value. In our case, customer demand and past demand attributes are related through a support vector regression function *F*. Such a regression function is needed in order to translate past demand attributes into accucate demand forecasts. In the next section, the derivation of the support vector regression function *F* is explained in full detail.

4.3 Support vector regression

In this section, we describe support vector regression (SVR) based on the statistical learning theory developed by Vapnik (Vapnik, 1998). Given training data

$$\{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_N, y_N)\} \subset R^Z \times R \tag{4.4}$$

where x_t is the input vector at time period t and y_t is the associated customer demand for every x_t, the goal is to find regression function function $F(x_t)$:

$$F(x_t) = w^T x_t + \beta \qquad w, x_t \in R^Z, \beta \in R \tag{4.5}$$

The main insight of the statistical learning theory is that in order to obtain a regression function with high generalisation behaviour, one needs to control both model complexity and training error tolerance (Chalimourda *et al.*, 2004). Model complexity is illustrated by the flatness of the function F which in turns means small **w** values. One way to ensure this is to minimise the Euclidean norm $\|w\|$. On the other hand, the regression function should not be too flat but rather complicated enough so as to fit closely with the demand training points. In order to control training error tolerance, the ε-insentitive loss-function $|\xi|_\varepsilon$ is employed:

$$|\xi|_\varepsilon = max(0,|F(x_t) - y_t| - \varepsilon) \quad (4.6)$$

The ε-insensitive loss function ensures that errors less than ε are not taken into consideration. However, we penalise any deviations larger than ε, meaning all training points that lie outside the ε-insensitive tube as shown graphically in the following figure.

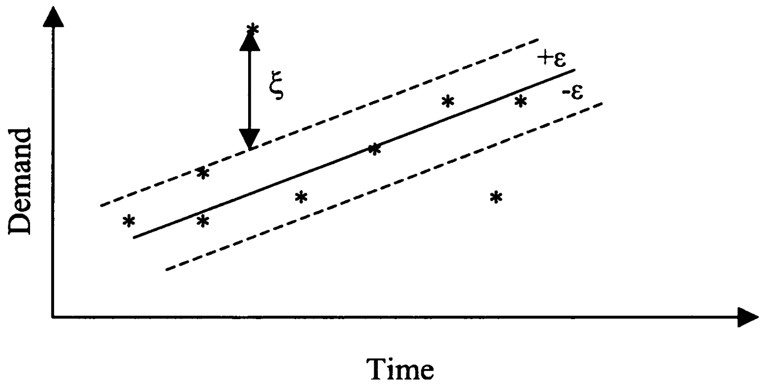

Figure 4.1: *Graphical representation of support vector regression (* depicts training points)*

Overall, support vector regression analysis takes the form of the following constrained optimisation problem (Vapnik, 1998):

[Problem ε-SVR]

$$\min_{w,\beta,\xi_t,\xi_t^*} \quad \frac{1}{2}\|w\|^2 + C\cdot\sum_{t=1}^{N}(\xi_t + \xi_t^*)$$

Subject to:

$$y_t - (w^T x_t + \beta) \le \varepsilon + \xi_t \quad \forall\, t = 1....N$$

$$(w^T x_t + \beta) - y_t \le \varepsilon + \xi_t^* \quad \forall\, t = 1....N$$

$$\xi_t \ge 0 \quad \forall\, t = 1....N$$

$$\xi_t^* \ge 0 \quad \forall\, t = 1....N$$

The first term in the objective function represents model complexity (flatness) while the second term represents model accuracy (error tolerance). The parameter C is a positive scalar determining the trade-off between flatness and error tolerance (regularisation parameter), while ξ_t and ξ_t^* represent the absolute deviations above and below the ε-insensitive tube.

From a mathematical point of view, the aforementioned nonlinear optimisation problem features a number of very interesting properties. Problem ε-SVR constitutes a convex NLP optimisation problem since it involves the minimisation of a quadratic function subject to a linear set of constraints, meaning that every local solution to the problem is also a global solution (Bertsekas, 1995), (Floudas, 1995). Furthermore, ε-SVR is a convex primal problem satisfying strong duality conditions. Therefore, instead of solving primal problem ε-SVR, we can obtain the exact same global minimum solution by solving its dual counterpart. Thanks to its reduced size both in terms of constraints and variables, the dual model formulation requires significantly less computational effort to solve. Without compromising the quality of the obtained solution, the dual problem formulation can also easily be extended to accommodate the general case of nonlinear regression through appropriately defined kernel functions as it is demonstrated later on in this section.

We can easily construct the Lagrangean function of the primal problem by bringing all constraints into the objective function with the use of appropriately defined Lagrange multipliers $\lambda_t, \lambda_t^*, \mu_t, \mu_t^*$ as follows:

$$L = \frac{1}{2}\|w\|^2 + C \cdot \sum_{t=1}^{N}(\xi_t + \xi_t^*) - \sum_{t=1}^{N}\lambda_t \cdot \left(\varepsilon + \xi_t - y_t + (w^T x_t + \beta)\right)$$
$$- \sum_{t=1}^{N}\lambda_t^* \cdot \left(\varepsilon + \xi_t^* + y_t - (w^T x_t + \beta)\right) - \sum_{t=1}^{N}\mu_t \cdot \xi_t - \sum_{t=1}^{N}\mu_t^* \cdot \xi_t^* \quad (4.7)$$

At the optimal point Karush-Kuhn-Tucker (KTT) conditions impose that the partial derivatives of L with respect to the primal variables (w, β, ξ_t, ξ_t^*) equal zero:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w - \sum_{t=1}^{N}(\lambda_t - \lambda_t^*) \cdot x_t = 0 \quad (4.8)$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \sum_{t=1}^{N}(\lambda_t - \lambda_t^*) = 0 \quad (4.9)$$

$$\frac{\partial L}{\partial \xi_t} = 0 \Rightarrow C - \lambda_t - \mu_t = 0 \quad \forall t = 1....N \quad (4.10)$$

$$\frac{\partial L}{\partial \xi_t^*} = 0 \Rightarrow C - \lambda_t^* - \mu_t^* = 0 \quad \forall t = 1....N \quad (4.11)$$

By substituting equations of (4.8)-(4.11) into (4.7), we obtain the dual optimisation problem:

[Problem D]

$$\max_{\lambda_t, \lambda_t^*} -\frac{1}{2}\sum_{t'=1}^{N}\sum_{t=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot (\lambda_t - \lambda_t^*) \cdot x_{t'}^T x_t - \varepsilon \sum_{t=1}^{N}(\lambda_t + \lambda_t^*) + \sum_{t=1}^{N} y_t \cdot (\lambda_t - \lambda_t^*)$$

Subject to:

$$\sum_{t=1}^{N}(\lambda_t - \lambda_t^*) = 0$$

$$0 \le \lambda_t \le C \quad \forall\, t = 1....N$$

$$0 \le \lambda_t^* \le C \quad \forall\, t = 1....N$$

The dual problem optimisation problem maximises a quadratic objective function with respect to Lagrange multipliers λ_t and λ_t^* which now play the role of dual variables. The solution of the dual problem derives the optimal vector w as well as the regression function $F(x_t)$ as follows:

$$w = \sum_{t=1}^{N} (\lambda_t - \lambda_t^*) \cdot x_t \qquad (4.12)$$

$$F(x_t) = w^T x_t + \beta \qquad (4.13)$$

Or by using equation (4.9), we obtain the following expression for the regression function:

$$F(x_t) = \sum_{t'=1}^{N} (\lambda_{t'} - \lambda_{t'}^*) \cdot x_{t'}^T x_t + \beta \qquad (4.14)$$

Parameter β can be calculated from the KKT complementarity conditions which state that the product between dual variables and constraints has to vanish as follows:

$$\lambda_t \cdot \left(\varepsilon + \xi_t - y_t + (w^T x_t + \beta)\right) = 0 \quad \forall t = 1.....N \qquad (4.15)$$

$$\lambda_t^* \cdot \left(\varepsilon + \xi_t^* + y_t - (w^T x_t + \beta)\right) = 0 \quad \forall t = 1.....N \qquad (4.16)$$

$$(C - \lambda_t) \cdot \xi_t = 0 \quad \forall t = 1.....N \qquad (4.17)$$

$$(C - \lambda_t^*) \cdot \xi_t^* = 0 \quad \forall t = 1.....N \qquad (4.18)$$

According to the aforementioned KKT complementarity conditions, training points lying outside the ε-insensitive tube have $\lambda_t = C$ (or $\lambda_t^* = C$) and $\xi_t \neq 0$ (or $\xi_t^* \neq 0$). Those points are called support vectors. Furthermore, there exists no set of dual vaiariables λ_t and λ_t^* which are both nonzero simultaneously as this would require nonzero slack variables in both directions. Finally, training points within the ε-insensitive tube have $\lambda_t \in (0,C)$ (or $\lambda_t^* \in (0,C)$) and also $\xi_t = 0$ (or $\xi_t^* = 0$) (Smola and Scholkopf, 1998).

Alternatively, a practical way of calculating β and slack variables ξ_t and ξ_t^* is by solving a slightly differentiated version of the primal problem:

[Problem P]

$$\min_{\beta,\xi_t,\xi_t^*} \quad \sum_{t=1}^{N}(\xi_t + \xi_t^*)$$

Subject to:

$$y_t - \sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)\cdot x_{t'}^T \cdot x_t - \beta \le \varepsilon + \xi_t \quad \forall\, t = 1....N$$

$$\sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)\cdot x_{t'}^T \cdot x_t + \beta - y_t \le \varepsilon + \xi_t^* \quad \forall\, t = 1....N$$

$$\xi_t \ge 0 \quad \forall\, t = 1....N$$

$$\xi_t^* \ge 0 \quad \forall\, t = 1....N$$

The solution of Problem P derives simultaneously the values for both β and variables ξ_t and ξ_t^*. It is worth mentioning that λ_t and λ_t^* are now treated as parameters whose values are given from the solution of the dual problem solved earlier on. Notice also that problem P is a simple linear programming (LP) model and therefore it can be solved with great computational efficiency even for large number of training points.

As shown in equation (4.14), function F is used to perform a linear regression in input space R^Z based on input vectors $x_{t'}$ and x_t. For nonlinear regression however, we need to exploit the way training data appears in our problem. More specifically, according to equation (4.14) regression function F depends only on the inner product of input vectors ($x_{t'}^T x_t$) and therefore we can employ the following kernel trick (Aizerman *et al.*, 1964) as described by Burges (Burges, 1998). We first map input vectors into a high-dimensional feature space via mapping function Φ as follows:

$$\Phi : R^Z \rightarrow R^{Z'} \tag{4.19}$$

Regression function F then takes the following form:

$$F(x_t) = \sum_{t'=1}^{N} (\lambda_{t'} - \lambda_{t'}^{*}) \cdot \Phi(x_{t'}^{T}) \cdot \Phi(x_t) + \beta \qquad (4.20)$$

The difference between equations (4.14) and (4.20) is that function F is used to perform a linear regression in different spaces, the input space R^Z and the feature space $R^{Z'}$, respectively.

However, there is no particular need to define mapping function Φ explicitly, since the inner product of vectors in the feature space can be represented with a kernel function K as follows:

$$K(x_{t'}, x_t) = \Phi(x_{t'}^{T}) \cdot \Phi(x_t) \qquad (4.21)$$

Table 4.1: *Different types of kernel functions*

No.	Name of Kernel	Expression
1	Polynomial	$K(x_i, x_j) = ((x_i \cdot x_j) + 1)^p \quad p = 1,2,....$
2	Gaussian Radial Basis Function	$K(x_i, x_j) = exp\left(\frac{\lVert x_i - x_j \rVert^2}{-2p^2}\right)$
3	Exponential Radial Basis Function	$K(x_i, x_j) = exp\left(\frac{\lvert x_i - x_j \rvert}{-2p^2}\right)$
4	Multi-layer Perception	$K(x_i, x_j) = tanh(b(x_i \cdot x_j) - c)$
5	Fourier Series	$K(x_i, x_j) = \frac{sin\left(N + \frac{1}{2}\right)(x_i - x_j)}{sin\left(\frac{1}{2}(x_i - x_j)\right)}$
6	Tensor Product Splines	$K(x_i, x_j) = \prod_{m=1}^{n} K_m(x_{im}, x_{jm})$

It is worth mentioning that kernel function K is defined as a function of vectors in the original input space R^Z. In that sense, the expression of regression function F can now easily be extended to accommodate the case of nonlinear regression in input space R^Z

by performing a linear regression in feature space $R^{Z'}$ via the kernel function transformation as follows:

$$F(x_t) = \sum_{t'=1}^{N} (\lambda_{t'} - \lambda_{t'}^*) \cdot K(x_{t'}, x_t) + \beta \tag{4.22}$$

Any function satisfying Mercer's theorem (Mercer, 1909) may be employed as a kernel function. The types of functions most commonly used in support vector machine literature as kernel functions are summarised in Table 4.1 (see for example, Kulkarni *et al.*, 2004).

Finally, it is very interesting to notice that the introduction of kernel functions in the expression of the regression function as shown in equation (4.19) does not affect any of the previous analysis on linear support vector regression. All previous considerations hold intact with the only difference being that linear regression in now performed in a high-dimensional feature space $R^{Z'}$ in order to create a nonlinear regression function in the original input space R^Z (Gunn, 1998). Based on the equation (4.19), dual and primal model formulations can now easily be extended for nonlinear support vector regression as follows:

[Problem D1]

$$\max_{\lambda_t, \lambda_t^*} -\frac{1}{2} \sum_{t=1}^{N} \sum_{t'=1}^{N} (\lambda_t - \lambda_t^*) \cdot (\lambda_{t'} - \lambda_{t'}^*) \cdot K(x_t, x_{t'}) - \varepsilon \sum_{t=1}^{N} (\lambda_t + \lambda_t^*) + \sum_{t=1}^{N} y_t \cdot (\lambda_t - \lambda_t^*)$$

Subject to:

$$\sum_{t=1}^{N} (\lambda_t - \lambda_t^*) = 0$$

$$0 \le \lambda_t \le C \quad \forall\, t = 1....N$$

$$0 \le \lambda_t^* \le C \quad \forall\, t = 1....N$$

and

[Problem P1]

$$\min_{\beta,\xi_t,\xi_t^*} \quad \sum_{t=1}^{N}(\xi_t + \xi_t^*)$$

Subject to:

$$y_t - \left(\sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot K(\boldsymbol{x}_{t'}, \boldsymbol{x}_t) + \beta \right) \le \varepsilon + \xi_t \quad \forall\, t = 1....N$$

$$\left(\sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot K(\boldsymbol{x}_{t'}, \boldsymbol{x}_t) + \beta \right) - y_t \le \varepsilon + \xi_t^* \quad \forall\, t = 1....N$$

$$\xi_t \ge 0 \quad \forall\, t = 1....N$$

$$\xi_t^* \ge 0 \quad \forall\, t = 1....N$$

The aforementioned mathematical models are used as part of our proposed forecasting algorithm presented in the next section.

4.4 Proposed forecasting algorithm

Based on the support vector regression analysis presented in the previous section, we propose the following three-step algorithm.

[Algorithm A1]

```
Step 1: Solve Problem D1 to determine variables λ_t and λ_t*.
Step 2: Fix variables λ_t, λ_t* and solve Problem P1 to determine β.
Step 3: For t:=N+1 to N+M, do:
     i. Calculate customer demand prediction
```

$$y_t = \sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot K(\boldsymbol{x}_{t'}, \boldsymbol{x}_t) + \beta$$

```
     ii. Update input vectors x_t.
```

The first two steps of algorithm A1 are used for determining regression function F from the available N training points. In particular, the first step determines optimal values for the λ_t and λ_t^* from the solution of the dual NLP problem (Problem D1). In Step 2, λ_t and λ_t^* are fixed to their optimal levels while we determine parameter β by solving a linear programming (LP) model (Problem P1). The final step of the algorithm constitutes a post-processing recursive forecasting methodology. Having identified regression function F from the training data points in steps 1 and 2, a customer demand prediction can be made for the next M time periods based on the semi-deterministic past demand attributes.

Given a time series in the form of N training points $(\mathbf{x}_1, y_1)$, $(\mathbf{x}_2, y_2)$, ...,$(\mathbf{x}_N, y_N)$ predicting value y_{N+1} is done by considering the last Δ elements of the time series as deterministic attributes of input vector $\mathbf{x}_{N+1}$. The very next prediction for demand y_{N+2} would normally require the *actual* past demand attribute of point $N+1$. Although such a deterministic past demand attribute is not available to us, it can be substituted with the predicted demand value we estimated previously. In other words, the newly predicted demand value is used as a stochastic attribute (since it is not an actual demand value but merely a prediction) in the input vector $\mathbf{x}_{N+2}$ for predicting value y_{N+2}. In that fashion, by sequentially adding newly obtained data as attributes in the input vector (and removing the earliest elements), we construct a recursive forecasting algorithm based on semi-deterministic past demand data. More specifically, we employ an iterative moving forecasting horizon approach where in each iteration we calculate customer demand for only the following one time period. The customer demand prediction is then used in a recursive fashion to update the input vector information before calculating customer demand for the very next time period. The algorithm terminates when the entire forecasting horizon M is scanned.

Needless of course to mention that the predictive capabilities of the proposed algorithm are restricted by the inevitable error propagation and accuracy deterioration due to inaccurate forecasted data enterting as attributes in our calculations. However, for the purposes of short to medium-term forecast (forecasting/training points ratio between 5%-15%) the proposed algorithm performs with great accuracy, as it is demonstrated by the illustrative examples presented in section 4.5.

4.4.1 Parameter tuning

Before algorithm A1 is implemented, a number of parameters should be determined a priori. Parameter selection is a notorious problem that usually hinders the wide acceptance of forecasting techniques. However, support vector regression only relies on a handful of parameters as listed below:

- regularisation parameter C
- width of ε-insensitive tube
- kernel function K
- number of attributes Δ included in the input vector

As mentioned earlier, parameter C controls the trade-off between model complexity and model accuracy. Underfitting occurs when C is too small since the model does not have enough detail to describe the training data. On the other hand, overfitting occurs when C is too high (Chiang *et al.*, 2004). The optimal value of parameter C is usually determined by employing a grid-search in either a n-fold cross validation or leave-one-out error estimate approach (Kulkarni *et al.*, 2004). However, an exhaustive grid search is a time-consuming and computationally expensive way for parameter selection. Alternative ways for determining SVR parameters is an ongoing research area. In our methodology, we determine parameter C based on a heuristic rule recently proposed by Cherkassky and Ma (2004) as follows:

$$C = max(\bar{y}_t + 3 \cdot \sigma_{y_t}, \bar{y}_t - 3 \cdot \sigma_{y_t}) \qquad (4.23)$$

Where $\bar{y}_t$ is the mean average and σ_{y_t} is the standard deviation of the customer demand training points. The proposed formula for determining C has been validated for a number of different cases (Cherkassky and Ma, 2004).

The size of ε influences the number of support vectors (training points lying outside the ε-insensitive tube) and therefore allows direct control over the complexity of the model. Therefore in practice, parameter ε is chosen so as to reflect our relative view towards error and noise through the implementation of different ε-insensitive loss functions. In our experience, values of ε equal to approximately one order of magnitude less than the mean average of the training points target values provide good performance for various data sets. Mathematically, we have:

$$\varepsilon = \frac{\bar{y}_t}{k} \tag{4.24}$$

where k is a constant scalar taking values in the range *[10,30]*.

The gaussian radial basis function (RBF) is chosen as the kernel function in our proposed algorithm since it is the most commonly used kernel for support vector machines. Mathematically, it is given by the following form:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left(\frac{\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2}{-2p^2} \right) = exp\left(-\gamma \cdot \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2 \right) \tag{4.25}$$

Cherkassky and Ma (2004) proposed an empirical rule for determining RBF kernel width parameter p. According to their formula for z-dimensional problems, where all z input attributes are scaled to *[0,1]*, p is determined approximately by the following formula:

$$p^Z \approx (0.2 - 0.5) \tag{4.26}$$

In our methodology, input vector attributes (Δ) are scaled between zero and one and therefore we calculate parameter p by employing the mean value of the proposed formula, resulting in the following mathematical expression:

$$p = 0.35^{1/z} \Rightarrow \gamma = \frac{1}{2} \cdot 0.35^{-2/z} \tag{4.27}$$

The number of attributes included in the input vector should reflect the seasonality pattern underlying customer demand training points. A heuristic rule for accurate predictions is to employ a number of past attributes equal to an integer multiple of the period points. Say for example one wishes to predict a daily customer demand pattern that repeats itself more or less every week, then it would be advisable to use 7 or 14 past demand attributes.

4.4.2 Forecasting assessment

The assessment of the proposed forecasting algorithm is performed though the employment of the following accuracy criteria:

- Prediction Accuracy (*P.A.*): derives from the MAPE criterion and is used to compare actual demand and predicted demand values over the forecasting horizon time periods. It is defined as follows:

$$P.A. = 100 - MAPE = 100 - 100 \cdot \frac{\sum_{t=N+1}^{N+M} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{M} \tag{4.28}$$

- Fitting Accuracy (F.A.): it is used to compare actual demand and predicted demand values over all training points time periods. It is defined as follows:

$$F.A. = 100 - 100 \cdot \frac{\sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{N} \tag{4.29}$$

- Overall Accuracy (O.A.): it is used to compare actual demand and predicted demand values over all time periods (both training and forecasting). It is defined as follows:

$$O.A. = 100 - 100 \cdot \frac{\sum_{t=1}^{N+M} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{N+M} \tag{4.30}$$

The applicability of the proposed forecasting algorithm along with the heuristic-based rules for parameter estimation and assessment criteria is demonstrated by a number of illustrative examples presented in the next section.

4.5 Illustrative examples

This section presents customer demand forecasting results for three illustrative examples. All runs were implemented in GAMS (General Algebraic Modelling System) (Brooke *et al.*, 1998) and solved with commercially available solvers CONOPT (Drud, 1985) for the nonlinear models and CPLEX (ILOG, 1999) for the linear models using an IBM RS/6000 workstation.

4.5.1 Example 1

Illustrative example 1 is concerned with an electrical appliances distribution company. The company has collected customer sales daily data from all points-of-sale for the last 15 weeks as shown in Table 4.2. Based on this historical data only, the company wishes to make an accurate customer demand forecast for the following two weeks. The additional 2 weeks data (week 16 and week 17) shown in italics in Table 4.2 are not used as training points but they are employed only for validating the accuracy of the proposed methodology.

Table 4.2: *Electrical appliances daily sales (in hundreds of pieces)*

Week	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	7.80	7.70	6.00	6.80	9.20	12.00	3.20
2	5.30	6.30	5.80	5.80	6.00	10.00	5.00
3	5.50	6.40	5.80	5.90	6.10	10.20	5.00
4	6.20	7.80	6.40	6.40	7.20	12.00	6.00
5	7.00	9.00	7.80	7.90	8.30	14.00	7.00
6	9.00	11.00	9.80	9.90	10.40	17.80	9.00
7	8.58	8.47	6.60	7.48	10.12	13.20	3.52
8	5.83	6.93	6.38	6.38	6.60	11.00	5.50
9	6.05	7.04	6.38	6.49	6.71	11.22	5.50
10	6.82	8.58	7.04	7.04	7.92	13.20	6.60
11	7.70	9.90	8.58	8.69	9.13	15.40	7.70
12	9.90	12.10	10.78	10.89	11.44	19.58	9.90
13	10.14	10.01	7.80	8.84	11.96	15.60	4.16
14	6.89	8.19	7.54	7.54	7.80	13.00	6.50
15	7.15	8.32	7.54	7.67	7.93	13.26	6.50
16	*8.06*	*10.14*	*8.32*	*8.32*	*9.36*	*15.60*	*7.80*
17	*9.10*	*11.70*	*10.14*	*10.27*	*10.79*	*18.20*	*9.10*

The SVR parameter values used for example 1 can be found in Table 4.3. The values for SVR parameters C, ε and γ derive from the mean average and standard deviation of customer demand training points by employing the heuristic rules presented in section 4.4. Notice that the number of past demand attributes (Δ) equals 14 days, meaning that for each prediction point we base our calculations on the previous 2 weeks data. This is done so as to reflect the periodic behaviour of customer demand.

Table 4.3: *SVR parameter values for example 1*

Parameter Symbol	Parameter Value
$\bar{y}_t$	8.39
σ_{y_t}	2.825
k	20
Δ	14
C	16.864
ε	0.419
γ	0.581

Customer demand forecasting results are shown graphically in Figures 4.2 and 4.3. According to the results, the proposed regression function is able to capture the basic customer demand pattern and derive accurate forecast for the 2 weeks forecasting horizon. It is very interesting to notice that the regression function can easily follow the actual customer demand even in extreme time periods where demand fluctuates much above average. For instance, increased customer demand during Saturdays is captured efficiently for both weeks in the forecasting horizon (see Figure 4.3, time periods 111 and 118).

Moreover, based on the training points, SVR can also capture the positive trend underlying customer demand and produce increased customer sales expectations for the forecasting horizon under investigation. Overall, the quality of the proposed forecast is determined based on the previously defined assessment criteria as shown in Table 4.4.

Table 4.4: *Forecasting assessment criteria for example 1*

Criterion Symbol	Criterion Value
P.A.	95.22 %
F.A.	92.48 %
O.A.	92.80 %

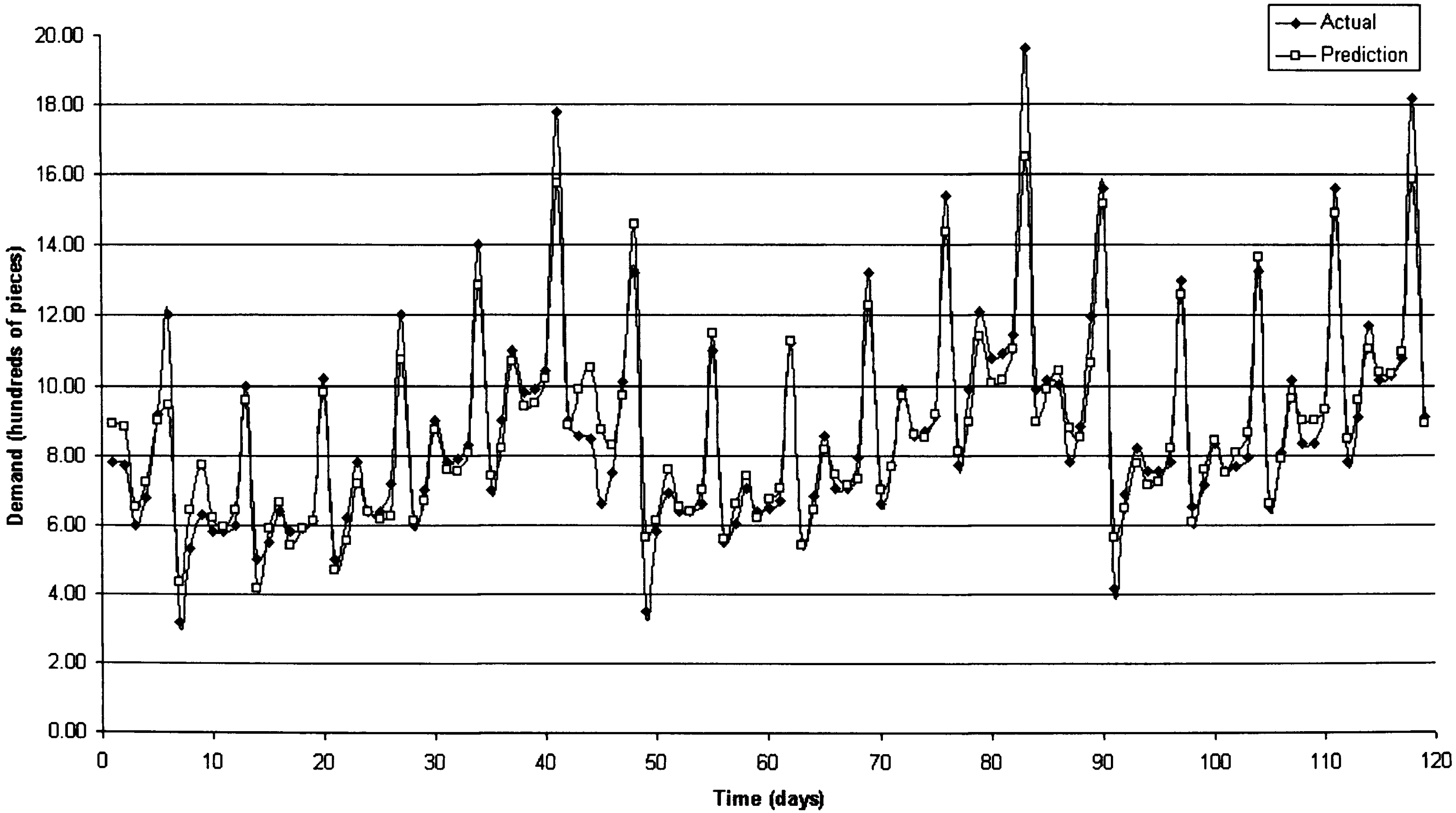

Figure 4.2: *Customer demand forecast for example 1*

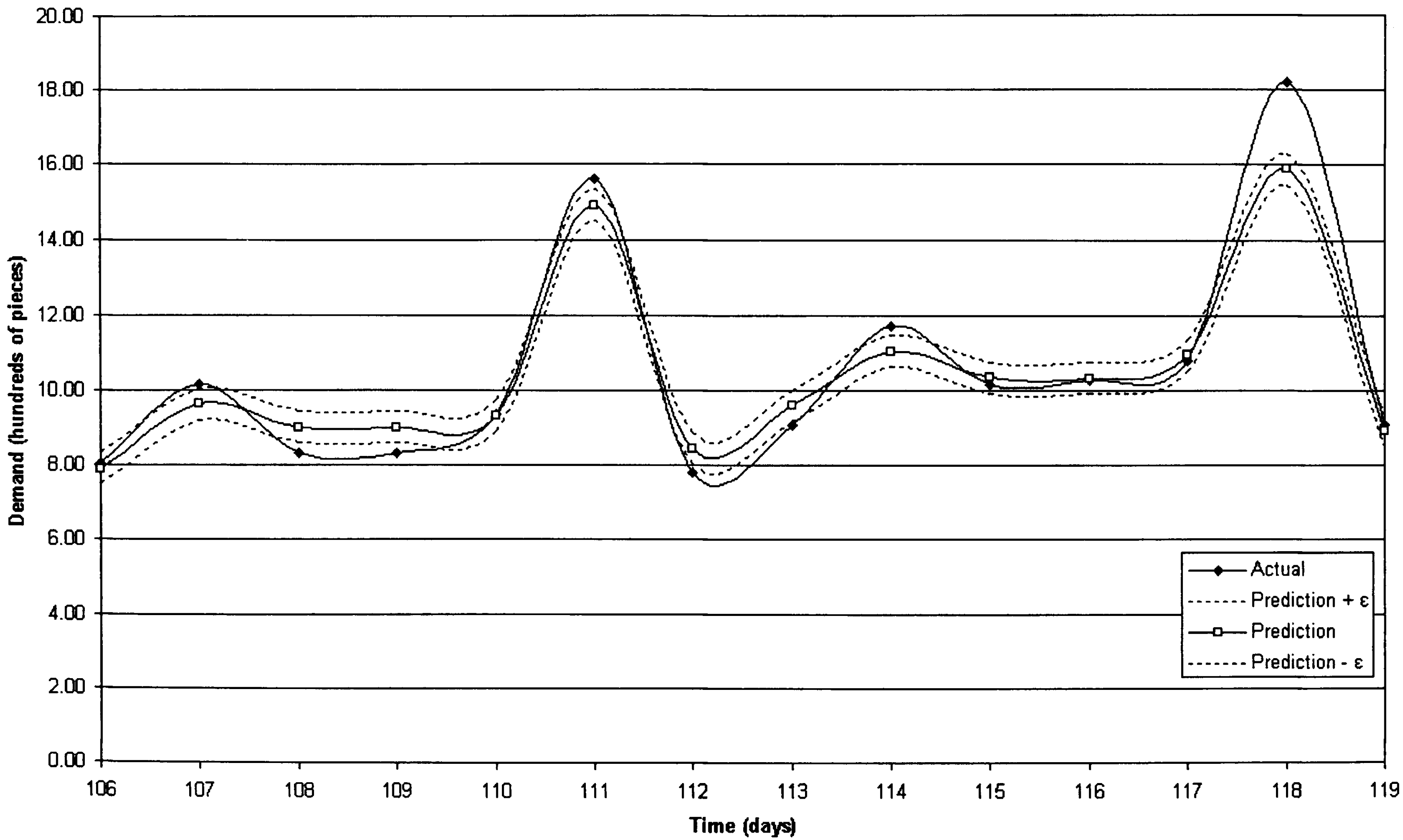

Figure 4.3: *Focus on customer demand forecast for example 1*

4.5.2 Example 2

Example 2 presents a customer demand forecasting example for a chemical process industry. Monthly sales data is provided for the last 9 years (see Table 4.5) and the question is to accurately forecast customer demand for the next 12 months. Sales data points for the last year, shown in italics in Table 4.5, are not used as training points but merely employed for assessing the predictive capabilities of the SVR.

Table 4.5: *Monthly chemical sales (in tonnes)*

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	5055	5904	6180	7073	7545	8037	6650	6947	7222	6132	6073	3724
2	3994	4326	5176	5900	7485	8034	6431	7215	7744	8017	7685	4950
3	5971	6816	7179	8338	8549	8773	7727	8295	8679	7275	5434	4270
4	5649	7188	7837	8631	9805	9760	7921	9094	9413	8755	8069	5174
5	6066	7084	7416	8487	9054	9644	7980	8336	8666	7358	7287	4468
6	4792	5191	6211	7080	8982	9640	7717	8658	9292	9620	9222	5940
7	7165	8179	8614	10005	10258	10527	9272	9954	10414	8730	6520	5124
8	6778	8625	9404	10357	11766	11712	9505	10912	11295	10506	9682	6208
9	7077	8265	8652	9902	10563	11251	9310	9725	10110	8584	8502	5213
10	*5591*	*6056*	*7246*	*8260*	*10479*	*11247*	*9003*	*10101*	*10841*	*11223*	*10759*	*6930*

The parameter values used for this example are given in Table 4.6 while customer demand forecasting results are depicted in Figures 4.4 and 4.5. Again, just as in example 1, SVR manages to capture successfully both the positive trend and the periodic pattern of demand.

Table 4.6: *SVR parameter values for example 2*

Parameter Symbol	Parameter Value
$\bar{y}_t$	7879.241
σ_{y_t}	1831.659
k	20
Δ	24
C	13374.218
ε	393.962
γ	0.546

Moreover in this example, every year contains an internal pattern resulting in spiky M-shaped time periods. By employing an increased number of attributes (Δ=24 months), SVR manages to overcome this difficulty in demand pattern recognition and provides accurate forecasts that follow closely the actual customer demand for both high and low demand time periods (see Figure 4.4). Furthermore, the proposed methodology is able to capture the effect of decreased customer demand occurring every July and derives a precise forecast for the seventh month of the forecasting horizon (see Figure 4.5, time period 115). According to the forecasting assessment criteria for example 2, SVR scores well above 95% in all accuracy measures as shown in Table 4.7

Table 4.7: *Forecasting assessment criteria for example 2*

Criterion Symbol	Criterion Value
P.A.	95.24 %
F.A.	95.38 %
O.A.	95.36 %

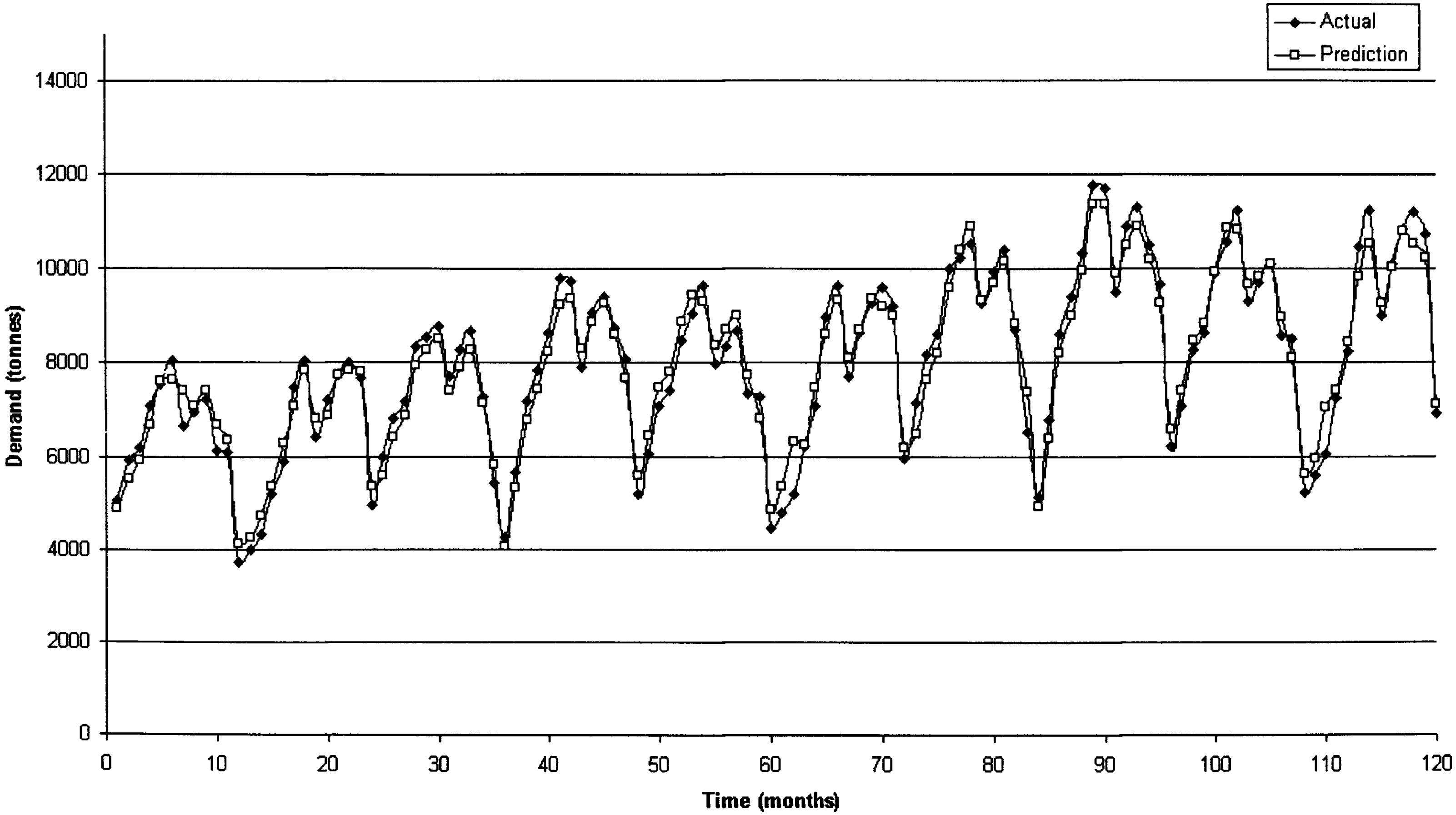

Figure 4.4: *Customer demand forecast for example 2*

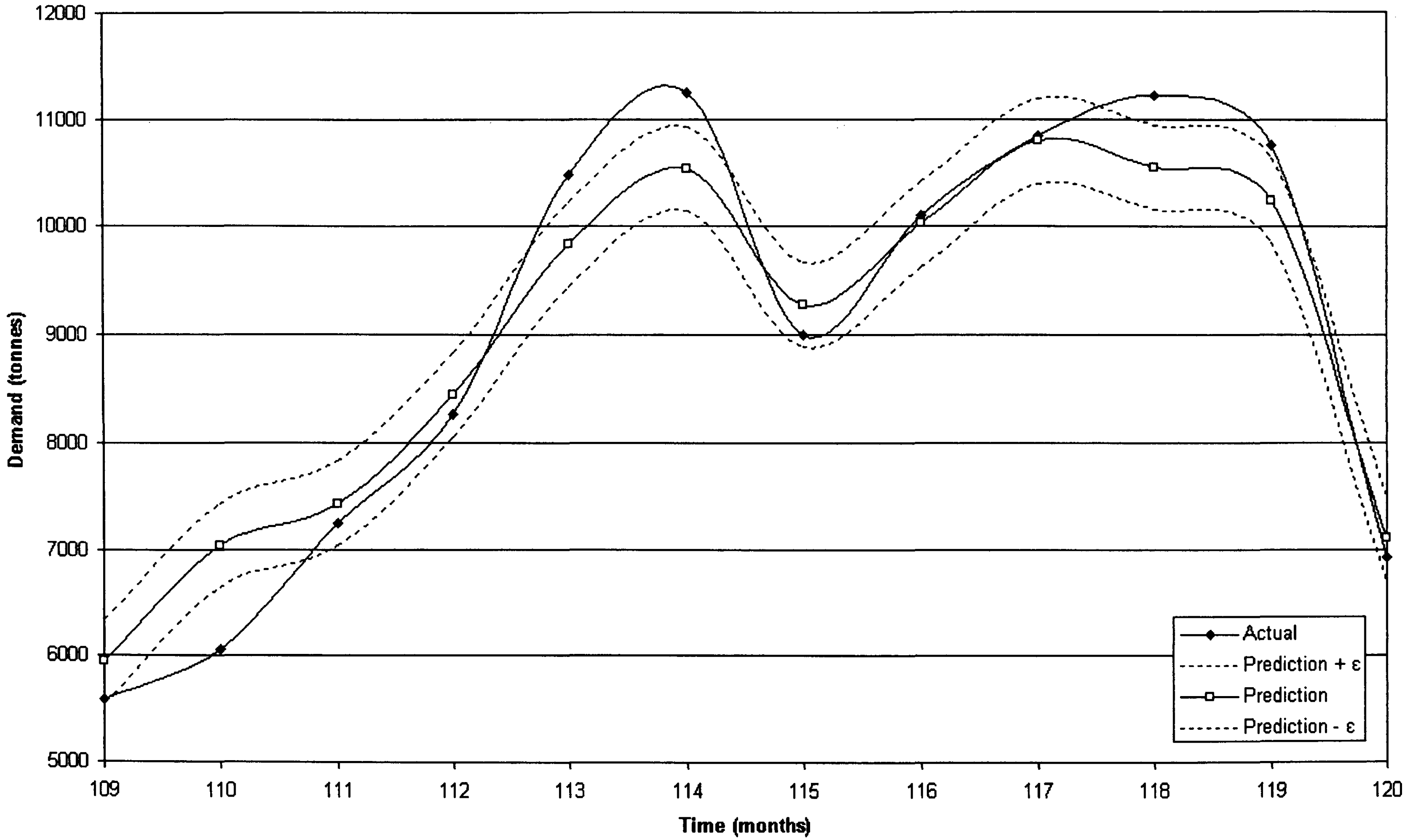

Figure 4.5: *Focus on customer demand forecast for example 2*

4.5.3 Example 3

Example 3 is a customer demand forecasting example in the food and drinks process industry. The monthly champagne sales data for the period between January 1964 and September 1972 has been collected from the Association of French Champagne Firms and published by Makridakis and Wheelwright (Makridakis and Wheelwright, 1978) as shown in Table 4.8. Based on the monthly champagne sales data only for the period between January 1964 and September 1971 (training points), we would like to forecast champagne sales for the following 12 months.

Table 4.8: *Monthly champagne sales – France (millions of bottles)*

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1964	2.815	2.672	2.755	2.721	2.946	3.036	2.282	2.212	2.922	4.301	5.494	7.312
1965	2.541	2.475	3.031	3.266	3.776	3.230	3.028	1.759	3.595	4.474	6.838	8.357
1966	3.113	3.006	4.047	3.523	3.937	3.986	3.260	1.573	3.528	5.211	7.614	9.254
1967	5.375	3.088	3.718	4.514	4.520	4.539	3.663	1.643	4.739	5.428	8.314	10.651
1968	3.633	4.292	4.154	4.121	4.647	4.753	3.965	1.723	5.048	6.922	9.858	11.331
1969	4.016	3.957	4.510	4.276	4.968	4.677	3.523	1.821	5.222	6.872	10.803	13.916
1970	2.639	2.899	3.370	3.740	2.927	3.986	4.217	1.738	5.221	6.424	9.842	13.076
1971	3.934	3.162	4.286	4.676	5.010	4.874	4.633	1.659	5.951	*6.981*	*9.851*	*12.670*
1972	*4.348*	*3.564*	*4.577*	*4.788*	*4.618*	*5.321*	*4.298*	*1.431*	*5.877*	-	-	-

From a preliminary study of the given historical sales data, we can clearly identify a customer demand pattern that consistently repeats itself each year. Customer demand for champagne is relatively low during summer months (low-peaks every August) while it steadily builds up during autumn before reaching its high-peak every December mainly due to holidays celebrations. Based on this observation, we choose to employ 12 past demand attributes in the SVR regression input vectors. All SVR parameters used for example 3 are given in Table 4.9.

Table 4.9: *SVR parameter values for example 3*

Parameter Symbol	Parameter Value
$\bar{y}_t$	4.638
σ_{y_t}	2.472
k	30
Δ	12
C	12.054
ε	0.155
γ	0.596

Forecasting results for example 3 are clearly illustrated in Figures 4.6 and 4.7. According to the results, the proposed methodology derives a very precise prediction for the entire forecasting horizon including the December high-peak and the August low-peak (see Figure 4.7, time periods 96 and 104 respectively).

It is also very interesting to notice that the regression function keeps very good track of the irregular customer demand and successfully reproduces the rising trend of December demands (see Figure 4.6). However, SVR does not naively mimic the training points but rather learns from them. The proposed forecasting algorithm manages to distinguish noisy data points from structural data points (see Figure 4.6, time period 72) and therefore derives a regression function that not only avoids *overfitting* but also interprets correctly the underlying customer demand pattern into a forecast of over 93% accuracy as shown in Table 4.10.

Table 4.10: *Forecasting assessment criteria for example 3*

Criterion Symbol	Criterion Value
P.A.	93.29 %
F.A.	91.92 %
O.A.	92.08 %

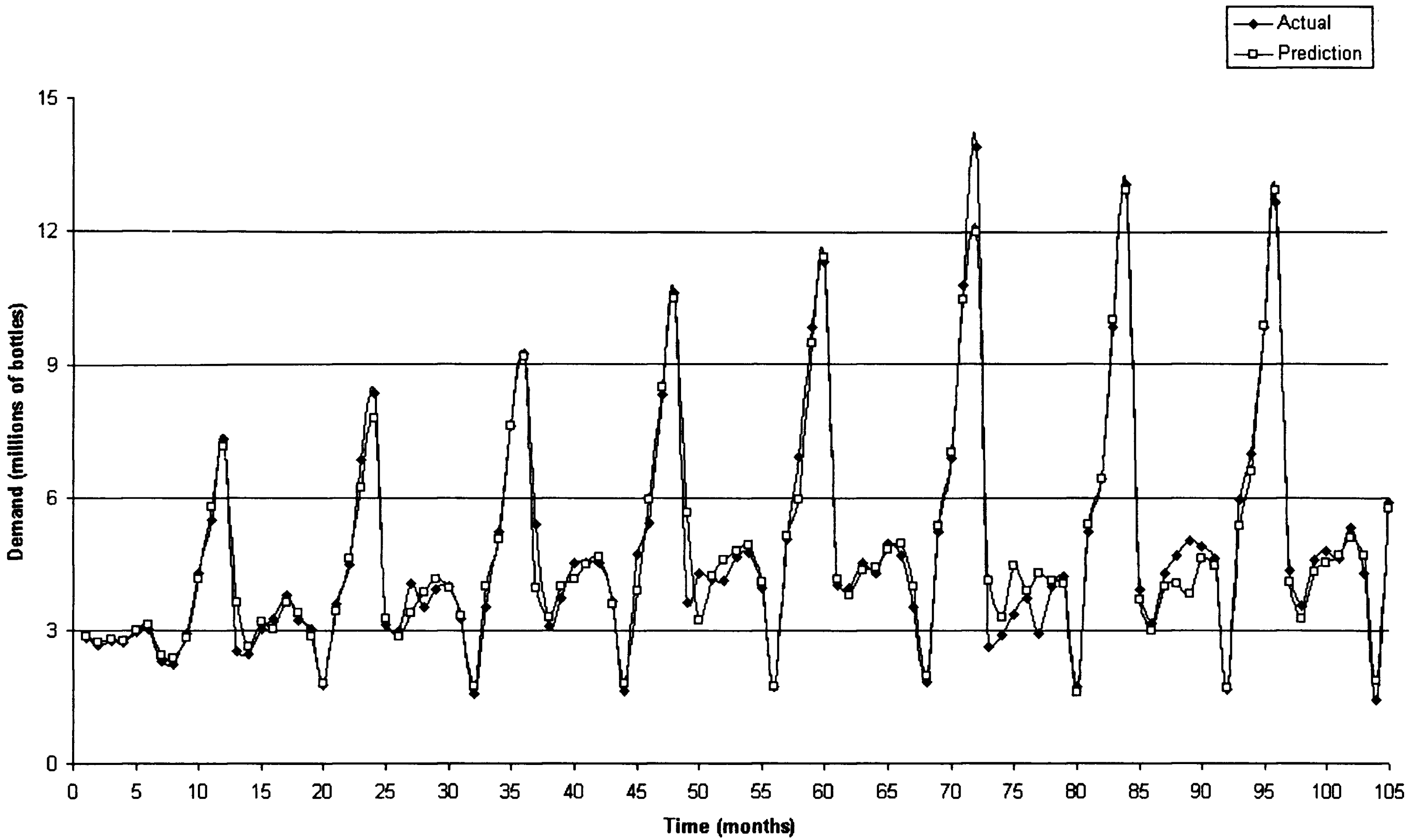

Figure 4.6: *Customer demand forecast for example 3*

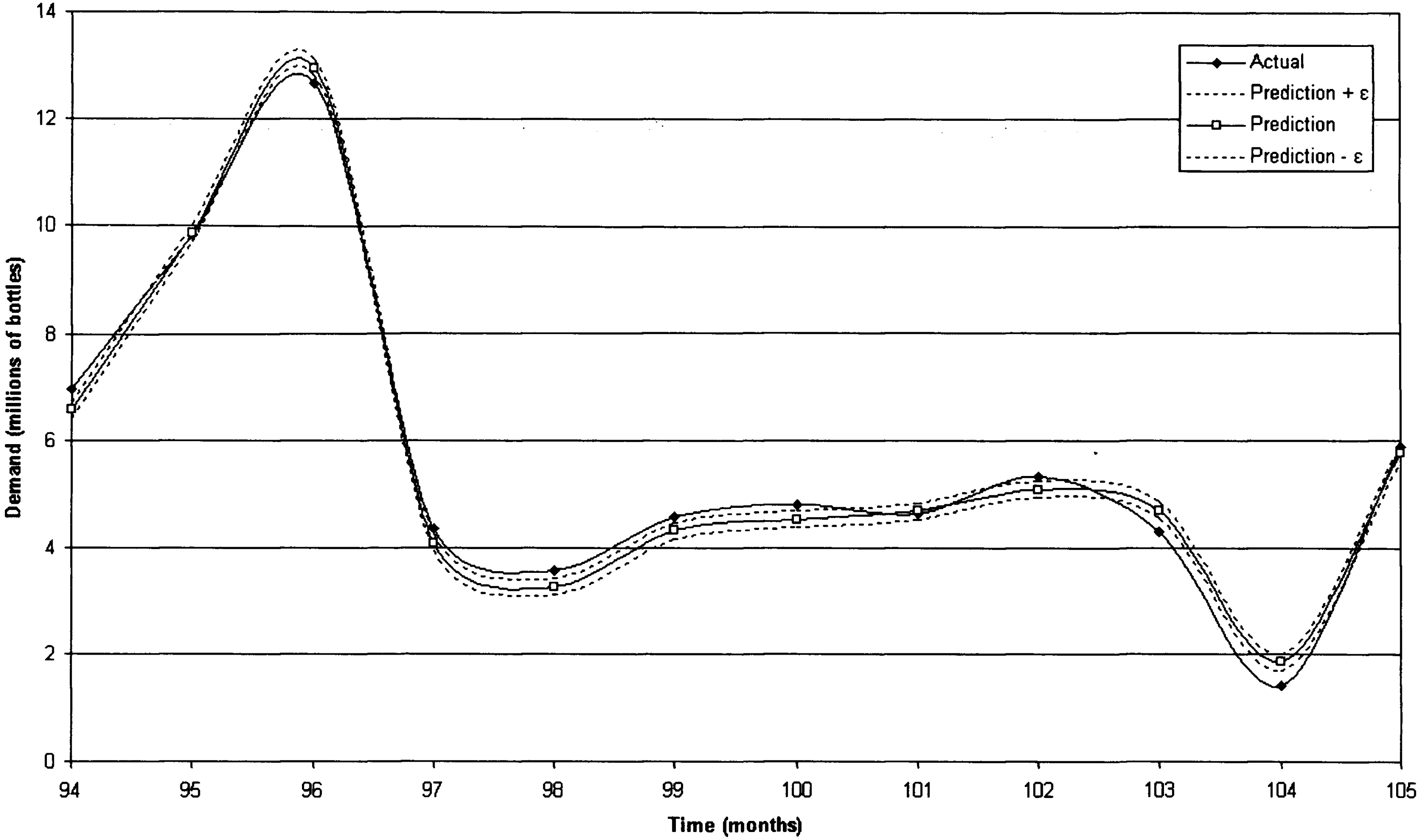

Figure 4.7: *Focus on customer demand forecast for example 3*

4.6 Concluding remarks

In this chapter, a systematic optimisation-based approach for customer demand forecasting was presented based on support vector regression (SVR) analysis. Historical customer demand patterns were used as training points attributes for the SVR. The proposed approach employed a three-step algorithm able to extract information from the training points in order to identify an adaptive basis regression function and perform a recursive methodology for customer demand forecasting.

The applicability of the proposed forecasting approach was validated through a number of illustrative customer demand forecasting examples. In all three examples, the proposed methodology handled successfully complex nonlinear customer demand patterns and derived forecasts with prediction accuracy of more than 93% in all cases. Although, future work should consider a more formal way of determining SVR parameters, support vector regression can still be regarded as a parsimonious alternative to complex artificial neural networks forecasting.

Chapter 5

Active demand management through price optimisation

5.1 Introduction and literature survey

In today's global marketplace, process industries no longer compete as independent entities but rather as integral part of supply chain links. The ultimate success of a firm depends on its managerial ability to integrate and co-ordinate the complex network of business relationships among supply chain members (Min and Zhou, 2002). The recent wave of mergers and acquisitions (M&S) has led a number of smaller companies to consolidate into a few giant supply chain firms (e.g. Unilever, Procter & Gamble) that provide close substitute products (e.g. fast moving consumer goods FMCG) to a wide range of customers. The intense competition among different companies is evident and occurs in almost every market sector nowadays.

In this competitive environment, customer demand is usually satisfied by a small number of companies, each one manufacturing and selling its individual subset of products. The goal of every company is to obtain the highest possible profit by determining optimal price levels for its portfolio products. In that case, an oligopolistic price competitive market environment is established that needs to account for both competitors' activities and customers' willingness to buy.

Duopolistic market competition is the natural starting point for investigating the behaviour of oligopolies. Consider two companies, Company A and Company B, each one manufacturing its own subset of substitute products. By substitute products we mean slightly differentiated product brands that belong to the same product-class (e.g. lubricants, detergents, cosmetics, carbonated soft drinks etc). In that case, increased sales of one product result in reduced sales of another, thus forming a market environment where products brands compete with each other over a common customer base.

The manufacturing of products usually takes place in production sites owned by the company (in-house manufacturing). Every site has a limited amount of available resources used for production. Alternatively, each company may have the option to allocate manufacturing of a certain amount of products to a third-party company (outsourcing). As shown in Figure 5.1, final products from each company (in-house manufactured and outsourced) are then transported to the marketplace in order to satisfy the anticipating customer demand at given product prices.

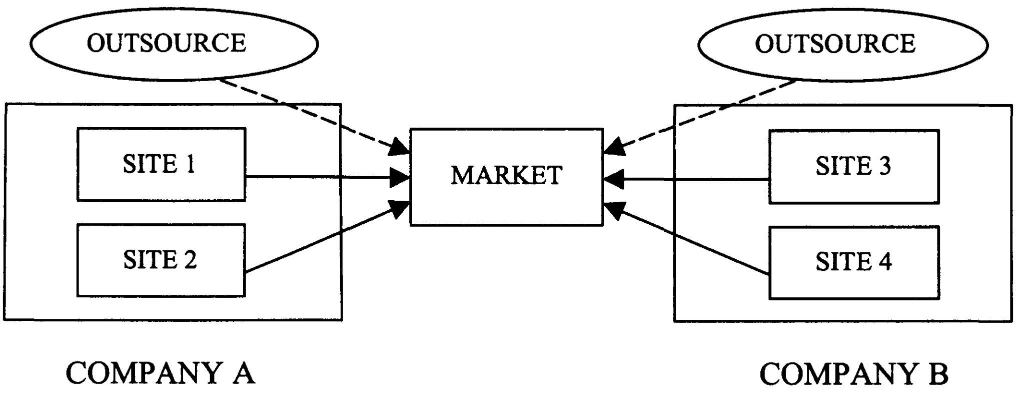

Figure 5.1: *Duopolistic market environment*

A crucial precondition of effective price competition is that customers are inclined to search for lower-priced substitute products. Low prices however can kill profit margins and jeopardise the overall company profitability. On the other hand, high prices will drive away potential customers and inevitably put company's market share at risk. Pricing decisions are of crucial importance and unless taken seriously, they can pose a major threat to the sustainability of the company.

Traditional approaches for customer demand management assume fixed product prices and usually rely on forecasting tools, trying to predict customer demand based on historical sales patterns (Markidakis and Wheelwright, 1982). Passive demand management (PDM) approaches ignore the importance of flexible product pricing and usually lead to poor customer demand satisfaction.

A new trend towards active demand management (ADM) has recently emerged focusing on how to actually *drive* customer demand away from traditional baseline forecasts so as to maximise both overall business performance and customer satisfaction. Many companies ranging from the automotive industry (Ford Motor Co.) to the Internet dotcoms (Amazon.com) have recently realised the potential benefits of adopting such a marketing-based concept of smart pricing (Coy, 2000). Some firms (e.g. those operating petrol forecourts) do not hesitate to go even a step further ahead and employ such clever "dynamic pricing" strategies almost on a daily basis. According to Manugistics (Manugistics, 2003), a leading company in pricing and revenue optimisation, pricing is the next battleground for competitiveness.

However, product pricing is not a trivial task. Successful pricing strategies should consider simultaneously rapidly changing customer expectations, fast-reacting competitors, complex product interactions and fluctuating manufacturing capacity constraints. Accelerating product lifecycles and increasing product mix diversity further magnify the complexity of capturing an accurate understating of the pricing environment and managing a comprehensive strategy around it (Rapt, 2003). Lanning *et al.* (2000) allow demand to be determined by prices via a constant-elasticity demand function. Prices are then optimised jointly with capacity investment decisions. Optimal capacity levels and prices for substitute products are considered by Birge *et al.* (1998) in a single-period model while joint co-ordination of production and marketing decisions are investigated by Eliashberg *et al.* (1987).

Although the problem of product pricing is not new in applied economics and operational research literature, previous studies adopt a somehow simplistic approach to the problem. They focus their attention on single-product firms and therefore cannot accommodate the nature of multi-product firms which are predominant nowadays. Another common drawback is that many studies consider product pricing in isolation of the market competition, thus ignoring the reaction effect of rival companies offering substitute products in the marketplace. Even in the case where competition between single-product firms is addressed, the number of firms is restricted to two (duopolistic competition) while joint production and pricing decision-making is based on unrealistic assumptions such as unlimited manufacturing resources that render the proposed solution of impractical value and inapplicable to real-life business problems.

There exists a clearly identified need to address product pricing issues in a more realistic context that is able not only to consider simultaneously multi-product firms competing in an oligopolistic market environment but also propose alternative pricing policies and production modes such as outsourcing options. Our proposed methodology of active demand management through price optimisation is able to capture the dominant trade-off between product price and product market share so as to deliver value to the customer while ensuring high profitability for the company.

The rest of the chapter is structured as follows. In the next section, the problem of active demand management is briefly described, while the main characteristics of an efficient pricing strategy are also discussed. Section 5.3 presents the case of single-product price competition between two firms. Analytical formulae are derived for determining Nash equilibrium in prices while we propose an iterative algorithm validated by a motivating example. In section 5.4 we extend the proposed algorithm in order to accommodate the case of multi-product firms operating in an oligopolistic market environment and address customer demand forecasts while also considering outsourcing options. Four cases of a comprehensive illustrative example are solved in section 5.5 while some concluding remarks are drawn in section 5.6.

5.2 Problem description

The marketing mix is defined as the set of controllable tactical marketing tools that the firm blends to produce the response it wants in the market place. The marketing mix consists of everything a firm can do to influence the demand for its product. The many possibilities gather into four groups of variables known as the "four P's": product, price, place and promotion (Kotler *et al.*, 1996).

In the narrowest sense, price is the amount of money charged for a product or service. More broadly, price is the sum of all the values that consumers exchange for the benefits of having or using a specific product or service. Price is the only element in the marketing mix that produces revenue, while all other elements represent costs (Kotler *et al.*, 1996). Product, promotion and place are value-creating activities while pricing can be viewed as the firm's attempt to capture some of the created value in the profits earned (Nagle and Holden, 1995). Therefore, pricing is identified as the most flexible element of the marketing mix, since it is the fastest and most cost-effective way to enhance company profits.

Every company nowadays is operating with a different set of business objectives. Many companies for example set profit maximisation as their ultimate goal. Other companies however, seek to increase their market share or even try to augment their customer satisfaction levels. Different business objectives can be achieved through the employment of alternative pricing strategies such as skim pricing, penetration pricing, neutral pricing etc.

Irrespective of the business objectives, an effective pricing strategy should consider simultaneously the following three main aspects: costs, customers and competition. Integrating cost management, customer behaviour and market competition into a unified framework is the key in developing a successful pricing strategy for active demand management.

5.2.1 Cost management

Costs play a significant role in formulating an efficient pricing strategy. There can be variable and/or fixed costs. Manufacturing costs are usually variable costs depending on the sales volume. Traditional pricing strategies are based on a cost-driven approach as shown in Figure 5.2. According to the cost-based pricing strategy, every product is priced so as to cover its own costs plus make a fair marginal profit. Although such a strategy seems as a simple guide to profitability, in practice it does not deliver the desired results. The fundamental problem with cost-driven pricing is that unit costs cannot be calculated before determining the product price. The reason for that is that pricing affects sales volumes and sales volumes in turn affect unit costs (Nagle and Holden, 1995).

5.2.2 Customer behaviour

In order to capture the trade-off between price and sales volume, a value-based pricing strategy can be employed as shown in Figure 5.2.

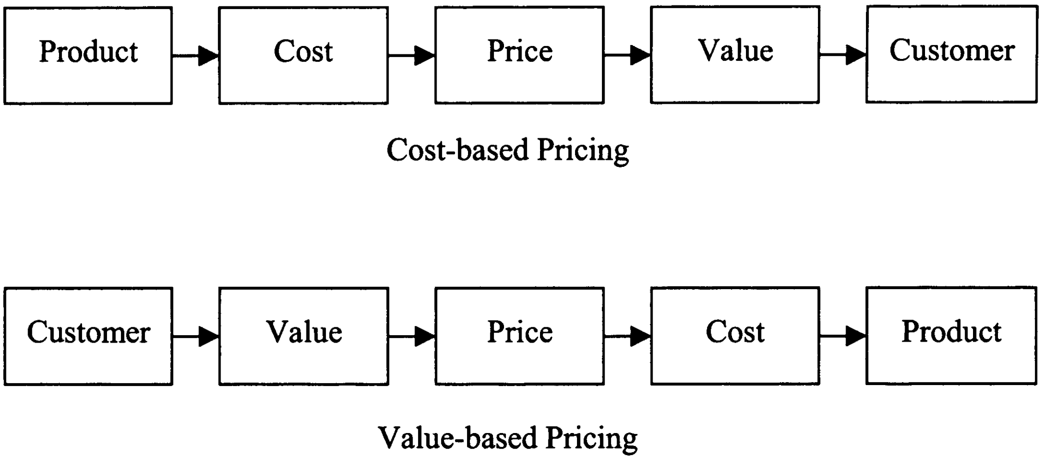

Figure 5.2: *Cost-based versus value-based pricing strategies*

The main difference in this case is the inverse order of decision-making allowing for a value-based pricing strategy that is more customer-oriented. Unlike, cost-based pricing, customer's perceived value is now the driving force for product pricing. Conjoint-analysis is a market research tool concerned with understanding how customers perceive product value and how they make choices between products based on their individual attributes. BPTO (brand-price trade-off) is a variation of the conjoint analysis used for testing price sensitivity in the context of brands available on the market so as to assess brand preference at any given price scenario.

Price sensitivity can be measured by using the concepts of demand elasticity and cross-elasticity. Price elasticity measures the percentage change in the quantity demanded relative to the percentage change in price (Pashigian, 1998). When there exists a certain degree of substitution between differentiated products, cross-elasticity can be used to measure the percentage change in the quantity demanded relative to the percentage change in price of another product, so as to quantify the competition effect between close substitutes brands (Pindyck and Rubinfield, 1992). Estimating elasticity and cross-elasticity parameters is an active research area while many market research companies are developing their own methodologies. A paper by Stavins (1997) adopted from the differentiated-product literature is an illustrative example of demand elasticity estimation in the personal computer (PC) market while Acutt and Dodgson (1996) present a method for calculating cross-elasticities between different public transport modes. Shankar *et al.* (1996) relate price sensitivity and price policy from a retailer point of view. Alves *et al.* (2003) estimate the cross-elasticity between gasoline and alcohol while Tellis (1988) confirms the negative sign of elasticity parameters.

5.2.3 Competition

Oligopolistic competition has received a great deal of attention in the research literature (Varian, 1992). However, the "oligopoly problem" has proved to be one of the most resilient problems in the history of economic thought (Vives, 1999). Very early models of oligopolistic competition include the models of Cournot (1838) and Bertrand (1883). According to those models competitng firms only act once and also act simultaneously to determine the outcome of competition among them. The Cournot model treats output (quantity) as the strategic decision variable of each firm while the Bertrand model focuses on price as the strategic decision variable to be determined by each firm.

Apart from the aforementioned Cournot and Bertrand models, oligopolist competition formulations include the repeated and the sequential games. The repeated games can be viewed as a series of Cournot-type or Bertrand-type models not related to each other and solved independently. The sequential games on the other hand, involve a sequence of decision-making between the firms where the outcome of competition derives from the interaction of logic-based firm policies. The Stackelberg model (Stackelberg, 1934), aslo known as the leader-follower model, constitutes an extension of the Cournot model that can be classified as a sequential game of oligopolistic competition. Output decisions are taken in turns with the leader-firm making the first move and the follower-firm acting upon observation of the previous move, resulting in a two-stage game.

A well-respected solution concept for non-cooperative games in oligopolistic competition is the Nash equilibrium point (Nash, 1951) which is defined as the point where all players in the game do their best given the choice of all other players. Sherali *et al.* (1983) study the supply side of an oligopolistic market supplying an homogeneous product noncooperatively. They characterise the nature of Stackelberg-Nash-Cournot equilibria and they prescribe methods for their computation. Sherali and Leleno (1988) present a mathematical programming approach for Nash-Cournot equilibrium analysis of oligopolies and derive equilibrium solutions in various market structures. Our proposed methodology can be classified as a sequential Bertrand-type price optimisation approach that aims to determine optimal price levels and production plans as described in the next section.

5.3 Single-product price optimisation

In this section, we focus our attention in the specific case of single-product firms operating in a duopolistic marketplace. Analyical formulae are derived for that special oligopolistic case, while we propose an iterative algorithm for determining optimal product prices. A motivating example is then solved in order to validate the applicability of both formulae and the proposed algorithm.

5.3.1 Analytical formulae

Consider two firms 1 and 2, each one offering a single product to the market. Suppose their products are close substitutes and compete with each other over the same customer base. However, there is at least some degree of differentiation between the two products and therefore each firm faces different demand curves ($Q1$, $Q2$) and different variable ($VC1$, $VC2$) and fixed ($FC1$, $FC2$) manufacturing costs while the products are sold for different prices ($P1$, $P2$). The sales volume for every firm is defined as a linear function of its own price ($P1$) and the competitor's price ($P2$):

$$\text{Firm 1: } Q1 = a1 - b1 \cdot P1 + c12 \cdot P2 \quad (5.1)$$

and

$$\text{Firm 2: } Q2 = a2 - b2 \cdot P2 + c21 \cdot P1 \quad (5.2)$$

Where $a1$, $a2$ are demand coefficients, $b1$, $b2$ are demand elasticity parameters and $c12$, $c21$ are demand cross-elasticity parameters. All parameters in our formulation take positive values.

Note that the quantity each firm can sell decreases when the firm raises its own price, but increases when its competitor charges a higher price. Each firm will choose its own price, taking the competitor's price as fixed. The profit of firm 1 equals its revenue minus the variable and fixed manufacturing costs:

$$\Pi 1 = (P1 - VC1) \cdot min(Q1, Cap1) - FC1 \tag{5.3}$$

Where $VC1$ is the variable unit cost, $Cap1$ is the available capacity and $FC1$ is the fixed cost for Firm 1. Depending on the values of $Q1$ and $Cap1$, the profit of Firm 1 equals to:

$$\Pi 1 = \begin{cases} (P1 - VC1) \cdot Q1 - FC1 & when\ Q1 < Cap1, \\ (P1 - VC1) \cdot Cap1 - FC1 & otherwise \end{cases} \tag{5.4}$$

If $Q1$ equals $Cap1$, then:

$$a1 - b1 \cdot P1 + c12 \cdot P2 = Cap1 \tag{5.5}$$

and the critical value for the price of Firm 1 equals to:

$$P1^c = \frac{a1 - Cap1 + c12 \cdot P2}{b1} \tag{5.6}$$

Similarly for Firm 2:

$$P2^c = \frac{a2 - Cap2 + c21 \cdot P1}{b2} \tag{5.7}$$

Case 1: Unconstrained-Unconstrained

In that case both firms have unlimited resource capacity, meaning that $Q1<Cap1$, $Q2<Cap2$, $P1>P1^c$ and $P2>P2^c$. The resulting profit for Firm 1 is calculated as follows:

$$\Pi 1 = (P1 - VC1) \cdot Q1 - FC1 \tag{5.8}$$

and substituting $Q1$ we get:

$$\Pi 1 = a1 \cdot P1 - b1 \cdot P1^2 + c12 \cdot P2 \cdot P1 - a1 \cdot VC1 + b1 \cdot P1 \cdot VC1 - c12 \cdot P2 \cdot VC1 - FC1 \tag{5.9}$$

Firm's 1 profit is maximised when the incremental profit from a very small increase in its own price is zero. Taking P2 as fixed, Firm 1's profit is concave in P1 and therefore the maximising price is given by:

$$\Delta\Pi 1 / \Delta P1 = a1 - 2 \cdot b1 \cdot P1 + b1 \cdot VC1 + c12 \cdot P2 = 0 \tag{5.10}$$

This can be rewritten to give the following pricing rule or reaction curve for Firm1:

$$P1 = \frac{a1 + b1 \cdot VC1 + c12 \cdot P2}{2 \cdot b1} \tag{5.11}$$

This equation dictates the price Firm 1 should set, given the price P2 that Firm 2 is setting. Similarly, we can derive the pricing rule (reaction curve) for Firm 2:

$$P2 = \frac{a2 + b2 \cdot VC2 + c21 \cdot P1}{2 \cdot b2} \tag{5.12}$$

The point where the two reactions curves cross determines the Nash equilibrium in prices. At that point each firm is doing the best it can, given the price its competitor has set and therefore, neither firm has the incentive to change its price.

By substituting $P2$ from equation (5.12) in Firm 1's reaction curve equation (5.11), we get:

$$P1^* = \frac{2 \cdot b2 \cdot (a1 + VC1 \cdot b1) + c12 \cdot (a2 + VC2 \cdot b2)}{4 \cdot b1 \cdot b2 - c12 \cdot c21} \tag{5.13}$$

Similarly, we get:

$$P2^* = \frac{a2 + b2 \cdot VC2 + c21 \cdot P1^*}{2 \cdot b2} \tag{5.14}$$

Nash equilibrium in prices is determined at point $(P1^*, P2^*)$.

Case 2: Constrained-Constrained

In that case both firms have limited capacity resources meaning that $Q1 \geq Cap1, Q2 \geq Cap2, P1 \leq P1^c \text{ and } P2 \leq P2^c$. The resulting profit for Firm 1 is calculated as follows:

$$\Pi 1 = (P1 - VC1) \cdot Cap1 - FC1 \qquad (5.15)$$

In that case, the firm's profit is a monotonically increasing function of price *P1*. Therefore, reaction curve for Firm 1 is given by the critical price $P1^c$.

$$P1 = \frac{a1 - Cap1 + c12 \cdot P2}{b1} \qquad (5.16)$$

and similarly for Firm 2 we have:

$$P2 = \frac{a2 - Cap2 + c21 \cdot P1}{b2} \qquad (5.17)$$

Again, Nash equilibrium in prices is determined at the point where the two reactions curves cross each other.

By substituting *P2* from equation (5.17) in Firm 1's reaction curve (5.16), we get:

$$P1^* = \frac{a1 \cdot b2 - Cap1 \cdot b2 + c12 \cdot a2 - c12 \cdot Cap2}{b1 \cdot b2 - c12 \cdot c21} \qquad (5.18)$$

and similarly:

$$P2^* = \frac{a2 - Cap2 + c21 \cdot P1^*}{b2} \qquad (5.19)$$

Nash equilibrium in prices is determined at point $(P1^*, P2^*)$.

Case 3: Unconstrained-Constrained

In that case it is assumed that Firm 1 has unlimited capacity resources (unconstrained) while Firm 2 has a limited amount of capacity resources (constrained), meaning that $Q1 < Cap1, Q2 \geq Cap2, P1 > P1^c \text{ and } P2 \leq P2^c$. The reaction curve for Firm 1 is calculated as in Case 1:

$$P1 = \frac{a1 + b1 \cdot VC1 + c12 \cdot P2}{2 \cdot b1} \tag{5.20}$$

While the reaction curve of Firm 2 is calculated as in Case 2:

$$P2 = \frac{a2 - Cap2 + c21 \cdot P1}{b2} \tag{5.21}$$

By substitution in Firm 1's reaction curve, we get:

$$P1^* = \frac{a1 \cdot b2 + b2 \cdot VC1 \cdot b1 + c12 \cdot a2 - c12 \cdot Cap2}{2 \cdot b1 \cdot b2 - c12 \cdot c21} \tag{5.22}$$

and similarly:

$$P2^* = \frac{a2 - Cap2 + c21 \cdot P1^*}{b2} \tag{5.23}$$

Therefore, Nash equilibrium in prices is determined at point $(P1^*, P2^*)$.

5.3.2 Iterative algorithm A1

In the previous section, the analytical form of the Nash equilibrium was derived for the case of price competition between two firms that manufacture and sell two substitute products. Based on the capacity resource levels of each company, three different cases were studied, namely the unconstrained-unconstrained, the constrained-constrained and the unconstrained-constrained case respectively. For each case, the closed form of the resulting Nash equilibrium in prices was calculated.

In this section, we propose an iterative algorithm [Algorithm A1] able to accommodate all the aforementioned cases and derive the Nash Equilibrium point by employing mathematical programming techniques. In every iteration of the algorithm, each company decides on its individual pricing policy while taking into account the price its competitor is currently charging (P^0). The pricing decision-making process for each firm is formulated as a non-linear programming (NLP) mathematical model that tries to maximise the company profit given the competitor's price. The optimisation problem for Firm 1 in iteration *m* is mathematically formulated as follows:

[Problem Firm 1]

$$\max \quad \Pi 1^m = (P1^m - VC1) \cdot Q1^m - FC1$$

Subject to:

Demand Constraints

$$Q1^m = a1 - b1 \cdot P1^m + c12 \cdot P2^m$$

$$Q2^m = a2 - b2 \cdot P2^m + c21 \cdot P1^m$$

Price Constraints

$$P2^m = P2^0$$

Capacity Constraints

$$Q1^m \leq Cap1$$

Similarly, the optimisation problem for Firm 2 in iteration m is formulated as follows:

[Problem Firm 2]

$$\max \quad \Pi 2^m = (P2^m - VC2) \cdot Q2^m - FC2$$

Subject to:

Demand Constraints

$$Q1^m = a1 - b1 \cdot P1^m + c12 \cdot P2^m$$

$$Q2^m = a2 - b2 \cdot P2^m + c21 \cdot P1^m$$

Price Constraints

$$P1^m = P1^0$$

Capacity Constraints

$$Q2^m \leq Cap2$$

The proposed algorithm [Algorithm A1] comprises the following steps:

[Algorithm A1]

Step 1. Set price levels to current market prices $P1=P1^0$ and $P2=P2^0$ and initialise iterations counter m:=0.

Step 2. Set iterations counter to m:=m+1. If $m>m^{max}$ then STOP.

Step 3. Solve [Problem Firm 1] & update $P1^0$ price level.

Step 4. Solve [Problem Firm 2] & update $P2^0$ price level.

Step 5. If $\frac{\Pi 1^m - \Pi 1^{m-1}}{\Pi 1^m} \leq \varepsilon$ and $\frac{\Pi 2^m - \Pi 2^{m-1}}{\Pi 2^m} \leq \varepsilon$ then STOP.

Otherwise, go to Step 2.

The proposed algorithm simulates the pricing decision-making process between two competing firms. Each firm decides on its optimal pricing policy while taking into account the observable current price charged by its competitor firm. Therefore, the algorithm is able to capture the game-theoretical nature of the pricing problem and successfully simulate the sequential decision-making process between the two firms. The algorithm terminates at a point where neither company wants to change its pricing policy given the price of its competitor. At that point both companies are doing their best, therefore neither company wants to deviate from that point and that is by definition, the Nash equilibrium point in prices. The applicability of the proposed algorithm is demonstrated by solving a motivating example as described in the following section.

5.3.3 Motivating example

Consider two firms that offer two differentiated products that are close substitutes to each other. Suppose that the two companies are facing the following demand curves:

Firm 1: $Q1 = 160 - 30 \cdot P1 + 4 \cdot P2$ and

Firm 2: $Q2 = 180 - 40 \cdot P2 + 3 \cdot P1$

The additional input data concerning the two companies is shown in Table 5.1.

Table 5.1: *Additional input data*

Parameter	Firm 1	Firm 2
Variable Cost (VC)	0.5	0.4
Fixed Cost (FC)	20	25
Capacity (Cap) in Case 1	Unlimited	Unlimited
Capacity (Cap) in Case 2	60	50
Capacity (Cap) in Case 3	Unlimited	50

Three different cases, namely Case 1, Case 2 and Case 3 are examined based on the capacity resource availability. In Case 1, both firms have unlimited capacity resources. According to the analytical form equations derived in the previous section, the Nash Equilibrium in prices is determined at point $(P1^*, P2^*) = (3.09, 2.57)$. Also, Algorithm A1 successfully predicts the exact same Equilibrium point within less than 3 iterations depending on the starting point (initial price vector) as illustrated in Figure 5.3. Most importantly, product prices converge to the same equilibrium point irrespective of the starting point, thus illustrating the robustness of the proposed methodology.

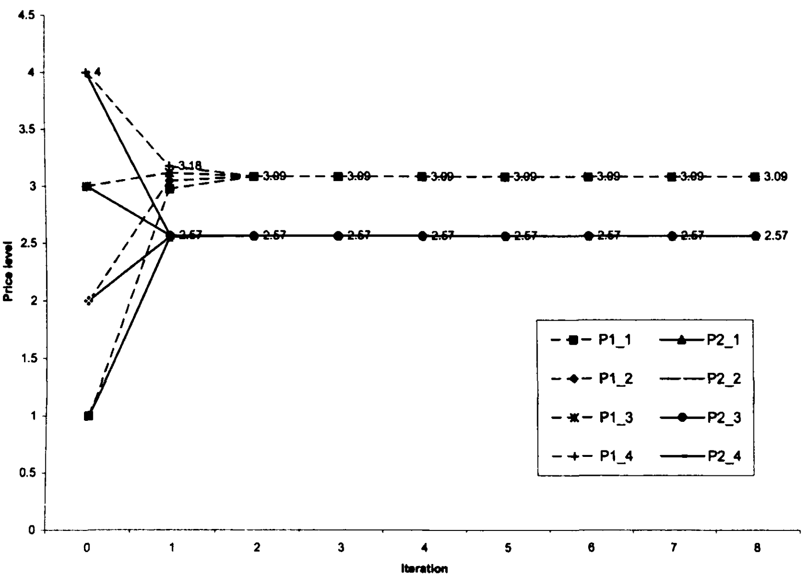

Figure 5.3: *Nash equilibrium - Case 1*

In Case 2, both firms have limited amounts of capacity resources, therefore their output levels are restricted by the resource availability of every firm. Consequently, the Nash equilibrium in prices is also influenced by the lack of unlimited resources.

According to the theoretically derived equations, Nash equilibrium in prices is now determined at point $(P1^*, P2^*) = (3.80, 3.54)$. The proposed algorithm derives the exact same equilibrium point irrespective of the initial price vector employed as shown in Figure 5.4. It is very interesting to notice that the equilibrium prices in this case are slightly higher than the equilibrium prices in Case 1. This is mainly attributed to the fact that the outputs in Case 2 are restricted to the available resource levels. At the equilibrium point, both companies make full utilisation of their resources, producing 50 and 60 units of product respectively which are less compared to the equilibrium outputs in Case 1 (77.6 and 86.6 respectively). In order to compensate for the decreased output levels, both firms are now forced to raise their prices so as to maximise their profits.

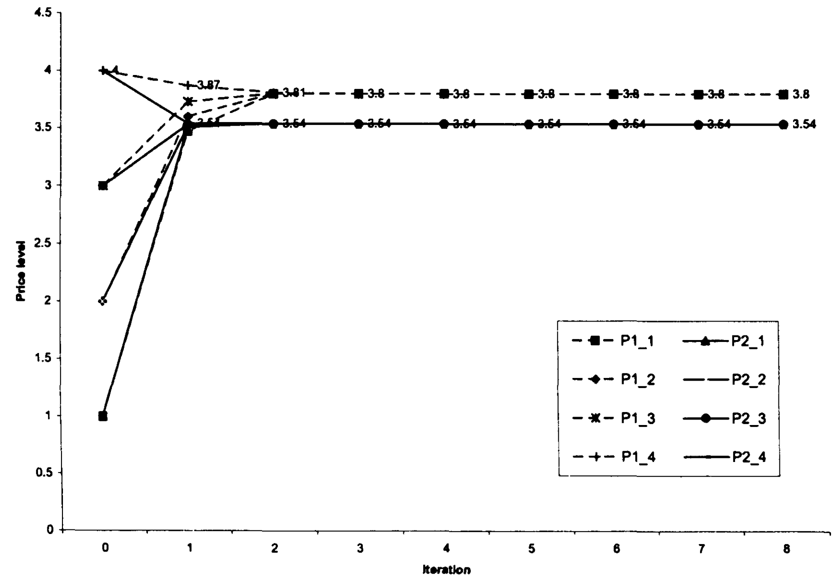

Figure 5.4: *Nash equilibrium - Case 2*

Finally in Case 3, Firm 1 has unlimited amount of capacity resource while Firm 2 has a finite level of capacity resource. The closed form equations predict that the Nash equilibrium in prices lies at point $(P1^*, P2^*) = (3.15, 3.49)$. The proposed algorithm converges at the exact same equilibrium point as shown in Figure 5.5. Unlike Cases 1 and 2, the equilibrium price for Firm 1 is now slightly lower than the price charged by Firm 2. Firm 2 has a limited capacity resource and therefore its equilibrium output is restricted to 50 product units. The lack of resources for Firm 2 is inevitably reflected on the resulting high price. On the other hand, Firm 1 is able to produce a larger output and charge a lower price for its product so as to benefit from the economy of scale. The results from Case 3 clearly illustrate what takes place in a real life marketplace, where it is very common that big companies supply large quantities in relatively low prices and consequently outrun the small companies who struggle to cover their costs by charging high prices.

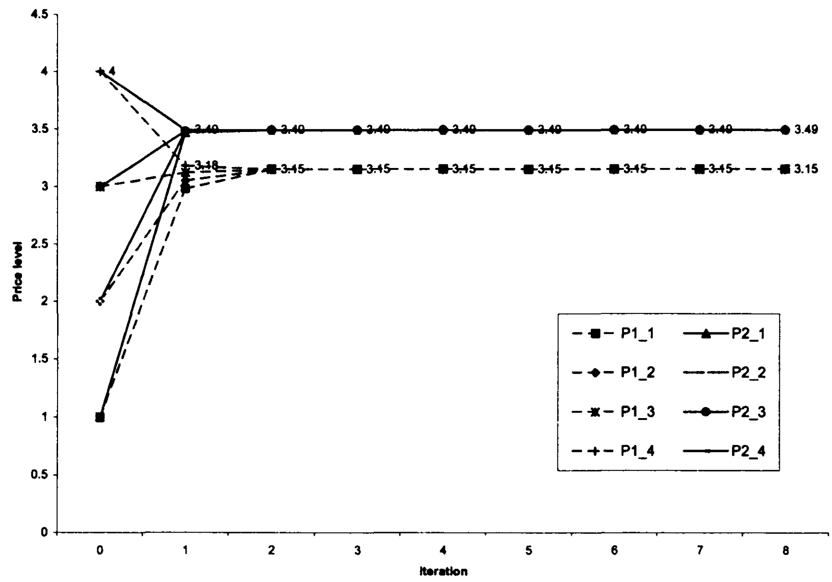

Figure 5.5: *Nash equilibrium - Case 3*

The Nash Equilibrium points for Cases 1, 2 and 3 are summarised in Table 5.2. In all three cases the proposed algorithm A1 successfully determines the exact same Nash Equilibrium point as the one predicted from the closed form equations.

Table 5.2: *Nash equilibrium points*

	Case 1		Case 2		Case 3	
	Firm 1	Firm 2	Firm 1	Firm 2	Firm 1	Firm 2
Price	3.09	2.57	3.80	3.54	3.15	3.49
Output	77.6	86.6	60	50	79.47	50
Profit	180.89	162.63	178.28	131.77	190.53	129.31

Algorithm A1 is further extended so as to accommodate the case of multi-product competing firms trying to satisfy the anticipated customer demand forecast while considering outsourcing options as described in the next section.

5.4 Multi-product price optimisation

In the previous sections, we investigated the case of price competition in a duopolistic market environment where each company is producing only one product. However, process industries nowadays usually operate multi-product plants producing a set of differentiated products (e.g. different paints, detergents, carbonated drinks, etc). These products belong to the same family of products (product class) and they share a number of common characteristics (e.g. water-based paints). On the other hand, products are differentiated from each other in such a way so as to cover a broad range of customer preferences (e.g. different paint colours/quality).

Market segmentation is a widely used marketing strategy that recognises the different ways customers perceive product value and make their purchase choices accordingly. In order to deliver value to the different existing customer segments, most companies decide to market launch a wide variety of slightly differentiated products so as to attract customers via a tailor-based marketing approach. Each one of the company products is a unique brand name with unique features that clearly differentiates itself from the rest of the family products. The unique attributes of each product appeal to a very distinct customer base that is choosing to buy that specific product over the entire range of products present in the marketplace. Customers are willing to buy their preferred product as long as the product price charged by the company reflects their perceived value of the product. Product brand loyalty is expressed by repeat purchases of the installed customer base. Alternatively, the customer may well switch to a lower-priced substitute product offered by the same company or a rival company.

Pricing in a multi-product competitive market environment is not an easy task. The analytical formulae presented in section 5.3 for the two-products pricing problem cannot be applied to the multi-product pricing problem so as to derive a meaningful Nash equilibrium in prices. In the multi-product case, price competition exists not only between company and non-company products but also between differentiated brands belonging to the very same company. Moreover, company products are manufactured by utilising a common pool of available resources. Products belonging to the same family compete with each other for scarce and shared manufacturing resources. Therefore, product cannibalisation effects have to be seriously taken into account when determining an optimal pricing policy. On the same time though, the company has to account for the market competition by considering the pricing policy adopted by the rival company for its products also present in the marketplace. On top of that modern process industries have recently realised the benefits of adopting outsource manufacturing policies in an attempt to drive manufacturing costs further down and avoid any unnecessary capacity expansion overheads. Such outsource options should be addressed in a proper manner before deciding on a comprehensive pricing strategy.

Reaction curve analysis cannot be applied in a straightforward way as in the previous two-products case. However, in order to capture the trade-off between product price and market share in a multi-product environment, an extension of the previously developed non-linear programming (NLP) mathematical model is proposed. Based on that mathematical model, an extended algorithm [Algorithm A2] is also proposed able to determine optimal pricing polices for multi-product competing companies.

5.4.1 Mathematical model

The following nomenclature is used in our mathematical model formulation:

Indices

i,j products

s production sites

r resources

Sets

P_C set of products i for company C

S_C set of production sites s for company C

R_C set of resources r for company C

Z_C set of products i using resource r at site s in company C

Parameters

a_i demand coefficient for product i

b_i demand elasticity coefficient for product i

c_{ij} demand cross-elasticity coefficient between products i and j

rvc_s relative variable cost coefficient for site s

rtc_s relative transportation cost coefficient for site s

VC_i variable manufacturing cost for product i

FC_i fixed manufacturing cost for product i

TC_i unit transportation cost for product i

OC_i unit outsource cost for product i

ρ_{ir} unit consumption coefficient for product i using resource r

A_{rs} availability level of resource r at site s

DF total market demand forecast

Variables

P_i price for product i

V_i sales volume for product i

Q_{is} amount of product i manufactured at site s

O_i amount of product i outsourced

Π_C total profit for company C

The derivation of the general mathematical model for Company C [Model M1] is described next.

Price Elasticity Constraints

The sales volume for every product i is a monotonically decreasing function of its price and a monotonically increasing function of the price of all other competing products, including substitute products belonging to Company C as well as non-company products. The sales volume for every product i is given by the following linear function:

$$V_i = a_i - b_i \cdot P_i + \sum_{j \neq i} c_{ij} \cdot P_j \qquad \forall\ i \tag{5.24}$$

Mass Balance Constraints

The total sales volume of every product i equals to the amount manufactured in-house at all production sites belonging to company C plus the amount manufacturing from outsourcing:

$$V_i = \sum_{s \in S_C} Q_{is} + O_i \qquad \forall\ i \in P_C \tag{5.25}$$

Resource Availability Constraints

The amount of company products manufactured in-house at every production site s is limited by the availability of the shared company resources. The following constraints safeguard that the resource availability levels are not exceeded.

$$\sum_{i \in Z_C} \rho_{ir} \cdot Q_{is} \leq A_{rs} \qquad \forall\ r \in R_C, s \in S_C \tag{5.26}$$

Demand Forecast Constraints

Market research surveys are conducted periodically so as to assess the current trends and predict future customer demand of a specific product class. The total sales volume of all products present in the marketplace should be greater or equal to the forecasted customer demand.

$$\sum_i V_i \geq DF \tag{5.27}$$

Objective function

The objective function employed in our mathematical model corresponds to the nett profit generated by the subset of the products belong to Company C. The nett profit is calculated as sales revenue minus the different costs, namely variable and fixed manufacturing costs, transportation and outsourcing costs. Mathematically we have:

max

$$\Pi_C = \sum_{i\in P_C} P_i \cdot V_i - \sum_{i\in P_C}\sum_{s\in S_C} rvc_s \cdot VC_i \cdot Q_{is} - \sum_{i\in P_C}\sum_{s\in S_C} rfc_s \cdot FC_i - \sum_{i\in P_C}\sum_{s\in S_C} rtc_s \cdot TC_i \cdot Q_{is} - \sum_{i\in P_C} OC_i \cdot O_i \qquad (5.28)$$

Summary of the mathematical model

In the general case, the optimisation problem for Company C is mathematically formulated as follows:

[Model M1]

max

$$\Pi_C = \sum_{i\in P_C} P_i \cdot V_i - \sum_{i\in P_C}\sum_{s\in S_C} rvc_s \cdot VC_i \cdot Q_{is} - \sum_{i\in P_C}\sum_{s\in S_C} rfc_s \cdot FC_i - \sum_{i\in P_C}\sum_{s\in S_C} rtc_s \cdot TC_i \cdot Q_{is} - \sum_{i\in P_C} OC_i \cdot O_i$$

Subject to:

$$V_i = a_i - b_i \cdot P_i + \sum_{j\neq i} c_{ij} \cdot P_j \qquad \forall\ i$$

$$V_i = \sum_{s\in S_C} Q_{is} + O_i \qquad \forall\ i \in P_C$$

$$\sum_{i\in Z_C} \rho_{ir} \cdot Q_{is} \le A_{rs} \qquad \forall\ r \in R_C, s \in S_C$$

$$\sum_i V_i \ge DF$$

5.4.2 Iterative algorithm A2

The extended algorithm [Algorithm A2] comprises the following steps:

```
[Algorithm A2]

Step 1. Set price levels to current market prices
```
$P_i = P_i^0$
```
     and initialise iterations counter m:=0.
```

Step 2. Set iterations counter to m:=m+1. If $m>m^{max}$ then STOP.

Step 3. For every company C present in the marketplace:

i. Solve [Model M1] for company C

ii. Update price levels for all products belonging to company C *only*.

Step 4. If $\frac{\Pi_C^m - \Pi_C^{m-1}}{\Pi_C^m} \leq \varepsilon$ for all companies then STOP.

Otherwise, go to Step 2.

Clearly, the restrictions imposed by the analytical formulae and Algorithm A1 are now alleviated. Extended Algorithm A2 is able to accommodate the case of oligopolistic market competition where more than two firms are competing. Furthermore, every company present in the marketplace is allowed to manufacture in-house and/or outsource more than one product. The proposed generalised algorithm is able to determine optimal production policies and prices for all products, as it is demonstrated by the illustrative example described in the next section.

5.4.3 Illustrative examples

Consider two firms, namely Company A and Company B that manufacture and sell products *P1, P2, P3, P4* and *P5, P6, P7* respectively as shown in Figure 5.1. Products *P1-P7* are close substitutes to each other and therefore each product has a unique demand function curve associated with it, as described by equation (5.24). Demand coefficients include parameter α_i, elasticity b_i and cross-elasticity c_{ij} parameters as shown in Tables 5.3 and 5.4. Every company has two available manufacturing sites (sites 1 & 2 for Company A and sites 3 & 4 for company B). The products can be manufactured in-house by using shared manufacturing resources available at each site (in-house manufacturing). Resource utilisation coefficients for every product are given at Table 5.5 while resource availability levels for every resource at each site are given at Table 5.6. Manufacturing sites are geographically distributed facilities, therefore relative manufacturing cost and transportation cost coefficient are used so as to capture the effect of different manufacturing locations (see Table 5.7). Final products are transported from the manufacturing sites to the end-customers at a given transportation cost TC_i. Alternatively, a certain amount of production can be outsourced by a third-party company at a given outsource cost OC_i. Note that since we are dealing with products belonging to the same product class, fixed costs are assumed to be the same for all products, therefore they are not considered explicitly in the illustrative example.

Given an initial price vector (current market prices P_i^0) for all products, the problem is to determine optimal product prices, output levels and outsource amounts so as to derive a comprehensive Nash equilibrium point for companies A and B that neither company would wish to deviate from.

Table 5.3: *Product input data*

Product	a_i	b_i	VC_i	TC_i	OC_i	P_i^0
P1	160	25	4	1	5.3	8
P2	200	30	3	1	5.3	10
P3	150	25	4	1	5.3	9
P4	120	20	5	1	5.3	11
P5	170	30	3	1	4.2	9
P6	110	25	4	1	4.2	10
P7	180	30	3	1	4.2	8

Table 5.4: *Cross-elasticity parameters (c_{ij})*

Product	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*
P1	-	4	3	2	6	7	4
P2	2	-	5	3	5	4	2
P3	3	3	-	2	3	2	5
P4	4	2	4	-	2	6	3
P5	2	4	3	5	-	3	2
P6	5	3	4	3	2	-	3
P7	3	3	2	2	4	3	-

Table 5.5: *Resource utilisation coefficients (ρ_{ir})*

Product	*res1*	*res2*	*res3*	*res4*	*res5*	*res6*
P1	1	1.1	0.8	-	-	-
P2	0.7	1.2	0.7	-	-	-
P3	1.2	1.4	0.9	-	-	-
P4	1.1	1.3	0.4	-	-	-
P5	-	-	-	0.9	1.2	0.7
P6	-	-	-	0.8	1.4	0.6
P7	-	-	-	1	1.6	0.8

Table 5.6: *Initial resource availability levels across manufacturing sites*

Resources	*Site 1*	*Site 2*	*Site 3*	*Site 4*
Res1	200	340	-	-
Res2	300	370	-	-
Res3	150	270	-	-
Res4	-	-	140	150
Res5	-	-	210	220
Res6	-	-	110	130

Table 5.7: *Manufacturing sites related data*

Manuf. Site	rvc_s	rtc_s
Site 1	1	1
Site 2	0.8	1.2
Site 3	0.7	1.4
Site 4	0.9	1.1

Initially, both companies A and B manufacture their products in-house by only relying on the manufacturing capabilities of their own production sites while no outsourcing is considered. In particular, the allocation of production between the different sites is shown in Tables 5.8 and 5.9. The total amount of sales for the specific product class equals the combined manufacturing volume of both companies (703 units). Given the initial price vector and output levels for all products, the initial profit is 2194 rmu* and 1129 rmu for company A and B, respectively.

* rmu = relative monetary units

Table 5.8: *Initial state of Company A*

Products	*Site 1*	*Site 2*	*Outsource*
P1	-	205	-
P2	95	-	-
P3	68	20	-
P4	-	90	-
Total	163	315	0

Table 5.9: *Initial state of Company B*

Products	*Site 3*	*Site 4*	*Outsource*
P5	84	-	-
P6	41	-	-
P7	32	68	-
Total	157	68	0

A recent market research survey has estimated future customer demand for the product family under investigation to be equal to 712 units and therefore the two companies are competing over the anticipated customer demand. Every company has the strategic choice to consider outsourcing options or rely entirely on its own in-house manufacturing capabilities, thus resulting in four distinct cases as explained in the following sections. All four cases were implemented in GAMS (Brooke *et al.*, 1998) using the CONOPT NLP solver (Drud, 1985) while all runs were performed on an IBM RS/6000 workstation.

Case 1: In-house/In-house

In this case both companies A and B manufacture their products in-house while no outsourcing is allowed to take place. Algorithm A2 is applied with the outsource variable fixed to zero for both companies. As shown in Figure 5.6, Nash equilibrium is reached after 5 iterations resulting in profits 2259 rmu and 1262 rmu for company A and B, respectively. Optimal product price levels are determined as illustrated in Figure 5.7. More specifically, equilibrium prices for *P1* and *P3* lie above their original levels while a price decline is suggested for products *P2*, *P5* and *P6*. Finally, the optimal prices of products *P4* and *P7* are very close to their original values. Nash Equilibrium profits and prices for both companies across all cases are given at Table 5.10.

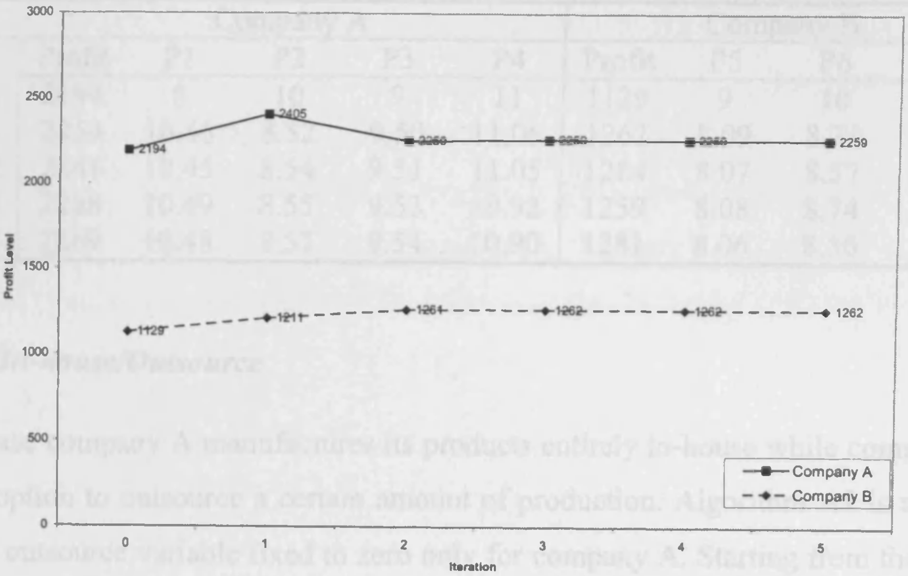

Figure 5.6: *Nash equilibrium - Case 1*

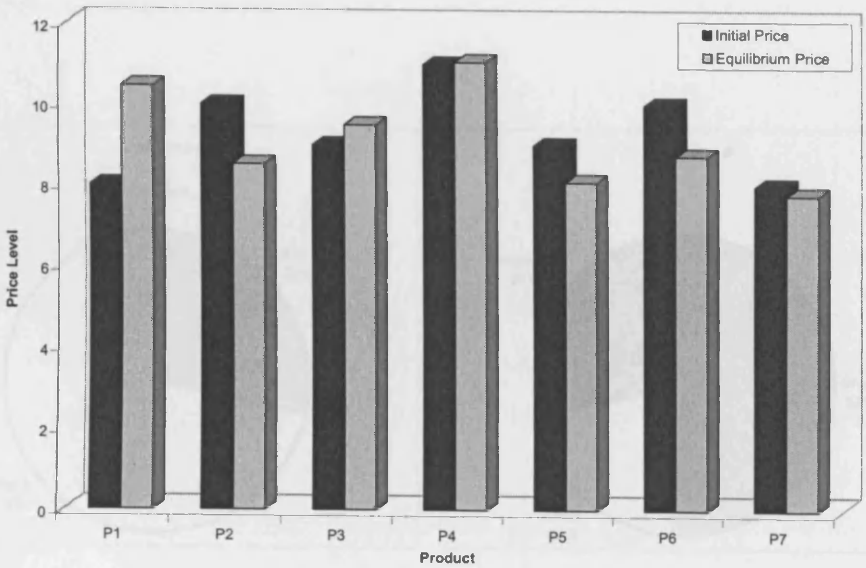

Figure 5.7: *Product prices - Case 1*

Table 5.10: *Nash equilibrium results across all cases*

	Company A					Company B			
	Profit	P1	P2	P3	P4	Profit	P5	P6	P7
Initial	2194	8	10	9	11	1129	9	10	8
Case 1	2259	10.46	8.52	9.50	11.06	1262	8.09	8.74	7.78
Case 2	2241	10.45	8.54	9.51	11.05	1284	8.07	8.57	7.78
Case 3	2288	10.49	8.55	9.53	10.92	1259	8.08	8.74	7.77
Case 4	2269	10.48	8.57	9.54	10.90	1281	8.06	8.56	7.77

Case 2: In-house/Outsource

In this case company A manufactures its products entirely in-house while company B has the option to outsource a certain amount of production. Algorithm A2 is applied with the outsource variable fixed to zero only for company A. Starting from the same initial state as in case 1, Nash equilibrium results in profits 2241 rmu and 1284 rmu for company A and B respectively while optimal product prices follow the same trends as in case 1 and their values are given in Table 5.10. However the main difference in this case is that Company B outsources 84 units of product *P6* while the rest of the production is taking place in its own manufacturing sites 3 and 4 as shown in Figure 5.8.

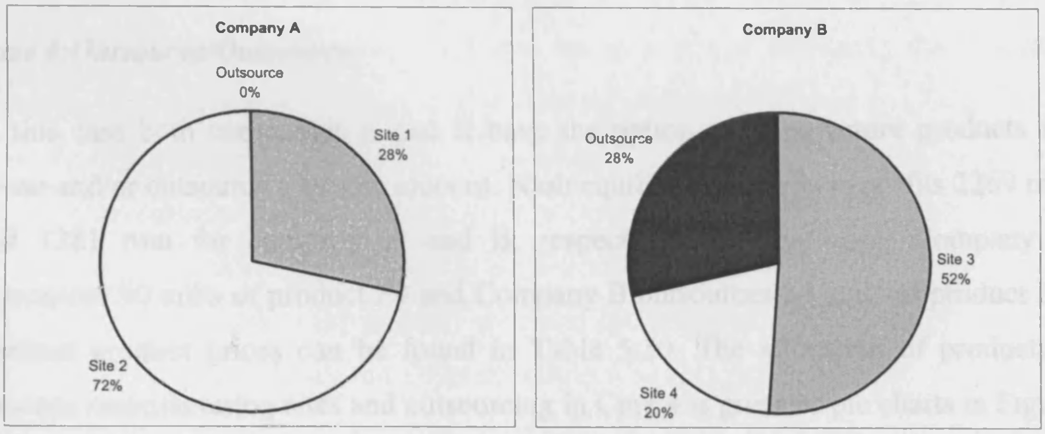

Figure 5.8: *Allocation of production - Case 2*

Case 3: Outsource/In-house

This case is the exact inverse of case 2. Company B manufactures all of its products in-house while company A has the option to outsource a certain amount of production. According to the results, profits of 2288 rmu and 1259 rmu for company A and B respectively are achieved at the Nash equilibrium point while optimal product prices are given in Table 5.10. Notice that in this case Company A outsources 91 units of product *P4* representing 22% of its total production as depicted in Figure 5.9.

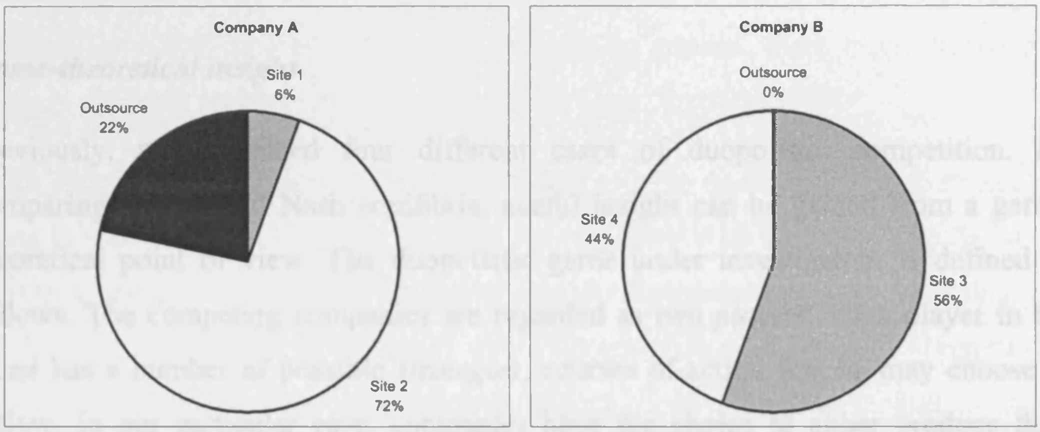

Figure 5.9: *Allocation of production - Case 3*

Case 4:Outsource/Outsource

In this case both companies A and B have the option to manufacture products in-house and/or outsource a certain amount. Nash equilibrium results in profits 2269 rmu and 1281 rmu for company A and B, respectively. In this case, Company A outsources 90 units of product *P4* and Company B outsources 84 units of product P6. Optimal product prices can be found in Table 5.10. The allocation of production between manufacturing sites and outsourcing in Case 4 is given as pie charts in Figure 5.10 for both companies. Notice that the largest share of production is allocated to site 2 and site 3 since they both offer low variable manufacturing cost compared with sites 1 and site 4 respectively. According to the obtained results, outsourcing activity constitutes over 20% of total production for both companies.

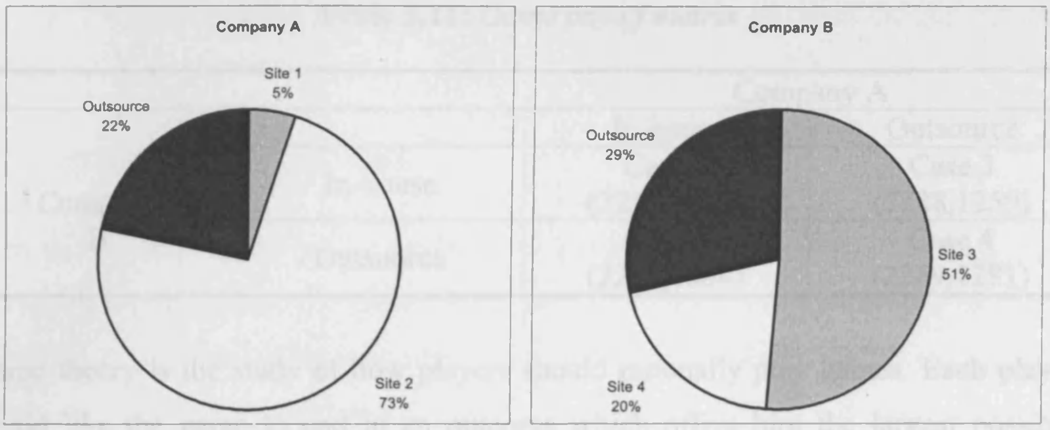

Figure 5.10: *Allocation of production - Case 4*

Game-theoretical insight

Previously, we examined four different cases of duopolistic competition. By comparing the derived Nash equilibria, useful insight can be gained from a game-theoretical point of view. The duopolistic game under investigation is defined as follows. The competing companies are regarded as two *players*. Each player in the game has a number of possible *strategies*, courses of action that he may choose to follow. In our particular case, companies have the choice to either produce their products entirely in their own manufacturing sites (in-house strategy) or produce a certain amount in-house and also outsource a certain percentage of production (outsource strategy). The strategies chosen by each player determine the so-called *outcomes* of the game. In our example, we end up with four different combinations of outcomes, namely in-house/in-house, in-house/outsource, outsource/in-house and outsource/outsource, each one representing a case examined in the previous sections. In every formally stated game, there is a collection of numerical *payoffs*, one to each player, associated with every possible outcome of the game. Those payoffs represent the value of the outcome to the different players. In this example, Nash equilibrium profits can play the role of companies payoffs for every particular case examined. Overall, we are dealing with a two-person game with two strategies per player and a *game payoff matrix* as shown in Table 5.11. The values in parentheses are the Nash equilibrium profits determined previously for all four cases, with the first number being the profit for Company A and the second one the profit of Company B.

Table 5.11: *Game payoff matrix*

		Company A	
		In-house	Outsource
Company B	In-house	Case 1 (2259,1262)	Case 3 (2288,1259)
	Outsource	Case 2 (2241,1284)	Case 4 (2269,1281)

Game theory is the study of how players should rationally play games. Each player would like the game to end in an outcome which offers him the largest possible payoff. He has some control over the outcome, since his choice of strategy will influence it. However, the outcome is not determined by his choice alone, but also depends upon the choices of all other players. In general, there might be conflict because different players value outcomes differently (Straffin, 1993).

In the illustrative example, companies are faced with the question of which strategy to adopt in order to reach the Nash Equilibrium associated with the highest profit for the company under investigation. First, let us consider Company A. Company A does not have any indication of which policy rival Company B will adopt. If Company B adopts a strictly in-house manufacturing policy, then Company A has a choice between Case 1 and 3. Since the profit for Company A in Case 3 is higher than the one in Case 1 (2288 vs. 2259), Company A decides to adopt an outsourcing strategy. In case Company B adopts an outsourcing policy, Company A has a choice between Case 2 and 4. Case 4 offers Company A with a profit of 2269 which is higher that the Case 2 profit (2241). So, in both scenarios, Company A is better off by choosing to outsource a certain amount of its production, irrespective of the production policy adopted by rival Company B. Similarly, we can prove that the exact same rule applies for Company B as well. Without any prior knowledge of the production policy adopted by company A, Company B always earns a higher profit by adopting an outsourcing strategy.

It is very interesting to notice that Case 3 provides the highest profit for Company A while Case 2 provides the highest profit for Company B. However, Case 4 is considered to be the most likely outcome of the game since the outsource/outsource policy guarantees higher profits for both companies no matter what policy the rival company decides to adopt, thus providing a robust Nash equilibrium for both companies.

5.5 Concluding remarks

A systematic mathematical programming approach for active demand management through price optimisation was presented in this chapter. First, we derived analytical formulae for calculating Nash equilibrium points in a duopolistic market environment where each company produces and sells only one product. An iterative algorithm was then proposed able to simulate the decision-making process and derive the exact same equilibrium points predicted by the closed-form formulae as demonstrated by a motivating example.

Following that, the proposed algorithm was further extended in order to accommodate the case of multi-product firms and also consider additional features such as customer demand forecast and mixed in-house and outsourcing production policies. An illustrative example was solved in order to demonstrate the applicability of the proposed methodology across four different cases. Finally, a comparison among the different cases provided us with valuable game-theoretical insight concerning the problem of duopolistic competition coupled with outsourcing options.

Chapter 6

Conclusions and future directions

The aim of the thesis was *to facilitate business decision-making by applying mathematical programming techniques for multi-site capacity planning and business optimisation in process industries*. Towards that goal, a number of mathematical models and solution algorithms have been developed in order to assist business decision-making in process industries. The key contributions of the thesis are summarised in the next section, while section 6.2 suggests promising new directions for future research work.

6.1 Contributions of the thesis

6.1.1 Multi-site capacity planning

Part I of the thesis was concerned with the problem of multi-site capacity planning under uncertainty in the pharmaceutical industry. First, a simultaneous approach for multi-site capacity planning was developed. An extensive literature survey was presented in order to familiarise the reader with the current status of the pharmaceutical industry and highlight recent academic contributions in the area of mathematical programming related to the problem of capacity planning. Following the literature survey, we focused on the specific characteristics of the multi-site capacity planning problem under uncertainty and the overall problem was formally stated. According to the simultaneous approach, the overall problem was formulated as a comprehensive two-stage, multi-scenario mixed-integer linear mathematical programming (MILP) model incorporating issues related to product management, clinical trials uncertainty, capacity management and trading structure of the company. The proposed MILP model was solved using traditional branch-and-bound techniques via the direct application of the commercially available XPRESS-MP (Dash Associates, 1999) MILP solver. The applicability of the simultaneous approach was illustrated by five illustrative examples of varying sizes. According to the results, small and modest size example problems were solved successfully by employing the proposed methodology. However, the solution of large-scale instances of the problem proved to be a very demanding task in terms of computational effort needed due to the inherent complexity of the problem.

The previously revealed computational limitations of the simultaneous approach necessitated the implementation of an alternative solution methodology. For that reason, a hierarchical approach was also developed able to accommodate the combinatorial nature of the problem. The proposed hierarchical approach was based on the decoupling of the strategic and operational decision-making levels identified in our problem via the employment of an efficient suite-aggregate model formulation. The aforementioned five illustrative examples were revisited in order to validate the applicability of the proposed hierarchical solution algorithms H1 and H2. According to the results, the proposed algorithms were able to alleviate the computational burden and yield near-optimal solutions within reasonable computational time even for large-scale MILP problems. A comparative study was also performed in order to assess the computational performance of the proposed hierarchical algorithms against the simultaneous approach and a previously developed hierarchical algorithm found in the literature (Rotstein *et al.*, 1999). In both cases, the proposed hierarchical algorithms outperformed both the simultaneous approach and the previously developed algorithm by achieving significant computational savings and most importantly, without compromising the solution quality.

6.1.2 Business optimisation

Part II of the thesis was concerned with the problem of business optimisation for customer demand management. Initially, we highlighted recent approaches in the area of demand forecasting and identified their major strengths and weaknesses. Based on the recently developed statistical learning theory (Vapnik, 1998), we then presented the essential mathematical background of support vector regression (SVR) analysis. A systematic mathematical programming approach for customer demand forecasting was proposed based on SVR analysis. The proposed approach employed a three-step algorithm able to extract information from historical data training points and identify an adaptive basis regression function before carrying out a recursive methodology for customer demand forecasting. Three illustrative customer demand forecasting examples were solved so as to validate the applicability of the proposed algorithm. According to the results, the proposed algorithm successfully handled complex nonlinear customer demand patterns and derived forecasts with prediction accuracy of more than 93% in all cases.

A mathematical programming approach for active demand management of close substitute products through price optimisation was also developed. The proposed methodology identified price as the ultimate driver behind customer purchasing behaviour and aimed to maximise company profits while taking into consideration manufacturing costs, resource availability, customer demand elasticity, outsourcing and market competition. Analytical formulae for calculating Nash equilibrium points in a duopolistic market environment were derived. An iterative optimisation-based algorithm was then proposed able to simulate the decision-making process and obtain the exact same Nash equilibrium points as the ones predicted by the closed-form formulae. The proposed algorithm was then extended so as to accommodate the case of multi-product firms and consider additional features such as customer demand forecasts and mixed in-house and outsourcing production policies. Four illustrative example cases were solved in order to demonstrate the applicability of the proposed methodology while a comparison among them provided us with some valuable game-theoretical insight concerning the problem of duopolistic competition coupled with outsourcing options.

6.2 Recommendations for future work

A number of promising future research directions related to multi-site capacity planning and business optimisation problems are presented in this section. Our aim is to provide the reader with some future insight in the area of multi-site capacity planning and business optimisation as well as highlight a number of emerging research issues that could benefit from the developed mathematical modelling frameworks presented in the thesis. Next, we consider those future research issues in detail.

6.2.1 Tax minimisation

In Chapter 2 of the thesis a mathematical model for multi-site capacity planning was proposed that successfully incorporated issues related to the trading structure of the company. The proposed model employed a priori given tariff percentages parameters for the cost-minus and profit-plus pricing formulas of the production sites and sales regions respectively. Alternatively, those tariffs could be treated as variables to be determined by the optimisation algorithm. By doing so, the mathematical model would become more flexible in terms of where and how much product to manufacture so as to maximise the after-tax expected nett present value (eNPV) by properly adjusting financial flows between the various business centres resulting in minimum overall taxes paid by the company.

From a mathematical point of view, the aforementioned modification to the proposed mathematical model would add more degrees of freedom into the optimisation problem and therefore the objective function (eNPV) would become less constrained. In that fashion, the optimisation algorithm would be able theoretically to derive a solution with increased after-tax profitability by performing a tax minimisation based on the derived internal trading structure of the company.

6.2.2 Multi-objective optimisation

The proposed mathematical model for multi-site capacity planning under uncertainty employed a profit maximisation objective function (eNPV). However, it would be very beneficial to incorporate multiple business performance criteria (see for example Hugo *et al.*, 2003). Apart from the profit-based eNPV, the proposed mathematical model could also incorporate financial risk assessment criteria so as to minimise the effect of adverse scenarios realisation.

Towards that direction, a multi-objective approach could leverage high-profit aspirations with low-risk capital investment and operating policies (see for example Barbaro and Bagajewicz, 2004). Solutions lying on a pareto optimal curve could be obtained in order to resolve the trade-off between profit and risk in an optimal manner. Depending on the managerial view towards risk, the company would then be able to choose from a wide range of risk-averse or risk-taking multi-site investment strategies.

6.2.3 Training of support vector machines

Chapter 4 of the thesis described a support vector regression-based methodology for customer demand forecasting while illustrative examples provided some very promising results. However, the field of support vector machines training constitutes a very active research area with numerous open issues still seeking for an answer.

In particular, parameter selection for support vector regression is an issue of crucial importance. In our proposed methodology, support vector regression parameters were determined based on heuristic rules recently proposed in the literature (Cherkassky and Ma, 2004). However, a more formal way of parameter selection is needed in order to make support vector regression less dependant on heuristic-based rules.

Scholkopf *et al.* (2000) have proposed a new class of support vector algorithms for classification and regression problems. Based on the original ε-SVR formulation, they developed the ν-SVR formulation. Instead of choosing the width parameter of the ε-insensitive tube, they introduced parameter ν that effectively allows control of the number of support vectors while eliminating the need to determine parameter ε explicitly. Chalimourda *et al.* (2004) have recently proposed theoretically optimal values for the ν parameter, although these values have derived under strong theoretical assumptions that are not satisfied in practical support vector machines.

According to a recent study by Smola and Scholkopf (2004), reduced set methods is an important topic for speeding up training of support vector machines. Platt (1999) recently proposed a sequential minimal optimisation (SMO) algorithm for fast training of support vector machines. However, at the moment, data mining application require algorithms that are able to deal with databases that are often at least one order of magnitude larger than the current practical size of support vector regression (Smola and Scholkopf, 2004).

6.2.4 Marketing-based pricing

Chapter 5 of the thesis presented the case of both single-product and multi-product firms competing in an oligopolistic marketplace. The proposed methodology took into account manufacturing costs, resource availability, customer demand elasticity, outsourcing and market competition to determine optimal pricing policies.

However, it would be very interesting to consider a marketing-based pricing approach for oligopolistic competition where issues related to product lifecycles are incorporated as well. Instead of determining static price levels, a marketing-based approach should adopt a dynamic perspective of oligopolistic competition where product price profiles could be proposed over a short to medium-term horizon. Such an approach could incorporate issues related to product advertising and promotions and also consider the special cases of new product development (NPD) and market product launching (MPL). Such a marketing-based pricing tool would help process industries not only to attract new customers, but on the same time, fairly reward existing customers via the deployment of properly envisaged loyalty brand schemes.

6.2.5 Customer demand uncertainty

The active demand management approach, presented in Chapter 5, employed deterministic values for customer demand elasticity and cross-elasticity parameters. Another interesting aspect of future research work would be to address the uncertain nature of those parameters and assess their impact in a more formal way. Towards that direction, different values of elasticity and cross-elasticity parameters could be used resulting into a large number of possible market scenarios so as to map the entire space of uncertainty.

Alternatively, chance-constrained programming (CCP) techniques (Charnes and Cooper, 1959) could be employed in order to incorporate parameter uncertainty in a more affordable way, in terms of computational effort, by deviating the need for discrete scenarios. In that fashion, chance constraints could capture efficiently the effect of uncertain elasticity and cross-elasticity parameters and assess company performance across a wide range of possible market conditions.

References

1. Acutt, M.Z. and Dodgson, J.S., (1996), Cross-elasticities of demand for travel, *Transport Policy* **4**, 271-277.

2. Agrawal, M., Jade, A.M., Jayaraman, V.K. and Kulkarni, B.D., (2003), Support vector machines: a useful tool for process engineering applications, *Chemical Engineering Progress* **99**, 57-62.

3. Aizerman, M.A., Braverman, E.M. and Rozoner, L.I., (1964), Theoretical foundations of the potential function method in pattern recognition learning, *Automation and Remote Control* **25**, 821-837.

4. Alves, D.C.O. and Bueno, R.D., (2003), Short-run, long-run and cross-elasticities of gasoline demand in Brasil, *Energy Economics* **25**, 191-199.

5. Arntzen, B.C., Brown, G.C., Harrison, T.P. and Trafton, L.L., (1995), Global supply chain management at Digital Equipment Corporation, *Interfaces* **25**, 69-93.

6. Barbaro, A.F. and Bagajewicz, M., (2004), Managing financial risk in planning under uncertainty, *AIChE Journal* **50**, 963-989.

7. Bertrand, J., (1983), Theorie mathematique de la richesse sociale, *Journal des Savants*, 499-508.

8. Bertsekas, D.P., *Nonlinear Programming*, Athena Scientific, 1995.

9. Bhat, N.V. and McAvoy, T.J., (1992), Determining model structure for neural models by network stripping, *Computers and Chemical Engineering* **16**, 271-181.

10. Birge, J.R., Drogosz, J. and Duenyas, I., (1998), Setting single-period optimal capacity levels and prices for substitutable products, *The International Journal of Flexible Manufacturing Systems* **10**, 407-430.

11. Blau, G., Mehta, B., Bose, S., Pekny, J.F., Sinclair, G., Keunker, K. and Bunch, P., (2000), Risk management in development of new products in highly regulated industries, *Computers and Chemical Engineering* **24**, 659-664.

12. Box, G.E.P. and Jenckins, G.M., *Time series analysis, forecasting and control*, Holden-Day, 1969.

13. Brooke, A., Kendrick, D., Meeraus, A. and Raman, R., *GAMS: A user's guide*, GAMS Development Corporation, Washington, 1998.

14. Burges, C.J.C., (1998), A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* **2**, 121-167.

15. Chalimourda, A., Scholkopf, B. and Smola, A.J., (2004), Experimentally optimal ν in support vector regression for different noise models and parameters settings, *Neural Networks* **17**, 127-141.

16. Chang, C.-C. and Lin C.-J., *LIBSVM: a library for support vector machines*, 2001. Software available at website: http://www.csie.ntu.edu.tw/~cjlin/libsvm

17. Charnes, A. and Cooper, W.W., (1959), Chance-constrained programming, *Management Science* **6**, 73-79.

18. Cheng, L., Subrahmanian, E. and Westerberg, A.W., (2003), Design and planning under uncertainty: issues on problem formulation and solution, *Computers and Chemical Engineering* **27**, 781-801.

19. Cherkassky, V. and Mulier, F., *Learning from data: concepts, theory and methods*, John Wiley and Sons, 1998.

20. Cherkassky, V. and Ma, Y.Q., (2004), Practical selection of SVM parameters and noise estimation for SVM regression, *Neural Networks* **17**, 113-126.

21. Chiang, L.H., Kotanchek, M.E. and Kordon, A.K., (2004), Fault diagnosis based on Fisher discriminant analysis and support vector machines, *Computers and Chemical Engineering* **28**, 1389-1401.

22. Cournot, A., *Recherches sur les principes mathematiques de la theorie des richesses*, Hachette, Paris, 1938.

23. Coy, P., (2000), The power of smart pricing, *Business Week*, Economics, 10th April.

24. Dash Associates XPRESS-MP v12.13, *User's Manual*, Dash Associates, UK, 1999.

25. Drud, A.S., (1985), CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems, *Mathematical Programming* **31**, 153-191.

26. EFPIA-European Federation of Pharmaceutical Industries and Associations, Website: http://www.efpia.org/

27. Eliashberg, J. and Steinberg, R., (1987), Marketing-production decisions in an industrial channel of distribution, *Management Science* **33**, 981-1000.

28. Floudas, C.A., *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*, Oxford University Press Inc., 1995.

29. Forrester, J.W., (1958), Industrial dynamics: A major breakthrough for decision makers, *Harvard Business Review* **38**, 37-66.

30. Foster, W.R., Collopy, F. and Ungar, L.H., (1992), Neural network forecasting of short noisy time series, *Computers and Chemical Engineering* **16**, 293-297.

31. Gatica, G., Shah, N. and Papageorgiou, L.G., (2001), Capacity planning under clinical trials uncertainty for the pharmaceutical industry, *In Proceedings of ESCAPE-11: 11th European Symposium on Computer-Aided Process Engineering*, Gani, R. and Jorgensen, S.B. (Editors), Elsevier, 865-870.

32. Gatica, G., Shah N. and Papageorgiou L.G., (2003), An aggregation approach for capacity planning under uncertainty for the pharmaceutical industry, *In Foundations of Computer-Aided Process Operations (FOCAPO)*, Vol. 4, Grossmann, I.E. and McDonald, C.M. (Editors), CACHE, 245-248.

33. Gunn, S.R., *Support vector machines for classification and regression*, Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton, 1998.

34. Gupta, A. and Maranas, C.D., (1999), A hierarchical lagrangean relaxation for solving midterm planning problems, *Industrial and Engineering Chemistry Research* **38**, 1937-1947.

35. Gupta, A. and Maranas, C.D., (2000), A two-stage modelling and solution framework for multi-site midterm planning under demand uncertainty, *Industrial and Engineering Chemistry Research* **39**, 3799-2813.

36. Hugo, A., Ciumei, C., Buxton, A. and Pistikopoulos, E.N., (2003), Environmental impact minimisation through material substitution: a multi-objective optimisation approach, *In Proceedings ESCAPE-13: 13th European Symposium on Computer-Aided Process Engineering*, Kraslawski A. and Turunen, I. (Editors), Elsevier, 683-688.

37. Ierapetritou, M.G. and Pistikopoulos, E.N., (1994), Novel optimization approach of stochastic planning models, *Industrial and Engineering Chemistry Research* **33**, 1930-1942.

38. ILOG CPLEX v6.5, *User's Manual*, ILOG, France, 1999.

39. Jain, V. and Grossmann, I.E., (1999), Resource-constrained scheduling of tasks in new product development, *Industrial and Engineering Chemistry Research* **38**, 3013-3026.

40. Kallrath, J. and Wilson, J.M., *Business optimisation using mathematical programming*, Macmillan Press Ltd., London, 1997.

41. Kotler, P., Armstrong, G., Saunders, J. and Wong, V., *Principles of Marketing*, Prentice Hall, Englewood Cliffs, 1996.

42. Kulkarni, A., Jayaraman, V.K. and Kulkarni, B.D., (2004), Support vector classification with parameter tuning assisted by agent-based technique, *Computers and Chemical Engineering* **28**, 311-318.

43. Lanning, S., Mitra, D., Wang, Q. and Wright, M., (2000), "Optimal planning for optical transport networks", *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* **358**, 2183-2196.

44. Lasschuit, W. and Thijssen, N., (2004), Supporting supply chain planning and scheduling decision in the oil and chemistry industry, *Computers and Chemical Engineering* **28**, 863-870.

45. Liu, M.L. and Sahinidis, S.V., (1996), Optimization in process planning and uncertainty, *Industrial and Engineering Chemistry Research* **35**, 4154-4165.

46. Makridakis, S. and Wheelwright, S.C., *Interactive Forecasting: univariate and multivariate methods*, Holden-Day Inc., 1978.

47. Makridakis, S. and Hibon, M., (1979), Accuracy of forecasting: An empirical investigation, *Journal of the Royal Statistical Society A* **142**, 97-125.

48. Makridakis, S. and Wheelwright, S.C., *The handbook of forecasting: A manager's guide*, John Wiley and Sons, 1982.

49. Manugistics Inc., Website: http://www.manugistics.com/

50. Maravelias, C.T. and Grossmann, I.E., (2001), Simultaneous planning for new product development and batch manufacturing facilities, *Industrial and Engineering Chemistry Research* **40**, 6147-6164.

51. Mercer, J., (1909), "Functions of positive and negative type and their connection with the theory of integral equations", *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* **209**, 415-446.

52. Min, H. and Zhou, G., (2002), Supply Chain Modeling: past, present and future, *Computers and Industrial Engineering* **43**, 231-249.

53. Mulvey, J.M. and Ruszczynski A.J., (1995), A new scenario decomposition method for large-scale stochastic optimization, *Operations Research* **43**, 477-490.

54. Munir, A. and Benson R., *Benchmarking in the process industries*, IChemE, Rugby, 1999.

55. Myasnikova, E., Samsonova, A., Samsonova, M. and Reinitz, J., (2002), Support vector regression applied to the determination of the developmental age of a Drosophila embryo from its segmentation gene expression patterns, *Bioinformatics* **18**, S87-S95.

56. Nagle, T. and Holden, R., The strategy and tactics of pricing: A guide to profitable decision making, 2nd Ed., Prentice Hall, Englewood Cliffs, 1995.

57. Nash, J.F., (1951), Non-cooperative games, Annals of Mathematics **54**, 286-295.

58. Papageorgiou, L.G., Rotstein, G.E. and Shah, N., (2001), Strategic supply chain optimization for the pharmaceutical industries, *Industrial and Engineering Chemistry Research* **40**, 275-286.

59. Pashigian, B.P., *Price theory and applications*, 2nd Ed., McGraw-Hill, Boston, 1998.

60. Pindyck, R.S. and Rubinfield, D.L., *Microeconomics*, Prentice Hall, Upper Saddle River, 1992.

61. Pistikopoulos, E.N., (1995), Uncertainty in process design and operations, *Computers and Chemical Engineering* **19**, S553-S563.

62. Platt, J.C., (1999), Fast training of support vector machines using sequential minimal opimization, *In Advances of Kernel Methods-Support Vector Learning*, Scholkopf, B., Burges, C.J.C. and Smola, A.J. (Editors), Cambridge, MA, MIT Press, 185-208.

63. Prakasvudhisarn, C., Trafalis, T.B. and Raman, S., (2003), Support vector regression for determination of minimum zone, *Journal of Manufacturing Science and Engineering* **125**, 736-739.

64. Rapt Inc., Website: http://www.rapt.com/

65. Reeves, C.R. (Editor), *Modern heuristic techniques for combinatorial problems*, McGraw-Hill, London, 1995.

66. Rogers, M.J., Gupta, A. and Maranas, C.D., (2002), Real options based analysis of optimal pharmaceutical research and development portfolios, *Industrial and Engineering Chemistry Research* **41**, 6607-6620.

67. Roslof, J., Harjunkoski, I., Westerlund, T. and Isaksson, J., (2002), Solving a large-scale industrial scheduling problem using MILP combined with a heuristic procedure, *European Journal of Operational Research* **138**, 29-42.

68. Rotstein, G.E., Papageorgiou, L.G., Shah, N., Murphy, D.C. and Mustafa, R., (1999), A product portfolio approach in the pharmaceutical industry, *Computers and Chemical Engineering* **23**, S883-S886.

69. Samuelsson, J., *Pharmaceutical capacity planning*, M.Sc. Thesis, Imperial College, University of London, 1999.

70. Sargent, R.W.H., (1977), The decomposition of systems of procedures and algebraic equations, *In Proceedings of Biennial Conference on Numerical Mathematics*, Watson, G.A (Editor), Spinger Verlag, Dundee.

71. Schmidt, C.W. and Grossmann, I.E., (1996), Optimization models for the scheduling of testing tasks in new product development, *Industrial and Engineering Chemistry Research* **35**, 3498-3510.

72. Scholkopf, B., Smola, A.J., Williamson, R.C. and Bartlett, P.L., (2000), New support vector algorithms, *Neural Computation* **12**, 1207-1245.

73. Shankar, V. and Krishnamurthi, L., (1996), Relative price sensitivity to retailer promotional variables and pricing policy: an empirical analysis, *Journal of Retailing* **72**, 249-272.

74. Sherali, H.D., Soyster, A.L. and Murphy, F.H., (1983), Stackelberg-Nash-Cournot equilibria:charactrizations and computations, *Operations Research* **31**, 253-276.

75. Sherali, H.D., Leleno, J.M., (1988), A mathematical programming approach to a Nash-Cournot equilibrium analysis for a two-stage network of oligopolies, *Operations Research* **36**, 682-702.

76. Smola, A.J. and Scholkopf, B., *A tutorial on support vector regression*, NeuroCOLT2 Technical Report Series, 1998. Available from website: http://www.neurocolt.com

77. Smola, A.J. and Scholkopf, B., (2004), A tutorial on support vector regression, *Statistics and Computing* **14**, 199-222.

78. Stackelberg, V.H., *Marktform und gleichgewicht*, Springer, Vienna, 1934.

79. Stavins, J., (1997), Estimating demand elasticities in a differentiated product industry: the personal computer market, *Journal of Economics and Business* **49**, 347-367.

80. Straffin, P.D., *Game Theory and Strategy*, The Mathematical Association of America, Washington, 1993.

81. Subrahmanyam, S., Pekny, J.F. and Reklaitis, G.V., (1996), Decomposition approaches to batch plant design and planning, *Industrial and Engineering Chemistry Research* **35**, 1866-1876.

82. Subramanian, D., Pekny, J.F. and Reklaitis, G.V., (2000), A simulation-optimization framework for addressing combinatorial and stochastic aspects of an R&D pipeline management problem, *Computers and Chemical Engineering* **24**, 1005-1011.

83. Subramanian, D., Pekny, J.F., Reklaitis, G.V. and Blau, G.E., (2003), Simulation-optimization framework for tochastic optimization of R&D pipeline management, *AIChE Journal* **49**, 96-112.

84. Tellis, G.J., (1988), The price elasticity of selective demand: a meta-analysis of econometric models of sales, *Journal of Marketing Research* **25**, 331-341.

85. Vapnik, V.N., *Statistical learning theory*, John Wiley and Sons, 1998.

86. Varian, H.R., *Microeconomic Analysis*, W.W. Norton & Company Inc., New York, 1992.

87. Vives, X., *Oligopoly Pricing*, The MIT Press, Cambridge, 1999.

88. Williams, H.P., *Model building in mathematical programming*, John Wiley, New York, 1999.

89. Wolsey, L.A., *Integer programming*, John Wiley & Sons, Chichester, 1998.

REFEREED RESEARCH PAPERS FROM THIS THESIS

CONFERENCE PAPERS

1. Levis, A.A. and Papageorgiou, L.G., A mathematical programming approach for multi-site capacity planning in the pharmaceutical industry, *In Proceedings of CHISA 2002: 15th International Congress of Chemical and Process Engineering*, 25-29 August 2002, Prague, Czech Republic.

2. Levis, A.A. and Papageorgiou, L.G., Multi-site capacity planning for the pharmaceutical industry using mathematical programming, *In Proceedings of ESCAPE-13: 13th European Symposium on Computer-Aided Process Engineering*, 1-4 June 2003, Lappenranta, Finland.

3. Georgiadis, M.C., Pistikopoulos, E.N., Levis, A.A., Papageorgiou, L.G., Sequeira, S.E., Espuña, A., Puigjaner L., Jernström, P., Westerlund, T., *et al.*, Virtual Plant-Wide Management and Optimisation of Responsive Manufacturing Networks (VIP-NET): An EU Collaborative Research Project, *In Proceedings of ESCAPE-14: 14th European Symposium on Computer-Aided Process Engineering*, 16-19 May 2004, Lisbon, Portugal.

4. Tsiakis, P., Pantelides, C.C., Levis, A.A., Papageorgiou, L.G., Georgiadis, M.C., and Sanidiotis, I., "Efficient Production Scheduling of a Large-Scale Industrial Facility", *In Proceedings of CHISA 2004: 16th International Congress of Chemical and Process Engineering*, 22-26 August 2004, Prague, Czech Republic.

JOURNAL PAPERS

5. Levis, A.A. and Papageorgiou, L.G., (2004), A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry, *Computers and Chemical Engineering* **28**, 707-725.

6. Levis, A.A. and Papageorgiou, L.G., Customer demand forecasting via support vector regression analysis, *(submitted for publication)*.

7. Georgiadis, M.C., Levis, A.A., Tsiakis, P., Sanidiotis, I., Pantelides, C.C. and Papageorgiou, L.G., Optimisation-based scheduling: A discrete manufacturing case study, *(submitted for publication)*.

8. Levis, A.A. and Papageorgiou, L.G., Active demand management for substitute products through price optimisation, *(submitted for publication)*.

Appendix A

Translation-fixing step for Algorithms H1 and H2

Based on the solution of the aggregate model [Problem A], the goal of the translation-fixing step is to map the derived aggregate investment decision and suite availability variables into the corresponding detailed decision variables [Problem D]. The proposed translation-fixing step features a *unique mapping* between integer and binary investment decision variables so as to mirror the information deriving from the solution of the aggregate model. A simple example is described next in order to elaborate the use of the translation-fixing step.

Say for example that the solution of the aggregate model [Problem A] results in the following integer investment decision variables for production site *B*.

$$h\widetilde{E}_{t1}^{B} = 2 \text{ and } n\widetilde{E}_{t2}^{B} = 4 \tag{A.1}$$

For the sake of the argument, we assume that initially there are no manufacturing suites available at site *B*. The aggregate solution suggests the investment of two header suites at time period $t=t1$ and four non-header suites at time period $t=t2$. Notice that the proposed solution satisfies aggregate constraints (3.5) and (3.6).

According to these constraints, the total number of invested non-header suites at time period $t=t2$ has to be greater or equal to 3 and lower or equal to 6. These lower and upper bounds are based on the assumption that each manufacturing block is made up of one header and three non-header suites ($M=4$).

Before we translate the aggregate investment plan into a detailed investment plan, we need to discriminate between header and non-header suites in the detailed model. The suite investment decision variables in the detailed model are binary variables E_{it}^{l} where suite index i represents both types of suites. The classification between header and non-header suites is done as follows:

$$Mod(ord(i)-1,M)=0 \Rightarrow i \equiv header\ suite \qquad (A.2)$$

$$Mod(ord(i)-1,M)\neq 0 \Rightarrow i \equiv non-header\ suite \qquad (A.3)$$

Therefore, the translation-fixing step disaggregates the integer investment decision variables into the following binary investment decisions variables for each individual manufacturing suite, used in the detailed MILP model:

$$E_{i1,t1}^{B}=1,\ E_{i2,t2}^{B}=1,\ E_{i3,t2}^{B}=1, E_{i4,t2}^{B}=1, E_{i5,t1}^{B}=1 \text{ and } E_{i6,t2}^{B}=1 \qquad (A.4)$$

$$E_{it}^{B}=0 \quad \forall\ i \notin \{i1,i2,i3,i4,i5,i6\},\ t \qquad (A.5)$$

In our case, the proposed investment plan consists of one complete manufacturing block with 4 suites and one semi-complete block with 2 suites. By taking into account the suite construction lead times δ_i (3 and 2 years for header and non-header suites, respectively), we can then determine the suite availability binary decision variables for the entire planning horizon by employing equation (2.3).

In our specific case, we end up with the following binary suite availability variable levels:

$$A_{it}^{B}=1 \quad \forall\ i \in \{i1,i2,i3,i4,i5,i6\},\ t \geq t4 \qquad (A.5)$$

$$A_{it}^{B}=0 \quad \forall\ i \in \{i1,i2,i3,i4,i5,i6\},\ t \leq t3 \qquad (A.6)$$

$$A_{it}^{B}=0 \quad \forall\ i \notin \{i1,i2,i3,i4,i5,i6\},\ t \qquad (A.7)$$

Having calculated the levels of binary variables E_{it}^{l} and A_{it}^{l}, there is no further need to consider detailed constraints related to suite investment decisions and suite availability. Instead, the derived levels for the binary variables are fed into the original detailed model [Problem D] as given parameters resulting in the reduced detailed MILP model as described in Chapter 3.