Design of Advanced Manufacturing Systems

Design of Advanced Manufacturing Systems

Models for Capacity Planning in Advanced Manufacturing Systems

Edited by

ANDREA MATTA
Politecnico di Milano,
Milan, Italy

and

QUIRICO SEMERARO
Politecnico di Milano,
Milan, Italy

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-2930-6 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-10 1-4020-2931-8 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-1-4020-2930-1 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-1-4020-2931-8 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Printed on acid-free paper

springeronline.com

Printed in the Netherlands.

Contents

List of Figures

List of Tables

Preface

Since manufacturing has acquired industrial relevance, the problem of adequately sizing manufacturing plants has always been discussed and has represented a difficult problem for the enterprises, which prepare strategic plans to competitively operate in the market. Manufacturing capacity is quite expensive and its exploitation and planning must be carefully designed in order to avoid large wastes, or to preserve the survival of enterprises in the market. Indeed a good choice of manufacturing capacity can result in improved performance in terms of cost, innovativeness, flexibility, quality and service delivery. Unfortunately the capacity planning problem is not easy to solve because of the lack of clarity in the decisional process, the large number of variables involved, the high correlation among variables and the high level of uncertainty that inevitably affects decisions.

The aim of this book is to provide a framework and specific methods and tools for the selection and configuration of capacity of Advanced Manufacturing Systems (AMS). In particular this book defines an architecture where the multidisciplinary aspects of the design of AMS are properly organized and addressed. The tool will support the decision-maker in the definition of the configuration of the system which is best suited for the particular competitive context where the firm operates or wants to operate.

This book is of interest for academic researchers in the field of industrial engineering and particularly indicated in the areas of operations and manufacturing strategy. Also we think that the content, even if it is very technical in some sections, is helpful for those mangers who want to know, and possibly to use in practice, a reference architecture for the strategic capacity planning problem in manufacturing.

The first chapter provides a complete view of the capacity planning problem and describes the reference architecture in which enterprise should plan their manufacturing capacity in the long term. This chapter focuses on structuring the main problem in many hierarchical sub-problems, each one described in detail by specifying the type of decision the firm has to make and the nature of information that is available at the moment of the decision. The following chapters contain the decision models this book proposes to support managers in the capacity planning problem, from the decision on the type of manufacturing systems to adopt to their detailed configuration in terms of resources (machines, buffers, transporters, etc.). Given the organization of the volume, the reading of Chapter 1 is particularly suggested in the reading of the book.

We want to acknowledge the MIUR (Ministero dell'Istruzione, dell' Università e della Ricerca) which funded this research. Indeed, the content represents the result of a two years Italian research project *Models for Capacity Planning in Advanced Manufacturing Systems* funded by MIUR in 2000. Five universities participated in the project: Politecnico di Milano and Torino, Università degli studi di Lecce, Palermo and Salerno.

ANDREA MATTA

QUIRICO SEMERARO

Contributing Authors

A. Alfieri
Dipartimento di Sistemi di Produzione ed Economia dell'Azienda, Politecnico di Torino, Torino, Italy.

A. Anglani
Dipartimento Ingegneria dell'Innovazione, Università degli Studi di Lecce, Lecce, Italy.

P. Brandimarte
Dipartimento di Sistemi di Produzione ed Economia dell'Azienda, Politecnico di Torino, Torino, Italy.

M. Bruccoleri
Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale, Università degli Studi di Palermo, Palermo, Italy.

F. Caiazzo
Dipartimento di Ingegneria Meccanica, Università degli Studi di Salerno, Fisciano (SA), Italy.

P. Caricato
Dipartimento Ingegneria dell'Innovazione, Università degli Studi di Lecce, Lecce, Italy.

A. Grieco
Dipartimento Ingegneria dell'Innovazione, Università degli Studi di Lecce, Lecce, Italy.

G. Lo Nigro
Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale, Università degli Studi di Palermo, Palermo, Italy.

A. Matta
Dipartimento di Meccanica, Politecnico di Milano, Milano, Italy.

S. Noto La Diega
Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale, Università degli Studi di Palermo, Palermo, Italy.

F. Nucci
Dipartimento Ingegneria dell'Innovazione, Università degli Studi di Lecce, Lecce, Italy.

R. Pasquino
Dipartimento di Ingegneria Meccanica, Università degli Studi di Salerno, Fisciano (SA), Italy.

G. Perrone
Dipartimento di Fisica Ambientale, Università degli Studi della Basilicata, Potenza, Italy.

P. Renna
Dipartimento di Fisica Ambientale, Università degli Studi della Basilicata, Potenza, Italy.

Q. Semeraro
Dipartimento di Meccanica, Politecnico di Milano, Milano, Italy.

V. Sergi
Dipartimento di Ingegneria Meccanica, Università degli Studi di Salerno, Fisciano (SA), Italy.

B. Spiezio
Dipartimento di Ingegneria Meccanica, Università degli Studi di Salerno, Fisciano (SA), Italy.

T. Tolio
Dipartimento di Meccanica, Politecnico di Milano, Milano, Italy.

Chapter 1

A FRAMEWORK FOR LONG TERM CAPACITY DECISIONS IN AMSS

A. Matta, Q. Semeraro, T. Tolio

Dipartimento di Meccanica

Politecnico di Milano

via Bonardi 9, 20133, Milano, Italy

andrea.matta@polimi.it, quirico.semeraro@polimi.it, tullio.tolio@polimi.it

Abstract Investment in Advanced Manufacturing Systems has a strategic impact that can affect the long term competitiveness of enterprises improving the ability of firms to create new markets, introduce new products, and to react quickly and effectively to competitors. Since the available methodologies to support strategic decisions are not easily applied, firms are in the unpleasant position of evaluating strategic decisions without any practical tool that is able to estimate the value of each specific action and its consequences at the strategic level. The aim of this chapter is to provide a reference framework for the selection and configuration of Advanced Manufacturing Systems. In particular, a framework is proposed in which the different aspects and evaluations that are involved in long term capacity planning are properly organized.

Keywords: Advanced Manufacturing Systems; capacity planning; technology choice.

1. Introduction

In recent years a relevant change has developed in manufacturing systems technology. Such change has been mainly due to the introduction into the market of new equipments able to combine microelectronic and programmable devices within mechanical machines. The so-called Advanced Manufacturing Systems (AMSs) are a result of such a revolution. At the moment different architectures of Advanced Manufacturing Systems are available: some of these architectures are well known and tested like flexible transfer lines, flexible manufacturing cells, flexible manufacturing systems, while others are rather new and they are being studied within national and international research projects (Koren et al., 1997; Matta et al., 2000; Matta et al., 2001) or directly proposed by machine tool builders. The problem of capacity acquisition when AMSs are considered is particularly complex for several reasons (Perrone and Diega, 1999; Naik and Chakravarty, 1992; Gerwin, 1992; Fine and Freund, 1990; Price et al., 1998). First of all, the high investment involved makes companies very sensitive to the risk factor thus precluding the adoption of AMSs. The consequent sensitivity of the management staff could therefore cause the improper evaluation of benefits of these systems, such as scalability and flexibility. Another reason is that flexible capacity enlarges the spectrum of possible future scenarios because many alternative strategies are viable, thus making the risk evaluation more difficult. Furthermore, many advantages of these manufacturing systems are not easy to quantify and therefore they are seldom evaluated properly. The strong interaction among the components of AMSs makes it necessary to carry out evaluations considering the system as a whole. Therefore, simple rules of thumb are normally quite misleading, and appropriate and sometimes rather sophisticated evaluation methods are required. The problem is further complicated by the fact that choices must also be evaluated in both strategic and economical terms. Indeed manufacturing systems can be a good competitive weapon for the strategy of the firm if the capacity choice is coherent with the overall strategy of the firm. To be profitable in the long term, a production system must be both efficient and aligned with the company strategy. In practice even a good manufacturing plant can have problems if its production system does not conform to the company strategy.

The selection of capacity is becoming more and more relevant for manufacturing companies because a good or bad decision can deeply affect the profitability of the company that invests in new capacity. In practice, since capacity has a cost, it is not possible to solve the problem simply by acquiring extra capacity to face all possible future requirements, but it is necessary somehow to weight the advantages of having enough capacity to front future needs with the cost of maintaining unused capacity. All these aspects tend to hinder the exploitation of the opportunities offered by AMSs. This is particularly true in SMEs (Small and Medium Enterprises) where structured approaches to the solution of capacity acquisition problem are not applied.

The problem of capacity planning in Advanced Manufacturing Systems has been deeply investigated in the last 20 years. However, the proposed methodologies did not reach the main goal. Most of them do not propose a solution to the whole problem, but limit their focus on some well-defined sub-problems. This approach has led to solutions that are not practical to apply to real problems since they treat only a portion of the problem. A software tool like a Decision Support System (DSS) that is able to treat such a complex multidimensional problem is needed in order to support people involved in long term capacity decision planning. To develop this Decision Support System it is necessary first to define the whole decisional process in all its steps and details so that different sub-problems can be identified and then solved, in an integrated way, by means of specific tools. The aim of this chapter is to provide a framework and specific tools for the selection and configuration of Advanced Manufacturing Systems in the long term capacity planning problem.

This chapter is organized as follows. The next section defines the concept of capacity in manufacturing while the basis of manufacturing strategy theory are summarily explained in Section 3. In Section 4 the AMSs investigated in this book are described and Section 5 structures the strategic problem by means of the IDEF0 language modelling.

2. Manufacturing capacity

Since manufacturing has become an industrial phenomenon, the problem of adequately sizing plants has always been discussed. As Whitmore wrote in the early 20th century, an important *work of organizations includes taking stock of the resources at one's command and planning the fullest use of them all* (Whitmore, 1907). Capacity in general can be defined as the set of any kind of resources that can be used to create value for the customer and, in general, the cost of capacity is lower than the value the customer pays to acquire the product or the service provided. Without capacity it is not possible to create value because at least a minimum amount of resources is necessary. Furthermore, manufacturing capacity is defined as the set of human resources and equipments that the company can use to produce goods or services to sell in the market.

The dimensions of manufacturing capacity are:

- **Type**. There are in practice many manufacturing systems that differ in terms of their characteristics and several keys of classification can be used, some of them are: standard or advanced, rigid or flexible, capital intensive or not, automated or manned, etc. In practice, the characteristics of the system to stress depend on the the type of analysis we want to carry out.
- **Amount**. The quantity of capacity acquired to create value to customers. Since capacity cannot be fully exploited, literature

generally uses the terms *theoretical* or *nominal* amount of capacity to refer to the purchased capacity. The amount of capacity can be expressed in machine time available in a period (e.g. hours per day, hours per week, etc.) or in number of pieces per period. Knowing the production rate of products on that system it is possible to move from time to part units.
The portion of capacity that is used to manufacture products is known as *utilized capacity.* Utilized capacity can vary day by day for several reasons, thus average and standard deviation are used to represent utilized capacity in a defined time period.

- **Cost**. The total economic value that is necessary to spend for acquiring, running, maintaining and dismissing a manufacturing system.

The above characteristics synthesize the main strategic issues of manufacturing systems. Let us discuss more in detail the capacity amount. The *available capacity* is the amount of production time the firm can effectively use to satisfy the market demand, that is the amount of theoretical capacity taken from the unused portion due to any reason except lack of demand. In practice, available capacity is normally compared with what customers demand. If the available capacity is greater than the capacity used to satisfy the customer demand there is a waste quantified by the difference between available capacity and utilized capacity; this waste is also known as *excess capacity* (Olhager et al., 2001). However, if the available capacity is lower than the capacity that would be necessary to fully satisfy the customer demand, there is a lack of capacity quantified by the difference between requested capacity and available capacity; this lack is also known as *demand surplus.* The ratio between utilized capacity and theoretical capacity is the *utilization* level of the plant. Among the several causes of the difference between utilized and theoretical capacity, the most frequent are: personnel scheduling, set-ups, maintenance and lack of demand. A quantitative analysis on the reasons for unused capacity should always be done before deciding to acquire new capacity.
Manufacturing capacity is characterized by the following issues:

- Capacity cannot be stored. If a manufacturing system is not used in a period because of lack of demand, the related portion of capacity is wasted and cannot be utilized in the future. An alternative is to produce even if there is no demand with the purpose of storing finished goods; in this case higher inventory costs are incurred.

- Capacity can be changed only in discrete steps. In practice, an increase or decrease of capacity corresponds to an acquisition or a dismission of a finite amount of resources (e.g. a machine or a human operator).
- Capacity can be changed with considerable lead times. In practice, ordering a new machine or a new production system can take several months.
- Generally capacity cannot be reduced but only expanded because it could be difficult in practice to sell used mechanical devices.
- Manufacturing systems have a long life cycle (from 5 to 20 years).
- Manufacturing systems have a ramp-up period in which the production level is lower than the theoretical one and all the efforts are devoted to reach the target value as soon as possible. The ramp-up period can be very critical because it can take several months, or years or in some cases the system never reaches the target production value.

3. Manufacturing strategy

The capacity problem is a decision related to the overall strategy defined by the company. Strategy is a term used in business planning that refers to the overall scheme of managing and governing the future course or direction of the company. Strategy implies careful selection and application of resources for the most advantageous position, in anticipation of future events. A company strategy is a set of plans and policies by which a company tries to gain advantages over its competitors. A company strategy is defined at corporate level and must consider several issues such as research and development (R&D), sales, marketing, finance and manufacturing. From the company strategy all the function strategies are then derived and, among these, we are interested in manufacturing strategy. In particular, *manufacturing strategy deals with the decisions concerning the specific role of manufacturing in order to achieve competitive advantage in the market.*

Skinner was the first to introduce the concept of manufacturing strategy in the '70s. In his numerous publications Skinner emphasizes the role of the manufacturing in the whole company strategy pointing out the need of coherence between company strategy and implemented manufacturing tasks. Manufacturing strategy can contribute to firms' success by supporting the implementation of the competitive strategy defined by the corporate. A company's competitive strategy at a given time should place particular demands on its manufacturing function, and, conversely, the company's manufacturing policy and operations should be specifically designed to fulfil the tasks demanded by strategic plans. A mismatch between company strategy and manufacturing strategy can be source of lack of competitiveness. The conceptual strategy model gener-

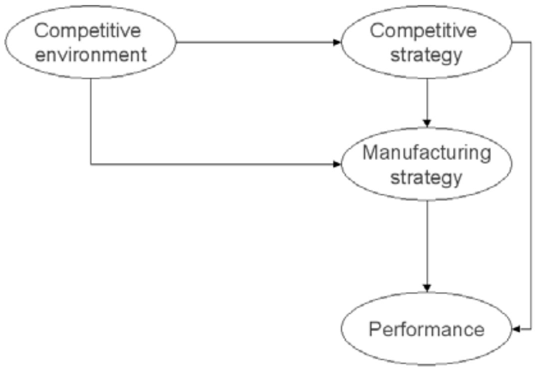


Figure 1.1. Links of manufacturing strategy with environment, competitive strategy and performance.

ally recognized in literature is shown in Figure 1.1. In the model, mainly derived from Skinner (Skinner, 1985), the competitive environment influences both the company and manufacturing strategies. The study of Ward and Duray (Ward and Duray, 2000) [1] empirically demonstrates there is a relationship between competitive environment and competitive strategy defined at corporate level. The link between competitive strategy and manufacturing strategy has been investigated in (Vickery et al., 1993; Williams et al., 1995; Ward and Duray, 2000). In particular Vickery et al. (Vickery et al., 1993) state there is a relationship between competitive strategy and productive competence with business performance while Ward and Duray (Ward and Duray, 2000) demonstrate that this link is valid only for high business performers. The analysis of Ward and Duray does not support the link between environment and manufacturing strategy and the authors explain that environment is mediated by competitive strategy. They also demonstrate the positive dependence between manufacturing strategy and performance for high performers. However the relation between competitive strategy and performance is not supported by empirical evidence, the reason is that manufacturing

[1] The analysis is based on an empirical study of 101 USA companies whose primary product is in one of three sectors: fabricated metal components, electrical devices and electronic controls.

strategy mediates between them. In conclusion, it appears (Ward and Duray, 2000) that the conceptual model shown in Figure 1.1 is valid for high performers and the link between competitive and manufacturing strategy is highly relevant, that is a competitive strategy works well when supported by coherent manufacturing tasks as Skinner wrote in the '70s.

Let us enter into more detail on what is a manufacturing strategy. According to the Hayes and Wheelwright's model (Hayes and Wheelwright, 1984), generally recognized as the reference model in literature, a manufacturing strategy is constituted by *competitive priorities* and *decision areas*. The competitive priorities are a consistent set of goals for manufacturing:

- Cost: production and distribution of the product at low cost. The lower the cost is the higher the profit or the possibility to operate an aggressive strategy of price competition in the market is.

- Delivery: reliability and speed of delivery. It is generally recognized in literature how important the level of deliveries is on the customer perception.

- Quality: manufacture of products with high quality and performance standards. Garvin states (Garvin, 1987) that quality, in all its multidimensional aspects, can be used to gain competitive advantage.

- Flexibility: product mix and volume. The ability to change the priorities of jobs, or the machine assignments of jobs in the shop floor, or the production volume can allow the firm a competitive advantage. See (Gerwin, 1992) for more details and (Sethi and Sethi, 1990) for a good survey on manufacturing flexibility types.

- Innovation: capability to introduce new products or product variations effectively. The presence of innovation in the list of manufacturing competitive priorities is not generally recognized. However, we agree with Hayes and Wheelwright in saying that innovation can be an important weapon in the market competition.

After specifying the competitive priorities coherently with the company's strategy, the manufacturing actions potentially adoptable to pursue the stated goals are classified into two categories: structural and infrastructural decision areas. The structural decision areas have generally a long term impact, are difficult to reverse and they require substantial capital investment. A brief comment for all decision areas is now reported and the reader is referred to (Hayes and Wheelwright, 1984) for more details.

- Facilities: the company should decide on the location, the size and the focus of facilities.
- Process technologies: the company should decide which process technologies to adopt to manufacture products. In addition the company has to choose between acquiring or developing the chosen technology, and other strategic issues such as the degree of automation, the layout, the scalability and flexibility of the process.
- Capacity: the company should decide on the type of capacity to use in manufacturing, the amount and timing, that is when to acquire and how much.
- Vertical integration: the company should decide on the relationships with its providers and customers.

The infra-structural decision areas affect the people and the systems that do manufacturing work. The infra-structural decision areas are generally more tactical, linked with specific operating aspects and do not require substantial capital investment:

- Vendors: the company should decide on the structure and size of the network of vendors and also the relationships with them.
- Human resources: the company should decide how human resources shall be selected, trained and payed. Also, the company should design the job and the skill levels.
- System practices: the company should decide the practices to be adopted for production and material planning, management of manufacturing systems, quality, standards, etc.
- Organization and management: the company should decide the nature of management. For instance, employees in manufacturing can be organized by product, function, or geographical areas.

It is very important that all decisions made in different areas are coherent and together contribute to reaching the defined competitive priorities. Indeed, the success of a company depends on the coherence of its strategy with the competitive environment and the level of integration of its strategies and decisions. The firms that do not maintain consistency between the pursued competitive priority and the manufacturing decisions they implement do not achieve superior business performance.

The selection of capacity is one of the strategic decisions of a firm's manufacturing strategy that has direct consequences on all the competitive priorities defined in the manufacturing strategy. First of all, the capacity choice deeply affects production costs. Indeed, different manufacturing systems have different costs because they may differ in the personnel involved, cost of devices, consumption of power and tools, reliability of equipment, etc.
It is also important to take into account the timing of the investment. If a 100% increase of the customer demand is forecasted in the immediate future it is necessary to have the additional capacity necessary as soon as possible to front the market expansion. If the increase of demand is forecasted to occur 5 years in the future, it is hopeful that the firm waits to expand its capacity unless large wastes are incurred. Therefore, a firm can fundamentally adopt two different policies: lead or follow the customer demand (see Figure 1.2). If there is a capacity demand surplus, i.e. the firm follows the market demand, the capacity utilization will be high but there is also a risk to loose customers due to long delivery lead times (Olhager et al., 2001). If there is an excess capacity, the firm anticipates the market demand, the system utilization will be low but it will easier to maintain high delivery reliability and flexibility. However, market demand is uncertain and it may occur that the capacity expansion of the firm is not followed by the increase of market demand thus causing capacity wastes. In other words there is a trade-off between utilization and delivery reliability.

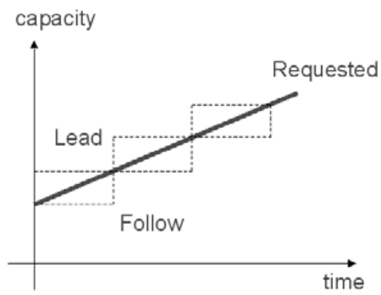


Figure 1.2. Timing of capacity: lead or follow the market demand ?

Also flexibility is affected by the capacity choice because an excess of capacity allows the firm more flexibility to react to changes in market demand. Depending on the type of equipment selected it will change the ability of the firm to modify the production mix; for instance a rigid machine forces the firm to run large lot sizes in order to avoid expensive set-up times. Furthermore, the type of capacity can influence the quality of products; for instance different machine tools reach different precisions and therefore products with different quality levels. Also innovation can be improved by properly selecting the type of capacity. If a firm has only dedicated systems in the shop floor, the frequency of the introduction of new products will probably be small because the launch of a new product involves the re-configuration, often very expensive, of the whole system.

4. Advanced Manufacturing Systems

Advanced Manufacturing Technology (AMT) has been subject of investigation since late the '70s when computer numerically controlled (CNC) machine tools started to be widely adopted on shop floors. AMT covers a large area of non traditional technologies that firms can use to maintain or improve their competitiveness. In practice production systems such as CNC machine tools, automated flow lines, cellular manufacturing systems, flexible manufacturing systems, or design tools such as CAM (Computer Aided Machining), CAPP (Computer Aided Process Planning), but also management tools such as MRP (Material Resource Planning) and ERP (Enterprise Resource Planning) are considered advanced manufacturing technology.

In this book, for simplicity of exposition, the authors deal with only a portion of the large set of production systems: chip removal manufacturing systems, that is systems having turning, or milling, or drilling, or grinding, or all those processes that obtain the finished part by cutting material (therefore deforming, casting and assembling are not considered).

Furthermore, in this book the attention is restricted to those systems with high degree of automation and large amount of capital involved. The reason is that completely automated systems are considered complex and it is difficult in practice to evaluate their performance, to manage and design them. Therefore the need of having adequate decision models for this class of systems seems to be evident. In addition, this necessity grows if these systems require large amounts of capital because a wrong choice could compromise the profitability of the investment and, in some cases, the survival of the companies in the market. In particular, systems like stand alone machine tools are not considered, even if they are expensive (e.g. machines are CNC type), because they are easy to evaluate in terms of production rate and utilization and adequate models are already available. Two classes of systems are taken into consideration throughout this book: Dedicated Manufacturing Systems (DMS) and Flexible Manufacturing Systems (FMS). These classes of systems are described in the following sub-sections.

4.1 Dedicated Manufacturing Systems

Dedicated Manufacturing Systems are those systems that are conceived, designed and managed appositely on the needs of a product or a very restricted family of products. The main characteristics of these systems are:

- Rigid equipment. The equipment is designed to satisfy the needs of the product, or the restricted family of products, to which the whole system is dedicated. Therefore the machines and devices such as transporters, grippers, etc. are designed to accomplish a very limited set of operations that cannot be normally expanded unless large costs are incurred. Stations in transfer lines are a typical example of equipment rigidity. Normally machine movements are not numerically controlled by a computer but mechanically by means of cams or other mechanical devices.

- High production rates. The equipment is dedicated and normally designed to minimize processing times. In order to cut processing times, one or more operations can be performed in parallel. As a consequence machines are generally fast allowing the system to reach higher production rates compared with other ones (e.g. Flexible Manufacturing Systems).

- Low skills. The skills needed to run the system are normally low since human jobs are reduced to loading and unloading parts and maintenance.

- Easy management. Given the limited number of products a DMS processes, and the simplicity of flows in the system, the scheduling of resources is quite easy.

- Low investment. The equipment is rigid and everything is designed to accomplish only the operations that are necessary to manufacture the products to which the DMS is dedicated. Therefore the investment cost of the system is not large if compared with that of more flexible systems with CNC machines.

- Excess capacity. The amount of capacity unused because of lack of demand cannot be used to manufacture different products. For this reason the residual value of the investment is very small.

There are fundamentally two categories of DMS in practice: dedicated machines and dedicated flow lines. Dedicated machines are those machines appositely designed to perform efficiently the product process cycle; these machines are generally conceived and developed in the firm because a high knowledge and experience is necessary on the process. Dedicated machines work in stand-alone mode and in general are completely automated except for the loading and unloading of parts; therefore they are also simple to manage not requiring any sophisticated tool and for this reason they are not considered in this book. We deal with Dedicated manufacturing flow lines that are an important and wide spread type of DMS. This type of manufacturing systems is described in detail in Chapter 4.

4.2 Flexible Manufacturing Systems

CECIMO (Commit Europeenne de Cooperation des Industries de la Machine Outil) defines an FMS *as an automated manufacturing system capable, with a minimal human action, of producing any part type belonging to a pre-defined family; these systems are generally adopted for the production at small or medium volumes, in variable lot sizes that differ also in their composition. The system flexibility is generally limited to the family of part types on which the system is conceived. The FMS has devices for planning the manufacturing, scheduling the resources and saving the production data.*
As the above definition points out, the main characteristics of FMSs are:

- Flexible equipment. The equipment is flexible enough to satisfy the needs of all the products belonging to the family. Indeed all the machines are CNC type and can be programmed to perform a large number of operations. In practice, it is only needed the ability to write a simple computer program to code the process cycle of a product into instructions that the numerical control of the machine can read, understand and operate to execute them.

- Low production rates. Machines have generally a spindle for executing operations in a sequential way. As a consequence machines are generally slow in comparison with machines of DMS. Recent innovations such as high spindle machines and linear motors are rapidly spreading in FMSs thus reducing processing times and inactive rapid movements respectively.

- Medium/High skills. The skills needed to run the system requires a minimum knowledge in programming and managing CNC machines.

- Complex management. The management of FMSs is complicated by the large number of products. Indeed, for each product it is necessary to schedule properly machines, fixtures and tools.

- Large investment. Machines are flexible and require large investments. Therefore, the investment cost of the system is very large if compared with that of dedicated systems.

- Excess capacity. The amount of capacity unused because of lack of demand can be used to manufacture different products. The residual value of the investment takes into account this issue.

FMSs are described in detail in Chapter 4.

5. A framework for capacity problems

Many factors have to be taken into account when a capacity investment decision is analyzed: firm's strategy, uncertainty of markets, competitors' strategy, available system architectures, types of technologies, etc. Investment in AMSs is like an umbrella covering different sub-problems that have to be analyzed and solved before making the final decision "which system to buy". These sub-problems are not independent since they are related one another and their relationships are not simple to formalize and to quantify. Some of the different sub-problems the firm has to solve when an investment in AMSs is analyzed are in the following.

- Market. The firm has to decide where to concentrate its efforts: niche or broad market. This decision is taken at corporate level and is generally already available when the capacity problem is analyzed. This subject is out of the scope of this book and it is not taken into consideration.

- Products. The firm has to decide which products will sell in the selected market. This decision is made at the corporate level which defines the market segment and the macro-characteristics of the products with which the firm wants to compete. Therefore the new potential products to launch in the market are known. At the manufacturing strategy level of detail the final choice will deal with the selection of product codes to launch in the market. The first decisional level of product selection is out of the scope of this book and it is not taken into consideration.

- Service level. The firm has to establish the level of service provided in the market. A high level of service typically involves large efforts, however a low level of service may be the cause of a loss of customers that, unsatisfied, change their supplier. This decision is part of the manufacturing strategy and it will be investigated in Chapter 2.

- Technology. The firm has to decide which technology is most appropriate for manufacturing the products to market in the future. The choice of technology can be fundamental in the market strategic position of the firm. An innovative process technology developed internally to the firm can put the firm in a leadership position. On the contrary, a standard technology process can be acquired by any competitor and cannot be a competitive weapon in the market. This decision is part of the manufacturing strategy but it is not subject of investigation in the models presented in this book.

- Make or buy. The firm has to establish if and how much production capacity can be acquired from subcontractors. The decision has to consider all the possible future consequences that can derive from this choice. Indeed, the outsourcing of a product can be loss of knowledge and skills and can decrease the innovation level of the firm in the long term. Also the contractual power with outsourcers is critical because it dynamically changes depending on the particular relationships that are defined and modified between seller and buyer. Guidelines of the outsourcing strategy are generally decided at the corporate level while details on outsourcers and the quantitative levels of externalization are part of the manufacturing strategy and will be faced in the following Chapters 2, 3 and 5.

- Flexibility. The firm has to decide the levels of flexibility the manufacturing capacity should have. The more flexibile the acquired capacity is the faster and cheaper the firm's reaction to any changes in the market is. This decision is part of the manufacturing strategy and will be faced in Chapter 2.

- System architecture. The firm has to decide the type of production systems. Indeed, given a type of technology selected at higher level in the decisional process, the firm has to choose among several potentially adoptable alternatives to manufacture products. This decision is part of the manufacturing strategy and will be faced in Chapter 2.

- Resources. The firm has to decide on the specific type and number of machines, carriers, fixtures, tools, etc. to use in the new system. In other words the firm has to decide on the detailed configuration of the manufacturing system, eventually supported by the builder of the production system. This decision is part of the manufacturing strategy and will be faced in Chapters 4 and 5.

A correct evaluation of the investment in AMSs should consider all the factors in an integrated and global risk-approach that analyzes the investment from different points of view. Frequently it occurs that an action to improve a specific key-factor of the firm can have a negative impact on other key-factors; for instance, an increase of flexibility often causes an increase of costs incurred by the firm. Therefore, it is necessary to quantify the impact that each single decision has on the whole problem in order to solve the numerous trade-offs that normally characterize strategic problems. Taking as a reference the manufacturing strategy model described in Section 3, the firm has to evaluate the impact that each alternative AMS has on the competitive priorities the firm defines at manufacturing strategy level. Also dependencies with the other decisional areas are very important because an incoherence between the value of each capacity choice depends also on the type of selected technology, facility position, current knowledge, etc. The strategic problem of planning the manufacturing capacity in Advanced Manufacturing Systems is described in the following sub-sections by means of the IDEF0 formalism (IDEF,) where inputs, outputs, controls and mechanisms are encoded using the ICOM approach.

5.1 A-0 Context diagram

The purpose of activity A0 *Planning production capacity in Advanced Manufacturing Systems* is to define the detailed configuration of Advanced Manufacturing Systems in the planning horizon. The viewpoint adopted in the diagram is that of decision-makers. Decision-makers are the managers that solve the capacity problem. Starting from input information regarding system architectures (i.e. type of production systems that are currently used and potential ones that could be acquired as additional capacity resources) and products (i.e. technical and economical data related to those products that are currently manufactured and potential ones that could be manufactured by the firm in the future), the outputs of the activity A0 are the definition of the detailed configuration of AMSs to be adopted in the different periods of future planned horizon and the selection of the product codes that will be produced by the firm per period. In particular, the defined plan is a timing of the estimated capacity in AMSs that will be required in the future planning horizon. The planning horizon is broken down into periods of three or six months depending on the level of detail of the analysis. An example of the main output of the activity is shown in Figure 1.3 where the minimum and maximum values of needed internal and external capacity are tabled for each product, on each manufacturing system, for every time period in the planning horizon.

	PERIOD 1		PERIOD 2		PERIOD ...	
AMS	Product code	Int/ext capacity	Product code	Int/ext capacity	Product code	Int/ext capacity
DMS 1	1	[100,500]/[0,0]	1	[100,700]/[0,0]	...	...
DMS 2	-	-	2	[500,1500]/[0,0]	...	...
FMS	4	[100,200]/[0,200]	4	[100,250]/[0,200]	...	...
	5	[100,500]/[0,250]	5	[100,500]/[0,250]	...	...
	6	[100,500]/[100,300]	6	[100,400]/[100,300]	...	...
	-	-	-	-	...	...
DMS 3	-	-	-	-	...	...

Figure 1.3. Example of the capacity problem solution.

The A-0 context diagram describing the long-term capacity planning in Advanced Manufacturing Systems is shown in Figure 1.4.

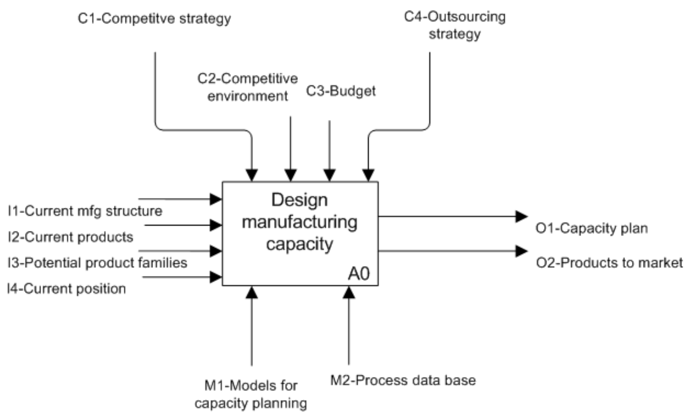


Figure 1.4. A0 context diagram.

More precisely, capacity planning requires a large set of information as explained in the following. The necessary input information for the proposed capacity planning models is:

- I1 **Current manufacturing structure**: information describing the whole set of hardware and management resources currently used to manufacture the products that are sold in the market. In particular with hardware we refer to production systems and equipment, while for management we refer to general practices that are necessary for the production such as production planning, quality control, etc.

- I2 **Current products**: information on products that are currently produced and marketed by the company. This information contains both technological (i.e. technical drawing, process cycle, ...) and management information (i.e. forecasted market demand, production cost, price, ...).
- I3 **Potential product families**: information on the new products potentially marketable in the future. This set of products is decided higher up at the corporate level, however the final choice can be made only after a complete and detailed product profitability analysis is carried out, that is after the manufacturing system to use for these products has been selected and consequently also estimates on future production costs become more reliable. Product input contains both technological and management information: features, process cycle, forecasted market demand, etc.
- I4 **Current position**: definition of the actual firm's market position, if there is any (e.g. the company could also be a new comer in the market).

Capacity planning is controlled by higher decisions made at the corporate level or structural characteristics of the market in which the firm operates in:

- C1 **Competitive strategy**: the whole strategy pursued by the company at corporate level. Porter developed the idea that all competitive strategies are variants of generic strategies characterized by a choice between differentiation and delivered cost, i.e. the product price (Porter, 1980). This choice should be completed with the information of the market focus: niche or broad market ? This information is necessary in order to plan capital investments coherently with the corporate policies. For instance, the aggressive policy of increasing market share deeply affects the capacity decision.
- C2 **Competitive environment**: this constraint follows the description of the market in which the firm will operate in the future. A model of competitors, customers, type of market, etc., represents the reference environment to be considered in decision making.
- C3 **Budget**: the profile of the budget available for investments in AMS in the long term. Also this constraint is decided at the cor-

porate level because it involves the analysis of the firm's financial position.

- C4 **Outsourcing strategy**: it is the decision, made at the corporate level, dealing with which products should be outsourced and fixing a qualitative level of externalization.

The main outputs of a capacity planning problem are:

- O1 **Capacity plan**: the decisional process leads to a plan of all the internal and external capacity that is necessary to have in the planned horizon.
- O2 **Products to market**: the decisional process leads to the final selection of product codes the firm will produce and market in the planned horizon.

In order to obtain the above outputs it is necessary to use the following mechanims:

- M1 **Decision models**: the decision process is supported by formalized models and tools that aid decision-makers in structuring the problem and in quantitatively evaluating, in terms of benefits and cost, the value of each alternative solution.
- M2 **Process & System data base**: decision-makers normally use technological information on product process cycles and production systems potentially adoptable in shop floor. We assume that this information is already available in a database or it is provided by technicians.

According to the description of the capacity problem, the decision is hierarchical and thus it is necessary to make first some important strategic decisions such as the quantitative level of provided service, the flexibility needs that future capacity should have, etc. After the main strategic variables have been fixed, it is possible to evaluate the production capacity that is necessary to have in order to reach the defined strategic objectives. At this step, a more detailed investigation about the required production capacity is needed: this means to evaluate the make or buy sub-problem in order to define a rough "internal" production capacity level per period. Starting from this information, alternative system configurations can be proposed and a performance evaluation of each one (in economic and productive terms) is required to select the best ones for each time period of the planning horizon. This hierarchical decision process is described by means of the IDEF0 modelling language in the following subsections.

5.2 A0 Level diagram

The goal of an A0 Level diagram is to have a more detailed definition of the overall architecture for the decision of capacity planning in Advanced Manufacturing Systems. In an A0 level diagram the different decisional steps of the whole problem are shown pointing out their interactions in terms of information. Activity A0 has been hierarchically decomposed into 4 sub-activities that are now described in detail. For each sub-activity of an A0 level diagram a decision model that supports decision-makers in the long term capacity planning problem is proposed. Interactions with decision-makers are also specified in the comments of the diagram. The first two activities (A1 and A2) deal more specifically with the strategic aspects involved in the investment in AMS, i.e. selection of investment amount, accepted risk level, production mix and system type, while the last two activities (A3 and A4) face the problem of the detailed AMS configuration, that means the generation of alternatives of production systems, the evaluation of their performance and finally the choice of the best ones to use in the final capacity plan. All these activities are strictly related by information and decision flows. The A0 level diagram is shown in Figure 1.5 and its functions are explained below.

[A1] Planning at strategic level. The purpose of this activity is to design strategic variables involved in the capacity acquisition problem when the company strategy, the competitive scenario and the competitive position are properly defined by decision-makers (Gerwin, 1992; Naik and Chakravarty, 1992; Perrone and Diega, 1999). Indeed, when firm plans the production capacity in AMSs in the next planning horizon, it becomes necessary to revise its manufacturing strategy on the basis of the estimated investment cost of new production systems to be acquired. This involves, among the others things, to select products that could be manufactured in the future. However, at strategic level other decisions are considered. Furthermore, at strategic level the type of production system has to be preliminarily defined since it can have long-term impacts; in fact, the introduction of a new technology can deeply affect a firm in the change management phase or in searching for new people with different skills. Therefore, the system architectures that seem to be the most promising at strategic level are indicated to the downstream decision-makers in the capacity planning problem.

More in details, input for activity A1 consists of the following set of information:

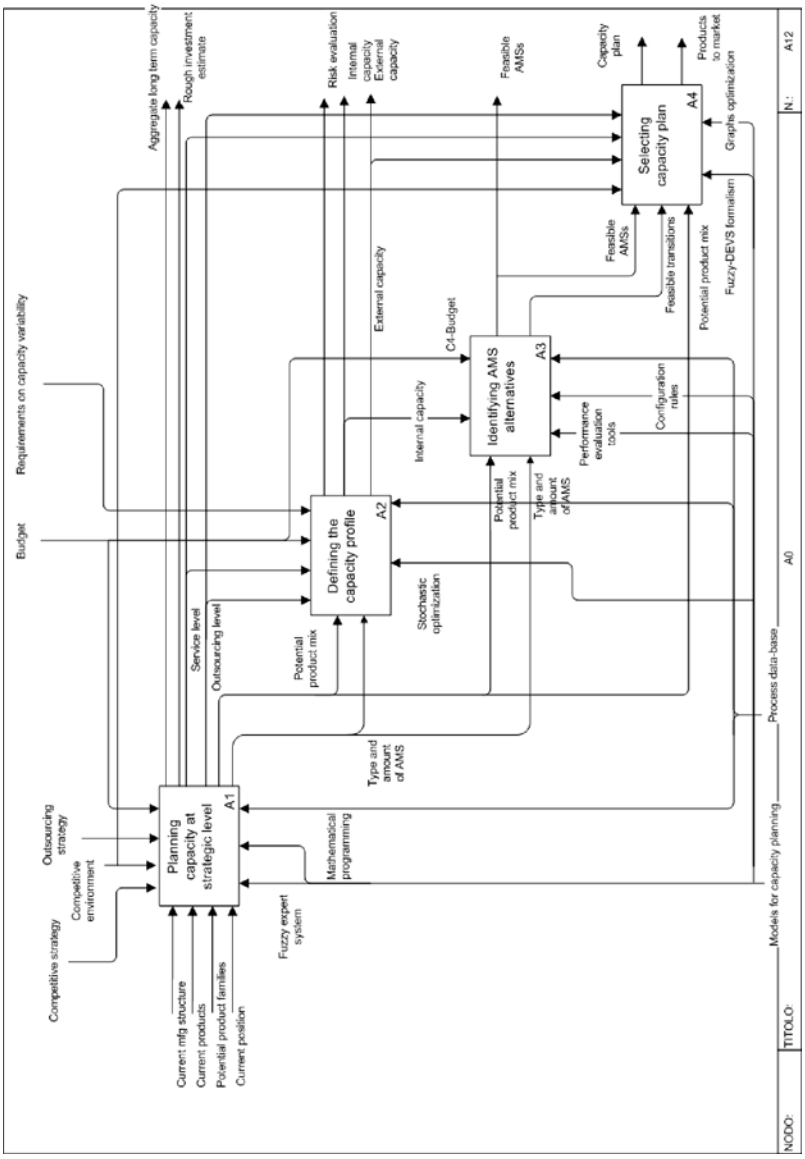


Figure 1.5. A0 level diagram.

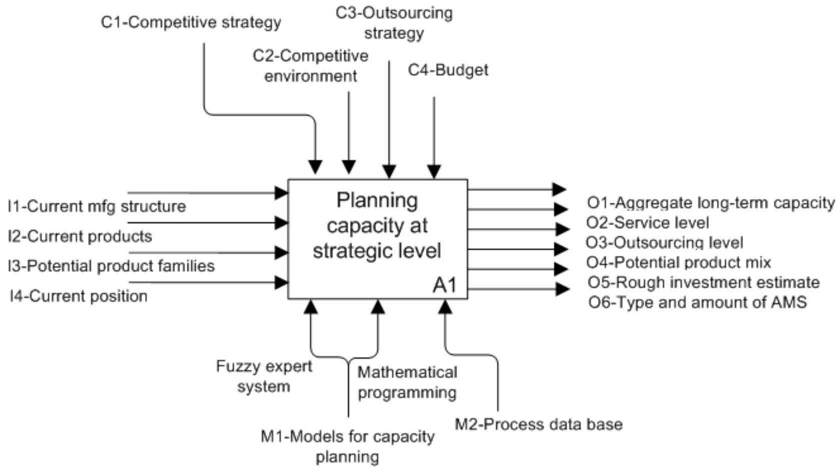


Figure 1.6. A1 context diagram.

- [A1]-I1 **Current manufacturing structure**: production systems owned by firms and described in terms of type, production rate, cost and availability. Details on system practices are not necessary at this level of the problem. This information arrives from the manufacturing and accounting areas.

- [A1]-I2 **Current products**: process cycle of products currently manufactured by the firm, historical production volumes, historical demand, forecasting on the average demand value in the long term, internal and eventually external production costs. This information arrives from manufacturing and accounting areas.

- [A1]-I3 **Potential product families**: rough process cycle of potential products the firm may manufacture in the long term, forecasting on the average demand value in the long term, estimates on internal, and eventually external, production costs. The indication of the products derives from corporate decisions while the detailed information arrives from R&D, marketing and manufacturing areas.

- [A1]-I4 **Current position**: definition of the actual firm's market position. Current position is characterized by the market share, or the growth rate for each actual product, etc. This information comes from the marketing area.

The constraints imposed by the company strategy and the global context are described in more details:

- [A1]-C1 **Competitive strategy**: the market strategy the company selects to pursue such as cost leadership, differentiation but also risk attitude of management, financial strategy of the company, etc. This information arrives from the corporate level and is qualitative.

- [A1]-C2 **Competitive environment**: information regarding the market in which the company intends to compete, that is the market uncertainty level, the market competition rate, the market innovation rate, the market concentration rate and so forth. This information derives from the marketing area.

- [A1]-C3 **Outsourcing strategy**: information about the policy decided at the corporate level and main supplier characteristics such as location, reliability, outsourcing prices and so on. In a few words the outsourcing conditions describe the market supplier network where the company usually does business. This is very important for defining capacity acquisition strategies because market supplier network constraints can affect the company "make or buy" strategy. This information arrives from the corporate level and the manufacturing area.

- [A1]-C4 **Budget**: negative cash flows available in the planned horizon for the investment in additional manufacturing capacity. A reasonable assumption is that portions of budget that are not invested in a time period can be used in the following ones. This information arrives from the corporate level.

The output will consist of the following indications:

- [A1]-O1 **Aggregate long term capacity**: amount of production capacity required to produce the potential production mix at the established service level. This information can be useful to decision-makers that evaluate the output of activity A1 and can decide to introduce some changes in the problem definition.

- [A1]-O2 **Service level**: definition of the minimum level of satisfaction of the market demand that is acceptable to achieve the strategic goals. This information is necessary to define an optimal capacity planning in the long term and represents a constraint for activities A2 and A4.

- [A1]-O3 **Outsourcing level**: detailed indications about "make or buy" strategies of the firm. In particular for each product a range of admissible levels of externalization is defined. This information is necessary to limit the outsourcing coherently with the company's strategic decisions already taken at the corporate level and represents a constraint for activities A2 and A4.

- [A1]-O4 **Potential production mix**: preliminary selection on the types of products the firm could manufacture in the planning horizon specifying long term volumes for each product. This information is an input for all downstream activities.
- [A1]-O5 **Rough investment estimates**: preliminary estimate on the investment cost that is necessary to acquire additional capacity. This information can be useful to decision-makers who evaluate the output of activity A1 and can decide to introduce changes in the problem definition.
- [A1]-O6 **Types and amount of AMS**: indications on the type of manufacturing system architectures potentially profitable to work the production mix. As already written in the previous sections, the models in this book will deal with Dedicated Manufacturing Systems and Flexible Manufacturing Systems. It is possible that the same product can be manufactured profitably on both types of AMS, in this case the final choice will be made downstream this activity after more refined analysis, or the firm decides to adopt, if possible, both systems to get more flexibility. This information is an input for activities A2 and A3.

To produce its outputs, activity A1 uses:

- [A1]-M1 **Mathematical Programming**: standard mathematical programming techniques are used to define a first capacity planning that is necessary to have for making the strategic decisions above described.
- [A1]-M2 **Expert Systems**: an expert system with fuzzy rules is proposed to make decisions at this level. Indeed the fuzzy approach seems proper to model the vagueness characterizing the strategic design input variables.
- [A1]-M3 **Process & System data base**: at this level aggregate and rough technical data are used to define and evaluate the production rate of the alternative AMS. What is important at this level is the total product processing time at a machine and the aggregate costs of the system such as investment and main variable and fixed costs. Other useful information includes the lead time

of the AMS, that is the time between the ordering of an AMS and the beginning of its running.

[A2] Defining the capacity profile. After the main strategic decisions have been made, the next step is to decide the production capacity that has to be acquired during the whole planning horizon. Indeed, the analysis developed at the previous step can suggest only an aggregate long-term production capacity level without taking into account dynamic features such as the market demand volatility or the possibility of delaying the investment during the planning horizon (Lim and Kim, 1998; Brandimarte and Villa, 1995; Dangl, 1999). The ability of the system to meet the demand can be assessed considering both capacity of owned production systems and corrective actions: for instance, demand peak can be smoothed out by acquiring some extra-capacity from a subcontractor or by stocking in advance some production items. Of course, these actions cannot be always taken, and even if they can, a careful optimization of their use is needed, as they have a cost. However, information about the possibility of using subcontractors (such as costs, reliability and accuracy of each sub-contractor, etc.) and storage (storage capacity, inventory costs, and so on) is needed. The purpose of activity A2 is to define the capacity timing introducing the multi-period and dynamic point of view. Starting from the designed strategic variables, A2 defines how to use the make or buy strategy, indicating how much capacity the firm has to acquire (see Figure 1.7) per period and how to proceed with the externalization, if there is any, suggesting more refined values. From activity A1, activity A2 gets as inputs:

- [A2]-I1 **Potential product mix**: the technical characteristics of products are necessary to plan capacity in the long term.
- [A2]-I2 **Types and amount of AMS**: the choice between rigid or flexible capacity affects the definition of the amount of internal capacity, as they have different investment and operating costs.

From activity A1, activity A2 receives as constraints:

- [A2]-C1 **Outsourcing level**: external capacity planning is limited by the decision made in activity A1. In such a way the strategic directions on outsourcing provided by the corporate flow through the decisional process.
- [A2]-C2 **Service level**: it is necessary, in order to avoid unsatisfactory or trivial solutions, to know the minimum level of demand fulfilment the firm should guarantee as decided in activity A1.

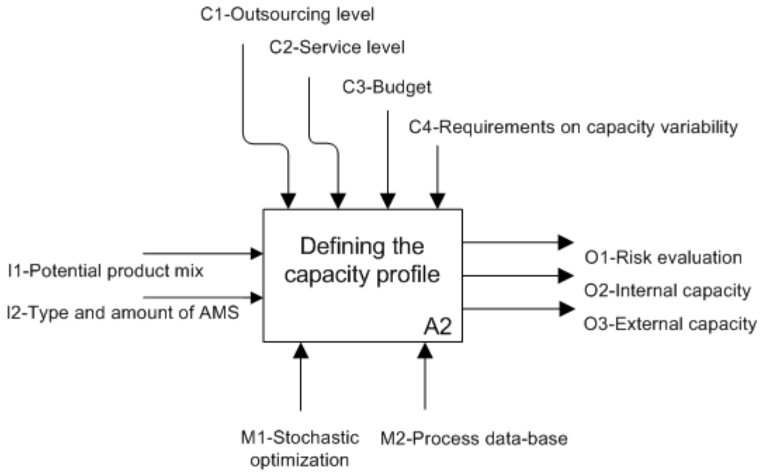


Figure 1.7. A2 context diagram.

From decision-makers activity A2 gets as constraints:

- [A2]-C3 **Budget**: the budget available limits the set of feasible solutions.
- [A2]-C4 **Requirements on capacity variability**: decision-makers can introduce more constraints during the planning horizon. For instance, it could be dangerous to double the capacity of the firm because serious operative problems could be incurred. Whatever the reason is, the decision-maker can introduce this type of constraint on the capacity of the firm. Because of the level of detail, this constraint is not present in the context diagram of Figure 1.4.

The output of activity A2 consists essentially of:

- [A2]-O1 **Risk evaluation**: a preliminary evaluation of the risk in terms of variability of cash flows. This information is useful for decision-makers that analyze the output of the activity.
- [A2]-O2 **Internal capacity**: information on the amount of capacity that is necessary to have in the future. It is represented as a time varying range of required capacity for each type of internal resource (i.e. DMS and FMS). This range of effective production capacity per period is used by the downstream activity A3 to select configurations feasible to the plan.
- [A2]-O3 **External capacity**: information on the amount of capacity that will be probably externalized in the future. It is rep-

resented as a time varying range of required capacity that it may be externalized. This range of external production capacity per period is used in activity A4 to select the best capacity plan.

The essential mechanism is based on:

- [A2]-M1 **Stochastic optimization**: optimization methods are used to select the time manufacturing capacity taking into consideration the uncertainty of the market demand and the different decision times in the planning horizon.

- [A2]-M2 **Process & System data base**: same information as in activity A1.

[A3] Identifying the AMSs alternatives. Function A3 is the activity that has the objective of accurately defining the potential configurations of AMSs during the planning horizon specifying all the allowable changes that can be introduced into the system to react to future market evolutions. These identified configurations are only potential because they are a preliminary selection of the production systems to be adopted in the future; the final selection will be done by the downstream module A4. In order to define a preliminary set of detailed configurations, it is necessary to consider the range of capacity established at higher level by activity A2. The example in Figure 1.8 shows the capacity profile the AMS has to respect. In such a way the sets of systems that do not fit with the range of capacity provided by activity A2 are discarded thus decreasing the number of potentially adoptable solutions.

The detailed alternative investment plans, which are the outputs of the activity, are modelled as possible paths in a graph in which nodes represent detailed configurations (i.e. type and number of machines, carriers, fixtures, etc.) and arcs represent feasible transitions for moving from a specific configuration to another one (i.e. in the case market demand increases, a production system can be enlarged by adding a new machine). The example in Figure 1.9 shows the identified alternative AMS configurations based on the internal capacity profile of Figure 1.8.

The inputs of activity A3 are:

- [A3]-I1 **Potential product mix**: detailed information on the potential set of products. At this level the analysis is more refined and the information on the single processing operations is necessary in order to correctly estimate machine processing times. This information arrives, enriched with more details, from activity A1.

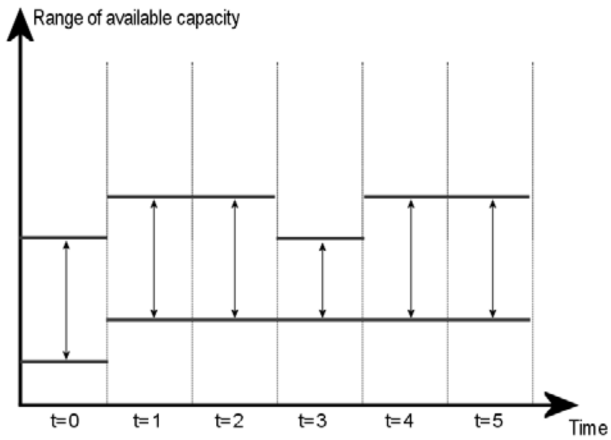


Figure 1.8. Example of internal capacity input for each AMS.

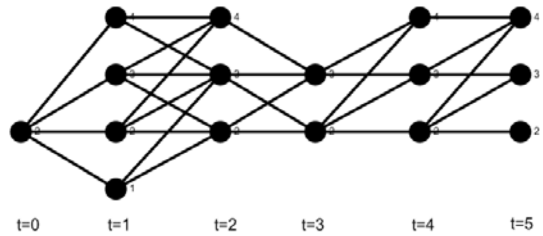


Figure 1.9. Example of graph of feasible alternatives.

- [A3]-I2 **Types and amount of AMSs**: set of manufacturing systems to dimension by allocating their resources: type and number of machines, carriers, tools and buffers. This is the same information that arrives from activity A1 to A2.

Activity A3 is controlled by:

- [A3]-C1 **Internal capacity**: the estimated needed internal capacity range expressed in number of pieces for each product. The capacity range depends on the time period. This information arrives from activity A2.
- [A3]-C2 **Budget**: the budget constraint inserted from decision-makers.

The outputs of activity A3 are:

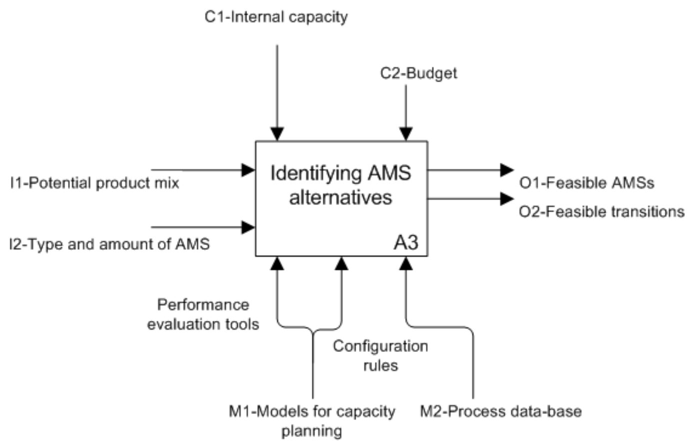


Figure 1.10. A3 context diagram.

- [A3]-O1 **Feasible AMSs**: feasible configurations to adopt in the planning horizon. The information on these configurations (i.e. the nodes of the graph) is very detailed because it specifies the type of systems with all their resources such as machine tools, buffers, part carriers, tool carriers, fixtures, load/unload stations. This information is used by activity A4.
- [A3]-O2 **Feasible transitions**: future allowable changes in configurations (i.e. the arcs of the graph) that can be introduced by the firm in the future. Costs and times to implement transitions on configurations are also provided as outputs of the activity. This information is used by activity A4.

Mechanisms used by activity A3 are essentially:

- [A3]-M1 **Performance evaluation tools**: analytical methods are used to evaluate the performance of configured manufacturing systems. In particular, simple and static equations model in an approximate way the behavior of AMS in a preliminary analysis, while queuing theory is used to dynamically evaluate the behavior of manufacturing systems.
- [A3]-M2 **Configuration rules**: set of technological rules that allows the proper selection of system devices coherently with the operations of potential products.
- [A3]-M3 **Process & system database**: detailed information on system devices: speed of machines, working cube, movement times, etc.

Given the type of problem and the long planning horizon , these tools must take into account uncertainty. In particular, within activity A3 two different performance evaluation modules are used, the first one based on approximate analytical techniques for cases where uncertainty can be expressed in stochastic terms (see Chapter 4) and the second one for cases where uncertainty must be evaluated in fuzzy terms (see Chapter 6).

[A4] Searching for the capacity acquisition plan. The goal of activity A4 is to find out the most profitable capacity plans in the planning horizon on the basis of detailed alternative configurations defined by activity A3. To do this it is necessary to "simulate" the described market environment selected at strategic level in order to estimate the value of the different capacity plans. Optimal plans have to consider outsourcing policies defined by activity A2; indeed a simulation of the market will allow calculating profitability of the outsourcing level of a specific product instead of acquiring more internal capacity. Furthermore, optimal plans have to fit with all the constraints established at strategic levels such as the profile of budget, or the maximum acceptable risk. Therefore, final plans of internal and external capacity are provided to decision-makers. Again, decision-makers will have the possibility of interacting with this module; for instance, a request from the user could be the evaluation of a particular plan that has in his mind, or to recalculate optimal plans after the introduction of more strict constraints. The representation by fuzzy set theory may be useful to represent market condition variability and its relative level of uncertainty. This is possible both in a long period view, in which it is very difficult to represent the probabilistic values, and in a short period view, where, for example, it is necessary to define the demand of a new product. In this case, in fact, no historical data are available to infer correctly the probability distributions.
Under activity A4 innovative algorithms to solve a Travel Salesmen Problem on fuzzy networks are devised. The model must consider not only the uncertainty in the input parameters provided by the configuration activity, but also all the evolutions that are consequences of the different adopted strategies.

The inputs of the module are made up of the following information:

- [A4]-I1 **Feasible AMSs**: the nodes of the graph, i.e. feasible configurations per period, their performance and optimal areas of exploitation, from activity A3.

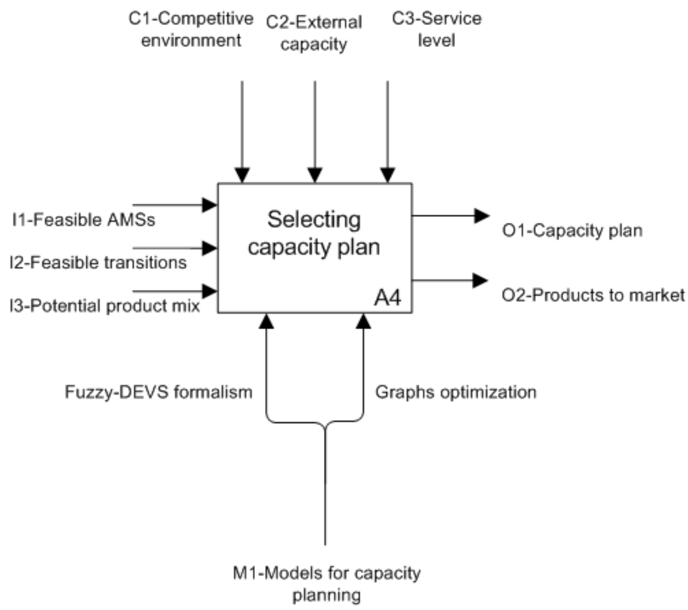


Figure 1.11. A4 context diagram.

- [A4]-I2 **Feasible transitions**: the list of the feasible transactions between the configurations of each successive sub-periods with the related cost and time, from activity A3.

- [A4]-I3 **Potential product mix**: the whole set of information about products. This information arrives from activity A1.

Constraints to the activity A4 are constituted by:

- [A4]-C1 **Competitive environment**: description of the environment in which the firm operates. This information, directly from decision-makers, is used to generate simulations for testing the feasible AMSs.

- [A4]-C2 **External capacity**: possible outsourced quantities for each product. In the case in which a system cannot satisfy at the desired service level the market demand the firm can outsource production in the quantities specified by this control. This control arrives from activity A2.

- [A4]-C3 **Service level**: the level of market demand satisfaction. If this constraint is not satisfied a plan is considered unfeasible. This control arrives from activity A1.

The output of activity A4 is:

- [A4]-O1 **Capacity plan**: the detailed capacity acquisition plan in the planning horizon. The information concerns the optimal evolution path among the many possible paths on the graph of the alternative configurations, in other words which type and how many resources for each type have to be acquired in every time period and which capacity size will be the recourse to externalization.

- [A4]-O2 **Products to market**: the final choice of products, among the potential ones in input, that are selected for the production in each time period, i.e. the products that can be profitably marketed.

The mechanisms used in activity A4 are:

- [A4]-M1 **Fuzzy-DEVS formalism**: given the vagueness of information, some parameters must be defined in fuzzy terms for each elementary period. Therefore, a model to represent production system dynamics under a fuzzy market representation is developed: to achieve this goal, Fuzzy-DEVS formalism [Anglani et al., 2000] is used.

- [A4]-M2 **Graphs optimization**: heuristics for finding the optimal path in the graph defined by activity A3.

The models described as mechanisms in the proposed framework will be described in the following chapters. These models implemented in software tools constitute SW modules which, integrated in a common software architecture, will constitute the specific packages of a Decision Support System (DSS) to long term capacity planning in AMSs:

Strategy planner Module to aid decision-makers in taking strategic decisions such as market segment, market share, growth rate, products to market, etc. This module will be based on an expert system incorporating top management rules. Vagueness of market will be modelled by means of fuzzy set theory. An aggregate capacity planning is also modelled. For details see models described in Chapter 2.

Risk planner Module to aid decision-maker in defining the profile of the estimated capacity that is necessary to pursue strategic goals. This module will be based on stochastic and dynamic optimization to find the required time capacity in the planning horizon. Once required capacity has been defined, another module will be able to aid decision-makers in selecting between internal (e.g. firm's shop floor) and external (e.g. outsourcing) resources. For details see the models described in Chapter 3.

Configurator Module to aid decision-makers in defining the detailed design of production system. This module will use a performance evaluator tool to find good solutions. For details see models described in Chapter 4 and Chapter 6.

Capacity selector Module to aid decision-makers in selecting "which" AMSs have to be acquired and "when" they will be acquired. This module will select also the type of capacity (i.e. internal or external) and will be based on optimal path's search in graphs where both arcs and nodes are weighted by costs. For details see models described in Chapter 5.

Decision models have been tested on a real case in the automotive metal-component sector. In this market the fierce competition leads firms to increase the flexibility of their facilities in order to react to the frequent market changes (Koren et al., 1997,Matta et al., 2000,Matta et al., 2001). The reason for this change in turn is motivated by the fact that automotive suppliers tend to increase the range of products to attract the consumer, launch new models of car and decrease the time to market; in other words more attractive products in shorter intervals they propose in the market and more competitive they are. Each component manufacturer tends to produce well defined types of products (e.g. outlet manifolds) that are supplied to different car manufacturers, therefore the market is composed by fewer and fewer focused suppliers. The whole market of final goods is subject to uncertainty: each single final product can be a success or a failure and the same is for components of which is made. Given the fact that stock reduction and just in time policies are normally adopted, the producer of components must follow, even in the short term, the fluctuation in the demand. Also the weak contractual power of producers of components reduces the profit per part. Car component suppliers suffer this trend. They have to face frequent changes in product demand, changes in mix, modifications on existing products and introduction of new products selecting the best production system in terms of profitableness. Existing production systems do not match with the above market trends. Traditionally DMS have been adopted for the production of a small family of part types (one or few part types) requested by the market in high volume (Matta et al., 2001). Since DMS scalability is low they are normally sized to reach from the beginning the maximum market demand the firm forecasts to satisfy in the future. But in many situations DMS do not operate at full capacity due to the lack of demand. Analyzed transfer lines operating in the sector of automotive components were saturated 53 % on average (Matta et al., 2000). In this case DMS profitableness is very low because the potential capacity of the system is not exploited. On the contrary FMS have been adopted for the production of a large part mix in small quantities. FMS are conceived to react to all the possible changes of the market, therefore their flexibility is too large and expensive for the needs of the firms (Perrone and Diega, 1999). In many cases car component suppliers partially exploit the flexibility offered by these systems, given the fact that it is rare that their part mix changes completely. Investment to acquire FMS is very high and it considerably affects the cost per part unit produced. Flexibility, customized to the potential future changes the products may undergo, would be fully exploited by the car component suppliers that would not buy unneeded flexibility. The production system would be designed with the desired level of flexibility so that it can face efficiently the future changes of the part families during their life cycle (Matta et al., 2000,Perrone and Diega, 1999). The solution is to have rapid adaptive machines to industrialize new parts in short times, to react to limited changes in demand and part features, and finally to produce with low cost per part (Perrone and Diega, 1999). All the proposed models have been validated on the data collected from an enterprise competing in the component automotive sector.

References

Anglani, A., Caricato, P., Grieco, A., Nucci, F., Matta, A., Semeraro, Q., and Tolio, T. (2000). Evaluation of capacity expansion by means of fuzzy-devs. *14nd European Simulation Multiconference, Ghent, Belgium*, pages 129–133.

Brandimarte, P. and Villa, A. (1995). *Advanced models for manufacturing systems management*. CRC Press Boca Raton (FL).

Dangl, T. (1999). Investment and capacity under uncertain demand. *European Journal of Operation Research*, 117:415–428.

Fine, C. and Freund, R. (1990). Optimal investment in product-flexible manufacturing capacity. *Management Science*, 36(4):449–464.

Garvin, D. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6):101–109.

Gerwin, D. (1992). Manufacturing flexibility: A strategic perspective. *Management Science*, 39(4):395–410.

Hayes, R. and Wheelwright, S. (1984). *Restoring our competitve edge*. Wiley, New York.

IDEF. Idef0 function modeling method. Website address: IDEF0 Function modeling method.

Koren, Y., Hu, S., and Weber, T. (1997). Impact of manufacturing system configuration on performance. *Annals of CIRP*, 47(1):369–372.

Lim, S. and Kim, Y. (1998). Capacity planning for phased implementation of flexible manufacturing systems under budget restrictions. *European Journal of Operation Research*, 104:175–186.

Matta, A., Tolio, T., Karaesmen, F., and Dallery, Y. (2000). A new system architecture compared with conventional production system architectures. *International Journal of Production Research*, 39(17):4159–4170.

Matta, A., Tolio, T., Karaesmen, F., and Dallery, Y. (2001). An integrated approach for the configuration of automated manufacturing systems. *Robotics and CIM*, 17(1-2):19–26.

Naik, B. and Chakravarty, A. (1992). Strategic acquisition of new manufacturing technology: a review and research framework. *International Journal of Production Research*, 30(7):1575–1601.

Olhager, J., Rudberg, M., and Vikner, J. (2001). Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning. *International Journal of Production Economics*, pages 215–225.

Perrone, G. and Diega, S. N. L. (1999). A reference scenario for ims strategic design. *Annals of CIRP*, 48:381–384.

Porter, M. (1980). *Competitive strategy*. The Free Press, New York.

Price, D., Beach, R., Muhlemann, A., Sharp, J., and Paterson, A. (1998). A system to support the enhancement of strategic flexibility in manufacturing enterprises. *European Journal of Operation Research*, 109:362–376.

Sethi, K. and Sethi, S. (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, 2:289–328.

Skinner, W. (1985). *Manufacturing: the formidable competitive weapon*. John Wiley and Sons.

Vickery, S., Droge, C., and Markland, R. (1993). Production competence and business strategy: do they affect business performance ? *Decision Sciences*, 24(4):435–456.

Ward, P. and Duray, R. (2000). Manufacturing strategy in context: environment, competitive strategy and manufacturing strategy. *Journal of Operations Management*, 18:123–138.

Whitmore, J. (1907). Factory accounting applied to machine shops. *Journal of Accountancy*, page 213.

Williams, F., D'Souza, D., Rosenfeldt, M., and Kassaee, M. (1995). Manufacturing strategy, business strategy and firm performance in a mature industry. *Journal of Operations Management*, 13:19–33.

Chapter 2

A DSS FOR STRATEGIC PLANNING

M. Bruccoleri, G. Lo Nigro, S. Noto La Diega
Dipartimento di Tecnologia Meccanica, Produzione e Ingegneria Gestionale (DTMPIG)
Università degli Studi di Palermo
Viale delle Scienze, 90128, Palermo, Italy
manbru@dtpm.unipa.it, glonigro@dtpm.unipa.it, nld@dtpm.unipa.it

P. Renna, G. Perrone
Dipartimento di Fisica Ambientale
Università degli Studi della Basilicata
C. da Macchia Romana, 85100, Potenza, Italy
perrone@unibas.it, renna@unibas.it

Abstract This chapter presents an innovative approach for assisting entrepreneurs in making long term capacity decisions in Advanced Manufacturing Systems (AMSs). AMSs require high investment costs in manufacturing equipments, human resources and technology knowledge. Such high investments together with the wideness and the variability of the competition scenario contribute to increase the perception of the risk for industrial entrepreneurs especially in SMEs. This problem could be approached by providing the entrepreneur with a Decision Support System (DSS) able to assist her/him in making long term capacity decision in AMS. The DSS proposed in this chapter allows the entrepreneur to plan its production strategy starting from company business strategy, market strategy, competition scenario and outsourcing scenario. Starting from such information, a Fuzzy Expert Systems allows defining the kind of strategic flexibility the company needs and how the company should compose its production mix between internal production and outsourced one. This strategic information represents the input of a Long Term Capacity Planning Model based on economy of scope models that constitutes the economic and financial hearth of the DSS.

Keywords: Advanced Manufacturing Systems, long-term capacity planning, fuzzy systems, economy of scope.

Introduction

This chapter is about the investigation of the major phases of a strategic planning process in advanced manufacturing systems (AMSs). These systems represent manufacturing technologies that embody all the advantages springing from industrial automation (Numerical Control, Robot, AGVs), integrated and computerized control (Industrial Local Area Network), distributed architecture (agents and holonic manufacturing) and distributed artificial intelligence techniques.

From a global competition point of view, especially if we refer to small and medium enterprises (SMEs) operating in highly dynamic and competitive industries, AMSs enable enterprises to acquire flexibility, i.e. the ability to react fast and with low costs to market changes (Agile Manufacturing).

This is the reason why both industrials and academics agree in assuming a strategic approach to evaluate AMSs investments (Naik et al., 1992). On the other hand, every decision making process related to manufacturing investments involves considerations regarding: risk evaluation, uncertainty estimation, investment planning and timing. In case of AMS investment, these factors heavily impact the final decision because: a) the enterprise perceives a risk that is higher if compared with other manufacturing investments; b) AMSs embody an high flexibility degree that enlarges the investment scenario making higher the investment uncertainty; c) the competitive scenario evolution needs to be also evaluated in order to carry out a correct investment planning and timing.

For such reasons, AMSs investment decisions are perceived, especially from SMEs, as high risk decisions in a very uncertain and complex environment. Many entrepreneurs and researchers have highlighted that this complexity and the related risk do not encourage the adoption of AMSs causing a looseness of competitiveness for SMEs; at the same time, it has been also stressed how the availability of proper Decision Support Systems (DDSs) able to assist enterprises in making decisions about AMSs investments, could reduce the risk and the complexity perception making SMEs more competitive and profitable (Price, et al., 1998). It should be known that AMS design is a complex process that can be hierarchically divided into three phases: a) strategic design, b) production system configuration, c) detailed design. The strategic design phase aims at providing suggestions and indications about AMS strategic variables such as flexibility forms (mix flexibility, technological flexibility, volume flexibility, expansion flexibility and so forth), competitive policies (production mix and volumes etc.), make or buy strategies, and of course, an estimation about the long term capacity to be installed in a time horizon equal to the AMS life cycle. The set of the above decisions puts some important architectural constraints that need to be considered in the following, and more detailed, design phases.

During the last decades, several researches focused on supporting the entrepreneurs in making right decisions, at strategic level, about AMSs investments. The result is a very rich and articulated literature. A detailed analysis of the literature concerning the AMSs strategic issues reveals a predominance of qualitative studies whose main objective is to stress the strategic impact of AMSs; in particular, it has been pointed out how important is to conceive a proper manufacturing strategy aligned with the market (Berry et al., 1999) and in the meantime, able to take into account for the influence of new technologies such as AMSs (Banerjee, 2000; Wu et al., 2000). From a more specific manufacturing point of view, several researchers have pointed out the strategic impact of manufacturing flexibility in changing times (Frazelle, 1986; De Meyer et al., 1989; Tombak, 1990), in order to improve the company ability in creating new markets, reacting faster to market changes, reducing time to market for new product developing.

On the other hand, quantitative studies at strategic level have principally focused on flexibility evaluation and measurement (Feurstein et al., 2000; Parker et al., 1999; Shewchuk, 1999; Bateman et al., 1999), strategic evaluation of AMS installation (Sarkis et al., 1999; Elango et al., 1994; Sheng et al., 1995), economic and financial justification of AMSs (Albayrakoglu, 1996; Mohanty, 1993; Parsei et al., 1989), design approaches at strategic levels (Chan et al., 2000; Babic, 1999; Perrone et al., 1999-a), optimal capacity models for flexible manufacturing systems (Fine et al., 1990) under constraints situations (Lim et al., 1998) and several market conditions (Chung et al., 1998), and finally, the analysis of uncertain impact on AMSs investments decisions (Dangl, 1999; Harrison et al., 1999).

However, from the analysis of the literature three paths that should be deeply investigated emerge: a) the formulation of a set of theoretical models able to highlight the real competitive advantage that several forms of AMSs can lead to a company; b) a deep analysis of the impact of the scenario uncertainty and vagueness on AMSs strategic design decisions; c) the development of an integrated and comprehensive support system able to assist the entrepreneur in all the aspects concerning the definition of AMSs investment decisions (Price et al., 1998). Investment decisions mainly concern the production strategy and the long-term capacity planning, and include decisions regarding the typologies of manufacturing systems to purchase throughout the planning horizon, eventually a mix of typologies such as dedicated manufacturing lines or flexible manufacturing systems. The development of an integrated decision support environment that puts into operation such features, cannot be obtained independently from the development of a theoretical framework and from a deep understanding about how scenario uncertainty can impact such decisions (Perrone et al., 1999-b).

The research presented here follows these directions, specifically, focusing on the development of a theoretical framework able to provide a general understanding of what are the market and competition conditions that have a critical impact on the strategic planning of manufacturing capacity. Then, a decision support system which implements this theoretical framework and suitable for assisting entrepreneurs in making the right strategic decision for manufacturing system design and planning, is presented.

1. The strategic planning process

The activity "A1 Planning at strategic level" of the IDEF0 context diagram, reported in Figure 1.6 of Chapter 1, can be decomposed in two macro-activities: production strategy planning and long-term capacity planning. Figure 2.1 shows the IDEF0 diagram of these activities.

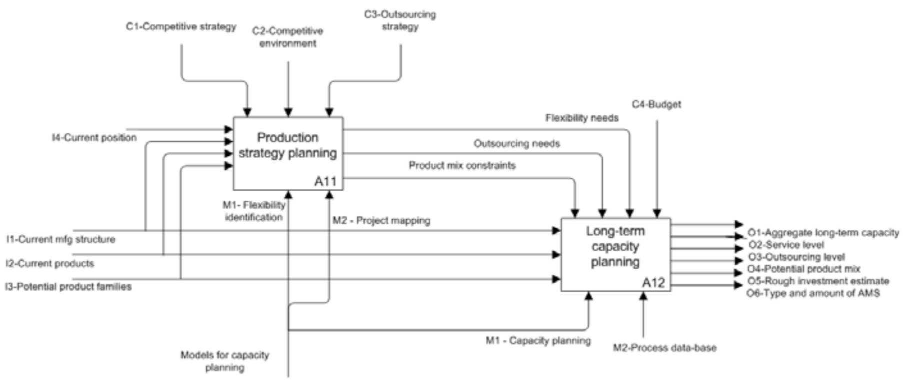


Figure 2.1. A1 level diagram: the strategic planning process.

According to the IDEF0 graphical notation, the inputs, outputs, constraints, and models reported in Figure 2.1, are fully explained in the following.

1.1 [A1-1] Production Strategy Planning

This activity aims at the identification of the enterprise production strategy. The production strategy involves three main decisions: the flexibility types to adopt (product, routing, expansion, and reconfiguration flexibility), the product mix constraints, and the outsourcing policy constraints.

Input

- [A1-1]-I1 **Current manufacturing structure**: such input concerns information regarding the production systems and manufacturing equipments that are currently used to manufacture the current products.
- [A1-1]-I2 **Current products**: the set of products the enterprise currently produces.
- [A1-1]-I3 **Potential products**: the set of the products the enterprise is going to produce in the future.
- [A1-1]-I4 **Current position**: this input describes the current market positioning of the enterprise.

Output

- [A1-1]-O1 **Flexibility types**: this output indicates which flexibility types (product, routing, expansion, and reconfiguration flexibility) are necessary for producing the potential product mix. The output is expressed by a linguistic term associated to each flexibility type, e.g. “the routing flexibility is very important”.
- [A1-1]-O2 **Outsourcing constraints**: this output summarizes the outsourcing policies related to every product and it is expressed by means of a linguistic term, which represents the level of suitability of outsourcing a given product, e.g. “the product AA1 should be strongly outsourced”.
- [A1-1]-O3 **Product mix constraints**: this output gives information regarding the competitive constraints that should be applied to the mix of products. In particular, the output is expressed by means of a linguistic term, which represents the product mix policy to be adopted, e.g. “the product mix should be amplified”.

Constraints

- [A1-1]-C1 **Competitive strategies**: this constraint indicates the business strategy the enterprise wants to pursue. The constraint is expressed by a linguistic term that synthesizes the business strategy itself, e.g. "the business strategy is oriented to the product differentiation".

- [A1-1]-C2 **Competitive environment**: this constraint includes information on the market scenario where the enterprise wants to compete, for example information on the uncertainty level, competition level, and innovation rate. The constraint is expressed by a linguistic term, which represents the competitive scenario, e.g. "the competitive scenario is strongly dynamic".

- [A1-1]-C3 **Outsourcing strategy**: such constraint concerns the market conditions related to a potential outsourcing activity of some products or components. This constraint includes a preliminary analysis of the suppliers of the products and components to be outsourced, in view of their availability, their reliability, and the outsourcing costs. The constraint is expressed in terms of a linguistic term related to the suitability of the outsourcing activity of a given product, e.g. "the cost of outsourcing product AA1 is low".

Models

- [A1-1]-M1 **Models for flexibility identification**: these models are based on expert systems, specifically fuzzy systems, that determine which flexibility types (product, routing, expansion, and reconfiguration flexibility) are strategic for the enterprise, as a result of considerations on current products, potential products, business strategy, and competitive scenario.

- [A1-1]-M2 **Models for project mapping**: these models are based on fuzzy systems that identify which outsourcing constraints and product mix constraints are strategic for the enterprise, as a result of considerations on current products, potential products, business strategy, and competitive scenario.

1.2 [A1-2] Long-term Capacity Planning

This activity involves the determination of the manufacturing resource mix (composition of the manufacturing system as a mix of dedicated, flexible, and reconfigurable resources) and the relative manufacturing capacity.

Input

- [A1-2]-I2 **Current products**: the set of products the enterprise currently produces.
- [A1-2]-I3 **Potential products**: the set of the products the enterprise is going to produce in the future.

Output

- [A1-2]-O1 **Aggregate long-term capacity**: this output indicates the number of manufacturing system for each type, which should be added to the current manufacturing system configuration.
- [A1-2]-O2 **Service level**: this output gives information about the level of demand fulfillment for every product and for every time bucket (2 years).
- [A1-2]-O3 **Outsourcing level**: this outputs indicates, for each product, the total volume percentage, which should be, outsourced.
- [A1-2]-O4 **Potential production mix**: this output indicates the products that should be part of the new product mix.
- [A1-2]-O5 **Rough investment estimates**: this output gives a preliminary estimates on the investment cost that is necessary to acquire additional capacity.
- [A1-2]-O6 **Types and amount of AMS**: this output defines the manufacturing system composition in terms of the all possible manufacturing system types (dedicated, flexible, or reconfigurable system).

Constraints

- [A1-2]-C1 **Flexibility types**: this constraint originates from the output [A1-1]-O1 and indicates which flexibility types (product, routing, expansion, and reconfiguration flexibility) are necessary for producing the potential product mix. It is expressed by a linguistic term associated to each flexibility type, e.g. "the routing flexibility is very important".
- [A1-2]-C2 **Outsourcing**: this constraint originates from the output [A1-1]-O2 and summarizes the outsourcing policies related to every product. It is expressed by means of a linguistic term, which represents the level of suitability of outsourcing a given product, e.g. "the product AA1 should be strongly outsourced".

- [A1-2]-C3 **Product mix**: this constraint originates from the output [A1-1]-O3 and gives information regarding the competitive constraints that should be applied to the mix of products. In particular, it is expressed by means of a linguistic term, which represents the product mix policy to be adopted, e.g. "the product mix should be amplified".
- [A1-2]-C4 **Budget**: this constraint defines the budget which is available for investments in AMSs in the long term. This constraint is decided at the corporate level because it involves the analysis of the firm financial position.

Models

- [A1-2]-M1 **Models for capacity planning**: these models are based on mathematical programming algorithms that, by elaborating fuzzy information from the prior production strategy planning activities, identify the strategic mix of manufacturing systems and manufacturing capacity.

2. Models for Production Strategy Planning

As already mentioned in the previous section, the first level of the strategic planning process concerns the process of determining the production strategy. The production strategy, in few words, consists on making three main decisions:

1 Decisions related to the type of flexibility to be implemented by the manufacturing system in order to be able to manufacture all the parts within the product mix. Product, routing, expansion, and reconfiguration flexibilities are example of flexibility types.

2 Decisions related to the outsourcing policy to be implemented, i.e. the identification of constraints which lead the policy of outsourcing of some manufacturing activities.

3 Decisions related to the competitive constraints, i.e. the constraints of the competitive strategy that influences the design of the production system. Concerning the first kind of decisions, the models for flexibility identification ([A1-1]-M1) have been developed, while regarding the second and third kinds of decisions, the so called models for project mapping ([A1-1]-M2) have been developed.

Concerning the first kind of decisions, the models for flexibility identification ([A1-1]-M1) have been developed, while regarding the second and third kinds of decisions, the so called models for project mapping ([A1-1]-M2) have been developed.

2.1 Models for flexibility identification

The four models for the identification of the flexibility type, one for each flexibility type, are based on expert systems, specifically fuzzy systems. These models are represented in Figure 2.2 where the input and output variables, according to the main IDEF0 diagram are shown.

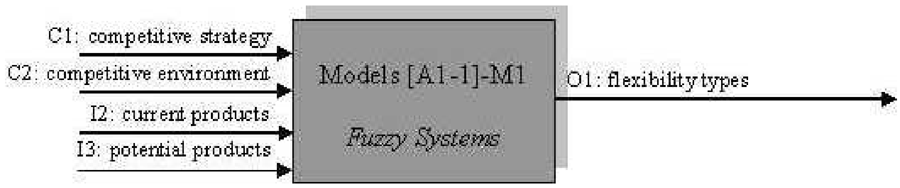


Figure 2.2. Inputs and outputs of the model [A1-1]-M1.

The production system flexibility can be defined as the system ability to rapidly and cost-effectively adapt to market (external) change requirements or enterprise (internal) change requirements. The flexibility types considered by the fuzzy systems are:

- Product (mix- change) flexibility. This is the ability to change the current mix of products by adding new products or substituting the existing ones.
- Routing flexibility. This is the ability to manufacture a product by different alternative process routings throughout the system.
- Expansion flexibility. This is the ability to expand the manufacturing capacity, by means of modular system architectures.
- Reconfiguration Flexibility. This is the ability to change the system configuration when necessary to face market changes for new models of the same product.

As already mentioned four fuzzy systems have been developed, one for each type of flexibility, and each of them uses the business strategy, the competitive scenario constraints, the current products, and potential products input as input variables depending on the considered flexibility type. As output, the fuzzy systems give the importance of each flexibility type. Moreover, these input and output are all linguistic variables (fuzzy variables) that can take the values “low”, “medium”, and “high”; thus, the term set *T(x)* associated to each variable x is *T(x)* = [*Low, Medium, High*], where each term is characterized by a fuzzy set in U=[0 , 1]. The membership functions of the three linguistic values related to every linguistic variable are shown in Figure 2.3.

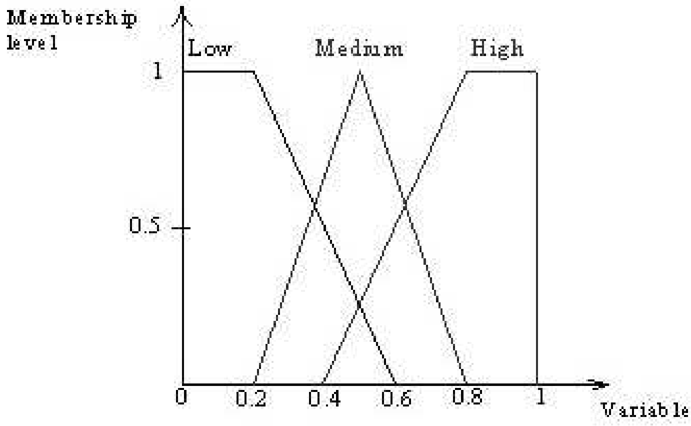


Figure 2.3. Membership functions for the values “low”, “medium”, and “high”.

In Figure 2.4, the input and output specific to the four fuzzy systems are graphically reported. As the reader can notice in Figure 2.4, the fuzzy systems take as input some variables related to the company business strategy and the competitive scenario. In particular, the decision maker should evaluate by means of a linguistic statement the importance of the following input variables:

- reactiveness to internal and external changes
- set-up cost and time reduction
- demand variation
- product life cycle reduction
- reactiveness to internal changes
- product variety

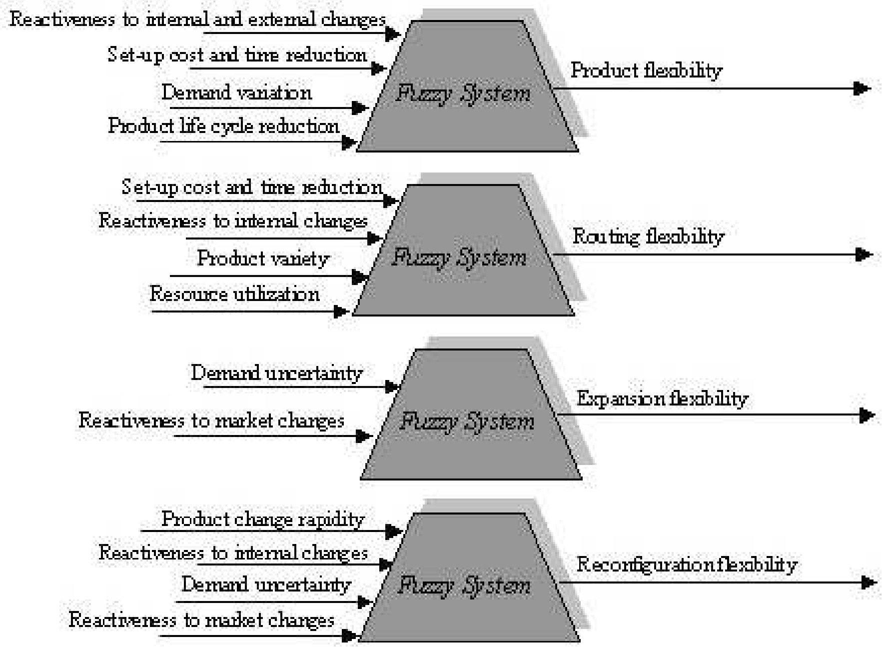


Figure 2.4. Inputs and outputs of the four fuzzy systems of the model [A1-1]-M1.

- resource utilization
- demand uncertainty
- reactiveness to market change
- product change rapidity.

The fuzzy system, by using a knowledge base which consists of a set of fuzzy rules, determines the linguistic values associated with the importance of the output variable of the fuzzy system itself, e.g. the importance of implementing product flexibility. If A_i is the i-th input strategic variable and $L_k(A_i)$ the linguistic value given by the decision maker for the importance of variable A_i, the fuzzy rule $R_{i,f}$, associated with the input variable i and the flexibility type f results:

$$R_{i,f} : \text{IF } A_i \text{ is } L_k(A_i) \text{ THEN } f \text{ is } V_s(f)$$

where $V_s(f)$ is the linguistic variable that expresses the importance associated with the flexibility type *f*.

2.2 Models for project mapping

The models for project mapping are depicted in Figure 2.5 and consist of two fuzzy systems. The former is used to determine the outsourcing conditions and the latter to identify the product mix constraints, as already described in Section 1.

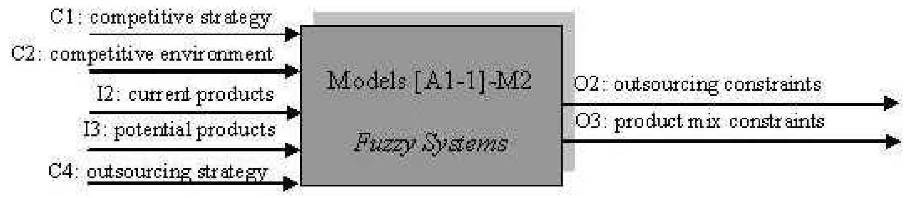


Figure 2.5. Inputs and outputs of the models [A1-1]-M2.

Specifically, Figure 2.6 reports the input and output variables of the two fuzzy systems.

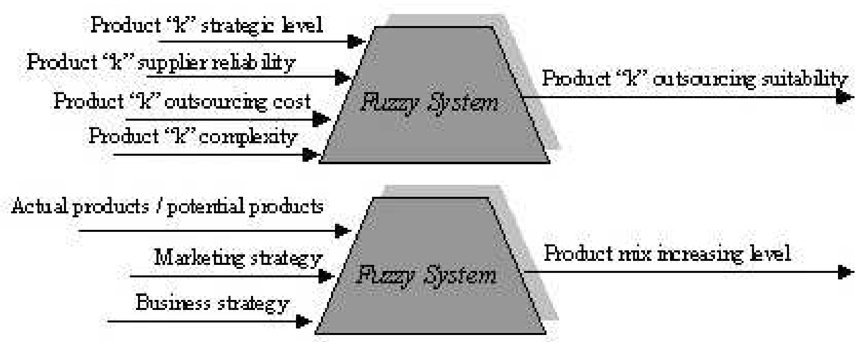


Figure 2.6. Inputs and outputs of the two fuzzy systems of the model [A1-1]-M2.

Specifically, the first fuzzy system, by inferring four linguistic variables describing the strategic level of a specific product "k", i.e. its "strategic level", its "supplier reliability", its "outsourcing cost", and its "complexity", gives as output its "outsourcing suitability". On the other hand, the second fuzzy system, by inferring three linguistic variables describing the business strategies, i.e. the ratio "actual products/potential products", the "marketing strategy", and the "business strategy", gives as output the "product mix increasing level".

3. Models for Long-term Capacity Planning

This model is intended for the identification of the manufacturing system composition in terms of which number of dedicated manufacturing lines (DML), flexible manufacturing systems (FMS), and reconfigurable manufacturing systems (RMS) need to be part of the manufacturing system itself. In other words, the model is able to determine the long term manufacturing capacity for each manufacturing system kind which is implemented. The model runs optimization algorithms based on mathematical programming, capable to deal with fuzzy information. From models [A1-1]-M1 e [A1-1]-M2, it is possible to determine the enterprise production strategy. The production strategy, as it has been defined, means specific choices on the flexibility type to adopt and on the product mix and outsourcing activity constraints to fit. In other words, the output of models [A1-1]-M1 e [A1-1]-M2 represents the input of the model [A1-2]-M1. Such a model is utilized in order to analyze, from an economic perspective, the suitability of a specific manufacturing system configuration (manufacturing mix and capacity) for a given market demand scenario. The next sub-sections present, first, the optimization model in which only the traditional manufacturing system types (dedicated lines and flexible manufacturing systems) are considered. Then, an innovative model, which also takes into consideration the new reconfigurable manufacturing system paradigm, is presented.

3.1 DML and FMS model

Notation:

i	product index, i= *1, ...I*;
j	time bucket index, j= *1, ...J*;
r	cost of capital;
A_j	time availability in j;
D_{ij}	market demand for product i in j;
m_{ij}	contribution margin for product i in j;
V_{ij}^{DML}	volume of product i to manufacture in j by DML;
V_{ij}^{FMS}	volume of product i to manufacture in j by FMS;

V_{ij}^{EST}	volume of product i to outsource in j;
V_{ij}^{TOT}	total volume of product i in j;
DML_i	Dedicate manufacturing line producing product i;
O_i	number of technological operations for product i;
o	technological operation index, $o = 1, \ldots O_i$ for product i;
t_{io}	processing time of operation o for product i;
BT_i	processing time of the bottleneck machine of DML_i;
CP_i^{DML}	total volume of product i in j;
C_i^{DML}	investment cost for DML_i;
L_{ij}^{DML}	number of DML_i to purchase in j;
FT_i	total processing time for manufacturing product i;
L_j^{FMS}	number of FMS to purchase in j;
WL_j^{FMS}	FMS workload for manufacturing I in j;
C_{FMS}	investment cost for purchasing the FMS;

By denoting with α^{FMS} and β^{FMS} the economy of scope parameters that put into relation DML and FMS, the FMS workload and its investment cost can be calculated respectively as in expressions (2.1) and (2.2):

$$WL_j^{FMS} = \sum_{i=1}^{I} \alpha^{FMS} \times FT_i \times V_{ij}^{DML} \tag{2.1}$$

$$C^{FMS} = \beta^{FMS} \times \sum_{i=1}^{I} DML_i \tag{2.2}$$

The economy of scope technological coefficient, α^{FMS}, takes into account that a flexible system takes less time to fulfill a set of operations than a dedicated line. For this reason α^{FMS} satisfies the condition expressed in equation (2.3):

$$BT_i \leq \alpha^{FMS} \times FT_i \leq FT_i \Leftrightarrow \frac{BT_i}{FT_1} \leq \alpha^{FMS} \leq 1 \tag{2.3}$$

On the other hand, the economy of scope cost coefficient, β^{FMS}, takes into consideration that a flexible system which processes a set of parts is less expensive than the set of dedicated lines needed for processing the same set of parts, although the flexible system is more expensive than each of the dedicated lines. For this reason β^{FMS} must satisfy the condition (2.4):

$$\max_i C_i^{DML} \leq C^{FMS} \leq \sum_i C_i^{DML} \Leftrightarrow \frac{\max_i C_i^{DML}}{\sum_i C_i^{DML}} \leq \beta^{FMS} \leq 1 \tag{2.4}$$

In order to set the optimal investment for DML and FMS, an optimization non-linear constrained programming model has been proposed. Such a model maximizes the return on investment (ROI) calculated as in expression (2.5) for what concerns the DML system and expression (2.6) for the FMS.

$$ROI(DML) = \frac{\sum_j \sum_i m_{ij} \times V_{ij}^{DML} \times (1+r)^{1-j}}{\sum_j \sum_i L_{ij}^{DML} \times C_i^{DML} \times (1+r)^{1-j}} \tag{2.5}$$

$$ROI(FMS) = \frac{\sum_j \sum_i m_{ij} \times V_{ij}^{FMS} \times (1+r)^{1-j}}{\sum_j \sum_i L_{ij}^{FMS} \times C_i^{FMS} \times (1+r)^{1-j}} \tag{2.6}$$

By defining the following functions which map the ROI measures into [0 , 1], both the ROI indexes can be taken into consideration in a single objective function. These mapping functions are reported in expressions (2.7) and (2.8):

$$\mu_{ROI(DML)} = \max\left[0, \min\left[1, \frac{ROI(DML) - ROI_{MIN}}{ROI_{MAX} - ROI_{MIN}}\right]\right] \tag{2.7}$$

$$\mu_{ROI(FMS)} = \max\left[0, \min\left[1, \frac{ROI(FMS) - ROI_{MIN}}{ROI_{MAX} - ROI_{MIN}}\right]\right] \tag{2.8}$$

where ROI_{MIN} and ROI_{MAX} are the minimum and the maximum that the decision maker expects to gain for the two economic variables. The following constraints need to be considered.

Model Constraints

1 *Volume composition constraint.* This constraint implies that the total volume needed for product i in the bucket j is given by the sum of the volume of the same product by producing it in the DML and in the FMS and by outsourcing the volume V_{ij}^{EST} as expressed by equation (2.9).

$$V_{ij}^{TOT} = V_{ij}^{DML} + V_{ij}^{FMS} + V_{ij}^{EST} \tag{2.9}$$

2 *Demand fulfillment constraint.* Condition (2.10) guarantees that the total volume of product i in j must be less than the demand of the same product in the same time bucket and greater than a minimum level depending on the strategic level of the product i itself, expressed through a variable $x_i \in [0;1]$.

$$x_i \times D_{ij} \leq V_{ij}^{TOT} \leq D_{ij} \tag{2.10}$$

3 *Outsourcing strategy constraint.* Relation (2.11) expresses that the volume of product i to outsource in the time bucket j depends on the total volume by means of the variable z_{ij} which, in this way, represents the total volume percentage to outsource. Such a variable, takes into account the information from models [A1-1]-M2, concerning the generic product "k" outsourcing suitability.

$$V_{ij}^{EST} = z_{ij} \times V_{ij}^{TOT} \tag{2.11}$$

In particular, as the output coming out from the model [A1-1]-M2 is a linguistic term like "the outsourcing level of product i is VARling" (where VARling can be "high" - H, "medium" - M, or "low" - L), the outsourcing constraint is expressed by the objective function (2.12):

$$\mu(V_{ij}^{EST}) = \max[0; \gamma_L \times \min[0; (1 - z_{ij})]; \gamma_M \times \tag{2.12}$$

$$\times \min\left[0; \frac{z_{ij}-0.5}{0.5}; \frac{0.5-z_{ij}}{0.5}\right]; \gamma_H \times \min[0; z_{ij}]]$$

where $\gamma_k = 1$, if $k =$ VARling, 0 otherwise ($k =$ L, M, H).

4 *Manufacturing capacity constraint.* This constraint guarantees that the workloads on the DML and FMS are less than the respective system capacities and is expressed in equations (2.13) and (2.14).

$$V_{ij}^{DML} \leq \sum_{k=1}^{j} L_{ij}^{DML} \times Cp^{DML} \quad \forall i, j \tag{2.13}$$

$$WL_j^{FMS} \leq \sum_{k=1}^{j} L_j^{FMS} \times A_k \quad \forall i, j \tag{2.14}$$

5 *Product mix constraint.* The product mix can be defined by introducing the binary variable y_i which is equal to 0 if product i is internally produced, and 0 otherwise. Of course, condition (2.15) must be satisfied.

$$MIX = \sum_i y_i \tag{2.15}$$

As it has been seen in the previous sections, the output of the model [A1-1]-M2 is a linguistic term representing the increasing level of the product mix (increasing, constant, or decreasing). As for the *Outsourcing strategy constraint*, the constraint on the product mix is translated into the objective function (2.16).

$$\mu_{MIX} = max\left[0; min\left[\frac{\sum_i y_i - a}{\frac{b-a}{2}}; \frac{b - \sum_i y_i}{\frac{b-a}{2}}\right]\right] \tag{2.16}$$

The coefficients a and b can be calculated as in expressions (2.17) and (2.18), given the current mix MIX_{curr}, and m, n the percentage values of the minimum and maximum values of the linguistic terms coming out from the model [A1-1]-M2.

$$a = (1 + m) \times MIX_{curr} \tag{2.17}$$

$$b = (1 + n) \times MIX_{curr} \tag{2.18}$$

Notice that, in the formulation of this model, the budget constraint (C4 in Figure 2.1) and the current manufacturing structure input (I1 in Figure 2.1) have not been considered.

Model Objective Function

After having translated some constraints into single objective functions, the final multi-objective function becomes the one represented in expression (2.19).

$$max[g_1[\frac{1}{I}\sum_i y_i \times \frac{1}{J}\sum_j \mu_{ij}^{EST}]+ \tag{2.19}$$

$$+g_2[k_F \times \mu_{ROI(FMS)} + (1 - k_F) \times \mu_{ROI(DML)}] + g_3[\mu_{MIX}]]$$

Where:

- g_1, g_2 e g_3 are the weights of the three main factors of the objective function and can be chosen by the manager for a specific case during the optimization phase;

- k_F is a flexibility parameter. Its value is obtained from model [A1-1]-M1, by giving proper weights to the crisp values of the *Product Flexibility*, *Routing Flexibility*, and *Expansion Flexibility*.

3.2 DML, FMS and RMS model

Taking into account RMS increases the complexity of the above described investment decision problem. Indeed, it gives even more manufacturing solutions to consider and, also, increases both the uncertainty and risk levels. Let's consider the objective function (2.20) determined in the already presented model by including the reconfigurable manufacturing system option.

$$max[g_1[\frac{1}{I}\sum_i y_i \times \frac{1}{J}\sum_j \mu_{ij}^{EST}]+ \tag{2.20}$$

$$+g_2[k_F \times \mu_{ROI(FMS)} + k_R \times \mu_{ROI(RMS)}+$$

$$+(1-k_F-k_R) \times \mu_{ROI(DML)}] + g_3[\mu_{MIX}]]$$

The objective function presents a new factor $\mu_{ROI(RMS)}$ (once again, the ROI mapping function which can be achieved by producing in the RMS) which is weighted by a flexibility parameter k_R gained by the crisp value of the Reconfiguration flexibility linguistic output from model [A1-1]-M1. Concerning the relationships among RMS with DML, the same considerations that have been exposed for FMS/DML can be formulated. Indeed, although RMS and FMS are two different manufacturing solutions in terms of hardware and software architecture, from an operation management perspective, they present the same feature with respect to DML. Thus, the workload of an RMS that produces all product I in j can be calculated as in equation (2.21), while the investment cost for purchasing the RMS as in equation (2.22).

$$WL_j^{RMS} = \alpha^{RMS} \times \sum_i (FT_i \times V_{ij}^{RMS}) \tag{2.21}$$

$$C^{RMS} = \beta^{RMS} \times \sum_i C_i^{DML} \tag{2.22}$$

Moreover, the conditions (2.23) and (2.24), analogous to (2.3) and (2.4) are still valid.

$$BT_i \leq \alpha^{RMS} \times FT_i \leq FT_i \Leftrightarrow \frac{BT_i}{FT_1} \leq \alpha^{RMS} \leq 1 \tag{2.23}$$

$$\max_i C_i^{DML} \leq C^{RMS} \leq \sum_i C_i^{DML} \Leftrightarrow \frac{\max_i C_i^{DML}}{\sum_i C_i^{DML}} \leq \beta^{RMS} \leq 1 \tag{2.24}$$

When it comes to the relationships among FMS and RMS, some more considerations need to be pointed out. If we think at the RMS as a system made up of a base structure where a number of different modules (each necessary for processing a part type) can be added and removed, then the reconfiguration time (the time to remove a module and add a new one) can be thought of as the FMS set-up time, even though it is surely greater. Then, condition (2.25) holds.

$$\frac{BT_i}{FT_i} \leq \alpha^{FMS} \leq \alpha^{RMS} \leq 1 \tag{2.25}$$

Regarding the investment costs, it can be stated the cost for purchasing an RMS is minor then the one for purchasing an FMS. This because of the different design, structure, and technological levels associated to the two manufacturing systems solutions. For this reason, from relations (2.4) and (2.24), condition (2.26) derives.

$$\frac{\max_i C_i^{DML}}{\sum_i C_i^{DML}} \leq \beta^{RMS} \leq \beta^{FMS} \leq 1 \tag{2.26}$$

4. DSS description

The Decision Support System (DSS), including the fuzzy systems and the interface forms, has been developed into the Visual Basic platform, while the optimization algorithms of the model [A1-2]-M1 has been implemented into LINGO optimization software. The initial menu form of the DSS (Figure 2.7) presents to the user three buttons, each activating the possible use cases. “Insert Data”, from which the user can insert data about the specific problem; “Models for flexibility identification”, by which the user can enter into the models [A1-1]-M1, “Models for project mapping”, by which the user can activate the models [A1-1]-M2.

By clicking the “Models for project mapping” button, the user enter the form represented in Figure 2.8 and can determinate the impact of the input variables *business strategies*, *competitive scenario*, *current products*, *potential products*, and *outsourcing conditions* on the output variables *outsourcing constraints* and *product mix constraints*. The form and its sub-forms (e.g. Figure 2.9), indeed are connected to the fuzzy systems discussed in Section 2. As the reader can notice in Figure 2.8, in order to have the output “outsourcing constraints”, the user needs to select the product, insert a linguistic evaluation on the input parameters, and to run the fuzzy engine by clicking on the button “Results”. It is also possible to visualize the externalization coefficients which will be used as input of the models [A1-2]-M2, i.e. “Long term capacity planning”. When it comes to the strategic evaluation of the flexibility types, by clicking the button “Models for flexibility identification” of the main menu, the form depicted in Figure 2.10 will appear. This form allows selecting a flexibility type among product flexibility, routing flexibility, expansion flexibility, reconfiguration flexibility and to enter the respec-

Figure 2.7. The initial menu form.

tive fuzzy model. For example, Figure 2.11 reports the form for product flexibility. The user expresses the importance level of the input variables and runs the fuzzy engine in order to obtain the evaluation of the importance level of the considered flexibility type. At this point, directly from the form of Figure 2.10, the user can access the models [A1-2]-M1, i.e. "Long term capacity planning" as showed in Figure 2.12 At this point, the user is required to insert the investment costs for purchasing the DML_i and the coefficients α and β which are necessary to run the optimization model. Also, the user must specify the minimum and maximum expected values of ROI and the weights g_1, g_2, g_3 of the objective function. Then for each product and time bucket, processing times and contribution margins, the expected demand, and the strategic level x_i of the particular product i. As soon as all these input are inserted, the user can run the optimization algorithm, automatically performed by the LINGO solver, just clicking the button "Solve". The optimization results are reported in a new form represented in figure 2.13. As it can be observed, from such a form the following information comes out.

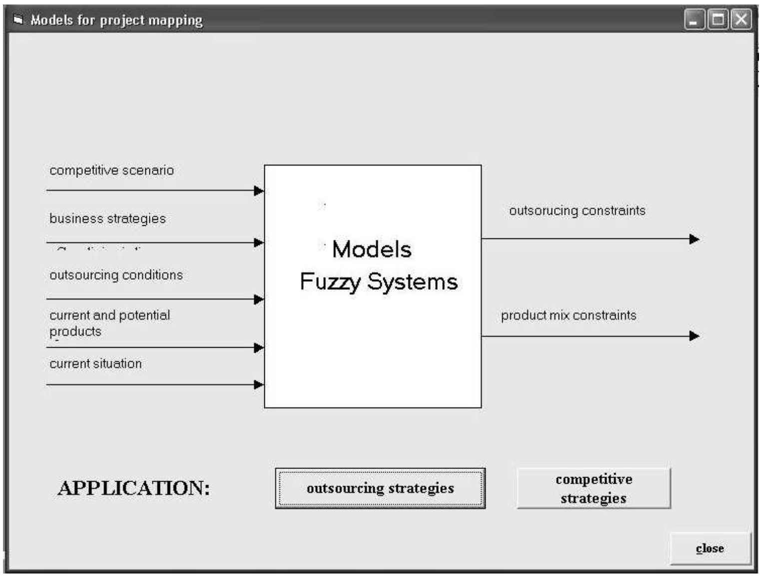


Figure 2.8. The form Models for project mapping.

- The number L_{ij}^{DML} of dedicated lines to purchase in the time bucket j for manufacturing product i;
- The number Lj^{FMS} of flexible manufacturing system to buy in the bucket j;
- The volumes V_{ij} of product i that have to be manufactured in the DML or FMS in the time bucket j;
- The total volume percentage to outsource for each product and time bucket.

4.1 DSS integration in the Strategic Planning DSS platform

One of the main requirements in the DSS design has been its easy integration with other software systems which support the entire process of strategic planning by implementing all the other planning models as described in the first chapter of this book. For this reason, the DSS, besides displaying the results in a Visual Basic form, generates an output

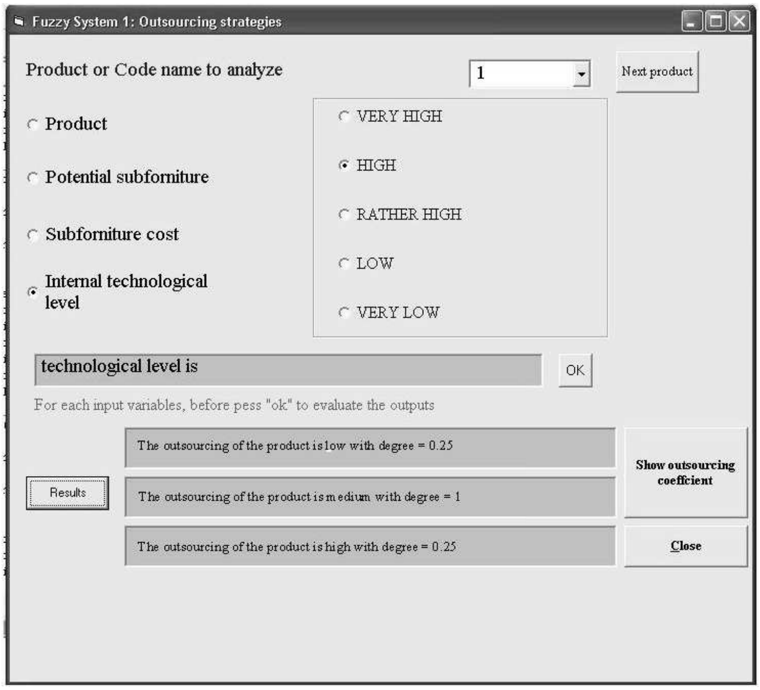


Figure 2.9. The form for outsourcing constraint determination.

file (a *txt* file) with the same data as in the optimization results form. Such a file can be automatically read by the other software systems that take such a data as input of their models. Figure 2.14 shows a sketch of this *txt* file.

5. Tests and results

In order to test the proposed methodologies, the developed DSS has been applied to the case study presented in Section **??** of Chapter 1. Specifically the DSS has been run in two different scenarios.

5.1 Scenario 1

Input

- Product 1

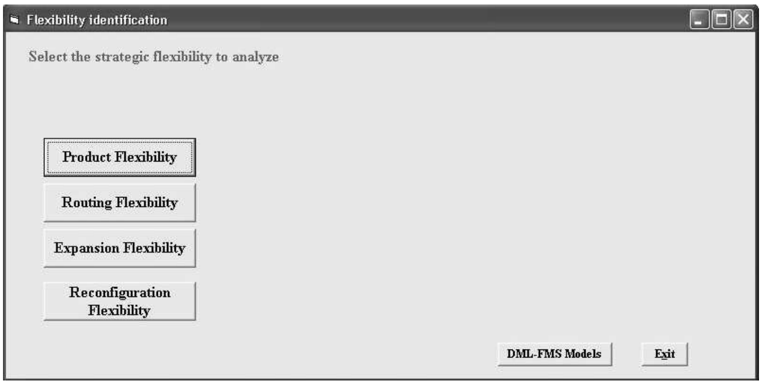


Figure 2.10. The form for flexibility identification.

- Strategic level: very strategic;
- Potential supplier: not very reliable;
- Outsourcing cost: not very suitable;
- Technological level: rather high;

■ Product 2

- Strategic level: strategic;
- Potential supplier: not very reliable;
- Outsourcing cost: not very suitable;
- Technological level: high;

■ Product 3

- Strategic level: strategic;
- Potential supplier: rather reliable;
- Outsourcing cost: not very suitable;
- Technological level: high;

■ Product 4

- Strategic level: not very strategic;
- Potential supplier: reliable;
- Outsourcing cost: suitable;

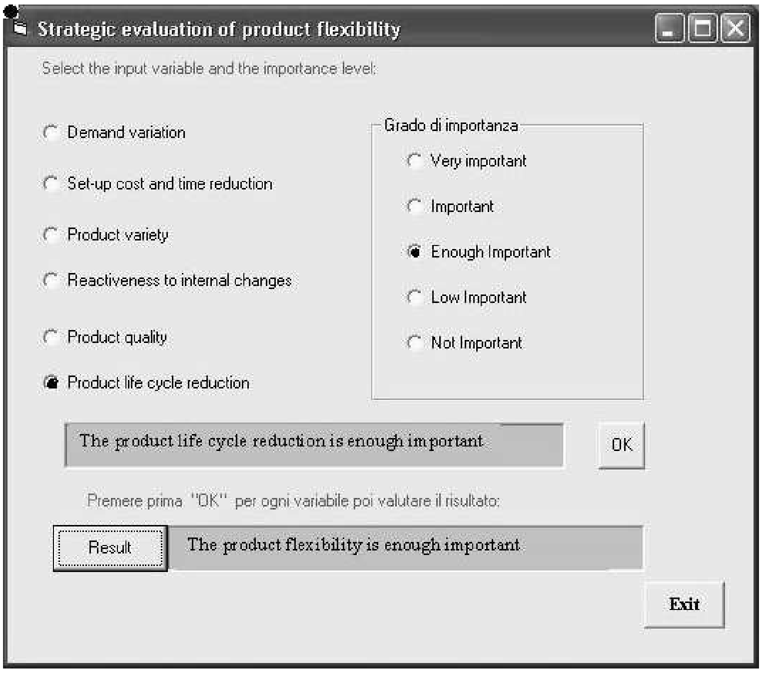


Figure 2.11. The form for the strategic evaluation of *product flexibility*.

  - Technological level: not very high;

- Product 5
  - Strategic level: very strategic;
  - Potential supplier: not very reliable;
  - Outsourcing cost: suitable;
  - Technological level: high;
- Current products/potential products: very high;
- Marketing strategy: market penetration;
- Business strategy: differentiation;
- Product flexibility
  - Reactiveness to internal and external changes: very important;

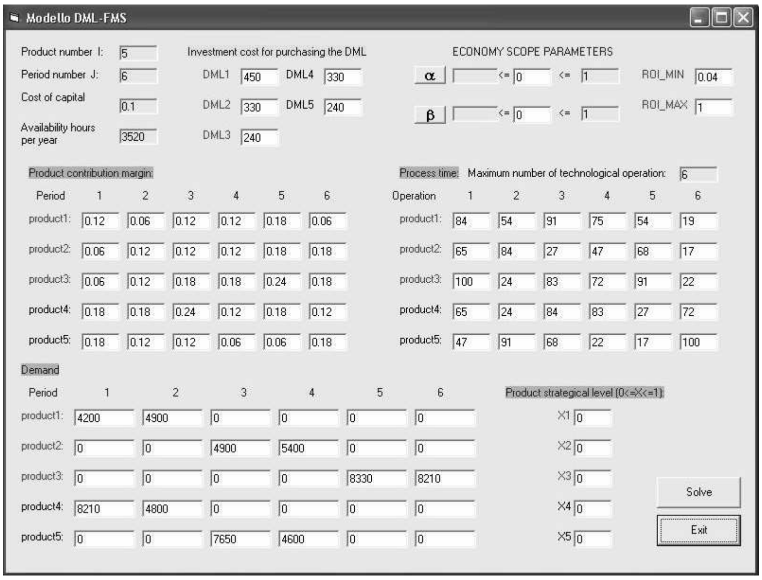


Figure 2.12. Form for the strategic evaluation of the manufacturing capacity.

  - Set-up cost and time reduction: not important;
  - Demand variation: rather important;
  - Product life cycle reduction: not important;

- Routing flexibility
  - Set-up cost and time reduction: not very important;
  - Product variety: not important;
  - Reactiveness to internal changes: rather important;
  - Resource utilization: important;
- Expansion flexibility
  - Reactiveness to market changes: very important;
  - Demand uncertainty: not important;
- Reconfiguration flexibility
  - Product change rapidity: important;
  - Reactiveness to internal changes: very important;

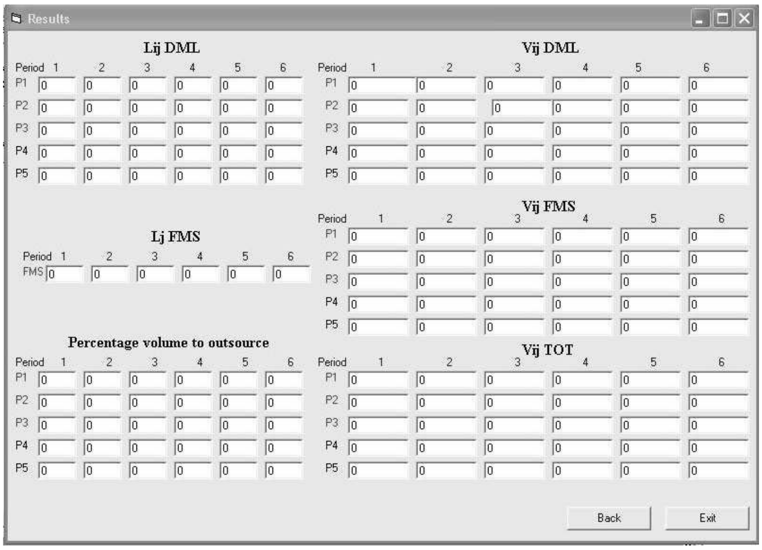


Figure 2.13. The form of the optimization results.

- Demand uncertainty: rather important;
- Reactiveness to market changes: not very important;

■ Economy of scope coefficients

- coefficient $\alpha = 0.4$
- coefficient $\beta = 0.35$

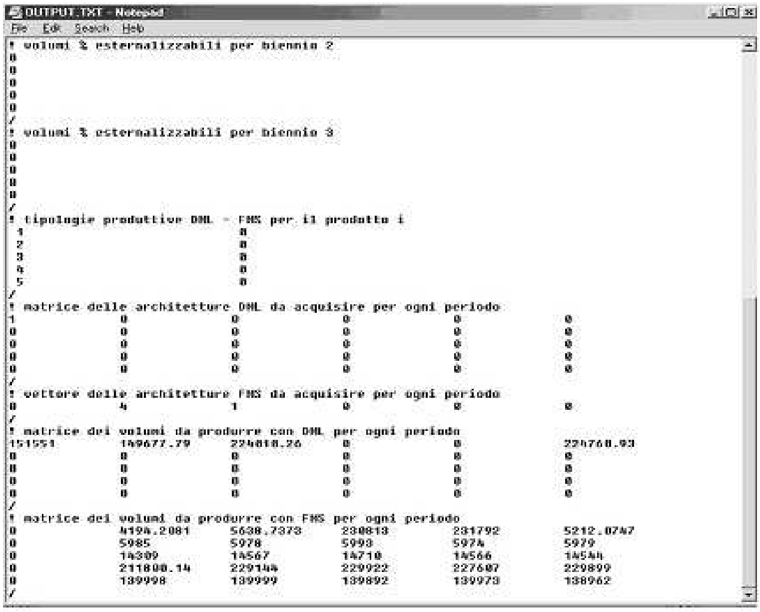
Output
The input data relative to the number of products to manufacture, product costs, contribution margins, outsourcing costs, DMLs costs and throughput, are reported in Figure 2.15. In the same figure the output results of the Scenario 1 are presented.

5.2 Scenario 2

Input

■ Product 1

- Strategic level: very strategic;



```
OUTPUT.TXT - Notepad
File  Edit  Search  Help
! volumi % esternalizzabili per biennio 2
0
0
0
0
0
/
! volumi % esternalizzabili per biennio 3
0
0
0
0
0
/
! tipologie produttive DML - FMS per il prodotto i
 1                          0
 2                          0
 3                          0
 4                          0
 5                          0
/
! matrice delle architetture DML da acquisire per ogni periodo
1             0             0             0             0             0
0             0             0             0             0             0
0             0             0             0             0             0
0             0             0             0             0             0
0             0             0             0             0             0
/
! vettore delle architetture FMS da acquisire per ogni periodo
0             4             1             0             0             0
/
! matrice dei volumi da produrre con DML per ogni periodo
151551        149677.79     224818.26     0             0             224768.93
0             0             0             0             0             0
0             0             0             0             0             0
0             0             0             0             0             0
0             0             0             0             0             0
/
! matrice dei volumi da produrre con FMS per ogni periodo
0             4194.2081     5638.7373     230813        231792        5212.0747
0             5985          5978          5993          5974          5979
0             14309         14567         14710         14566         14544
0             211808.14     229144        229922        227607        229899
0             139998        139999        139892        139973        138962
/
```

Figure 2.14. A sketch of the output *txt* file.

  - Potential supplier: not very reliable;
  - Outsourcing cost: not very suitable;
  - Technological level: rather high;

- Product 2
  - Strategic level: strategic;
  - Potential supplier: reliable;
  - Outsourcing cost: suitable;
  - Technological level: high;

- Product 3
  - Strategic level: not very strategic;
  - Potential supplier: rather reliable;
  - Outsourcing cost: suitable;
  - Technological level: not very high;

Product#	Product Code	Contribution Margin	Outsourcing cost
1	9623784580	2.8	5.6
2	9629560780	0.44	0.88
3	1461645080	2.9	5.8
4	507912	1.38	2.76
5	9620867780	0.44	2.76

	DML_i throughput	DML_i cost	FMS throughput	FMS cost
1	0.909	2200	0.871	3220
2	0.317	1800	0.257	
3	0.144	2800	0.103	
4	0.141	1400	0.072	
5	0.298	1000	0.110	

%Total Volumes to outsource

	Year 1 and 2	Year 3 and 4	Year 5 and 6
1	0	0	0
2	0	0	0
3	0	0	0
4	70.6%	83.6%	83.4%
5	0	0	0

Number of DML_i to purchase

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	1	0	0	0	0	0

Number of FMS to purchase

Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
3	1	0	0	0	0

Volumes to be manufactured using DMLi

1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	121791	121791	121791	121791	121791	121791

Volumes to be manufactured using FMS

1	151551	123872	229649	230813	231792	229973
2	5990	5985	5978	5993	5974	5979
3	13576	14309	14567	14710	14266	12544
4	53007	81852	0	75238	75210	0
5	18185	18207	182	18101	18182	17171

Total Volumes

1	151551	153872	230813	230813	231792	229973
2	5990	5985	5993	5993	5974	5979
3	13576	14309	14710	14710	14566	14544
4	228781	229967	229144	229922	227607	229899
5	139976	139999	139999	139892	139973	138962

DMLi utilization

Year 1 and 2	Year 3 and 4	Year 5 and 6
0.895	0.895	0.895

FMS utilization

Year 1 and 2	Year 3 and 4	Year 5 and 6
0.875	0.681	0.678

Demand fulfillment %

	Year 1 and 2	Year 3 and 4	Year 5 and 6
1	100	100	100
2	100	100	100
3	100	100	100
4	100	100	100
5	100	100	100

Figure 2.15. Scenario 1 Output.

- Product 4
  - Strategic level: not very strategic;
  - Potential supplier: not very reliable;
  - Outsourcing cost: not suitable;
  - Technological level: not very high;
- Product 5
  - Strategic level: not strategic;
  - Potential supplier: very reliable;
  - Outsourcing cost: suitable;
  - Technological level: not high;
- Current products/potential products: very high;
- Marketing strategy: market penetration;
- Business strategy: differentiation;
- Product flexibility
  - Reactiveness to internal and external changes: very important;
  - Set-up cost and time reduction: not very important;
  - Demand variation: rather important;
  - Product life cycle reduction: not important;
- Routing flexibility
  - Set-up cost and time reduction: not very important;
  - Product variety: not important;
  - Reactiveness to internal changes: rather important;
  - Resource utilization: important;
- Expansion flexibility
  - Reactiveness to market changes: very important;
  - Demand uncertainty: not important;
- Reconfiguration flexibility
  - Product change rapidity: important;
  - Reactiveness to internal changes: very important;

  - Demand uncertainty: rather important;
  - Reactiveness to market changes: not very important;

- Economy of scope coefficients
  - coefficient $\alpha = 0.6$
  - coefficient $\beta = 0.5$

Output
The input data relative to the number of products to manufacture, product costs, contribution margins, outsourcing costs, DMLs costs and throughput, are reported in Figure 2.16. In the same figure the output results of the Scenario 2 are presented.

5.3 Test results

The main comments related to the test phase of the developed DSS on the case study data can be summarized as follows for the two supposed scenarios:

- Scenario 1: the DSS suggests outsourcing a great percentage of the production of product 4. This is pretty reasonable considered that such a product is not very strategic, its potential supplier is reliable and its outsourcing cost is suitable. None of the other products has these outsourcing fitting conditions and indeed the system suggests not outsourcing any of them. Also, the only DML that needs to be bought is DML5. This result was also expectable, given that DML5 has the minimum investment cost such as DML3. But, DML3 is not suggested for product 3 which is on the contrary produced by FMS. This can be explained by looking at the product demand configuration. Indeed, product 3 is required on periods 5 and 6, i.e. when the already purchased FMS are not busy for producing all of the other products. So, product 3 will be produced on the FMS which is, during those time buckets, idle. On the contrary, for the production of product 5, which is required on periods 3 and 4, the FMS already available cannot be used because busy due to the other products.

- Scenario 2: the DSS suggests outsourcing almost the total production of product 4 and 5, and a great part of the production of product 2. A quick view at the strategic and outsourcing conditions of these products makes clearer such a result. Of course the total number of FMS and DML purchased is minor of that in scenario 1 due to the more substantial outsourcing volume. Also, product 1 will be produced using DML and this probably because of product 1 high volume respect to product 2 and product 3 (this will be outsourced for more than 50 percent).

6. Conclusions

This chapter presents an innovative approach for assisting entrepreneurs in making long term capacity decisions in Advanced Manufacturing Systems (AMS). Indeed, concepts as flexibility and reconfigurability have introduced important strategic and risk issues in making long term capacity decisions when dealing with AMSs. Indeed, AMS allows reactions to internal and external changes making the company more reactive and this is a basic strategic issue in nowadays global competition. On the other hand, such AMS requires high investment in manufacturing equipments, human resources and technology knowledge. Such high investments together with the wideness and the variability of the competition scenario contribute to increase the perception of the risk for industrial entrepreneurs. This is especially true in SME, where the risk perception reduces the propensity to invest in AMS, and this contributes to increase the technological and competition gap in some manufacturing SMEs.

This problem could be approached by providing the entrepreneur with a Decision Support System able to assist her/him in making long term capacity decision in AMS. As often suggested in the scientific and industrial literature, such a DSS should be able to address both strategic and economical-financial issues of AMS such as flexible manufacturing systems (FMS) and Reconfigurable Manufacturing Systems (RMS).
The DSS proposed in this chapter goes toward this direction. It allows the entrepreneur to plan its production strategy starting from company business strategy, market strategy, competition scenario and outsourcing scenario. Starting from such information, a Fuzzy Expert Systems allows to define the kind of strategic flexibility the company needs and how the company should compose its production mix between internal production and outsourced one.

This strategic information represents the input of a Long Term Capacity Planning Model that constitutes the economic and financial hearth of the DSS. The Decision Model allows entrepreneur to make investment decisions in long term fashion by mixing up different manufacturing systems such as Dedicated Machining Lines (DML), Flexible Manufacturing Systems (FMS) and Reconfigurable Manufacturing Systems (RMS). In order to do that the Decision Support System employs an innovative parametric approach in comparing different manufacturing systems. The suggested approach consists in modeling long term capacity characteristics of FMS and RMS starting from the DML ones and accounting for the differences through proper parameters which take into account for scope and reconfiguration economies.

A prototype of the DSS has been built by using Microsoft Access and Microsoft Visual Basic. The Strategic Planning Model has been implemented by using a Visual Basic engine incorporating Fuzzy Rules. The Long Term Capacity Planning Model consists of a Visual Basic applications that, starting from the results of the Strategic Planning Model, builds up a multi objective economic-financial optimization model for making capacity decision regarding three kind of manufacturing system type, i.e. DML, FMS and RMS. The optimization model is solved by using the Lingo Solver. The DSS results suggest the entrepreneur how many DML, FMS and RMS to buy for each planning period, the optimal volumes to be produced and to be outsourced.

The DSS has been tested under two different scenarios. The results confirm, in both the cases, that the DSS provides important strategic information for making long term production capacity decisions in manufacturing enterprises. Authors believe that SME can obtain great advantages in term of decision making consistency by using a commercial evolution of the DSS here presented.

References

Albayrakoglu M. (1996). Justification of new manufacturing technology: a strategic approach using analytical hierarchy process. *Production and Inventory Management Journal*, 37: 71-77.

Babic, B. (1999).Axiomatic design of flexible manufacturing systems. *International Journal of Production Research*, 37: 1159-1173.

Banerjee S. K. (2000). Developing manufacturing management strategies: Influence of technology and other issues.*International Journal of Production Economics*, 64: 79-90.

Bateman N., Stockton D. J., and Lawrences P. (1999). Measuring the mix response flexibility of manufacturing systems. *International Journal of Production Research*, 37: 871-880.

Berry W. L., Hill T., and Klompmaker J. E. (1999). Aligning marketing and manufacturing strategies with the market. *International Journal of Production Research*, 37: 3599-3618.

Chan, F. T.S., Jiang B., and Tang, N. K.H. (2000). The development of intelligent decision support tools to aid the design of flexible manufacturing systems. *International Journal of Production Economics*, 65:73-84.

Chung K.J, and Tsou C.S. (1998). Analysis and algorithm for optimal investment times of new manufacturing technologies in a duopoly. *European Journal of Operational Research*, 109: 632-645.

Dangl T. (1999). Investment and capacity under uncertain demand. *European Journal of Operational Research*, 117: 415-428.

De Meyer, A., Nakane, J., and Ferdows, K. (1989). Flexibility: the next competitive battle - the manufacturing futures survey. *Strategic Management Journal*, 10: 135-144.

Elango, B, and Meinhart, W. A. (1994). Selecting a flexible manufacturing system - a strategic approach". *Long Range Planning*, 27: 118-126.

Feurstein M., and Natter M. (2000). Fast high precision decision rules for valuing manufacturing flexibility. *European Journal of Operational Research*, 120: 108-117.

Fine, C. H., and Freund R. M. (1990). Optimal Investment in Product-Flexible Manufacturing Capacity. *Management Science*, 36(4): 449-464.

Frazelle, E. H. (1986). Flexibility: a strategic response in changing times. *Industrial Engineering*, March: 17-20.

Harrison, J. M., and Mieghem, J. A. (1999). Multi-Resource investment strategies: Operational hedging under demand uncertainty. *European Journal of Operational Research*, 113: 17-29.

Lim S. K., and Kim, Y. D. (1998). Capacity planning for phased implementation of flexible manufacturing systems under budget restrictions. *European Journal of Operational Research*, 104: 175-186.

Mohanty, R. P. (1993). Analysis of justification problems in CIMS: review and projection. *IEE Transactions*, 19: 222-232.

Naik B., and Chakravarty, A. K. (1992). Strategic acquisition of new manufacturing technology: a review and research framework. *International Journal of Production Research*, 30(7): 1575-1601.

Parker, R. P., and Wirth, A. (1999). Manufacturing flexibility: Measures and relationships". *European Journal of Operational Research*, 118: 429-449.

Parsei, H. R., and Wilhelm, M. R. (1989). A justification methodology for automated manufacturing technologies. *Computers and Industrials Engineering*, 16: 363-373.

Perrone, G., and Noto La Diega, S. (1999-a). A Reference Scenario for IMS Strategic Design. *Annals of CIRP*, 48, 381-384.

Perrone, G., and Noto La Diega S. (1999-b). A Simultaneous Approach for IMS Design: a Possibility Based Approach. *Modeling Manufacturing Systems*, Springer Verlag.

Price, D.H.R., Beach, R., Muhlemann, A.P., Sharp, J.A., and Paterson, A. (1998). A system to support the enhancement of strategic flexibility in manufacturing enterprises. *European Journal of Operational Research*, 109: 362-376.

Sarkis J., and Talluri S. (1999). A decision model for evaluation of flexible manufacturing systems in the presence of both cardinal and ordinal factors. *International Journal of Production Research*, 37: 2927-2938.

Sheng, J., and Sueyoshi, T. (1995). A unified framework for the selection of flexible manufacturing systems. *European Journal of Operational Research*, 85: 297-315.

Shewchuk J. P. (1999). A set of flexibility measures for manufacturing applications. *International Journal of Production Research*, 37: 3017-3042.

Tombak, M. M. (1990). A strategic analysis of flexible manufacturing systems. *European Journal of Operational Research*, 47: 225-238.

Wu B., and Ellis R. (2000). Manufacturing strategy analysis and manufacturing system design: Process and Application. *International Journal of Production Economics*, 65: 55-77.

Product#	Product Code	Contribution Margin	Outsourcing cost
1	9623784580	2.8	5.6
2	9629560780	0.44	0.88
3	1461645080	2.9	5.8
4	507912	1.38	2.76
5	9620867780	0.44	2.76

	DML_i throughput	DML_i cost	FMS throughput	FMS cost
1	0.909	2200	0.58072	4600
2	0.317	1800	0.17182	
3	0.144	2800	0.06896	
4	0.141	1400	0.04831	
5	0.298	1000	0.07391	

% Total Volumes to outsource

	Year 1 and 2	Year 3 and 4	Year 5 and 6
1	0	0	0
2	50	50	84.9
3	0	0	0
4	96.9%	100	100
5	100	95.7787	100

Number of DML_i to purchase

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
1	1	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Number of FMS to purchase

Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
1	0	0	0	0	0

Volumes to be manufactured using DMLi

1	99766	153872	115392	230813	123612	115523
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Volumes to be manufactured using FMS

1	51784	0	114256	0	108179	114450
2	0	5985	0	5993	1800	0
3	13576	14309	14567	14710	14566	14544
4	5891	8002	0	0	0	0
5	0	0	0	11810	0	0

Total Volumes

1	151551	153872	230813	230813	231792	229973
2	5990	5985	5993	5993	5974	5979
3	13576	14309	14710	14710	14566	14544
4	228781	229967	229144	229922	227607	229899
5	139976	139998	139999	139892	139973	138962

DMLi utilization

Year 1 and 2	Year 3 and 4	Year 5 and 6
0.310	0.424	0.293

FMS utilization

Year 1 and 2	Year 3 and 4	Year 5 and 6
1	1	1

Demand fulfillment %

	Year 1 and 2	Year 3 and 4	Year 5 and 6
1	100	100	100
2	100	100	100
3	100	100	100
4	100	100	100
5	100	100	100

Figure 2.16. Scenario 2 Output.

Chapter 3

STOCHASTIC PROGRAMMING MODELS FOR MANUFACTURING APPLICATIONS

A tutorial introduction

A. Alfieri, P. Brandimarte

Dipartimento di Sistemi di Produzione ed Economia dell'Azienda (DSPEA)

Politecnico di Torino

Corso Duca degli Abruzzi 24, 10129 Torino, Italy

arianna.alfieri@polito.it, paolo.brandimarte@polito.it

Abstract Stochastic programming models have been proposed for capacity planning problems in different environments, including energy, telecommunication networks, distribution networks, and manufacturing systems. In this chapter we give an introductory tutorial to stochastic linear programming models, with emphasis on modeling techniques, rather than specialized solution methods. We consider two-stage and multi-stage stochastic programming models with recourse for manufacturing related applications, such as production planning and capacity planning with uncertainty on demand. We stress the importance of proper model formulation from two points of view: the first one is building strong mixed-integer formulations; the second one is generating scenario trees in order to suitably represent uncertainty while keeping them to a manageable size. We also compare the stochastic programming approach to traditional dynamic programming and to robust optimization.

Keywords: stochastic programming; production planning; capacity planning.

1. Introduction

Quite often one has to take decisions with incomplete information about problem data. The longer the planning horizon, the larger the uncertainty; hence, capacity planning models are a natural candidate for optimization methods allowing for an explicit representation of uncertainty. Uncertainty may take different forms: in the relatively lucky case, one has sufficient information to assume some probability distribution; in other settings, one has to deal with a brand new situation, whereby very little is known and the best one can do is coming up with a set of plausible scenarios, possibly obtained by interviewing a panel of domain experts. An example of the first situation could be a capacity expansion problem in the case of production of electrical energy or in the case of telecommunication networks; in both cases, availability of historical data is a valuable help. An typical example of the second situation is represented by brand new fashion products. In this case, uncertainty is not only linked to the realization of a random variable with a known distribution, since we may have no idea what such a distribution looks like or, to the very least, we have considerable uncertainty about its parameters. The two situations may be somehow reconciled by the concept of scenarios. Scenarios may result both from sampling a probability distribution or by assembling judgmental forecasts. In fact, in (Fisher and Raman, 1996) it is emphasized that differences in expert opinions should be leveraged in order to point out inherent uncertainty, rather than averaged out through some group forecasting method in order to come up with a point estimate. Different opinions may be used to build plausible scenarios, and methods able to yield robust decisions for a set of alternative scenarios may be applied to both types of situation.

The real difference is likely to be in terms of attitude towards risk. In repeated experiments, which is more typical of a probabilistic representation of uncertainty, it is reasonable to optimize the expected value of some performance measure, as this is related to optimal average performance in the long run. However, considering only the expected value implies a risk neutral attitude. In principle, it is possible to model different risk attitudes by using some utility function, as typical in financial decision making. From a conceptual point of view this is not too different from optimizing an expected value, and the main complicating issue is of computational nature, as dealing with a risk-averse utility function brings us into the realm of nonlinear programming. In practice, eliciting a utility function from a decision maker is quite difficult, and alternative ways have been devised to measure risk; risk measures (such as Value at Risk, or Conditional Value at Risk) are quite common in financial applications. In the manufacturing domain, a counterpart of such risk measures may be found in service levels commonly adopted in statistical inventory theory. However, there are settings in which we might prefer a radically different way of representing robustness requirements. This can be addressed by robust optimization models. Again, robust optimization models can be related to stochastic programming by the use of scenarios. Hence, due to the introductory nature of this contribution, we will concentrate our attention essentially on stochastic linear programming models, having in mind that this might be only the first step towards a satisfactory decision support tool.

One common approach to deal with sets of scenarios is analyzing how the optimal solution would change with respect to different scenarios; scenario analysis can be carried out by solving a set of optimization models with alternative sets of data. However, it is not at all clear how to blend all of the scenario dependent solutions into one robust solution. Stochastic programming is one method to carry out this task. Stochastic programming models have been proposed for capacity planning in the past. A well-known example in the manufacturing domain is (Eppen et al., 1989), where capacity planning issues for General Motors plants were considered. In fact, stochastic programming models are far from recent; they date back to the mid-fifties with the first contributions by George Dantzig (Dantzig, 1955). However, due to severe computational difficulties, only recently they have been proposed as a really practical tool.

Since dealing with uncertainty is so difficult, it is tempting to ignore it and solve the problem by assuming average values for the data, maybe adding some slack to the optimal solution, such as safety stock or capacity buffers. Indeed, we should not take for granted that this simple approach will work much worse than a sophisticated model in practice. So, in order to gain some feeling for the effect of uncertainty, and to pave the way for complex models, in Section 2 we consider a simple and well-known model, the newsvendor model.

Then we introduce alternative stochastic programming approaches in Section 3, together with a numerical toy example related to assembly-to-order environments. Two-stage linear programming models with recourse are fully formalized in Section 4, and numerical solution methods to exploit their peculiar structure are briefly outlined in Section 5. Multi-stage models generalize two-stage formulations and are the subject of Section 6. Such models are definitely hard to solve; so we need proper model formulations, which are dealt with in Section 7, and clever scenario generation approaches, described in Section 8. Some models for capacity planning are illustrated in Section 9. Finally, we outline alternative approaches based on robust optimization in Section 10 and we draw conclusions in Section 11.

2. The newsvendor problem

A simple illustrative example of production/purchasing planning under demand uncertainty is the newsvendor problem, the prototype model for manufacturing problems in which we have to meet demand within a single time window. This is a single-period problem, typical of fashion or perishable items; however, it can be generalized to some multiple periods problems (Nahmias, 2000); it is also the conceptual basis of the solution procedure for the real-life case described in (Fisher and Raman, 1996). It may also be considered as a prototype capacity planning problem, if we assume that we purchase capacity to meet some aggregate demand.

We have to decide how many items to order before knowing actual demand; it is assumed that its probability distribution is known. Each item costs c and is sold at a price $p > c$; after the sale time window, unsold items are either scrapped or sold at a discounted (markdown) price $c_d < c$. A naive approach would be assessing a point forecast for the demand and ordering that quantity, but this amounts to neglecting demand variability completely. Should we really order the expected demand? A naive answer is that, due to uncertainty, we could add some safety stock to average demand. But can we really rule out the possibility that the right quantity is smaller than average demand? To solve the problem we must model the effect of demand uncertainty explicitly.

Consider the following toy data.

- demand is discrete and uniformly distributed between 5 and 15; each value has probability 1/11 and the expected value is 10;
- each item costs $c = 20$ and is sold at price $p = 25$, with a profit margin $p - c = 5$;
- unsold items are scrapped and there is no salvage value.

Let Q be the order quantity and D the random demand. Net profit (revenue minus cost), is a function $P(Q; D)$ of the controllable parameter Q and of the random variable D. Conditional on the realized demand d, net profit is:

$$P(Q; d) = \begin{cases} (p-c)Q & \text{if } Q \le d \\ (p-c)d - c(Q-d) = pd - cQ & \text{if } Q > d \end{cases}$$

Hence, we may express expected profit as a function of Q:

$$\mathrm{E}[P(Q; D)] = \frac{1}{11}\left[\sum_{d=5}^{Q}(pd - cQ) + \sum_{d=Q+1}^{15}(p-c)Q\right]$$

It is important to note that by choosing order quantity Q we *do not* get a certain profit; rather, we select a *probability distribution* for profit. Then we must select the preferred probability distribution based on a

Table 3.1. Expected newsvendor's profit as a function of order quantity Q.

Q	5	6	7	8	9	10
E[P(Q;D)]	25.00	27.73	28.18	26.36	22.27	15.91

Q	11	12	13	14	15
E[P(Q;D)]	7.27	-3.64	-16.82	-32.27	-50.00

way of translating that distribution to a number. In our simple case this is expected value, but it could be expected utility or even some sophisticated measure of risk.

To search for the optimal solution, a brute force approach is tabulating the values of the expected profit for the reasonable values of Q, as shown in Table 3.1. We see that when we order only five units, the profit is actually deterministic, since we will certainly sell all of the items with a total profit of 25. The expected profit if we order $Q = \mathrm{E}[D] = 10$ is 15.91, whereas the optimal solution corresponds to $Q^* = 7$, with an expected profit 28.18. Hence, by ignoring uncertainty we have a loss, in terms of expected profit, given by $28.18 - 15.91 = 9.27$. Note that the optimal quantity is less than average demand. If we order a large amount of items, we will incur an expected *loss*. This happens because we scrap unsold items and the profit margin is not too large. If profit margins were higher and/or we could sell residual items at a markdown price the optimal solution would be different. Another important factor that would have an impact is the type of distribution; in this toy example we have a symmetric distribution; often a normal distribution is assumed for large volume items, but skewness would certainly affect the results as well.

Luckily, in a realistic newsvendor problem, there is no need to tabulate a large amount of values, since this simple problem can be solved analytically. Let us define a shortage cost $c_s = p - c$ and an overage cost $c_o = c - c_d$. Assume a continuous distribution of demand, with density $f_D(x)$ and distribution function $G_D(x) \equiv \mathbb{P}\{D \leq x\}$. Then, expected cost can be expressed as:

$$\mathrm{E}[C(Q;D)] = c_s \int_0^Q (Q-x) f_D(x)\, dx + c_o \int_Q^{+\infty} (x-Q) f_D(x)\, dx. \quad (3.1)$$

Writing down the first-order optimality condition (see, e.g., Hopp and Spearman, 2000), we get an equation for the optimal order quantity Q^*:

$$G(Q^*) = \frac{c_s}{c_o + c_s}. \tag{3.2}$$

So, we should order a quantity Q^* such that the probability of having a demand less than Q^* is equal to the critical ratio $c_s/(c_o + c_s)$. Note that, due to the monotonicity of the cumulative distribution function, the optimal order quantity tends to increase if c_s is large and to decrease is c_o is large. This is basically a nonlinear equation that can be solved numerically. It is easy to see that, for a symmetric distribution like the uniform or the normal one, we will order a quantity smaller than expected demand whenever the critical ratio in equation (3.2) is less than 0.5, and a larger quantity in the complementary case. In the case of discrete demand, due to the provable convexity of expected cost (3.1), we may solve the problem by finding a fractional solution and then rounding it up or down.

This is a very simple example: it is single-item, uncapacitated, and single-period. Of course, to deal with more complex situations, we need a more powerful modeling framework. A less obvious but important point is that here we are somewhat *passive* with respect to uncertainty. We are looking for a reliable solution, i.e., a solution which may not be optimal for a specific demand scenario, but for all of them on the average. Still, after discovering the true value of demand, we just fill customer orders the best we can. There is no real decision after resolving uncertainty. However, in many practical cases we may have clever ways to react to demand variability, such as subcontracting, negotiating discounts to customers willing to accept delayed deliveries, assembling to order, etc. This leads us to a dynamic decision process which can be modeled by stochastic programming with recourse.

3. Stochastic linear programming

The newsvendor problem we have just considered is a prototypical stylized example. Linear programming (LP) models are a powerful and flexible modeling framework to cope with complex problems. Due to the astonishing progress not only in computing hardware, but also in solvers speed and reliability, LP models are now a practical decision making tools also in quite complex settings. Algorithmic improvements both in the classical simplex algorithm and in interior point solvers have been matched by software engineering progress, as solvers can now be embedded in an object-oriented architecture allowing for database access and powerful user interaction.

However, a persisting difficulty is represented by the quality of input data, most notably when uncertainty cannot be avoided. The typical textbook approach to cope with data uncertainty is sensitivity analysis. However, even though sensitivity analysis is certainly a useful tool, it may be of little help when a solution which is optimal under one scenario is not only sub-optimal, but even unfeasible in another one. Indeed, its practical usefulness has been questioned, e.g., in (Wallace, 2000). An alternative is modeling uncertainty directly within the LP framework.

Consider the following deterministic LP model (in canonical form):

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

We may try to deal with uncertainty by making randomness in the data explicit. In the most general case, we may have randomness in all of our data, which could be represented by random variables $\mathbf{c}(\omega)$, $\mathbf{A}(\omega)$, and $\mathbf{b}(\omega)$, depending on an underlying event ω. However, we cannot simply translate the model above to something like:

$$\begin{aligned} \min \quad & \mathbf{c}(\omega)^T\mathbf{x} && (3.3) \\ \text{s.t.} \quad & \mathbf{A}(\omega)\mathbf{x} \geq \mathbf{b}(\omega) && (3.4) \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

To begin with, the objective function (3.3) does not make sense, since minimizing a random variable has no clear meaning. Still, we could solve this issue simply by considering its expected value. The real issue is that we should not require that the constraints (3.4) are satisfied for every event ω. In some cases, doing so would yield a so-called "fat" solution, which is expected to be quite costly. In other cases, it would be simply impossible to do so. To see this, consider a simple inventory control system operating under a reorder point policy: if demand is assumed normal, 100% service level would imply setting the reorder point to infinity. By the same token, in the framework above it is obviously impossible to require satisfaction of a set of inconsistent equality constraints, one per scenario, which must be somehow relaxed.

One possible approach to relax our requirements is to settle for a probabilistic constraint, stating that there must be a high probability to satisfy constraints. The *chance-constrained* approach deals with models of the form:

$$\min \quad \mathbf{c}^T\mathbf{x}$$

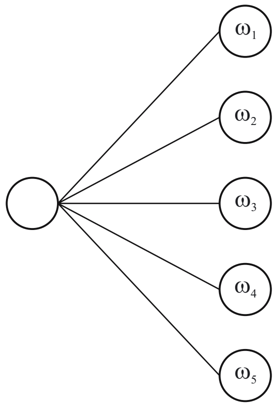


Figure 3.1. Scenario tree for a two-stage problem.

$$\begin{aligned} \text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbb{P}\{\mathbf{G}(\omega)\mathbf{x} \geq \mathbf{h}(\omega)\} \geq \alpha \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Note that we require a high probability of satisfying the *joint* set of constraints; alternatively, one may require this for each constraint separately. This modeling framework has a clear interpretation in terms of reliability of the solution, and it sounds certainly familiar to material managers used to enforce constraints on service levels. Nevertheless, there are a few difficulties:

- from a technical point of view, there is no guarantee that the resulting optimization problem is convex in general; non-convexity may arise with discrete probability distributions (the reason is that the union of convex sets is non convex in general);
- from a more practical point of view, we do not say anything about what will happen if constraints are violated; hence, corrective actions are left outside the model;
- finally, we do not account for a dynamic decision process whereby decisions are adapted when uncertainty is progressively resolved.

This is why we do not consider chance-constrained models in the following, referring the interested reader to (Prékopa, 2003). Another way to deal with stochastic optimization is stochastic programming *with recourse*. To get a feeling for this modeling framework, consider Figure 3.1.

- We have a scenario tree. The left node represents the current state, here and now; the right nodes represent different future states of nature, or scenarios. Each scenario has some probability of occurrence, which can be an objective measure derived by statistical information or a subjective measure of likelihood. As we have noted before, scenarios can be the result of a discretization of a continuous probability distribution or a set of plausible forecasts by a pool of experts.

- We should take a set of decisions now, but in the future, when uncertainty is at least partially resolved, we might take some action in order to "adjust" our previous decisions given additional information. These adjustments are called *recourse* actions.

- We want to find a set of decisions *now* in order to optimize immediate costs, which are certain, and the expected cost of the future, uncertain, recourse actions.

In order to get acquainted with this modeling framework, we consider next a simple production planning problem, which can be considered as a generalization of the simple newsvendor problem. For an introductory survey on modeling by stochastic programming, including further approaches we have not mentioned, please refer to (Sen, 1999).

Table 3.2. Bill of material for the ATO example.

	c_1	c_2	c_3	c_4	c_5
A_1	1	1	1	0	0
A_2	1	1	0	1	0
A_3	1	1	0	0	1

Table 3.3. Process plan, available capacity, and component cost for the ATO example.

	M_1	M_2	M_3	cost
c_1	1	2	1	20
c_2	1	2	2	30
c_3	2	2	0	10
c_4	1	2	0	10
c_5	3	2	0	10
cap.	800	700	600	

3.1 An example of two-stage stochastic programming: production planning in an assemble to order environment

In order to illustrate the logic underlying two-stage programming with recourse, we consider a toy production planning model for an assemble

Table 3.4. Demand scenarios, average demand, sale price.

	S_1	S_2	S_3	ave.	price
A_1	100	50	120	90	80
A_2	50	25	60	45	70
A_3	100	110	60	90	90

to order (ATO) problem. We have three end items (A_1, A_2, A_3), which are assembled out of five components (c_1, c_2, c_3, c_4, c_5). The bill of materials is flat and illustrated in Table 3.2. Note that we have two common components, c_1 and c_2, whereas the remaining three components are specific of each end item. We assume assembly is fast and hence it is not a bottleneck, which is reasonable for ATO. Components are manufactured using three machining centers (M_1, M_2, M_3). In Table 3.3 we list the processing times of each component on each machine; the last row reports the available capacity for each machine; the last column reports the cost of each component (such costs include material cost).

For what concerns end items, we need to have some demand forecast. Rather than a single (point) forecast, we consider a set of three possible scenarios (S_1, S_2, S_3) depicted in Table 3.4. The three scenarios have the same probability (1/3). We also list the sale price in the last column; the sale price is larger than the total component cost, which is 60 for each end item. Note that from a purely commercial point of view A_3 is in some sense the most attractive end item, as it has the largest profit margin (90-60=30), whereas A_2 is the worst one; such a simplistic consideration disregards capacity requirements linked to component manufacturing. In our simplified problem we do not consider interactions among sales of different end items (in practice, items with little or no profit margin may be useful anyway, in order to support sales of more profitable items). This problem is a generalization of the newsvendor problem, since there is a single sale window and unused components are scrapped (this is a somewhat simplified view of what might happen in fashion products).

Since dealing with uncertainty is complex, it is tempting to consider a simplified model in which uncertainty is ignored and we assume average demand will be realized; expected demand is listed in the second-to-last column of Table 3.4. Assuming continuous decision variables, which is actually a matter of scale since rounding effects are negligible for high volumes, we may build the following LP model:

$$\max \quad -\sum_{i=1}^{5} C_i x_i + \sum_{j=1}^{3} P_j y_j \tag{3.5}$$

$$\text{s.t.} \quad \sum_{i=1}^{5} T_{im} x_i \le L_m \qquad m = 1, 2, 3 \tag{3.6}$$

$$y_j \le \bar{d}_j \qquad j = 1, 2, 3 \tag{3.7}$$

$$\sum_{j=1}^{3} G_{ij} y_j \le x_i \qquad i = 1, 2, 3, 4, 5 \tag{3.8}$$

$$y_j, x_i \ge 0$$

In this model, subscript i refers to components, j to end items, and m to machining centers. The model data, corresponding to the information listed in the previous tables, are:

- the component cost C_i;
- the end item sale price P_j;
- machine availability L_m;
- processing time T_{im}, for item i on machine m;
- the gozinto factor G_{ij} given in the bill of materials;
- the average demand $\bar{d}_j$ which is assumed certain.

The decision variables are x_i, the amount of component i we produce, and y_j, the amount of end item j we sell; note that we pretend that we will *actually* sell what we plan to assemble. The model maximizes net profit (3.5), subject to capacity constraints (3.6). Equation (3.7) states that we cannot sell more than what is demanded, and (3.8) says that we cannot assemble end items if we do not have enough components.

Solving this model we get the following optimal solution:

$$\begin{array}{lll} x_1^* = 116.67 & x_2^* = 116.67 & \\ x_3^* = 26.67 & x_4^* = 0.00 & x_5^* = 90.00 \\ y_1^* = 26.67 & y_2^* = 0.00 & y_3^* = 90.00\ . \end{array}$$

In this specific case, it is easy to see what the model tries to accomplish. We sell the maximum quantity of the attractive item A_3, meeting all the demand; this requires production capacity for producing common components c_1 and c_2 and the specific component c_5. The residual capacity is used to produce some amount of component c_3 which is required for assembling A_1, plus the required common components; the low margin end item A_2, and the associated specific component c_4 are disregarded. It should be noted that, in general, high margin items may not be so attractive if they have high capacity requirements. In this case the optimal solution is quite readable, but also a bit extreme. A real-life production planner would immediately see the risk of this production plan, which is essentially a bet on high sales of the most profitable item.

The "optimal profit" is 3233.33, but this is actually an illusion, since we do not really know what the demand would be. In order to tackle uncertainty properly, we must complicate the model a bit:

$$\max \quad -\sum_{i=1}^{5} C_i x_i + \sum_{s=1}^{3}\sum_{j=1}^{3} \pi^s P_j y_j^s \tag{3.9}$$

$$\text{s.t.} \quad \sum_{i=1}^{5} T_{im} x_i \leq L_m \qquad m = 1, 2, 3 \tag{3.10}$$

$$y_j^s \leq d_j^s \qquad j = 1, 2, 3; \quad s = 1, 2, 3 \tag{3.11}$$

$$\sum_{j=1}^{3} G_{ij} y_j^s \leq x_i \qquad i = 1, 2, 3, 4, 5; \quad s = 1, 2, 3 \tag{3.12}$$

$$y_j^s, x_i \geq 0$$

The big change here is in the new set of decision variables y_j^s; this is the amount of item j we *would* sell under scenario s. They are *contingent* plans, conditional on the realization of a specific scenario, where end item demand is d_j^s. They are second-stage (recourse) decision variables; the real model output is the set of first-stage decision variables x_i, which are the decisions we would really implement here and now. The decision of how to use components would be postponed to a second stage, when we discover end item demand and we use the available components as best as we can.

Objective function (3.9) includes a term linked to expected revenue, where π^s is the probability of scenario s. Capacity constraint (3.10) is not changed at all, as this is a deterministic constraint related to first-stage variables only. The second-stage demand constraint (3.11) is now contingent on stochastic data. Finally, constraint (3.12) links the two decisional stages.

This is a typical example of two-stage stochastic programming with recourse. Solving it, we get the following optimal solution:

$$
\begin{array}{lll}
x_1^* = 115.72 & x_2^* = 115.72 & \\
x_3^* = 52.86 & x_4^* = 2.86 & x_5^* = 62.86 \\
y_1^{1*} = 52.86 & y_2^{1*} = 0.00 & y_3^{1*} = 62.86 \\
y_1^{2*} = 50.00 & y_2^{2*} = 2.86 & y_3^{2*} = 62.86 \\
y_1^{3*} = 52.86 & y_2^{3*} = 2.86 & y_3^{3*} = 60.00 .
\end{array}
$$

We see a qualitative difference with respect to the deterministic solution. The plan is much less extreme and possibly more robust. We do not take all of our chances by placing a bet on high sales of end item A_3, since this item has a low sale in scenario S_3. So, more specific components c_3 is produced in order to be able to sell item A_1; even a small amount if specific component c_4 is produced, since selling item A_2 is useful to exploit the availability of the common components in the case the demand for the more attractive end items is low.

It is also important to note that the two solutions do not differ significantly in terms of production of the common components. Common components represent a form of flexibility. Indeed, the opportunity of using common components to better hedge for uncertainty is a commonly considered practice, even if they are more expensive (Gerchack and Henig, 1986; Jönsson et al., 1993). In terms of capacity planning, a similar role is played by flexible capacity as opposed to efficient but dedicated machines; trading off efficiency for flexibility may be of strategic value in the case of demand uncertainty.

The expected net profit under the stochastic solution is 2885.71; comparing this value with the profit 3233.33 of the first solution does not make any sense. What we must do is computing the expected revenue of the first production plan, by plugging first-stage decision variables in the second-stage model, optimizing with respect to sales decision variables under each of the three scenarios. For instance, if scenario S_1 occurs, the optimal sales plan will be:

$$
y_1^* = 26.67 \qquad y_2^* = 0.00 \qquad y_3^* = 90.00 .
$$

The same applies under scenario S_2, but if scenario S_3 occurs, we are in trouble, as we lack component flexibility to react to low sales of end item A_3. Under this scenario the optimal sales plan is:

$$
y_1^* = 26.67 \qquad y_2^* = 0.00 \qquad y_3^* = 60.00 .
$$

The overall expected profit is actually 2333.33. So, the relative improvement of the stochastic solution with respect to the deterministic solution is:

$$\frac{2885.71 - 2333.33}{2333.33} \approx 23.67\%.$$

Of course this is just a toy example, and the improvement is so high due to the high impact of specific components. As shown in (Alfieri and Brandimarte, 1999), in this type of problems the number of specific components is the most important determinant of the benefit of using a stochastic formulation, together with demand variability and available capacity. It should also be noted that in practice three scenarios are a crude representation of uncertainty, and that in order to really assess robustness of solutions one should estimate expected revenue on the basis of *out-of-sample* scenarios, i.e., scenarios which are not included in the optimization models to keep it computationally tractable.

4. General structure of two-stage stochastic linear programs

In this section we formalize and generalize the modeling approach illustrated in the last example. A two-stage stochastic linear program with recourse has the form:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} + \mathrm{E}_\omega[h(\mathbf{x}, \omega)] \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

where the first-stage decisions $\mathbf{x}$ must be taken here and now, subject to deterministic constraints, taking into account the expected value of future recourse cost. The second-stage cost depends on our first-stage decisions, but also on random events ω. The notation E_ω is used to emphasize that the expected value is to be taken with respect to the probability measure of the underlying probability space with respect to events ω. After uncertainty is resolved, i.e., when we know the realized event ω and the value of the uncertain data, we take second-stage (recourse) decisions $\mathbf{y}$ by solving the second-stage optimization problem:

$$\begin{aligned} h(\mathbf{x}, \omega) \equiv \min \quad & \mathbf{q}(\omega)^T\mathbf{y} \\ \text{s.t.} \quad & \mathbf{W}(\omega)\mathbf{y} = \mathbf{r}(\omega) - \mathbf{T}(\omega)\mathbf{x} \\ & \mathbf{y} \geq \mathbf{0}. \end{aligned}$$

The general formulation above seems hopelessly complicated. The recourse function $H(\mathbf{x}) \equiv \mathrm{E}_\omega[h(\mathbf{x}, \omega)]$ is an expected value, hence, a possibly high-dimensional integral, of a function which is only defined implicitly by an optimization problem. Indeed, the resulting optimization problem is actually nonlinear. However, it is not too difficult to prove that $H(\mathbf{x})$ is, under fairly mild conditions, a convex function (Birge and Louveaux, 1997; Kall and Wallace, 1994), which paves the way to computational algorithms based on statistical sampling. An example is the stochastic decomposition algorithm of (Higle and Sen, 1996).

However, the common practical approach is to approximate the underlying continuous distribution of the parameters by a discrete one, represented by a finite set S of scenarios. This leads to the following deterministic equivalent, where π^s is the probability of scenario s:

$$\begin{aligned}
\min \quad & \mathbf{c}^T\mathbf{x} + \sum_{s \in S} \pi^s \mathbf{q}_s^T \mathbf{y}_s \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{T}_s\mathbf{x} + \mathbf{W}_s\mathbf{y}_s = \mathbf{r}_s \qquad \forall s \in S \\
& \mathbf{x}, \mathbf{y}_s \geq \mathbf{0}.
\end{aligned}$$

This is actually a standard LP problem, but possibly a large-scale one, which may call for clever solution methods (see next section).

A most important point in practice is checking if the additional complexity of such a model is really warranted, given the alternative of dealing with a much simpler model based on expected values. One way to assess this issue is by evaluating the so-called *Value of the Stochastic Solution* (VSS; see Birge and Louveaux, 1997), which is basically what we did with the ATO production planning example. Formally, let us consider a single scenario problem, assuming we know that the future scenario represented by event ω will certainly occur:

$$\begin{aligned}
\min \quad & z(\mathbf{x}, \omega) = \mathbf{c}^T\mathbf{x} + \min\{\mathbf{q}_\omega \mathbf{y} \mid \mathbf{W}_\omega = \mathbf{r}_\omega - \mathbf{T}_\omega \mathbf{x}, \mathbf{y} \geq \mathbf{0}\} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{x} \geq \mathbf{0}.
\end{aligned}$$

The two-stage problem with recourse can be formalized as:

$$\mathrm{RP} = \min_{\mathbf{x}} \mathrm{E}_\omega[z(\mathbf{x}, \omega)].$$

Using a somewhat sloppy notation, let us denote by $\bar{\omega} = \mathrm{E}[\omega]$ the expected value of the problem data; then, disregarding uncertainty and solving the deterministic expected value problem corresponds to:

$$\mathrm{EV} = \min_{\mathbf{x}} z(\mathbf{x}, \bar{\omega}),$$

which yields the "expected value solution" $\bar{\mathbf{x}}(\bar{\omega})$. This solution, however, must be evaluated in the real uncertain setting, yielding the expected value of the expected value solution:

$$\mathrm{EEV} = \mathrm{E}_\omega[z(\bar{\mathbf{x}}(\bar{\omega}), \omega)].$$

VSS is defined as:

$$\mathrm{VSS} = \mathrm{EEV} - \mathrm{RP}.$$

It can be shown that VSS is non-negative (Birge and Louveaux, 1997). The practical consequence is that ignoring uncertainty has some cost; the issue is exactly how much. For problems with low VSS, it is safe to solve a deterministic problem; otherwise, a stochastic problem should be solved. It should be also noted that in the definitions above, we assume that the description of uncertainty that we use in solving the model matches the "real" uncertainty exactly. In practice, when we base our model on a scenario tree, we should also test the robustness of our solution against out-of-sample scenarios.

5. Solution methods

In this section we hint at solution methods for solving two-stage stochastic programming problems. Readers interested only in modeling issues may skip this section with no loss of continuity.

In principle, after building the deterministic equivalent model, we may apply standard LP solvers based on the simplex method. However, two complicating issues should be taken into consideration. On the one hand, if we build a rich scenario model in order to represent uncertainty, the sheer size of the resulting optimization problem may be difficult to deal with. On the other one, a less obvious and often overlooked difficulty is that even moderate size stochastic programming problems may be numerically hard to solve. We should mention that the worst-case complexity of the simplex algorithm is exponential, and one alternative is using commercially available interior point codes.

If we are willing to step outside the standard commercial solvers, there is a rich set of algorithms that have been proposed. Some are variations of interior point methods, others are based on different forms of decomposition. We refer the interested reader, e.g., to (Birge, 1997), and we just give a hint of the oldest specialized method for stochastic programming, the L-shaped decomposition method, which is basically an adaptation of Benders decomposition (Van Slyke and Wets, 1969).

Consider the following deterministic equivalent:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} + \sum_{s\in S} \pi^s \mathbf{q}_s^T \mathbf{y}^s \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

$$\begin{aligned} &\mathbf{W}\mathbf{y}_s + \mathbf{T}_s\mathbf{x} = \mathbf{r}_s \qquad \forall s \in S \\ &\mathbf{x}, \mathbf{y}_s \geq \mathbf{0} \end{aligned}$$

Here the recourse matrix $\mathbf{W}$ is fixed, i.e., it does not depend on the realized scenario; we also assume, for simplicity, that whatever first-stage decision $\mathbf{x}$ we take, the second-stage problem for every scenario s is feasible. Technically, this corresponds to a *full* recourse model. Cases in which feasibility of the second-stage problem is not granted can be dealt with, but in most manufacturing problems we may obtain full recourse structure simply by allowing for suitably penalized lost sales.

The technological matrix of the overall problem has the following block-diagonal structure:

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{T}_1 & \mathbf{W} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{T}_2 & \mathbf{0} & \mathbf{W} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_S & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W} \end{bmatrix}.$$

First-stage decision variables are associated to the first group of columns of this matrix; then a set of scenario dependent second-stage variables is associated to each set of columns. We see that if we imagine fixing the value of first-stage variables $\mathbf{x}$, the remaining problem can be easily decomposed by scenarios. This decomposition principle may be exploited as follows.

We recall the convexity of the recourse function:

$$H(\mathbf{x}) = \sum_{s \in S} \pi^s h_s(\mathbf{x}),$$

where

$$\begin{aligned} h_s(\mathbf{x}) \equiv \quad \min \quad & \mathbf{q}_s^T \mathbf{y}^s && (3.13) \\ \text{s.t.} \quad & \mathbf{W}\mathbf{y}_s = \mathbf{r}_s - \mathbf{T}_s\mathbf{x} && (3.14) \\ & \mathbf{y}_s \geq \mathbf{0}. \end{aligned}$$

Convexity can be proved also for the general case in which we do not discretize uncertainty into a set of discrete scenarios. In this case, the recourse function will be also differentiable, as depicted in Figure 3.2. Convexity implies that at each point the function can be linearized by a support hyperplane. Using the upper envelope of a suitable set of support hyperplanes we may approximate the recourse function by a piecewise linear function. It turns out that in the case of a discrete

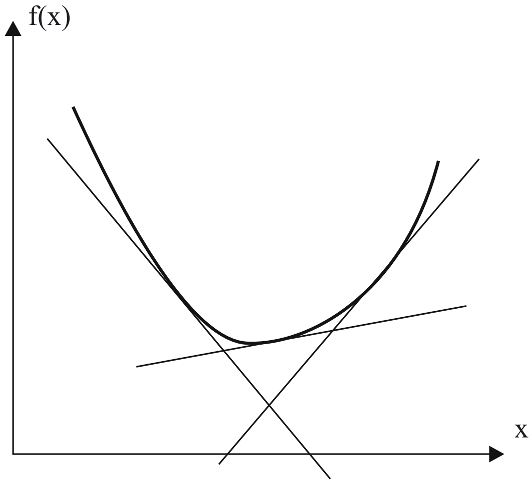


Figure 3.2. Convexity of the recourse function and approximation by support hyperplanes.

set of scenarios the recourse function is a piecewise linear function with kinks, i.e., it is not everywhere differentiable. Nevertheless, it is convex and it can be approximated by a set of support hyperplanes; technically, support hyperplanes are not associated to gradients of the recourse function, as in the case where differentiability applies, but to subgradients (a convex function is subdifferentiable in the interior of its domain).

Putting all of this together, we may rewrite the deterministic equivalent as:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} + \theta && (3.15) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \theta \geq H(\mathbf{x}) && (3.16) \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Computationally, constraint (3.16) is relaxed and approximated by a set of cuts:

$$\theta \geq \boldsymbol{\alpha}^T\mathbf{x} + \beta.$$

Each cut corresponds to a support hyperplane in Figure 3.2.

Under our full recourse hypothesis it is easy to come up with support hyperplanes. We solve the approximation (3.15), yielding a proposed first-stage decision $\hat{\mathbf{x}}$. Then we pass this first-stage decision as a proposal to the individual second-stage problems, one for each scenario. Let $\hat{\boldsymbol{\mu}}_s$ be the vector of optimal dual variables for constraint (3.14), depending on scenario s. Collecting all the information from second-stage problems we may build the following cut:

$$\theta \geq \sum_{s \in S} \pi^s \left(\mathbf{r}_s - \mathbf{T}_s \mathbf{x}\right)^T \hat{\boldsymbol{\mu}}_s.$$

This cut is added to the relaxed problem (3.15), where it cuts off the last proposal $\hat{\mathbf{x}}$. The process is repeated to convergence. Intuitively and (very) informally, we may think that the dual variables give us information about the derivative of the optimal value of the second-stage problem with respect to first-stage proposals, and this provides us with first-order information to enrich our piecewise linear approximation (this is actually not exact in the non-differentiable case, where dual variables give us subgradient rather than gradient information).

The procedure may be easily extended to cases in which full recourse does not apply and not all the second-stage problems are feasible given a first-stage solution; in such a case we may still find a cut which should be added to approximation (3.15).

6. Multi-stage stochastic programming models

Multi-stage stochastic programming formulations arise naturally as a generalization of two-stage models. At each stage we gather information and we take decisions accordingly, taking into account immediate costs and expected future recourse cost. The resulting decision process may be summarized as follows (Ruszczyński and Shapiro, 2003):

- at the beginning of first time period (at time $t = 0$) we take decisions $\mathbf{x}_0$; these decisions have deterministic immediate cost $\mathbf{c}_0$, and must satisfy constraints

  $$\mathbf{A}_{00}\mathbf{x}_0 = \mathbf{b}_0;$$

- at the beginning of second time period we observe random data $(\mathbf{A}_{10}, \mathbf{A}_{11}, \mathbf{c}_1, \mathbf{b}_1)$ depending on event ω_1; then, based on this information we take decisions $\mathbf{x}_1$; such decisions have immediate cost $\mathbf{c}_1$ and must satisfy constraints

  $$\mathbf{A}_{10}\mathbf{x}_0 + \mathbf{A}_{11}\mathbf{x}_1 = \mathbf{b}_1;$$

  note that these data are not known at time $t = 0$, but only at time $t = 1$; the new decisions depend on the realization of these random data and on the previous decisions;

- ...

- at the beginning of the last time period T we observe random data $(\mathbf{A}_{T,T-1}, \mathbf{A}_{TT}, \mathbf{c}_T, \mathbf{b}_T)$ depending on event ω_T; then, based on this information we take decisions $\mathbf{x}_T$; such decisions have immediate cost $\mathbf{c}_T$ and must satisfy constraints

$$\mathbf{A}_{T,T-1}\mathbf{x}_{T-1} + \mathbf{A}_{TT}\mathbf{x}_T = \mathbf{b}_T.$$

We see that we have a dynamic decision process in which we use the available information to adapt to new circumstances. This point may be appreciated by looking at the following recursive formulation of the multi-stage problem:

$$\min_{\substack{\mathbf{A}_{00}\mathbf{x}_0 = \mathbf{b}_0 \\ \mathbf{x}_0 \geq \mathbf{0}}} \mathbf{c}_0^T \mathbf{x}_0 + \mathrm{E}\left[\min_{\substack{\mathbf{A}_{10}\mathbf{x}_0 + \mathbf{A}_{11}\mathbf{x}_1 = \mathbf{b}_1 \\ \mathbf{x}_1 \geq \mathbf{0}}} \mathbf{c}_1^T \mathbf{x}_1 + \right.$$
$$\left. \mathrm{E}\left[\ldots + \mathrm{E}\left[\min_{\substack{\mathbf{A}_{T,T-1}\mathbf{x}_{T-1} + \mathbf{A}_{TT}\mathbf{x}_T = \mathbf{b}_T \\ \mathbf{x}_t \geq \mathbf{0}}} \mathbf{c}_T^T \mathbf{x}_T \right]\right]\right]$$

In this formulation we see that decisions $\mathbf{x}_t$ depends directly only on the previous decisions $\mathbf{x}_{t-1}$. In general, decisions may depend on all of the past history, leading to a slightly more complicated model. However, often we may introduce additional decision variables, playing the role of state variables in a Markov process, such that the above formulation applies. For instance, in a production planning model we may “forget” the past produced quantities if we know the current inventory levels. It should be noted that, in practice, the real output of the above model is the set of immediate decisions $\mathbf{x}_0$. The remaining decision variables could be thought of as contingent plans, which are implemented in time much in the vein of a feedback control policy, but in practical setting it is more likely that the model will be solved again and again according to a rolling horizon logic.

While this formulation points out the dynamic optimization nature of multi-stage problems, we usually resort to deterministic equivalents based on discrete scenario trees. A small scenario tree is depicted in Figure 3.3. This is a straightforward generalization of the two-stage tree depicted in Figure 3.1. A scenario is a path in the tree; in the example, we have eight scenarios, and scenario 6 corresponds to the sample path (n_0, n_2, n_5, n_{12}). Immediate decisions, here and now, must be taken in the root of the tree, node n_0. The scenario tree is a computationally viable way of discretizing the underlying probability distributions. It is

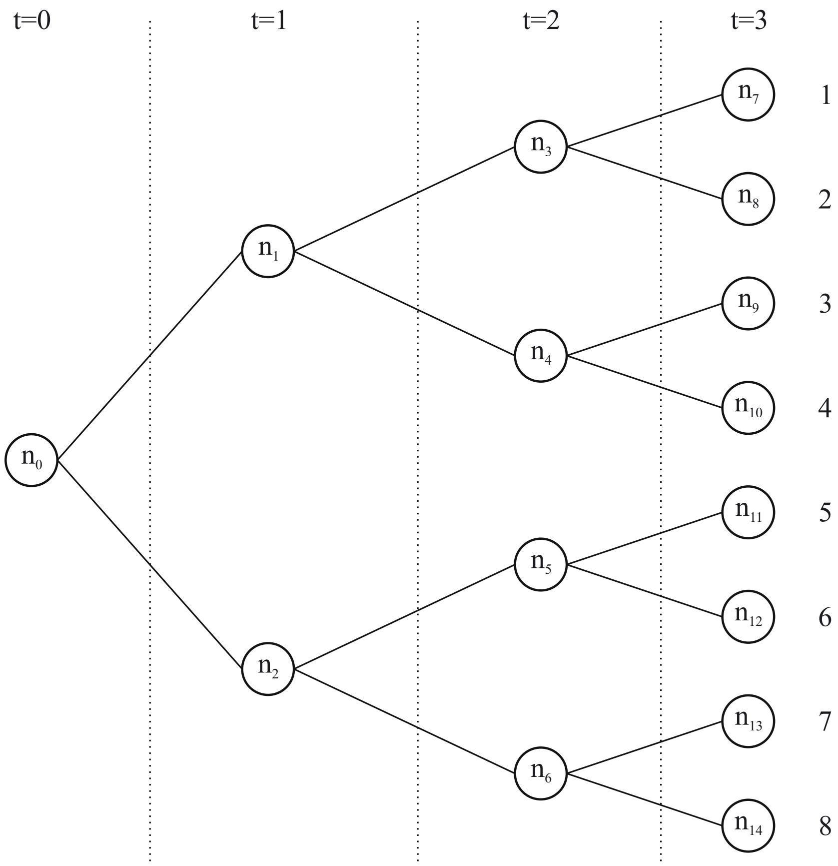


Figure 3.3. Scenario tree for multi-stage stochastic programming.

also useful to point out the role of information in decision making. At each node of the tree, we must take a decision, which will be conditional on the information we have gathered so far. In the Figure, if we are on node n_1, we do not know exactly which scenario we are living, since we cannot distinguish between scenarios 1, 2, 3, and 4. If at the next time step we find ourselves in node n_4, we have a more refined information, as we know that the alternative is between scenarios 3 and 4.

Our decision process must conform to the flow of available information, which basically means that decisions must be non-anticipative (or implementable).[1] There are two ways to impose this non-anticipativity requirement:

[1] Readers with a background in measure theoretic probability would probably prefer to say that the decision process is adapted to the filtration generated by the data process.

1 in *split variable* formulations we introduce a set of decision variables for each time period and each scenario; then we enforce non-anticipativity constraints explicitly based on the shape of the scenario tree;

2 in *compact* formulations we associate decision variables to the nodes of the tree and we build non-anticipativity in a more direct way.

In the next sections we give examples of the two formulation styles. To see this point in concrete, we will refer to the following simple production planning problem.

We give a deterministic formulation first. We have a discrete-time planning horizon, and we must decide production lot sizes x_{it} in order to meet demand d_{it} for item i in time period t. We have production capacity constraints, as each resource m has availability R_{mt} (e.g., expressed in hours) and each item of type i requires r_{im} units of resource m. Thus, it may be useful to build up inventories when demand is low (we neglect setup costs and times here); I_{it} is the inventory level for item i at the end of time period t, and unit inventory cost is h_i. We rule out backlog: in the case we cannot meet demand immediately, we have a lost sale z_{it} with unit penalty p_i. The resulting LP formulation is:

$$\min \quad \sum_t \sum_i (h_i I_{it} + p_i z_{it}) \tag{3.17}$$

$$\text{s.t.} \quad I_{it} = I_{i,t-1} + x_{it} - d_{it} + z_{it} \qquad \forall i, t \tag{3.18}$$

$$\sum_i r_{im} x_{it} \leq R_{mt} \qquad \forall m, t \tag{3.19}$$

$$I_{it}, x_{it}, z_{it} \geq 0. \tag{3.20}$$

We want to generalize this model to the case of stochastic demand. In Section 6.1 we analyze a split-variable formulation. In Section 6.2 we analyze a compact formulation.

6.1 Split variable model formulation

Assume we have a scenario tree for the demand process and that d_{it}^s is the demand for item i at time t in scenario s (recall that a scenario is a sequence of nodes, or events, in the tree). In the split-variable approach, we introduce scenario dependent decision variables:

- x_{it}^s is the amount of item i produced during time period t in scenario s;
- I_{it}^s is the corresponding inventory level;

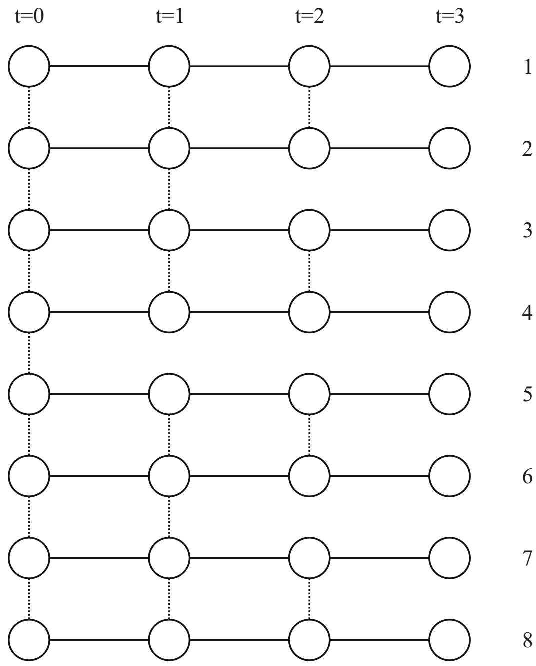


Figure 3.4. Split-variable view of an event tree.

- z_{it}^s is the lost sale.

Note that this way of defining decision variables has some redundancy. At time $t = 0$, we must take *one* decision, whereas in the above definition we have one group of variables x_{i0}^s for each scenario. Indeed, it is important to understand that if we define the decision variables in this way, we must enforce a non-anticipativity constraint explicitly. The issue may be understood by looking at Figure 3.4. This is the same scenario tree of Figure 3.3, where some nodes have been replicated. We should have a set of decision variables for each node however, and the decision variables corresponding to different scenarios at the same time t must be equal if the two scenarios are indistinguishable at time t. This is represented by the dotted lines in Figure 3.4. To begin with, the production in the first period must be the same for all scenarios. Hence:

$$x_{i0}^s = x_{i0}^{s'} \qquad i = 1, \ldots, N;\ s, s' = 1, \ldots, S.$$

Now consider time $t = 1$ and node n_1 of the original event tree as depicted in Figure 3.3; the scenarios $s = 1, 2, 3, 4$ pass through this node and are indistinguishable at time $t = 1$. Hence, we must have

$$x_{i1}^1 = x_{i1}^2 = x_{i1}^3 = x_{i1}^4 \qquad i = 1, \ldots, N.$$

In fact, node n_1 corresponds to the uppermost four nodes in the second column of the array representing the split view of the tree (see figure 3.4). By the same token, at time $t = 2$ we have constraints like

$$x_{i2}^5 = x_{i2}^6 \qquad i = 1, \ldots, N.$$

More generally, it is customary to denote by $\{s\}_t$ the set of scenarios which are not distinguishable from s up to time t. For instance:

$$\begin{aligned} \{1\}_0 &= \{1,2,3,4,5,6,7,8\} \\ \{2\}_1 &= \{1,2,3,4\} \\ \{5\}_2 &= \{5,6\}. \end{aligned}$$

Non-anticipativity constraints may be written in general as

$$x_{it}^s = x_{it}^{s'} \qquad \forall i, t, s, s' \in \{s\}_t.$$

This is not the only way of expressing the non-anticipativity requirement, and selection of the best approach depends on the solution algorithm one wants to select.

Now it is rather straightforward to adapt the deterministic formulation above to the case of stochastic demand:

$$\begin{aligned}
\min \quad & \sum_s \sum_t \sum_i \pi^s \left(h_i I_{it}^s + p_i z_{it}^s \right) && (3.21) \\
\text{s.t.} \quad & I_{it}^s = I_{i,t-1}^s + x_{it}^s - d_{it}^s + z_{it}^s \forall i, t, s && (3.22) \\
& \sum_i r_{im} x_{it}^s \le R_{mt} \qquad \forall m, t, s && (3.23) \\
& x_{it}^s = x_{it}^{s'} \qquad \forall i, t, s, s' \in \{s\}_t && (3.24) \\
& I_{it}^s = I_{it}^{s'} \qquad \forall i, t, s, s' \in \{s\}_t && (3.25) \\
& z_{it}^s = z_{it}^{s'} \qquad \forall i, t, s, s' \in \{s\}_t && (3.26) \\
& I_{it}^s, x_{it}^s, z_{it}^s \ge 0. && (3.27)
\end{aligned}$$

Actually, given the balance equation linking inventory levels, lot sizes and lost sales, only two out of the three sets of non-anticipativity constraints (3.24), (3.25), and (3.26) are needed, as the third one is redundant.

The split-variable formulation may look a bit nonsensical, as we build a redundant set of variables just to get rid of them by enforcing equality constraints. However, making the non-anticipativity constraints explicit has some merit, both from a conceptual and a computational point of view.

From a conceptual point of view, it is instructive to wonder what happens if we enforce non-anticipativity only for the first-stage decision variables. This results in a multi-period, but *two-stage* problem. Indeed, periods and stages need not coincide. In this case we take a first-stage decision under uncertainty, but then we assume perfect foresight. So all of the time periods but the first one collapse in a single second stage, from the information point of view. Many multi-period production planning models under uncertainty, which have been proposed in the literature, are actually two-stage (see, e.g., Bakir and Byrne, 1998). On the one hand, this may be considered a wrong way to build a stochastic model, but on the other one it may be a sensible simplification. The difference between stages and periods is further made clear in the capacity planning model we deal with in Section 9.1.

From a computational point of view, some specialized solution algorithms are based on the dualization of non-anticipativity constraints. One example is progressive hedging, also known as scenario aggregation; see, e.g., (Birge, 1997; Kall and Wallace, 1994) for a tutorial treatment and (Haugen et al., 2001) for an application to lot-sizing with setup costs. Another manufacturing application of scenario aggregation is proposed in (Jönsson et al., 1993). A split-variable model formulation for capacity planning has been proposed in (Chen et al., 2002), and is briefly outlined in Section 9.2.

6.2 Compact model formulation

The split-variable formulation is based on a large number of variables, which are then linked together by the non-anticipativity constraints. A more compact formulation may be obtained by associating decision variables directly to the nodes in the tree. Let us introduce the following notation:

- N is the set of event nodes; in the case of Figure 3.3:

$$N = \{n_0, n_1, n_2, \dots, n_{14}\}.$$

- Each node $n \in N$, apart from the root node n_0, has a unique direct ancestor, denoted by $a(n)$: for instance, $a(n_3) = n_1$ and $a(n_{12}) = n_5$.

Now we should associate decision variables and stochastic data to nodes in the scenario tree: x_i^n, I_i^n, z_i^n, and d_i^n. Note that we have no direct dependence on time, as this is implicit in the correspondence between nodes and time periods. We also need the unconditional probability π^n of getting to node n; this is simply the product of the conditional probabilities associated to each branching in the tree.

This results in the following formulation:

$$\min \quad \sum_n \pi^n \left[\sum_i (h_i I_i^n + p_i z_i^n) \right] \tag{3.28}$$

$$\text{s.t.} \quad I_i^n = I_i^{a(n)} + x_i^n - d_i^n + z_i^n \qquad \forall i, n \tag{3.29}$$

$$\sum_i r_{im} x_i^n \le R_m^n \qquad \forall m, n \tag{3.30}$$

$$I_i^n, x_i^n, z_i^n \ge 0. \tag{3.31}$$

Compact formulations like this one lend themselves to generalizations of the L-shaped methods outlined in Section 5, such as nested Benders decomposition (Birge and Louveaux, 1997). They are also computationally cheaper when using standard solvers.

A compact model formulation for capacity planning is described in Section 9.3.

7. Strong mixed-integer model formulations

The multi-stage planning models we have just considered are continuous LP models. However, in manufacturing applications we need to introduce binary or general integer decision variables quite often. This is the case both in production planning, due to setup times and costs, and in capacity planning, when additional resources can only be acquired in discrete units. This results in mixed-integer stochastic programming models, which are indeed a hard nut to crack. Considerable research effort is undergoing on methods to solve this class of problems (Sen, 2003), and solution strategies have been proposed for specific cases (see, e.g., Guan et al., 2004, for an application of branch-and-cut and Lulli and Sen., 2002, for an application of branch-and-price). In general models, state-of-the art principles used in branch-and-bound can also be combined with specific knowledge about the structure of stochastic programming models in order to use commercial solvers (Parija et al., 2004). Alternatively, heuristic solution algorithms could be devised, with preference for approaches which are not too application specific. A good way to achieve generality is using continuous relaxations of the MILP models, which is the basis of commercial, general purpose branch-and-bound codes and LP-based heuristics. In (Alfieri et al., 2002) a simple rounding strategy is found to be rather effective in solving deterministic lot sizing problems. Whatever approach we take, we must come up with suitably strong formulations.

It is well-known that problem size is not necessarily the main determinant of computational effort in mixed-integer programming. The strength of the formulation, i.e., the gap between the optimal integer solution and the solution of the continuously relaxed problem has a strong impact in the quality of the bounds we use within branch-and-bound methods (Wolsey, 1998). As a practical example, we consider here the formulation of a stochastic version of the classical multi-item Capacitated Lot Sizing Problem (CLSP). We start with the typical, and natural, formulation of the problem:

$$\begin{aligned}
\min \quad & \sum_t \sum_i (h_i I_{it} + f_i s_{it}) \\
\text{s.t.} \quad & I_{it} = I_{i,t-1} + x_{it} - d_{it} \forall i, t \\
& \sum_i (r_i x_{it} + r_i s_{it}) \leq R \qquad \forall t \\
& x_{it} \leq \sum_{\tau=t}^{T} d_{i\tau} s_{it} && (3.32) \\
& I_{it}, x_{it} \geq 0; \qquad s_{it} \in \{0, 1\}, && (3.33)
\end{aligned}$$

where h_i and f_i are the inventory and setup costs, respectively, for item i; R is the availability of the capacitated resource (we assume one, bottleneck, resource; the model can be trivially extended to multiple capacitated resources); r_i and r'_i are the processing and setup time for item i. This formulation finds the optimal tradeoff between inventory and fixed setup costs, under capacity constraints, assuming deterministic demand (which is the reason why lost sales are not considered). The model is similar to the previous planning model we have considered. The main difference is the binary setup variable s_{it}, which is set to 1 if production of item i takes place during time period t, and to 0 otherwise. The setup variable enters the objective function by a fixed cost and the capacity constraints by a setup time. This model can be solved by standard branch-and-bound, but the main issue is constraint (3.32) linking production and setup variables. This is a typical case of "big-M" constraints, whose general form is $x \leq Ms$, linking a continuous variable x and a binary variable s. The trouble is that if the constant M is big, by relaxing binary variables to continuous values ($s_{it} \in [0, 1]$) we get a weak lower bound on optimal cost, resulting in slow solution times and huge search trees. Alternative model formulations have been developed; see, e.g., (Brandimarte and Villa, 1995) for a tutorial treatment and a list of references. Such strong model formulations are also useful in developing LP-based heuristics (Alfieri et al., 2002). In this section we outline the use of a plant-location based reformulation, originally due to (Krarup and Bilde, 1977), which is commonly used in the lot-sizing literature. A similar model formulation for stochastic lot-sizing is also used

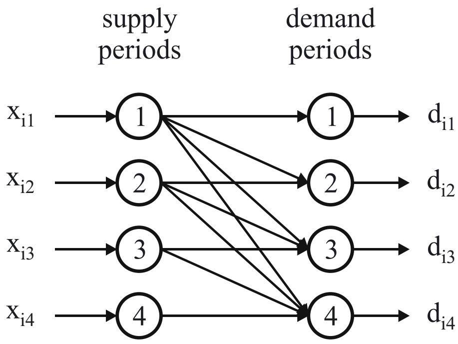


Figure 3.5. Graphical interpretation of plant location formulation.

in (Ahmed et al., 2003). For a survey on stochastic lot-sizing models, see (Sox et al., 1999).

The basic idea of the plant location formulation is disaggregating the production variable x_{it} into decision variables y_{itp}, denoting the amount of item i produced during time bucket t in order to meet demand in the current or in a future time bucket p ($p \geq t$). The model can be visualized as the network flow depicted Figure 3.5 for a single item. Supply and demand nodes are indexed by time periods. The variable x_{it} is the flow entering a supply node, which is routed to meet the demand which flows out of demand nodes. The disaggregate y_{itp} variables are just the flows on the intermediate arcs between the two arrays of nodes. This formulation can be interpreted as a plant location problem where the fixed cost of opening a plant corresponds to the setup cost during a time bucket, and the commodities are shipped in time rather than in space, incurring inventory instead of transportation costs. Assuming a deterministic demand d_{it}, this results the following model:

$$\begin{aligned}
\min \quad & \sum_i \sum_t \sum_{p \geq t} h_i (p-t) y_{itp} + \sum_i \sum_t f_i s_{it} \\
\text{s.t.} \quad & \sum_{t \leq p} y_{itp} = d_{ip} \qquad \forall i, p \\
& y_{itp} \leq d_{ip} s_{it} \qquad \forall i, t, p \geq t \qquad\qquad (3.34) \\
& \sum_i \sum_{p \geq t} r_i y_{itp} + \sum_i r'_i s_{it} \leq R \qquad \forall t \\
& y_{itp} \geq 0 \qquad s_{it} \in \{0, 1\}.
\end{aligned}$$

The key advantage in this model formulation is that the big-M in constraint (3.34) linking production and setup variables is smaller than in the natural formulation (it is the demand in one time bucket, rather than the sum of future demands). This results in remarkable advantages when using LP-based solution algorithms, but it has the unpleasing effect that coming up with a stochastic model is more difficult than with the natural formulation, especially if a compact formulation is adopted.

To write the stochastic multi-stage model, let us introduce the following notation (see Brandimarte, 2004, for more details):

- $n \in \mathcal{N}$ is a generic node of the scenario tree; assume 0 is the root node; Let $\mathcal{T}$ be the set of terminal nodes (the leaves of the tree);
- $T(n)$ is the time period for node n;
- $a(n)$ is the immediate predecessor for node n, $n \neq 0$;
- $\Omega(n, t)$ is the (unique) ancestor of node n at time period t ($n \neq 0,\ t < T(n)$);
- $\Sigma(n, t)$ is the *set* of successor nodes of n at time period t ($n \notin \mathcal{T},\ t > T(n)$).

The demand data are represented by d_i^n, the demand for item i in node n. We denote the unconditional probability of node n by π^n. Other data are just as in deterministic model. We introduce the following decision variables:

- $s_i^n \in \{0, 1\}$ setup variable for item i in node n;
- $y_{it}^n \geq 0$ is the amount of i produced in node n to meet demand in time period $t \geq T(n)$; note that, because of non-anticipativity, what we ship from supply node n is the same for all its successor nodes in the scenario tree corresponding to each time layer; this is why the y variable is not indexed by the demand node, but by the corresponding time index t;
- $I_i^n \geq 0$ is leftover inventory, i.e., the surplus amount of item i not consumed at node n and passed for use to *immediate* successor nodes in the tree;
- $z_i^n \geq 0$ is the lost sale, i.e., the amount of item i demand we fail to meet at node n.

Note that we leave room for the possibility of not satisfying part of the demand; lost sales for item i are penalized by a cost g_i. As we have noted previously, we could require that the "worst-case" demand is met under any circumstance, but this would lead to a fat solution, which may be too costly. Allowing for lost sales, we get full recourse. The resulting model formulation of SCLSP (Stochastic CLSP) is:

$$\begin{aligned}
\min \quad & \sum_{n\in\mathcal{N}} \pi^n \left(\sum_i \left(f_i s_i^n + h_i I_i^n + g_i z_i^n \right) \right) + \\
& \sum_{n\in\mathcal{N}\setminus\mathcal{T}} \pi^n \left(\sum_i \sum_{t>T(n)} h_i(t-T(n)) y_{it}^n \right) && (3.35) \\
\text{s.t.} \quad & I_i^{a(n)} + \sum_{t<T(n)} y_{i,T(n)}^{\Omega(n,t)} + y_{i,T(n)}^n = d_i^n + I_i^n - z_i^n \qquad \forall i,n \\
& && (3.36) \\
& y_{it}^n \le \left(\max_{j\in\Sigma(n,t)} d_i^j \right) s_i^n \qquad \forall i,n,t>T(n) && (3.37) \\
& y_{i,T(n)}^n \le d_i^n s_i^n \qquad \forall i,n && (3.38) \\
& \sum_i \sum_{t\ge T(n)} r_i y_{it}^n + \sum_i r_i' s_i^n \le R \qquad \forall n && (3.39) \\
& y_{it}^n, I_i^n, z_i^n \ge 0; \quad s_i^n \in \{0,1\}
\end{aligned}$$

The first term of the objective function (3.35) accounts for the expected value of setup costs, leftover inventory (which is the surplus of production with respect to demand and stays in inventory for one time period), and lost sales. The second term is the expected inventory cost, just like in the deterministic model: it corresponds to flows from supply to demand nodes. It involves nonterminal nodes only, since terminal nodes do not ship anything to the future intentionally (apart from leftover inventory at the end of the planning horizon, which is zero for the deterministic model, but not necessarily in the stochastic case).

Constraint (3.36) is just a flow balance constraint in each demand node. The inflow is the sum of the leftover inventory from the immediate predecessor node in the tree, plus the sum of all the shipments from ancestor nodes, plus the production for immediate consumption. The outflow is the demand, plus leftover inventory to immediate successor nodes, minus lost sales. Strictly speaking, we should write a different balance constraint for the initial root node, since its predecessor is not defined. In this case, $I_i^{a(n)}$ should be considered as the starting inventory, which is part of the problem data.

When linking setup and production variables, some care must be taken in selecting the right big-M. In (3.37) we must take the maximum over the possible demands of the successor nodes we are shipping items to. In the case of (3.38), we are dealing with production for immediate use, and in this case the demand is known (given our assumptions). The capacity constraint and the variable definition constraints are self-explanatory.

It is obvious that, for a sufficiently rich scenario tree, this model is not practically solvable at optimality. In (Brandimarte, 2004) computational experiments obtained by a heuristic solution strategy of this model are reported. We should note here that using strong formulations is just one side of the coin; the other side is the ability of representing uncertainty at a suitable level by a parsimonious scenario tree. Clever scenario generation is the topic of next section.

8. Scenario generation

Scenario trees are a powerful and flexible way of representing uncertainty. However, especially in multi-stage problems, they are prone to an exponential growth in size. Hence, due attention must be paid to scenario generation. In this section we review clever mechanisms that have been proposed to keep the size of the tree limited. We should bear in mind that the purpose of scenario trees is not really to yield a 100% faithful representation of the underlying uncertainty over the whole planning horizon, as there is little hope to achieve this goal while keeping the optimization model to a computationally tractable size. The real aim is to get robust first-stage decisions. Such robustness may be analyzed by theoretical or experimental tools. From a theoretical point of view, one should analyze the stability of the solution with respect to changes in the underlying probability measure, which are induced by errors in the approximation. Theoretical results are surveyed in (Römisch, 2003); a sensitivity analysis approach based on "contamination" between different scenario trees is described in (Dupačová, 1990); a practical simulation-based analysis is described for production planning problems in (Brandimarte, 2004).

In the following we assume that the shape of the scenario tree is given. Actually, the shape of the scenario tree can be optimized. For instance, we can use larger branching factors at the beginning of the planning horizon, on the basis that first-stage decisions are the important ones and that uncertainty must be represented more accurately at the beginning. A further observation is that the time step need not be the same for each stage; it may be reasonable to use larger time steps in later time periods, where aggregate decisions may be considered. It is also worth mentioning that setting suitable terminal conditions and or costs on the decision variables may help in cutting down the time horizon, therefore easing the difficulty.

8.1 Monte Carlo sampling and variance reduction methods

Sampling-based Monte Carlo simulation is the first approach that comes to mind in order to generate scenario trees. It is relatively easy to sample from a probability distribution, or from a time-series model. However, it is well-known that Monte Carlo methods require a large number of samples to achieve acceptable accuracy. This is feasible when evaluating system performance by simulation, but not in an optimization setting, especially when dealing with a multi-stage model.

It may be the case that due to variance in sampling, we get highly variable solutions (an issue referred to as sampling uncertainty). Actually, this may not be a problem if the solutions are almost equivalent in terms of the objective function, but sampling uncertainty is likely to undermine the trust in model recommendations. One possibility to overcome sampling uncertainty is to exploit standard variance reduction methods. Here we list the main approaches that can be used in stochastic programming.

Antithetic sampling. Antithetic sampling is based on the idea of inducing some correlation in the samples, in such a way to reduce overall variance. In the case of normal variates, this means that if we take one sample, say, on the right tail of the distribution, we should also take a corresponding sample on the left tail, symmetric with respect to the expected value. The two samples form an antithetic pair. In the normal case, and in general for symmetric distributions, it is easy to see that the sample will match the odd moments of the distribution. For arbitrary distributions, we may recall that the primary input to a stochastic simulation is a sequence of pseudo-random numbers from the uniform distribution on the interval $(0, 1)$; such numbers are then transformed into random variates with the desired distribution (Law and Kelton, 1999). To apply antithetic sampling, if we use a pseudo-random number U_i is a sample, we should use $1 - U_i$ in the corresponding antithetic sample.

Stratified sampling and Latin hypercubes. In the easiest version of stratified sampling, we partition the underlying distribution into "slices" (or strata), and take a suitable number of samples within each stratum. This makes sure that samples are distributed evenly, and it can be shown that this leads to reduced variance and improved estimates. The method is not easy to apply to multi-dimensional problems, as the number of strata gets easily out of hand (we should take the Cartesian product of the individual strata along each dimension). Latin hypercubes are a way to stratify, while keeping the number of samples low.

Importance sampling. Importance sampling is a more sophisticated strategy, based on the idea of changing (or twisting) the underlying probability measure in such a way to reduce variance. Of course, the estimator must be corrected by a suitable factor (called likelihood ratio) in order to compensate for this distortion. Importance sampling is often used to simulate rare events; in this case, in order to avoid wasting most of the samples, we should twist the distribution in such a way to make rare events more likely. Importance sampling has been proposed in the context of stochastic programming in (Dempster and Thompson, 1999) and (Infanger, 1998).

For more information on variance reduction strategies, the reader is referred, e.g., to (Brandimarte, 2001) for a tutorial treatment, or to (Glasserman, 2003) for a more advanced one.

8.2 Numerical integration methods: Gaussian formulae and low-discrepancy sequences

Monte Carlo simulation methods are essentially statistical approaches for numerical integration. To see this, consider the problem of approximating the value of a definite integral like

$$I = \int_0^1 h(x)\, dx,$$

where h is a function of a single variable. This integral can be thought of as an expected value:

$$I = \int_0^1 h(x) \cdot 1\, dx = \int_0^1 h(x) f_U(x)\, dx = \mathrm{E}[h(U)],$$

where $f_U(x) \equiv 1$ is the density of a uniform random variable U. Note that considering the unit interval is actually no loss of generality, as Monte Carlo simulation is actually based on uniform pseudo-random numbers. Everything we do with Monte Carlo is, conceptually, an integral over a unit hypercube in a m-dimensional space, where m is the number of random variates we need to carry out one replication of the simulation.

Statistical approaches are often used to overcome the curse of dimensionality, which makes traditional quadrature formulae ineffective when integrating in high dimensions. However, we should not forget that high-quality quadrature formulae are indeed available. Since the integration is a linear operator, it is natural to look for quadrature formulae preserving this property, such as:

$$\sum_{i=0}^{n} w_i h(x_i),$$

where w_i, $i = 0, 1, \ldots, n$, is a set of *weights* and $x_i \in [0, 1]$ is a set of *nodes*. Gaussian quadrature formulae are a clever approach for finding $n+1$ nodes *and* weights in such a way that the resulting formula is exact for integrating polynomials of degree up to $2n+1$ (see, e.g., Kincaid and Cheney, 2001). Gaussian quadrature is quite useful in converting continuous expectations into discrete ones. For an application to stochastic dynamic optimization see, e.g., (Miranda and Fackler, 2002). However, this approach is still restricted to cases in which the number of random factors is not too large.

Another possibility can be found by reflecting a bit on the nature of pseudo-random numbers. They are really not random at all, even if we pretend they are when we apply variance reduction strategies. In fact, they are generated by deterministic formulae, such as those used in standard Linear Congruential Generators. This suggests that there are sequences of numbers in the unit interval (or vectors in the unit hypercube), that can be successfully used for numerical integration. This leads to the idea of using *low-discrepancy* sequences (also known by the somewhat misleading name of *quasi-random* numbers, due to the appeal of the term "quasi-Monte Carlo integration").

Intuitively, we should find a sequence of points that fill the unit hypercube as uniformly as possible. This idea may be made more precise by defining the *discrepancy* of a sequence of numbers. Assume that we want to generate a sequence of N vectors $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N$ in the m-dimensional hypercube $I^m = [0, 1]^m \subset \mathbb{R}^m$. Now, given a sequence of such vectors, if they are well distributed, the number of points included in any subset G of I^m should be roughly proportional to its volume, vol(G).

Given a vector $\mathbf{X} = (x_1, x_2, \ldots, x_m)$, consider the rectangular subset $G_{\mathbf{x}}$ defined as

$$G_{\mathbf{X}} = [0, x_1) \times [0, x_2) \times \cdots \times [0, x_m),$$

which has a volume $x_1 x_2 \cdots x_m$. If we denote by $S_N(G)$ the function counting the number of points in the sequence, which are contained in a subset $G \subset I^m$, a possible definition of discrepancy is

$$D^*(\mathbf{x}^1, \ldots, \mathbf{x}^N) = \sup_{\mathbf{X} \in I^m} \mid S_N(G_{\mathbf{X}}) - N x_1 x_2 \cdots x_m \mid .$$

To be more precise, this is the concept of *star* discrepancy, as alternative definitions are also used. The role of discrepancy can be appreciated by referring to the Koksma-Hlawka bound (see, e.g., Niederreiter, 1992):

$$\left| \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{x}_i) - \int_{[0,1]^m} h(\mathbf{x}) \, d\mathbf{x} \right| \leq V(h) \cdot D^*(\mathbf{x}^1, \dots, \mathbf{x}^N),$$

where m is the dimension of the space over which we are integrating and $V(h)$ is a measure of the variation of the function h. This theorem yields an upper bound on the integration error we have if we use a set of sample points; the error is bounded by the product of a term linked to the difficulty of the function itself, the variation, and a term linked to discrepancy. Hence, a low value of discrepancy is desirable. We speak of low-discrepancy sequences when the discrepancy is something like $O(\ln N)^m/N$. This contribution to the integration error should be compared with the estimation error with Monte Carlo simulation which is of order $O(1/\sqrt{N})$, where N is the number of random samples.

Low-discrepancy sequences are sequences in the unit hypercube $[0,1]^m$, which is just what we need to simulate according to most probability distributions. We recall that, given a random variable X with cumulative distribution function $G_X(x) \equiv \mathbb{P}\{X \leq x\}$, we may generate samples of X by generating first a pseudo-random number U from the uniform distribution over the interval $(0,1)$, then we invert the distribution function and return $X = G_X^{-1}(U)$ (for alternative approaches, see, e.g., Law and Kelton, 1999).

In order to get a grasp of how low-discrepancy sequences work, we may consider the typical building block of such sequences, i.e., the Van der Corput sequence. This is a sequence in the unit interval $[0,1]$, and it based on a simple recipe:

- Representing an integer number n in a base b, where b is a prime number:
$$n = (\cdots d_4 d_3 d_2 d_1 d_0)_b$$
- Reflecting the digits and adding a radix point to obtain a number within the unit interval:
$$h = (0.d_0 d_1 d_2 d_3 d_4 \cdots)_b$$

More formally, if we represent an integer number n as

$$n = \sum_{k=0}^{\infty} d_k b^k,$$

where only a finite set of digits d_k is non-zero, the nth number in the sequence with base b is

$$h(n,b) = \sum_{k=0}^{\infty} d_k b^{-(k+1)}.$$

Here are the first numbers with base 2:

$$\begin{aligned}
n = 1 = (0001)_2 \quad &\Rightarrow \quad h(1,2) = (0.1000)_2 = 0.5000 \\
n = 2 = (0010)_2 \quad &\Rightarrow \quad h(2,2) = (0.0100)_2 = 0.2500 \\
n = 3 = (0011)_2 \quad &\Rightarrow \quad h(3,2) = (0.1100)_2 = 0.7500 \\
n = 4 = (0100)_2 \quad &\Rightarrow \quad h(4,2) = (0.0010)_2 = 0.1250 \\
n = 5 = (0101)_2 \quad &\Rightarrow \quad h(5,2) = (0.1010)_2 = 0.6250 \\
n = 6 = (0110)_2 \quad &\Rightarrow \quad h(6,2) = (0.0110)_2 = 0.3750 \\
n = 7 = (0111)_2 \quad &\Rightarrow \quad h(7,2) = (0.1110)_2 = 0.8750 \\
n = 8 = (1000)_2 \quad &\Rightarrow \quad h(8,2) = (0.0001)_2 = 0.0625 \\
n = 9 = (1001)_2 \quad &\Rightarrow \quad h(9,2) = (0.1001)_2 = 0.5625
\end{aligned}$$

We see that such numbers fill the unit interval in a nice way. Different m-dimensional low-discrepancy sequences have been proposed using Van der Corput sequences as building blocks. Halton's sequence in a space of dimension m is built by using Van der Corput sequences with a different base along each dimension; typically, the m bases are the first m prime numbers. More sophisticated approaches lead, e.g., to Faure and Sobol sequences (Glasserman, 2003).

8.3 Moment matching

Antithetic sampling, in the case of symmetric distributions, leads to a sample that matches odd moments of the underlying density; for instance, expected value is matched, and the symmetric sampling leads to zero skewness. It is natural to consider sampling in such a way that other moments are matched as well, such as variances, covariances, and kurtosis.

In general, matching all moments exactly will be impossible with a limited number of samples, but we can try to match them as well as possible, in a least squares sense, This leads to an approach to generate a set of “optimized” scenarios (Hoyland and Wallace, 2001), which is intuitively appealing, even though counterexamples have been provided showing that completely different distributions may match the first few moments (Hochreiter and Pflug, 2002).

To illustrate the approach, consider a demand which has a multivariate normal distribution within a time period. We know the expected values μ_i of demand for item i, as well as the variance σ_i^2 and the set of covariances σ_{ij} for each pair (i, j) of items $(\sigma_{ii} = \sigma_i^2)$. Furthermore, since we assume a normal distribution, we know that skewness $\xi = \mathrm{E}[(\tilde{d} - \mu)^3/\sigma^3]$ should be zero and that kurtosis $\chi = \mathrm{E}[(\tilde{d} - \mu)^4/\sigma^4]$ should be 3 (here we are considering the marginal distribution of the demand for each item).

If we also assume that demands in different time periods are independent, generating the scenario tree is relatively easy, since we must just get a sample of demands at each branching on the tree. We may also set the conditional probabilities of each node within a branching to equal values. In general, one may consider the probabilities of each branch as decision variables for the tree optimization model. Let us denote by d_i^s the demand for item i in node s belonging to a certain branching of size S. Natural requirements are:

$$
\begin{aligned}
&\frac{1}{S}\sum_s d_i^s \approx \mu_i \qquad \forall i \\
&\frac{1}{S}\sum_s (d_i^s - \mu_i)(d_j^s - \mu_j) \approx \sigma_{ij} \qquad \forall i, j \\
&\frac{1}{S}\sum_s \frac{(d_i^s - \mu_i)^3}{\sigma_i^3} \approx 0 \qquad \forall i \\
&\frac{1}{S}\sum_s \frac{(d_i^s - \mu_i)^4}{\sigma_i^4} \approx 3 \qquad \forall i.
\end{aligned}
$$

Note that we divide by S since the parameters are known a priori and not estimated from the data. Approximate moment matching is obtained by minimizing the following squared error:

$$
\begin{aligned}
&w_1 \sum_i \left[\frac{1}{S}\sum_s d_i^s - \mu_i\right]^2 \\
&+w_2 \sum_{i,j} \left[\frac{1}{S}\sum_s \left(d_i^s - \mu_i\right)\left(d_j^s - \mu_j\right) - \sigma_{ij}\right]^2 \\
&+w_3 \sum_i \left[\frac{1}{S}\sum_s \left(\frac{d_i^s - \mu_i}{\sigma_i}\right)^3\right]^2 \\
&+w_4 \sum_i \left[\frac{1}{S}\sum_s \left(\frac{d_i^s - \mu_i}{\sigma_i}\right)^4 - 3\right]^2 \qquad (3.40)
\end{aligned}
$$

over nonnegative demand values d_i^s. The objective function includes four weights w_k which may be used to fine tune performance.

It should be mentioned that the resulting scenario optimization problem need not be convex. However, if we manage to find any solution with a low value of the "error" objective function, this is arguably a satisfactory solution, even though it is not necessarily the globally optimal one (Hoyland and Wallace, 2001).

8.4 Optimal approximation of probability measures

The moment matching approach is a flexible and intuitively appealing way of generating scenarios. Nevertheless, it has been argued that it lacks a sound theoretical background. In order to find a scenario generation approach resting on a sound basis, some researchers have proposed formal approaches relying on stability concepts and the definition of probability metrics. These methods require a high level of mathematical sophistication; hence, in this introductory chapter, we limit ourselves to provide the reader with a basic feeling for the overall idea (see, e.g., Römisch, 2003, for a thorough treatment).

To begin with, we should try to formalize the concept of stability. To this aim, let us consider an abstract view of a stochastic optimization problem:

$$v(P) \equiv \inf_{\mathbf{x} \in \mathbf{X}} \int_{\Xi} f_0(\mathbf{x}, \xi) P(d\xi)$$

Here $\mathbf{x}$ is the set of decision variables, constrained on a set $\mathbf{X}$. The random data are represented by ξ, which belongs to set Ξ on which a probability measure P is defined. The optimal value of this stochastic program depends on the probability measure P, as pointed out by the notation $v(P)$. What happens if we perturb the measure P? A possible reason for the perturbation is that we have unreliable data, which means that we actually ignore the "true" measure P and we consider another measure Q instead. Alternatively, we may be forced to resort to an approximate measure Q, in the sense that we use a scenario tree which approximates the true measure P. Whatever the reason, we must first define a probability metric in order to quantify the distance between two probability measures.

There are many ways to do so. One possibility has its roots in the Monge transportation problem, which asks for the optimal way of transporting mass (e.g., soil, when we are building a road). The problem has a probabilistic interpretation, which was pointed out by Kantorovich, when we interpret mass in a probabilistic sense (see Rachev, 1991, for more details). In order to define a concept of distance between two probability measures, we may define a transportation functional:

$$\mu_c(P,Q) \equiv \inf \left\{ \int_{\Xi\times\Xi} c(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \pi_1\eta = P,\ \pi_2\eta = Q. \right\}$$

Here $c(\cdot,\cdot)$ is a suitably chosen cost function; the problem calls for finding the minimum of the integral over all joint measures η, defined on the Cartesian product $\Xi \times \Xi$, whose marginals coincide with P and Q, respectively (π_1 and π_2 represent projection operators). In the case of two discrete measures P and Q, this boils down to the classical transportation problem with a linear programming formulation. It can be shown, under some technical conditions, a form of Lipschitz continuity:

$$| v(P) - v(Q) | \leq L\mu_c(P,Q).$$

In practical terms, what one can do is selecting a cost function $c : \Xi\times\Xi \to \mathbb{R}$ in order to define a probability metric. Then we look for an approximate distribution P_{tree}, i.e., the scenario tree, such that $\mu_c(P, P_{tree}) < \epsilon$.

This leads to algorithms to reduce the scenario tree. In (Heitsch and Roemisch, 2003) a scenario reduction procedure is described, based on the theoretical concepts above. The idea is sampling a large tree, and then reducing its size to a manageable level.

9. Models for capacity planning

Capacity planning problems are clearly affected by uncertainty on demand; further uncertain factors, due to the strategic character of these problems, may be the cost of acquiring capacity, the selling price of end items, and the actual productivity of the plants. Hence, they are natural candidate for the application of stochastic programming methods. Indeed, papers applying stochastic programming for capacity planning have been published in diverse fields. For instance, capacity expansion in telecommunication networks is considered in (Sen et al., 1994). Applications to energy are described in (Wallace and Fleten, 2003). In this section we describe a few capacity planning models in the manufacturing context that have appeared in the literature.

Probably, the best known model in this vein has been described in (Eppen et al., 1989), where a problem faced by General Motors is considered. Here we do not attempt an exhaustive survey of the literature, which is likely to be quickly out-dated; rather, we outline a few illustrative examples in order to point out the general features of stochastic programming models that we have treated in the previous sections.

9.1 A two-stage model for capacity expansion

We begin with a basic two-stage model for capacity expansion. This model is a simplified version of a problem illustrated in (Higle and Sen, 1996); it is a benchmark problem in the stochastic programming literature, known as CEP1. The original version includes the effect of machine maintenance on available capacity. For the sake of simplicity we avoid modeling such issues.

We have a set of flexible machining centers indexed by $j \in J$. They are flexible in the sense that they are capable of producing a set of items $i \in I$, but at a different cost g_{ij} and at a different rate a_{ij}, possibly set to 0 if center j is not able to produce item i. Demand $d_i(\omega)$ is stochastic, and there is a penalty p_i for lost sales.

The currently available capacity of center j is R_j, expressed, e.g., in available hours per week, and we consider acquiring additional capacity x_j (again, hours per week). The cost of additional capacity must be expressed in a suitable way. In fact it is critical to understand the way we can express capacity planning in a two-stage framework. The two stages do not really correspond to time periods; also in multi-stage models, stages do not necessarily correspond to time periods (see the remark at the end of Section 6.1). The first stage corresponds to capacity acquisition which is carried out once, whereas the second stage is a sort of repeated experiment, once per week, corresponding to meeting demand at minimum cost with the expanded capacity. The idea of repeated experiments actually justifies the use of stochastic programming, but we cannot compare the cost of acquiring capacity for a possibly long time horizon with weekly penalties for not meeting demand. Hence, we should express capacity cost as an amortized cost c_j per week, per unit additional capacity. Also the budget B is expressed in consistent unit.

Keeping such observations in mind, the first-stage problem can be formulated as:

$$\min \quad \sum_{j \in J} c_j x_j + \mathrm{E}[h(\mathbf{x}, \omega)] \tag{3.41}$$

$$\text{s.t.} \quad \sum_{j \in J} c_j x_j \leq B \tag{3.42}$$

$$x_j \geq 0. \tag{3.43}$$

The second-stage model is:

$$h(\mathbf{x}, \omega) \equiv \min \quad \sum_{i \in I} \sum_{j \in J} g_{ij} y_{ij} + \sum_{i \in I} p_i z_i \tag{3.44}$$

$$\text{s.t.} \quad \sum_{j \in J} a_{ij} y_{ij} + z_i \geq d_i(\omega) \qquad \forall i \tag{3.45}$$

$$\sum_{i \in I} y_{ij} \leq R_j + x_j \qquad \forall j \tag{3.46}$$

$$y_{ij}, z_i \geq 0, \tag{3.47}$$

where y_{ij} expresses the number of hours (per week) dedicated on j to produce i, and z_i is the lost sale for item i.

This model is quite instructive in its separation between stages and time periods. It assumes continuous decision variables, which is debatable for capacity that can be only acquired in discrete units, but an integer programming version of the problem is easy to write down, though not so easy to solve. A bit more critical is the fact that the model implicitly assumes independently and identically distributed demand for all items. If trends or temporal patterns in product sales need to be modeled, we must resort to multi-stage models.

9.2 A multi-stage model for capacity planning

We outline here a model proposed in (Chen et al., 2002), with the aim of gathering managerial insight into the tradeoffs between flexible and dedicated capacity. Multi-stage models are needed when demand for items in different time periods is characterized by some form of dependency. This is the case, e.g., when product sales are characterized by cycles, from introduction, through maturity, to decline. The point is that sales may have two sources of uncertainty: one is linked, e.g., to product success, and the other one may be characterized by noise. Clearly, the two-stage model of the previous section is only able to cope with the second one.

The authors consider N products, indexed by $i = 1, \ldots, N$ and $N+1$ technologies. By convention, technology 0 is flexible, and the remaining technologies are dedicated and correspond to one product each. The planning horizon has length T and S scenarios are modeled by a tree; each scenario s has probability π^s.

In the paper a split-variable formulation is proposed, which is then solved by a form of Lagrangian decomposition. One advantage of this computational approach is that it may be easier to deal with non-linear costs, whereas compact formulations may be better suited to exploit large-scale linear programming methods. So, stochastic demand is modeled by a data process d^s_{it}, and non-anticipativity constraints are explicit.

Available capacity at the beginning of the planning period is C_{i0}, $i = 0, 1, \ldots, N$. The decision variables are related to capacity acquisitions and capacity allocations:

- X_{it}^s is the amount of capacity of type $i = 0, 1, \ldots, N$ added in time period t, in scenario s; this includes both flexible and dedicated technology;
- Y_{it}^s is the amount of dedicated technology of type $i = 1, \ldots, N$ dedicated to product i, at cost U_{it}; using a time-dependent cost allows for discounting;
- Z_{it}^s is the amount of flexible technology dedicated to product $i = 1, \ldots, N$, at cost V_{it}.

As to the cost of adding capacity, the authors model economies of scale and discounting by considering cost functions $F_{it}(X_{it}^s)$.

This model includes inventories; clearly, this would not make sense for the two-stage formulation we have just considered. Initial inventory is denoted by I_{i0}; inventories then are denoted by I_{it}^s; inventory holding cost is H_{it}. Considering inventories or not is a modeling choice which depends on the type of manufacturing environment. One observation is that even in a make-to-stock environment, modeling inventories may be debatable when stages correspond to large time periods. Another point worth mentioning, is that the model below has not full recourse, in the sense that it does not allow for lost sales; if extreme scenarios are included, the model may yield a very costly solution.

The resulting model is:

$$\min \quad \sum_{s=1}^{S} \pi^s \left\{ \left[\sum_{i=0}^{N} \sum_{t=1}^{T} F_{it}(X_{it}^s) \right] + \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (U_{it} Y_{it}^s + V_{it} Z_{it}^s) \right] + \left[\sum_{i=1}^{N} \sum_{t=1}^{T} H_{it} I_{it}^s \right] \right\} \tag{3.48}$$

$$\text{s.t.} \quad Y_{it}^s \leq C_{i0} + \sum_{\tau=1}^{t} X_{i\tau}^s \qquad i = 1, \ldots, N, \ \forall t, s \tag{3.49}$$

$$\sum_{i=1}^{N} Z_{it}^s \leq C_{00} + \sum_{\tau=1}^{t} X_{0t}^s \qquad \forall t, s \tag{3.50}$$

$$I_{i0} + Y_{i1}^s + Z_{i1}^s = d_{i1}^s + I_{i1}^s \qquad \forall i, s \tag{3.51}$$

$$I_{i,t-1} + Y_{it}^s + Z_{it}^s = d_{it}^s + I_{it}^s \tag{3.52}$$

$$i = 1, \ldots, N, \ t = 2, \ldots, T, \ \forall s \tag{3.53}$$

$$\text{non} - \text{anticipativity constraints} \tag{3.54}$$

$$X_{it}^s, Y_{it}^s, Z_{it}^s, I_{it}^s \geq 0. \tag{3.55}$$

The objective function (3.48) is the expected value of future costs due to capacity acquisition, capacity allocation, and inventory holding. Constraint (3.49) states that the amount of dedicated capacity allocated to product i in time period t cannot exceed what was available at the beginning, plus what we have acquired so far; constraint (3.50) states a similar requirement for the allocation of flexible capacity. Equations (3.52) and (3.53) are customary inventory equilibrium constraints. We see that in this model formulation, apparently, the production rate is the same for all items; actually this is just a matter of scale, since we may assume that inventory is expressed in units of consumed capacity, and making production rates explicit is easy. The non-anticipativity constraints (3.54) are similar to constraints (3.24), (3.25), and (3.26). We do not state them explicitly here because in the original paper the authors introduce a slightly more complicated information flow, such that nodes and arcs in the scenario tree must be distinguished; since this is level detail is not needed in this introductory survey, the reader is referred to the original reference for further information. Finally, from non-negativity constraints (3.55) we see that the model is based on continuous variables. In some circumstances, integer variables should be used for capacity acquisition, but if the aim of the model is just gaining managerial insights from suitable computational experiments, the additional complication of integer programming is not necessarily warranted.

9.3 A multi-stage integer programming model for capacity planning

In the last sections we have considered stochastic linear programming models for capacity planning. These models assume continuous decision variables; however, in capacity planning, integer variables may arise. One obvious reason is that some capacities may only be acquired in discrete quantities, and a model of this type has been described in (Barahona et al., 2001) for capacity planning in semiconductor manufacturing.

Another reason is that fixed costs may be introduced in order to represent, possibly in an approximate way, economies of scale, and this calls for the introduction of binary variables. A model in this vein has been proposed in (Ahmed et al., 2003). We outline their model here, which is also interesting in that it allows a reformulation in terms of a lot sizing problem.

Unlike the model of Section 9.2, this multi-stage model has been proposed in compact form; let us denote by $n \in \mathcal{N}$ a node of the scenario tree. The set of predecessors, in different time periods, of node n is denoted by $\mathcal{P}(n)$. The model considers the production of one (possibly aggregate) item, using different resource types $i \in \mathcal{I}$. Let d^n be the demand of the item in node n. By acquiring a resource of type i in node n we incur a variable cost α_{in} and a fixed cost β_{in}; the dependence of costs on the node allows not only to account for discounting, but also for uncertainty in capacity acquisition costs. Using decision variables $x_{in} \geq 0$ to denote the amount of resource i acquired in node n and $y_{in} \in \{0,1\}$ to denote the decision of acquiring additional capacity of type i, the model can be formulated as:

$$\min \quad \sum_{n\in\mathcal{N}} \pi^n \left\{ \sum_{i\in\mathcal{I}} (\alpha_{in}x_{in} + \beta_{in}y_{in}) \right\} \tag{3.56}$$

$$\text{s.t.} \quad 0 \leq x_{in} \leq M_{in}y_{in} \qquad \forall n, i \tag{3.57}$$

$$\sum_{m\in\mathcal{P}(n)} \sum_{i\in\mathcal{I}} x_{im} \geq d_n \qquad \forall n \tag{3.58}$$

$$y_{in} \in \{0,1\}. \tag{3.59}$$

This model uses the big-M modeling approach to account for fixed costs. We have already argued in Section 7 that this results in weak model formulations.

A very interesting fact is that the above model, for the case of a single resource, $|\mathcal{I}| = 1$, can be shown to be equivalent to a single-item lot-sizing model. Then, using a plant location formulation in the same vein as what we have described in Section 7, the authors are able to find a strong reformulation of their problem, paving the way for efficient solution algorithms.

10. An alternative approach to cope with uncertainty: robust optimization

All the stochastic programming models we have seen so far share some common features which may be rather questionable:

1 they assume risk neutrality, since they are aimed at optimizing the expected value of cost or profit; in principle, different attitudes towards risk may be modeled by introducing a concave utility function; apart from the increased computational derived by a nonlinear objective, finding the right utility function for a decision maker is not trivial;

2 they rely on some probability distribution, which is arguably unknown; this criticism may be somehow mitigated if we do not interpret scenarios in a strictly probabilistic sense;

3 since an expected value is optimized, stochastic programming models implicitly assume a sequence of repeated experiments; clearly, in a capacity planning model this assumption is not realistic.

In order to address these issues, different ways of formulating optimization models under uncertainty have been proposed in the literature. Broadly speaking, such models may be put under the label of *robust* optimization models, but the term is not so standard. In (Mulvey et al., 1995) a general framework is proposed, which includes many approaches as specific cases. Robustness issues are related to both solution optimality and feasibility; the first issue is called "solution robustness", whereas the term "model robustness" is used for the second one. If we assume that lost sales are suitably penalized, feasibility issues may be neglected. In order to help the reader in finding her way through this class of models, we outline a few modeling approaches that have been proposed to address the limitations of basic stochastic programming models.

One approach is based on the consideration of higher moments of the objective function. An apparently straightforward way to do so is to consider variance as well. Referring to the two-stage formulation of Section 4, and assuming a finite set of S scenarios, we may formulate the following model:

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} + \psi\left(h(\mathbf{x}, \omega_1), \ldots, h(\mathbf{x}, \omega_S)\right) \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{array}$$

where the function $\psi(\mathbf{z})$ is defined by

$$\psi(\mathbf{z}) \equiv \sum_{s=1}^{S} \pi^s z_s + \alpha \left[\sum_{s=1}^{S} \pi^s z_s^2 - \left(\sum_{s=1}^{S} \pi^s z_s \right)^2 \right].$$

The function ψ includes both expected value and variance, and for different choices of the coefficient α it may take into account different risk aversion levels on the part of the decision maker. This approach is clearly influenced by the theory of mean-variance efficient portfolios in finance. Actually, this model formulation does have some pitfalls, since it does not necessarily lead to a convex programming problem. However alternative formulations in a similar vein yield convex optimization problems (Ruszczyński and Shapiro, 2003). Models of this type are called *mean-risk* models, as they trade off a mean value against a risk measure. Variance is not the only nor the best way to express risk. Considerable research work is being carried out on suitable functions measuring risk (Artzner et al., 1999).

Mean-risk models rely on a probabilistic representation of uncertainty. Since the exact knowledge of the underlying distribution is a rare commodity, another line of research deals with the *min-max* approach to stochastic optimization. The idea is that we should find a solution which is robust in a worst-case sense. For instance, suppose that we have uncertainty about the probability distribution of the problem data, and that this is expressed by dealing not with a single distribution but with a family of probability measures $\mathcal{S}$. We would like to find solution which is satisfactory under all of the measures $P \in \mathcal{S}$. This may be expressed, in abstract terms, as the following optimization problem:

$$\min_{\mathbf{x}\in\mathbf{X}} \max_{P\in\mathcal{S}} \mathrm{E}_P \left[f(\mathbf{x}, \omega) \right].$$

This statement clarifies the use of min-max; see, e.g., (Shapiro and Kleywegt, 2002) for an analysis of this type of approach.

A somewhat more radical approach to robust optimization is based on the idea of disregarding any probabilistic interpretation of scenarios, or even to associate plain intervals to uncertain data. The last possibility is appealing, though not trivial from a computational point of view; see, e.g., (Mausser and Laguna, 1999) for an example in which a heuristic approach is proposed for LP problems with interval uncertainty on objective function coefficients. A well selected set of scenarios is arguably a more flexible approach, which is also able to capture some dependence in the data, which the decision maker may be aware of, even though it cannot be captured in the form of a joint probability distribution.

A reasonable objective is to optimize performance in the worst case. Following (Kouvelis and Yu, 1996), we may formalize different min-max objectives. Let us denote by D^s the set of data associated to scenario $s \in S$, and by F^s the set of feasible solutions for those data (assuming that also feasibility is an issue, and not only optimality). The optimal solution for scenario s would be $\mathbf{x}_s^*$ with cost:

$$z^s = f(\mathbf{x}_s^*, D^s) = \min_{\mathbf{x}\in F^s} f(\mathbf{x}, D^s).$$

The *absolute robust decision* $\mathbf{x}^A$ is the solution of the min-max problem:

$$z_A = \min_{\mathbf{x}\in\cap_{s\in S} F^s} \max_{s\in S} f(\mathbf{x}, D^s).$$

Clearly, the output of such an optimization problem will tend to be rather conservative. It is worth mentioning that sometimes the min-max objective may be associated to the concept of *regret*: this means that, after discovering the real value of the problem data, one can find a posteriori what the optimal solution should have been. We can measure a form of regret by comparing the cost of the selected (a priori) solution and the cost of the (a posteriori) optimal solution. This idea is formalized in (Kouvelis and Yu, 1996) as the *robust deviation decision* $\mathbf{x}_D$ solving:

$$z_D = \min_{\mathbf{x} \in \cap_{s \in S} F^s} \max_{s \in S} \left\{ f(\mathbf{x}, D^s) - f(\mathbf{x}_s^*, D^s) \right\}.$$

Minimizing the maximum (a priori) regret makes clear managerial sense, if one thinks of being evaluated by a panel of top managers a posteriori. One such model for capacity expansion is proposed in (Laguna, 1998).

11. Conclusions

In this chapter we have given a tutorial introduction to stochastic programming models with recourse as a valuable tool for planning under uncertainty. We have considered production planning and capacity planning models under demand uncertainty. As we have seen, there is a strong relationship between the two problems in terms of model formulation; furthermore, techniques for generating scenarios may be similar, although in capacity planning the longer time horizon calls for addressing a larger degree of uncertainty.

Due to the dynamic nature of the resulting decision process, one should wonder about the relationship between stochastic programming and stochastic optimal control. Indeed, the concept of recourse function looks quite similar to the concept of value function or cost-to-go in dynamic programming (Bertsekas, 2001). We should also note that nested Benders decomposition, which is the multi-stage generalization of what we have illustrated in Section 5, is based on a mechanism whereby primal variables are passed forwards and dual variables proceed backwards, much in the same vein as classical optimal control problems yielding a two-boundary value problem relating state and co-state variables (Kamien and Schwartz, 1991). While the two approaches are clearly related, they are actually complementary.

- Dynamic programming approaches require finding the value function, as a function of state variables, for each decision stage. Stochastic programming methods based on L-shaped decomposition aim at finding only a *local* approximation of the recourse function.

- Dynamic programming methods, after computing the value functions, allow for a simulation of the whole decision process over the planning horizon. Stochastic programming methods aim at finding the solution for the first stage only, even though in principle further stage decision variables represent a feedback policy. In this sense stochastic programming is a more operational approach. Indeed, the use of dynamic programming models is the rule whenever one wants to use an optimization model to gain insights in a problem, rather than actually solving it. This is quite common in economics (see, e.g., Campbell and Viceira, 2002, for an application to strategic asset management).

- Dynamic programming methods are able to cope with infinite-horizon problems, whereas stochastic programming methods are not. Again, this is typical of dynamic models in economics.
- Dynamic programming models, in some cases, may be solved analytically, maybe approximately. This is useful to gain insights, whereas stochastic programming approaches are numerical in nature; nevertheless, managerial insights may be gained by proper numerical experimentation as well.
- Dynamic programming models assume some condition on the underlying uncertainty, since the disturbance process should be Markovian (actually, often one can get around the difficulty by augmenting the set of state variables). In principle, any type of uncertainty and any type of intertemporal dependence can be tackled by stochastic programming, provided we are able to generate a scenario tree (this is important in capacity planning, see Chen et al., 2002).

We have also hinted on the related topic of robust optimization. Again, stochastic programming has both advantages and disadvantages with respect to this alternative, but the two fields are related.

As a final note, we have essentially ignored the financial side of capacity planning problems. We have assumed some implicit form of discounting in objective functions, but finding the proper way to discount while taking risk into account is not trivial. We should mention here that an alternative approach, inspired by financial engineering, is represented by real options. Since valuing real options is often an exercise in dynamic optimization, it should come as no surprise that real options and stochastic programming may be linked to together, as illustrated by (Birge, 2000) for capacity planning.

We have seen that stochastic and robust programming are a flexible and powerful modeling tool. The resulting models may be rather hard to solve, but with proper model formulations and scenario generation strategies, a good near-optimal solution can actually be found. Since capacity planning problems need not be solved in real time, this may be a computationally feasible and valuable approach.

References

Ahmed, S., King, A.J., and Parija, G. (2003). A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization*, **26**: 3-24.

Alfieri, A. and Brandimarte, P. (1999). Production planning in an assembly-to-order environment by two-stage stochastic linear programming. *Proceedings of the IV Aitem Conference*, Brescia, 151-158.

Alfieri, A., Brandimarte, P., and d'Orazio, S. (2002). LP-based heuristics for the capacitated lot-sizing problem: the interaction of model formulation and solution algorithm. *International Journal of Production Research*, **40**: 441-458.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**: 203-228.

Bakir, M.A. and Byrne, M.D. (1998). Stochastic linear optimisation of an MPMP production planning model. *International Journal of Production Economics*, **55**: 87-96.

Barahona, F., Bermon, S., Günlük, O., and Hood, S., Robust capacity planning in semiconductor manufacturing. IBM T.J. Watson Research Center, report RC22196 (also downloadable from `http://www.optimization-online.org`).

Bertsekas, D.P. (2001). *Dynamic programming and optimal control*. 2nd ed. Cambridge, MA: Athena Scientific.

Birge, J.R. (1997). Stochastic programming computation and applications. *INFORMS Journal of Computing*, **9**: 111-133.

Birge, J.R. (2000). Option methods for incorporating risk into linear capacity planning models. *Manufacturing and Service Operations Management*, **2**: 19-31.

Birge, J.R. and Louveaux, F. (1997). *Introduction to stochastic programming*. Berlin: Springer-Verlag.

Brandimarte, P. (2001). *Numerical methods in finance: a MATLAB-based introduction*. New York: Wiley.

Brandimarte, P. (2004). Multi-item capacitated lot-sizing with demand uncertainty. Submitted for publication; paper available from the author upon request.

Brandimarte, P., and Villa, A. (1995). *Advanced models for manufacturing systems management*. Boca Raton, FL: CRC Press.

Campbell, J.Y. and Viceira, L.M. (2002). *Strategic asset allocation*. Oxford: Oxford University Press.

Chen, Z.-L., Li, S., and Tirupati, D. (2002). A scenario-based stochastic programming approach for technology and capacity planning. *Computers and Operations Research*, **29**: 781-806.

Dantzig, G.B. (1955). Linear programming under uncertainty. *Management Science*, **1**: 197-206.

Dempster, M.A.H. and Thompson, R.T. (1999). EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. *Annals of Operations Research*, **90**: 161-184.

Dupačová, J. (1990). Stability and sensitivity analysis for stochastic programming. *Annals of Operations Research*, **27**: 115-142.

Eppen, G.D., Martin, R.K., and Schrage, L. (1989). A scenario approach to capacity planning. *Operations Research*, **37**: 517-527.

Fisher, M. and Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, **44**: 87-99.

Gerchack, Y. and Henig, M. (1986). An inventory problem with component commonality. *Operations Research Letters*, **5**: 157-160.

Glasserman, P. (2003) *Monte Carlo methods in financial engineering*. New York: Springer-Verlag.

Guan, Y., Ahmed, S., and Nemhauser, G.L. (2004). A branch-and-cut algorithm for the stochastic uncapacitated lot-sizing problem. Technical Report, School of Industrial & Systems Engineering, Georgia Institute of Technology (paper downloadable from `http://www.optimization-online.org`).

Haugen, K.K., Løkketangen, A., and Woodruff, D.L. (2001). Progressive hedging as a meta-heuristic applied to stochastic lot-sizing. *European Journal of Operational Research*, **132**: 116-122.

Heitsch, H. and Roemisch, W. (2003). Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, **24**: 187-206.

Higle, J.L. and Sen., S. (1996). *Stochastic decomposition*. Dordrecht: Kluwer Academic Publishers.

Hochreiter, R. and Pflug, G.Ch. (2002). Scenario tree generation as a multidimensional facility location problem. Aurora Technical Report, University of Wien (paper downloadable from `http://www.vcpc.univie.ac.at/aurora/publications/`).

Hopp, W. and Spearman, M. (2000). *Factory physics*. 2nd ed. New York: McGraw-Hill.

Hoyland, K. and Wallace, S.W. (2001). Generating scenario trees for multistage decision problems. *Management Science*, **47**: 296-307.

Infanger, G. (1998). *Planning under uncertainty: solving large-scale stochastic linear programs*. Danvers, MA: Boyd & Fraser.

Jönsson, E., Jörnsten, K., and Silver, E.A. (1993). Application of the scenario aggregation approach to a two-stage, stochastic, common component inventory problem with a budget constraint. *European Journal of Operational Research*, **68**: 196-211.

Kall, P. and Wallace, S.W. (1994). *Stochastic programming*. Chichester: Wiley.

Kamien, M.I. and Schwartz, N.L. (1991). *Dynamic optimization*. 2nd ed. Amsterdam: North Holland.

Kincaid, D. and Cheney, W. (2001). *Numerical analysis: mathematics of scientific computing*. Pacific Grove, CA: Brooks Cole.

Kouvelis, P. and Yu, G. (1996). *Robust discrete optimization and its applications*. Dordrecht: Kluwer Academic Publishers.

Krarup, J. and Bilde, O. (1977). Plant location, set covering, and economic lot size: an $O(mn)$ algorithm for structured problems. In: Collatz et al. (eds.), *Optimierung bei Graphentheoretischen und Ganzzahligen Problemen*, Numerische Methoden bei Optimierungsaufgaben, Band 3: 155-180. International Series of Numerical Mathematics, Birkhauser-Verlag.

Laguna, M. (1998). Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. *Management Science*, **44**: S101-S110.

Law, A.M. and Kelton, D.W. (1999). *Simulation modeling and analysis*. 3rd ed. New York: McGraw-Hill.

Lulli, G., and Sen., S. (2002). Stochastic batch-sizing: models and algorithms In: Woodruff, D.L. (ed.), *Network interdiction and stochastic integer programming*. Boston: Kluwer Academic Publishers.

Mausser, H.E. and Laguna, M. (1999). A heuristic to minimax absolute regret for linear programs with interval objective function coefficients. *European Journal of Operational Research*, **117**: 157-174.

Miranda, M.J. and Fackler, P.L. (2002). *Applied computational economics and finance*. Cambridge, MA: MIT Press.

Mulvey, J.M., Vanderbei, R.J., and Zenios, S.A. (1995). Robust optimization of large-scale systems. *Operations Research*, **43**: 264-281.

Nahmias, S. (2000). *Production and operations analysis*. 4th ed. New York: McGraw-Hill/Irwin.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Philadelphia, PA: Society for Industrial & Applied Mathematics.

Parija, G.R., Ahmed, S., and King, A.J. (2004). On bridging the gap between stochastic integer programming and MIP solver technologies. *INFORMS Journal on Computing*, **16**: 73-83.

Prékopa, A. (2003). Probabilistic programming. In: Ruszczynski, A. and Shapiro, A., (eds.), *Stochastic programming.* Vol. 10 of Handbooks in Operations Research and Management Science, (Amsterdam: Elsevier).

Rachev, S.T. (1991). *Probability metrics and the stability of stochastic models.* Chichester: Wiley.

Römisch, W. (2003). Stability of stochastic programming problems. In: Ruszczynski A., and Shapiro, A., (eds.), *Stochastic programming.* Vol. 10 of Handbooks in Operations Research and Management Science. Amsterdam: Elsevier.

Ruszczyński A. and Shapiro, A. (2003). Stochastic programming models. In: Ruszczynski A., and Shapiro, A., (eds.), *Stochastic programming.* Vol. 10 of Handbooks in Operations Research and Management Science. Amsterdam: Elsevier.

Sen, S. (2003). Algorithms for stochastic mixed-integer programming models. Preprint, University of Arizona, Tucson (downloadable from `http://tucson.sie.arizona.edu/MORE/papers.html`).

Sen, S., Doverspike, R.D., and Cosares, S. (1994). Network planning with random demand. *Telecommunications Systems*, **3**: 11-30.

Sen, S. and Higle, J.L. (1999). An introductory tutorial on stochastic programming models. *Interfaces*, **29**: 33-61.

Shapiro, A. and Kleywegt, A.J. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, **17**: 523-542.

Sox, C.R., Jackson, P.L., Bowman, A. and Muckstadt J.A. (1999). A review of the stochastic lot scheduling problem. *International Journal of Production Economics*, **62**: 181-200.

Van Slyke, R. and Wets, R.J-B. (1969). L-shaped linear programs with application to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, **17**: 638-663.

Wallace, S.W. (2000). Decision making under uncertainty: is sensitivity analysis of any use? *Operation Research*, **48**: 20-25.

Wallace, S.W., and Fleten, S.-E. (2003). Stochastic programming models in energy. In: Ruszczynski A., and Shapiro, A., (eds.), *Stochastic programming.* Vol. 10 of Handbooks in Operations Research and Management Science. Amsterdam: Elsevier.

Wolsey, L.A. (1998). *Integer programming.* Chichester: Wiley.

Chapter 4

CONFIGURATION OF AMSS

Stochastic models for performance evaluation of Automated Manufacturing Systems

A. Matta, Q. Semeraro, T. Tolio
Dipartimento di Meccanica
Politecnico di Milano
via Bonardi 9, 20133, Milano, Italy
andrea.matta@polimi.it, quirico.semeraro@polimi.it, tullio.tolio@polimi.it

Abstract This chapter proposes an integrated approach for supporting firms in the phase of detailed dimensioning of Advanced Manufacturing Systems. The problem is closely related to the performance evaluation of the system since discriminating indicators are necessary to rank different alternatives. Approximate stochastic models are proposed to estimate the performance of Advanced Manufacturing Systems. The accuracy of the proposed models is shown by applying them to several test and real cases.

Keywords: configuration; queuing networks; decomposition; FMS; flow lines

1. Introduction

The detailed configuration of manufacturing systems is a difficult task that heavily affects the strategic capacity decision that many firms have to make every time they acquire a new system or modify an existing one. Regardless of the cause, firms have to solve the challenging problem of selecting the resources that best fit their current and future needs. This is a very critical phase since each decision made at this level will directly affect the performance of the new system and therefore its profitability over the following years. In addition most of information available is not detailed and is often also uncertain. In particular uncertainty of the market demand must be considered during the configuration of the production system since unexpected variations of the volumes required by the market, or the introduction of new products, can make the solution unsuitable to fulfil the market requests. These types of uncertainty have been faced in Chapters 2 and 3. At the same time the decision must consider many system variables such as the number of machines, fixtures, carriers and tools and moreover dependencies among system variables are often unknown and are not easy to evaluate. In other terms the complexity of the problem is very large and hierarchical approaches are often used to deal with it.

In this chapter we provide models and tools for configuring a set of alternative manufacturing systems that can be adopted by the firm to manufacture the products in the planning horizon. The detail level of the analysis is deeper than that of the previous chapters since the single components of the system such as machines, buffers, carriers, etc. are considered for the first time in this book. How the final choice of manufacturing systems is made is described in Chapter 5.

This chapter is organized as follows. The configuration problem is described in Section 2 pointing out the links with the upstream and downstream modules as described in the main capacity problem framework presented in Chapter 1. Then a description of the AMSs considered in this book follows, i.e. Flexible Manufacturing Systems and Dedicated Manufacturing Systems. A solution methodology for designing alternative manufacturing systems is proposed in Section 4. Furthermore, since the efficiency of configuration techniques is mainly based on the speed and accuracy of models used to evaluate the performance of potential acquirable manufacturing systems, stochastic models for performance evaluation of AMSs are described in Sections 5 and 6.

2. Problem description

According to the whole capacity problem framework presented in Chapter 1 and shown in Figure 1.5, the objective of the activity A3 *Identifying AMS alternatives* is to define the set of feasible alternatives that the firm can decide to adopt for facing the future customer demand in the planning horizon. The activity is described in this section and the models are presented in the remainder of the chapter.

Figure 4.1 shows again inputs, controls, outputs and mechanisms of the activity A3.

The inputs of activity A3 are:

- [A3]-I1 **Potential product mix**: detailed technological information on the potential set of products. This information arrives, with more details, from activity A1 and contains:

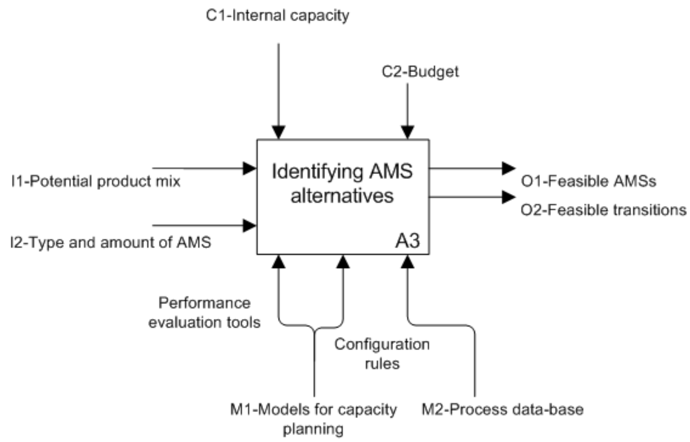


Figure 4.1. A3 context diagram.

- product codes
- the number of manufacturing operations for each product
- estimated processing times for each operation.

- [A3]-I2 **Types and amount of AMSs**: set of manufacturing systems to configure in all their details. This is the same information that arrives from activity A1 to A2 and contains:
  - system code, denoted by s
  - type of the system: FMS or DMS
  - product codes the system s must manufacture, denoted by the set $\Psi(s)$.

Activity A3 is controlled by:

- [A3]-C1 **Internal capacity**: the estimated needed internal capacity range expressed in number of pieces for each product to be manufactured on system s. The capacity range depends on the specific time period t in the planning horizon. This information arrives from activity A2 and contains:
  - minimum number of pieces of product type i the firm must be capable to produce internally on the shop floor on system s at period t. This quantity is denoted by $l_{i,s}(t)$ and should be respected in order to avoid demand surplus.
  - maximum number of pieces of product type i the firm must be capable to produce internally the shop floor on system s

at period t. This quantity is denoted by $u_{i,s}(t)$ and should be respected in order to avoid capacity wastes.

- [A3]-C2 **Budget**: the budget constraint inserted from decision-makers and depending on the specific time period t in the planning horizon. This quantity is denoted by $B(t)$.

The main output of activity A3 is a graph representing the set of alternative AMSs potentially adoptable in each specific time period t in the planning horizon (see an example in Figure 4.2). In detail, the outputs of activity A3 are:

- [A3]-O1 **Feasible AMSs**: feasible configurations to adopt in the planning horizon. The information on these configurations (i.e. the nodes of the graph) is very detailed because it specifies the type of systems with all their resources such as machine tools, buffers, part carriers, tool carriers, fixtures, load/unload stations. Notice that the cost to acquire the above described resources is also given in order to allow the calculation of the total investment cost associated to a particular system configuration. This information is used by activity A4.
- [A3]-O2 **Feasible transitions**: allowable changes in configurations (i.e. the arcs of the graph) that can be introduced by the firm in the future. Cost and times to implement transitions on configurations are also provided as outputs of the activity. This information is used by activity A4.

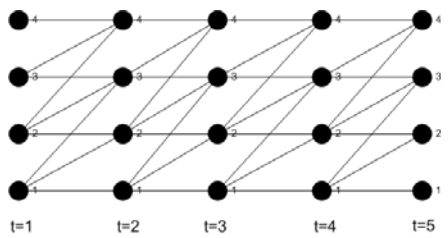


Figure 4.2. Example of the output of A3.

An illustrative example of the graph provided by the activity is shown in Figure 4.2. In this example for each time period t (with $t = 1, \ldots, 5$) four different alternatives are identified with an increasing number of machines from nodes $n_1(t)$ to nodes $n_4(t)$. Assume also, for simplicity's sake, that a node is identified by the number of machines: one machine for nodes $n_1(t)$, two machines for nodes $n_2(t)$ and so forth. The company must select for each time period the configuration of the AMS, i.e. one single node for each time period, trying to minimize the total cost overall the whole planning horizon. A further assumption is that a system can be gradually expanded by acquiring one or two machines maximum; for instance in the graph the nodes $n_1(t)$ of time period t are linked only with nodes $n_2(t+1)$ and $n_3(t+1)$ of time period $t+1$. Further, if we assume that machines cannot be sold we see in the graph that a system cannot be reduced during its time evolution. An expansion of the system involves an additional investment cost modelled by the weight of the edges between nodes $n_k(t)$ and $n_{k+1}(t)$ or $n_{k+2}(t)$. How to select the optimal path in this graph is explained in Chapter 5.

Mechanisms used by activity A3 are essentially:

- [A3]-M1 **Performance evaluation tools**: approximate analytical methods are used to evaluate the performance of manufacturing systems. In particular, simple and static equations model approximately the behavior of AMSs in a preliminary analysis, and then queuing theory is used to evaluate dynamically the behavior of manufacturing systems.

- [A3]-M2 **Configuration rules**: set of technological rules that allows the proper selection of system devices coherently with the manufacturing operations of potential products.

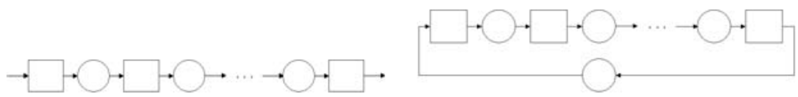
- [A3]-M3 **Process & system database**: detailed information on system devices: speed of machines, working cube, feed rates, etc.

3. Description of Automated Manufacturing Systems

3.1 Dedicated Manufacturing Flow Lines

Description. A dedicated manufacturing flow line is defined in literature as a serial production system in which pieces are worked sequentially by rigid machines: pieces flow from the first machine, in which they are still raw parts, to the last machine where the process cycle is completed and finished parts leave the system. Equipment is rigid and can be used to operate only on the product, or a limited family of products, for which the dedicated system has been designed. For simplicity of exposition we omit the term *dedicated* in the remainder of this section.

When a machine is not available to start a processing operation, pieces wait in its buffer located immediately upstream the machine. If the number of parts flowing in the system is constant during the production, these systems are also called closed flow lines (see Figure 4.3 where rectangles and circles represent machines and buffers of the system respectively) to distinguish them from open flow lines where the number of parts is not maintained constant. See Gershwin (Gershwin, 1994) for a complete description of flow lines in manufacturing. The production rate of manufacturing flow lines is clearly a function of speed and reliability of machines: the faster and more reliable the machines are the higher the production rate is. However, since machines can have different speeds and may be affected by random failures, the part flow can be interrupted at a certain point of the system causing blocking and starvation of machines. In particular, blocking occurs when at least one machine cannot move the parts just machined (BAS, Blocking After Service) or still to work (BBS, Blocking Before Service) to the next station. In open flow lines blocking of a machine can be caused only by a long processing time or a failure of a downstream machine. Analogously, starvation occurs when one or more machines cannot be operational because they have no input part to work; in this case the machine cannot work and it is said to be starved. In open production lines the starvation of a machine can be caused only by a long processing time or a failure of an upstream machine.



(a) Open (b) Closed.

Figure 4.3. Flow lines.

Therefore, buffer capacity affects the system behavior since the part flow interrupts when a buffer is empty (starvation) or full (blocking). If there is no area in which to store pieces between two adjacent machines, the behavior of machines is strongly correlated. Indeed, in flow lines the state of a machine affects the rest of the system because of blocking and starvation phenomena that propagate upstream and downstream respectively the source of flow interruption in the line.

In order to decrease blocking and starvation phenomena in flow lines, buffers between two adjacent machines are normally included to decouple the machines behavior. Indeed, buffers allow the absorption of the impact of a failure or a long processing time because (a) the presence of parts in buffers decreases the starvation of machines and (b) the possibility of storing parts in buffers decreases the blocking of machines. Therefore, the production rate of flow lines is also a function of buffer capacities; more precisely, production rate is a monotone positive function of the total buffer capacity of the system. See (Gershwin and Schor,

2000) for a list of works focused on the properties of production rate in flow lines.

Stochastic models. There are several factors which motivate the use of stochastic modelling for DMFLs. The machine breakdowns deeply affect the performance of the system since it may cause the interruption of flow. Processing times at a machine, load/unload of parts, tool breakdowns, scraps, etc. are other factors that can be taken into account by stochastic models.
There is a vast literature about stochastic models for performance evaluation of transfer lines. The most interesting technique is the decomposition introduced by Gershwin in (Gershwin, 1987) which, based on a Markov representation of the system, makes possible the analysis of long lines by means of two-machine lines simpler to study. A complete state of the art until 1992 is in the survey of Dallery and Gershwin on performance evaluation of flow lines (Dallery and Gershwin, 1992). Actually decomposition techniques have been developed to analyze:

- one product-open flow lines with deterministic processing times and discrete part flow. Refer to (Gershwin, 1994; Matta and Tolio, 1998; Dallery et al., 1988).

- one product-open flow lines with stochastic processing times and discrete part flow. Refer to (Gershwin, 1994; Levantesi et al., 2003a).

- one product-open flow lines with deterministic processing times and continuous part flow. Refer to (Gershwin, 1994; Levantesi et al., 2003b; Le Bihan and Dallery, 1997).

- one product-closed flow lines with deterministic processing times and discrete part flow. Refer to (Frein et al., 1996; Gershwin et al., 2001; Levantesi, 2002).

- two products-open flow lines with deterministic processing times and discrete part flow. Refer to (Colledani et al., 2003).

- one product-assembly/disassembly flow lines with deterministic processing times and discrete part flow. Refer to (Gershwin, 1994; Levantesi et al., 2000).

Stochastic models of flow lines considering scraps have been developed by Helber (Helber, 1999). All the above cited models assume that failures and repairs are exponentially (or geometrically in case of discrete time) distributed. However, when the Markovian memoryless assumption does not hold, e.g. when failures depend on the wear of machine's components, simulation is used to estimate the main performance measures considering the main features of the system. The main drawback of a simulation model is the total cost related to develop and run simulation experiments. In practice it takes long time to develop a detailed simulation model that captures the main characteristics related to part flow. In addition, the output of a simulation run is the outcome of a statistical experiment since a simulation run generates one of infinitely many possible realizations of the system's dynamic behavior. Therefore, the outputs of simulation experiments have to be treated as statistical experiments which imply running and analyzing the results of a large number of experiments to obtain statistically reliable estimates of the performance measures.

3.2 Flexible Manufacturing Systems

Description. FMSs are production systems composed of computer numerically controlled (CNC) machining centers that process prismatic metal components. A process cycle defining all the technological information (e.g. type of operations, tools, feed movements, working speeds, etc.) is available for each product so that the system has the complete knowledge for transforming *raw parts*, the state in which a piece enters into the system, into *finished parts*, the state in which a piece has completed the process cycle. The main components of FMSs are described in the following:

CNC machines perform the operations on raw parts. A machining operation consists of a tool, fixed in the machine, that removes the material from the raw part. The machines are CNC type in the sense that their movements during the machining operations are locally controlled by a computer. Machines can differ in their size, power, spindle speed, feed rates and number of controlled axes.

Pallets are the hardware standard physical interfaces between the system's components and the pieces. Pieces are clamped on pallets by means of automated fixtures that have the function of providing stability to parts during the machining operation. Generally, but not always, fixtures are dedicated to products.

Load/unload stations execute the operations of clamping raw parts onto pallets before entering into the system, and removing finished parts after their process cycle has been completed by the machines of the system. Stations can be manned, i.e. an operator accomplishes the task, or unmanned, i.e. a gantry robot accomplishes the task.

Part handling sub-system is the set of devices that move parts through the system. Different mechanical devices are adopted in reality: automated guided vehicles, carriers, conveyors, etc.

Tools perform the cutting operations on raw parts. Since tools are expensive resources their number is limited and as a consequence they are moved through the system when requested by machines.

Tool handling sub-system is the set of devices that move tools through the system. The most adopted solution is a carrier moving on tracks.

Central part buffer is the place where parts wait for the availability of system's resources (i.e. machines, carriers, load/unload stations).

Central tool buffer is the place where tools can be stored when they are not used.

Supervisor is the software that controls the resources at a system level by assigning pallets to machines and load/unload stations and scheduling tool and pallet transports.

Tool room is the place where tools are reconditioned after the end of their life.

Let us now describe the flow of pieces in the system. Generally more than one piece is loaded on pallets at the load/unload station of the system. The type and the number of pieces on the same pallet depends on products and system's components. In particular the number depends on the physical dimensions of pieces and machines while the types of products depend on the technical feasibility of clamping different products with the same fixture. Most of the time the pieces loaded in the same pallet are of the same typology, however the frequency of cases in which different part types are loaded on the same pallet is increasing. If the loading operation is executed manually by operators, then the corresponding time can be considered a random variable according to some estimated distribution; otherwise, the assumption of deterministic loading/unloading times holds if automated devices (e.g. a gantry robot) perform operations and no source of uncertainty is present in the task.

After parts are loaded on pallets, they are managed by the supervisor which decides the path each pallet has to follow to complete the process cycle of all its pieces. In order to complete the process cycle, pallets must visit at least one machine; if machines are busy, pallets wait in the central buffer. Each machine has at least two pallet positions: the first one is the *pallet working position*, where the pallet is machined by the tool, while the other positions, known as *pallet waiting positions*, are used for decoupling the machine from the part handling sub-system. Indeed, the pallet in the waiting position waits for the availability of the machine, or of the carrier if the pallet has already been worked. The machine is equipped with a pallet changer for moving a pallet from a waiting position to the working position and vice versa; this movement is executed by automated devices and can be considered deterministic since no source of variability is present in the operation. After the pallet has been blocked in the working position and the tools necessary for the operations are available to the machine, the processing operations can be executed. Processing times of machines can reasonably be assumed deterministic. In fact, the trajectory of the tool during the cutting of material is computer numerically controlled and therefore the sources of variability are eventually negligible.

The part handling sub-system moves pallets in the system among machines, load/unload stations and central buffer. After the completion of the process cycle, the finished pallet is moved to the load/unload station where the parts are unloaded and the empty pallet can be used again to load new raw parts if requested. The routing of pallets in the system is managed by the supervisor and it depends on the particular state in which the system is at a certain moment. A variety of dispatching rules are used in reality for managing the part flow in FMSs. See also the scheme of an FMS with four machining centers and a load/unload station in Figure 4.4.

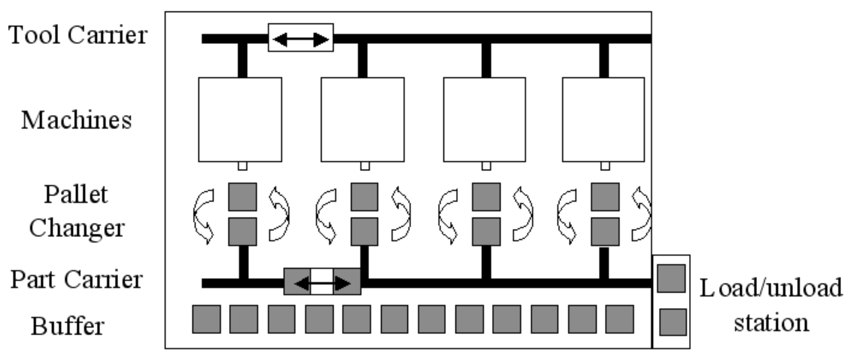


Figure 4.4. Lay-out of FMS.

Let us now describe the flow of tools in the system. During a machining operation the tool diminishes its life because of the wear phenomenon. The *tool life* is the amount of time for which a tool can be used for processing operations. After the tool life is finished, the tool has to be reconditioned in the tool room where it obtains a "new life". Furthermore, tools are unreliable in the sense that they can break thus becoming no longer utilizable in the system.
Tools are normally stored in the central tool storage waiting for a request from machines. The tool handling sub-system is responsible for moving them among the different machines and the central tool storage. When a machine requests a tool, the tool handling sub-system moves the tool from the central tool storage to the machine, where the tool is used for machining the raw parts. Each machine is equipped with a local tool storage where tools can be temporarily stored to decouple the machine from the tool handling sub-system. Then, after the completion of the processing operation, if the tool is no longer needed it is moved again to the central tool storage or possibly to another machine that has requested the same tool. In order to decrease the interdependency among machines, more than one instance of the same tool is present in the system. If a tool ends its life, or breaks, it is moved by the tool handling sub-system to the tool room where it is reconditioned or changed with a new tool.

Stochastic models. An important line of research is devoted to stochastic models of FMSs. There are certain inherent factors which motivate the use of stochastic modelling for an FMS. The typical example is a machine breakdown which is an unforeseen disruption to the behavior of system, this is typically captured by a stochastic model. Other less important factors with a stochastic nature are load/unload (if not automated), tool breakdowns, the releasing of orders, etc. The presence of the above factors which require modelling the random components of an FMS is only a partial justification for the use of stochastic modelling techniques. In fact, the key argument for the justification of a stochastic model seems to be the lack of exact information on the part mix to be produced on the system.
Buzacott and Yao present a literature review of the analytical models of FMSs covering the works of different groups until the mid eighties (Buzacott and Yao, 1986). Solot and van Vliet provide an updated account of the analytical modelling literature for FMS systems (Solot and van Vliet, 1994). They classify the analytical models according to the corresponding problems addressed. Five major classes are identified: processing capacity, buffer capacity, facility layout, pallet quantity and

material handling system. It is recognized that the most frequently studied problem is that of optimizing processing capacity in terms of machine allocation and grouping. Actually, it is possible to use analytical methods to model dynamically only the part flow without considering tools in the FMS or vice versa, i.e. only the tool flow without considering parts in the FMS. Therefore analytical models are used first to allocate machines, load/unload stations, central part storage and transporters of parts assuming that tools are always available and then, in a second step, to allocate the tools and the central tool storage not modelling explicitly the dynamics of parts. Refer to the book of Buzacott and Yao (Buzacott and Shantikumar, 1993) and Padopoulos et al. (Papadopoulos et al., 1993) for more details on queuing theory applied to FMSs.

Discrete event simulation tools enable a very detailed analysis of the underlying system to be analyzed. A discrete event simulation model can almost mimic the dynamic behavior of the FMS by representing explicitly machine operations, pallet movement, part carrier, tool changing and operations etc. The drawbacks of simulation used to estimate the performance of FMSs are the same as described for DMFLs.

4. Design of Automated Manufacturing Systems

Referring to the main output of activity A3, a node in the graph represents a potential adoptable set of AMSs the company can choose to manufacture products. This choice has to take into account the whole planning horizon, since a node that is dominated at time period t may prove to be the best solution in the rest of the horizon because of, for instance, a modification in the internal capacity constraint related to a change in the market demand. In other words, to solve independently several configuration problems, one for each time period, may lead to non optimal decisions. For this reason the approach described in this chapter is focused on identifying all the potential feasible alternatives in each period. An immediate drawback of the approach is the inevitable explosion of the number of alternatives; however the tools used for evaluating these alternatives, as also the method used for selecting the optimal capacity plan (see Chapter 5), are very fast in terms of computational time thus allowing in practice to deal with the complexity of the approach. The method for constructing the graph of AMSs in the planning horizon is now described. Figure 4.5 illustrates this process. First the set of alternative configurations is generated for each type s of AMS and each time period t (see A31 in Figure 4.5). An FMS configuration is defined by specifying the number and types of machines, load/unload stations, transporters and pallets; notice that the central part storage and the tool

flow is not considered at this stage. A DMFL configuration is defined by specifying the type and number of machines, the buffer capacities and pallets in the case of a closed system. Performance evaluation models are used during the generation phase for calculating the production rate of dimensioned systems (see A32 in Figure 4.5).
The generated configurations are feasible in the sense that the budget and internal capacity constraints are satisfied and the systems are technically admissible in their components. Notice that the budget constraint is verified on the single alternatives; this means that it is possible that a mix of AMSs obtained by grouping different alternatives violates the budget constraint.

Then nodes and arcs of the graph are identified in activities A33 and A34 respectively in Figure 4.5. In detail the nodes are obtained by combining the feasible alternatives of each AMS type and discarding the nodes that do not satisfy the budget constraint. At this step the budget is verified on the mix of AMSs belonging to that node.
The arcs are obtained by checking the feasibility of the transition in terms of cost and practicability of the change introduced.

4.1 [A31] generate configurations

The goal of the activity A31 is to generate, for each time period t (with $t = 1, \ldots, T$) of the planning horizon and for each type s of AMS (with $s \in \Omega(t)$), all the possible manufacturing systems that satisfy the constraints of budget and internal capacity (C1 and C2 in Figure 4.5). In detail, any AMS generated in this phase must respect:

- Budget constraints. The total investment cost $TC_s(t)$ necessary to acquire at period t a specific manufacturing system must be lower than the budget available until the same period:

$$TC_s(t) \leq \sum_{k=1}^{t} B(t) \quad \forall s \in \Omega(t), \quad t = 1, \ldots, T \tag{4.1}$$

  where $TC_s(t)$ is the total cost associated with the system s and $\Omega(t)$ the set of AMSs to design at time period t. This is because it could happen that a manufacturing system at time t is the result of several investments in the previous periods. To consider only the budget at period t would be too restrictive.

- Capacity constraints. For each product type i the number of pieces manufactured in period t by the system s, denoted by $x_{i,s}(t)$, must be greater and less than the minimum and maximum number of

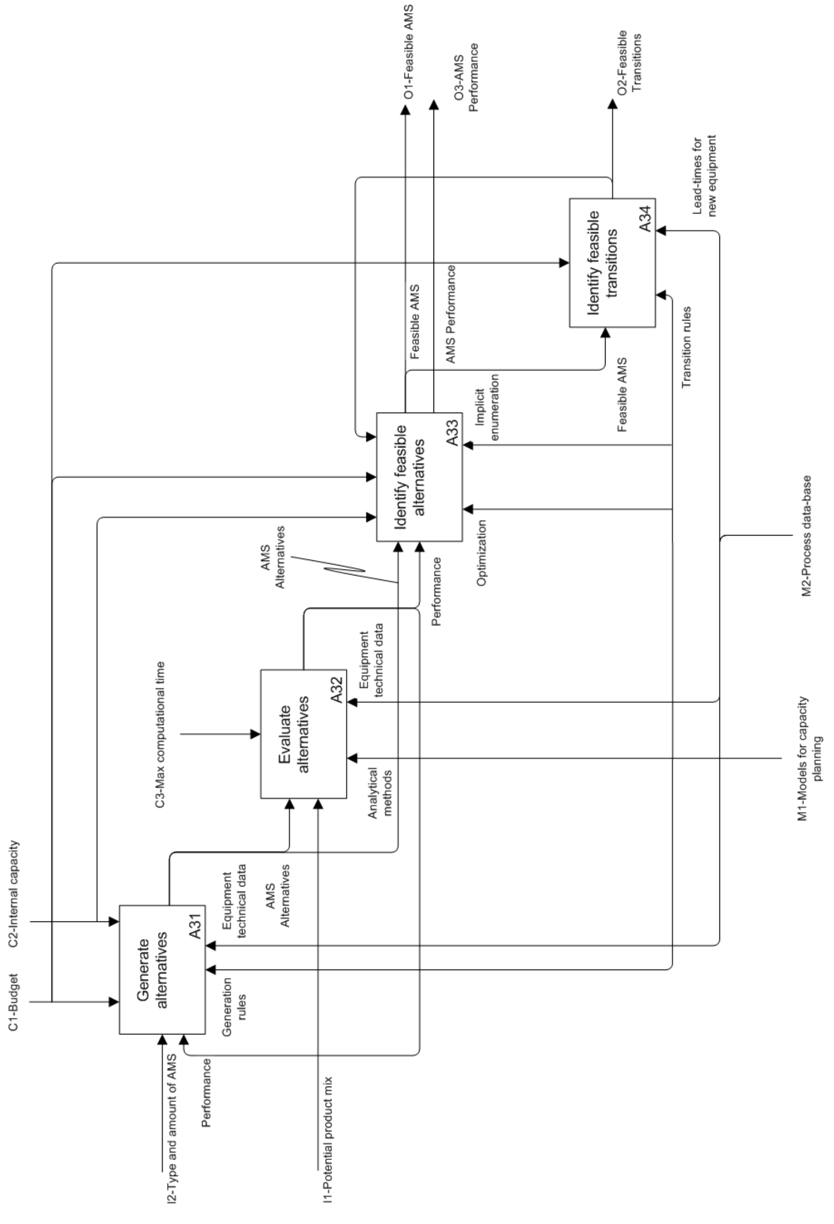


Figure 4.5. A3 level diagram.

pieces respectively as decided by activity A2:

$$l_{i,s}(t) \leq x_{i,s}(t) \leq u_{i,s}(t), \quad \begin{array}{l} \forall i \in \Psi(s) \\ \forall s \in \Omega(t), \\ t = 1, \ldots, T \end{array} \tag{4.2}$$

All the possible configurations are generated by using technological rules for grouping the different components. A necessary condition is that when the single components are grouped to form a system type s, the so generated system must be able to produce all the product types belonging to the set of products the system s has to work as decided at higher level by activities A1 and A2, that is any product type i with $i \in \Psi(s)$. To do this it is assumed to know a technological matrix of machines and products where the elements indicate if a machine type can manufacture a certain product type. The technological matrix can be obtained, in a simplified way, by matching the process plan of the product to the machine's technical characteristics such as power, size, CNC axes, etc. A limiting condition of the generation of alternatives is the maximum size of manufacturing systems. For instance, FMSs with more than eight machines are not considered because they are quite uncommon and their design typically involves several types of problems.

Therefore, the output of activity A31 is a list of alternatives for each type s of AMS in time period t. An FMS configuration is defined by specifying the number and types of machines, load/unload stations, transporters and pallets; as already said the tool handling system is not considered. FMS alternatives are obtained by combining and dimensioning all the combinations of machine types, load/unload station types and transporter types. There is a wide set of machines potentially acquirable in the market. Indeed, machines can differ in their dimensions, power, spindle, number of CNC axis, spindle orientation, spindle speed, tool changer, pallet changer, etc. For the purposes of our work, we consider the following characteristics of a machining center:

- Working cube: the volume space the machine is capable to machine. Generally, the bigger the volume space is the larger the number of pieces clamped on the same pallet is.
- Power: the energy per time unit available at machine's spindle for machining operations.
- Number of Computer Numerically Controlled axis. Generally this number is equal to 4 for standard pieces and 5 for more complex features.
- Tool changing time: it is the total time needed to change a tool from chip to chip. This time is generally related to the dimensions of the local tool storage of the machine.
- Pallet changing time: it is the total time needed to change a pallet from chip to chip. This time is normally related to the working

cube of the machine: the larger the working cube is the thicker the pallet is and therefore the time to execute the movements for transferring the pallet is longer.

The cost of a machining center depends on its characteristics, that is a machine with high performance is more expensive than a standard one. Transporters move pallets among different positions of the system. Theoretically the devices adopted in practice for moving pallets through the system are different: conveyors, Automated Guided Vehicles, carts, etc. In addition the same device may differ in its characteristics such as maximum velocity, acceleration, loading time, maximum weight loadable, etc.
Load/unload stations can be fundamentally of two types: automated or manual. Automated stations generally have robots for the operations of loading and unloading of pieces and require higher investment costs if compared to manual stations.
Pallets are normally dedicated to part types. We assume the design of the pallet is made off-line by technicians and depends on the working cube of machines composing the system. The bigger the machine is the larger the possibility of saturating the pallet by loading many pieces is, however the cost is larger too. For any generated FMS a certain number of pallets is allocated.
Assuming the sets of machines, transporters and load/unload stations are known it is possible to generate all the combinations of systems satisfying the budget and internal capacity constraints. To simplify the problem, we impose the condition that all machining centers in the same FMS require the same pallet dimensions, that is all the machines must have the same working cube size. This simplification is not stringent since it is very common in shop floors.

A DMFL configuration is defined by specifying the type and number of machines, buffer capacities and pallets in the case of a closed system. Machine types strongly depend on the specific product type and also their number and ordering in the line. For this reason we do not enter into the details of DMFL generation and we assume that a list of alternative systems is available for each product type for which a dedicated system is requested from activities A1 and A2.

4.2 [A32] evaluate alternatives

Performance evaluation of alternative manufacturing systems provides the main performance measures such as average throughput, work in progress, machine utilization, etc. A performance evaluation model should be accurate, complete, easy to use and cheap. In more detail,

we say a model is complete if represents adequately all the most important phenomena and variables of the real system we want to model. However the levels of adequacy and importance are difficult to establish and are related to the level of the decision. A validation of the model is necessary to test its completeness.
Accurateness depends on the completeness of the model. The more significant the variables considered are in the model the more precise the provided results are. We consider that a performance evaluation model is accurate if the measures provided differ slightly to the real modelled system. However there is a trade off between accuracy and cost of the model. Indeed, an accurate model may require a long time to be developed and validated. The cost of a performance evaluation model is related to the total cost necessary to obtain the estimated measures. Generally, the longer the computational time is the larger the cost sustained by a firm for assessing a manufacturing system is. In addition the cost of input data acquisition increases as the modelling detail increases too.

Any performance evaluation tool is characterized by its completeness, accuracy and cost. The techniques most used in practice are simulation and analytical methods whose characteristics are shown in Table 4.1. A simulation model can represent a complex manufacturing system in all its details. The only problem incurred by adding details into the simulation model is the long development time of the simulation code and the computational time to run experiments. Thus, the level of adequacy of a simulation model is decided by the user coherently with the objectives and the budget of the analysis. The accuracy of performance measures estimated by simulation greatly depend on the completeness of the model but also on the length of simulation experiments.
The most interesting feature of analytical methods is their synthesis of the main behavior of complex systems in a few related variables, most of the time implicitly, by dynamic equations. However, it is often necessary to introduce restrictive assumptions for simplifying the mathematical treatment of the model that reduce its applicability. A classical example is the memoryless property of Markovian models that do not fit with the phenomena of mechanical wear of machines and tools. On the other hand analytical methods generally perform fast and the provided results have average accuracy. Generally the error on production rates ranges from 1 to 15 percentage points depending on the specific model adopted. Given a set of assumptions on the model, we say an analytical model is *exact* if no further simplifications are introduced, i.e. the results provided by the model are coherent with the model's assumptions. On the contrary, we say an analytical model is *approximate* if

some simplifications are introduced to deal with the complexity of the model, i.e. there is another approximation in addition to the assumption of the model and the results provided are not exact. In many cases it is preferable to increase the complexity of the model by removing stringent assumptions even if the model is solved in an approximate way.

The choice of the most appropriate performance evaluation method depends on the level of the analysis. An interesting approach is to use analytical methods for an initial selection of good configuration candidates and then refining the choice among the selected candidates by performing a few simulation experiments. This two-step approach benefits from the speed of analytical methods in the initial phase for eliminating distinctively poor configurations, and in the second phase makes use of the detailed modelling feature of simulation to search for the best candidate; an application of the two-step approach is presented by Starr (Starr, 1991), Dekleva and Gaberc (Dekleva and Gaberc, 1994). The authors present the implementation of the integrated analytical/simulation performance evaluation tool in a software. In this way, initial design iterations can be performed rapidly using the analytical module and the simulation module can provide refinements.

At strategic capacity decision level it is important to consider as often as possible a large number of alternative solutions even if the estimates on performance measures are not very accurate. This is because it is fundamental to explore at the beginning of the analysis all the possibilities. At a later stage of the analysis, e.g. implementation, it could be very expensive to consider a new alternative never considered previously. Therefore only analytical methods are used to calculate performance measures of alternative solutions. The analytical methods adopted in our work for both FMS and DMS are of two types: static and dynamic (Matta et al., 2001). These models are used to calculate the production rate when resources are allocated during the generation of alternatives of activity A31. Static models are a rough approximation of the modelled system since they do not consider the dynamic behavior of machines,

Table 4.1. Simulation vs analytical methods.

Characteristics	*Simulation*	*Exact analytical methods*	*Approximate analytical methods*
Completeness	User specified	Low	Average
Accuracy	User specified	Average	High-average
Cost	High	Low	Low-average

transporters and parts. However, static models are fast and accurate enough for a preliminary evaluation of alternatives. In detail, static models take into account only the time available for machining products without considering non productive times due to starvation and blocking of machines. Dynamic analytical methods are then used to refine the analysis in a detailed way. See Sections 5 and 6 for a description of approximate analytical methods proposed for assessing the performance of FMSs and DMSs respectively.

4.3 [A33] identify feasible alternatives

Once all the feasible alternatives for each AMS type have been identified, the firm has to select for each time period the combination of AMSs that minimizes the total cost function calculated over all time periods. Thus, the solution in each time period is represented by an alternative manufacturing system for each type s (with $s \in \Omega(t)$) of AMSs. We graphically represent this solution with a node in a graph, denoted by $n_k(t)$ (with $k = 1, 2, \ldots$), that is a combination of feasible different AMS types. In lack of some particular assumptions on the mix of AMS belonging to the same node, all the possible nodes are considered for each time period except those violating the budget constraint. The idea is always the same: to investigate all the alternatives potentially adoptable by the firm.

Since the internal capacity constraint has been verified in activity A31, the only filtering to do on alternatives is to check if the budget constraint is verified on the whole mix of AMSs belonging to the same node. The total investment cost necessary to acquire at period t all the types s of AMSs in the same node must be lower than the budget available until period t. Notice also that the budget available may be larger than the value decided by the management if in the previous periods the firm spends less than the the assigned budget. The nodes which do not satisfy the budget constraint are discarded from the analysis. Summarizing, for each node in the graph it is known:

- system components, that is the main components of each AMS type (machines, load/unload stations, transporters, pallets for FMS and stations, buffers and pallets for DMS).
- investment cost $TC_s(t)$ for each AMS type in time period t.
- production rate $x_{i,s}(t)$ for each product type i and each s AMS type in time period t.

This information will be used to select the optimal capacity path in the planning horizon as described in Chapter 5.

4.4 [A34] identify feasible transitions

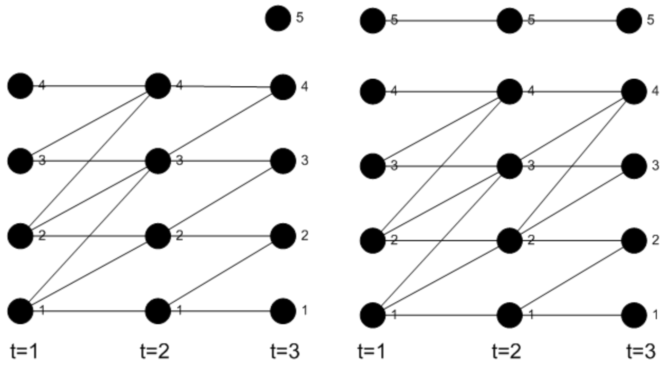
The production needs may change from period to period and the firm could be forced to modify the mix of manufacturing systems to react to changes in the market demand or the birth or death of some products. In that case it may occur that the best node in the following period $t+1$ will differ from that adopted in period t, i.e. the firm will introduce some changes on the shop floor. These changes on the shop floor should be feasible in practice from both technical and organization points of view. For instance, it is often not reasonable to quadruplicate an existing system due to the several problems the firm may encounter both from technical (e.g. the supervisor of the system cannot be scaled) and organizational points of view (e.g. no skill to manage this big change).

The goal of activity A34 is to identify all the feasible transitions among all adjacent nodes.
Any two nodes $n_k(t)$ and $n_l(t+1)$ belonging to adjacent time periods are linked by edges if it is possible to change the mix of systems represented by node $n_k(t)$ into $n_l(t+1)$. The weight of the edge is the total cost to introduce the changes and is denoted by $w_{k,l}(t, t+1)$. The time period unit selected is equal to six months, a reasonable time to introduce a relevant change. In fact the maximum lead time requested to order a machining center is approximately six months while all the other changes require a smaller time. However this assumption can be relaxed and different time periods can be considered thus having also non adjacent nodes linked by feasible edges. Finally, an edge must respect also the budget constraints.

In such a way it is easy to construct a graph with feasible nodes and arcs. At this point two particular cases may be encountered:

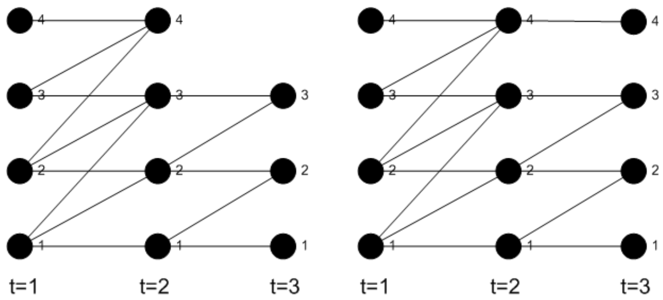
- a node $n_k(t)$ cannot be reached by any upstream node. This means that a firm will never have the possibility to adopt the mix of manufacturing systems of node n_t. This situation is illustrated in Figure 4.6a and can be conservatively adjusted by forcing the node $n_5(3)$ in all the previous periods (Figure 4.6b). An alternative is to eliminate the node $n_5(3)$ from the analysis.

- a node $n_k(t)$ does not have any downstream node. This may happen when a firm that chooses to adopt the mix of manufacturing systems of node $n_k(t)$ cannot introduce actions to reach a feasible solution in the next time period. This situation is illustrated in Figure 4.7a and can be conservatively adjusted by forcing the node $n_4(2)$ in all the following periods (Figure 4.7b). Again an alternative is to eliminate the node $n_4(2)$ from the analysis.



(a) (b)

Figure 4.6. Example of unreachable node.



(a) (b)

Figure 4.7. Example of unleaveble node.

5. Performance evaluation of Dedicated Manufacturing Flow Lines

In this section we present an approximate analytical tool for evaluating the performance of Dedicated Manufacturing Flow Lines. This tool requires a small computational effort and can model unreliable machines with deterministic processing times and finite buffer capacity. The approach presented in this section is based on the work developed by Matta and Tolio in (Matta and Tolio, 1998).

The main improvement of the proposed method over existing techniques is the ability to take into account explicitly the different sources of interruption of flow in the line. The sources of flow interruption are the failures of machines that can fail in different ways with consequences with different degrees of severity. Therefore, it can happen that the the generic $i-$th machine cannot work because its input buffer is empty due to an interruption of flow caused by a failure at machine $i-1$, or machine

$i-2$, or any upstream machine. Normally for each type of failure there is a specific MTTF (Mean Time to Failure) and MTTR (Mean Time to Repair). All the types of unforeseen events happen with different frequencies and stop the station under analysis for periods of different lengths.
A method that takes into account different MTTFs and MTTRs is now presented.

5.1 Outline of the method

The idea of the method is to reduce the complexity of the analysis of a long line by studying a set of smaller systems that, in their whole, have the same behavior of the long line but, at the same time, are easier to analyze. In order to evaluate the performance of the system a decomposition method is proposed (Gershwin, 1994)(Matta and Tolio, 1998). The method is based on the exact analytical solution for a flow line with two machines and one finite capacity buffer between them (Tolio et al., 2002). The decomposition consists in decomposing the original line into a set of two-machine lines, in detail a two machine line for each buffer of the original line (see Figure 4.8). To estimate the performance of the line, the flow of parts in and out of the buffer of the two-machine line should be equal to the flow in the corresponding buffer of the original line. The flow of parts in the buffer of the original line can

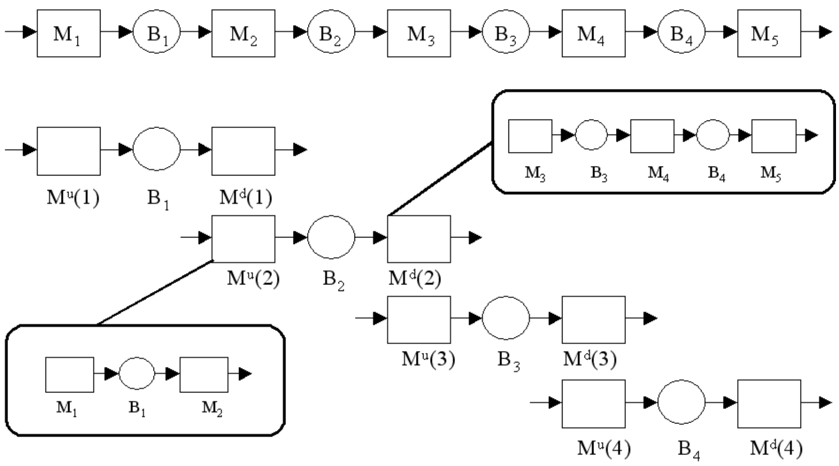


Figure 4.8. Decomposition method of a flow line with 5 machines.

stop because the upstream machine fails or because it becomes starved (i.e. starvation is due to a failure of one of the upstream machines). In the two-machine line however starvation of the upstream machine cannot occur since this is the first machine of the line which for assumptions can never be starved. Therefore, to mimic the effect of starvation in the two-machine line additional failure modes for the upstream machine are considered. By appropriately tuning the parameters of these additional failure modes it is possible to make the flow of parts in the buffer of the two-machine line resemble the flow in the corresponding buffer of the original line. The additional failure modes represent in fact the different causes of starvation of the corresponding machine in the original line. A similar reasoning applied to blocking leads to the introduction of additional failure modes to the downstream machine of a two-machine line. In the following section a formal description of the method is proposed. The decomposition equations are introduced with the goal of assessing the characteristic values for the failure and repair rates of the additional failure modes introduced in the various two-machine lines so that the coherence of the decomposed systems with the original one is preserved.

5.2 Analytical model: assumptions and notations

The method considers DMFL with K unreliable machines separated by $K-1$ buffers of finite capacity. Parts flow from outside the system to M_1, then to B_1, then to M_2 and so on until they reach the last machine M_K, after which they leave the system. Processing times are assumed to be deterministic and identical for all the machines of the line. Since DMFLs are very well balanced this hypothesis is normally acceptable.

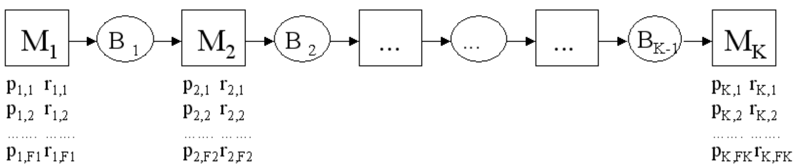


Figure 4.9. Flow line with K machines.

The processing time is assumed as time unit and relevant events happen at the beginning or at the end of a time unit. Therefore the time is modelled as a discrete variable. At the beginning of a time unit an operational machine can fail in different modes. Also at the beginning of a time unit a failed machine can get repaired.

An operational machine normally takes one part from the upstream buffer and puts it in the downstream buffer (i.e. it processes a part); the buffers variation occurs at the end of the time units when the machine operation has been completed. Blocking and starvation can occur as a result of machine failures. In fact when a given machine goes down the upstream buffers tend to become full because the upstream machine keeps on putting parts in the buffer. Eventually the buffer gets full and the upstream machine becomes blocked and therefore, even if it is operational, it cannot produce parts. This in turn propagates the effect of blocking upstream. Similarly, starvation arises and propagates through the line. Another assumption is that machines can fail only if they are operative, i.e. if they are not starved or not blocked (ODFs, Operation Dependent Failures). Also, a machine cannot fail in more than one mode, that is once a machine is down it cannot fail in other modes. Finally MTTFs and MTTRs are geometrically distributed.

Let us consider a flow line L with K machines and $K-1$ buffers in which each machine M_i can fail in F_i different modes. We denote the generic machine and buffer with M_i and B_i respectively, with $i = 1, \ldots, K-1, K$. The buffer B_i of the line has size N_i. At the beginning of each time unit, an operational machine M_i has a probability $p_{i,j}$ of failing in failure mode j (with $j = 1, ..., F_i$); at the beginning of a time unit if machine M_i is down because of failure type j, it has a probability $r_{i,j}$ of getting repaired. Given the failure and repair rates of a generic machine M_i, the maximum production rate of the machine is obtained when it is never starved or blocked and is known as the efficiency in isolation mode of the machine:

$$e_i = \frac{1}{1 + \sum_{j=1}^{F_i} \frac{p_{i,j}}{r_{i,j}}} \tag{4.3}$$

where F_i is the number of failure modes of M_i. In case of only one failure mode the equation reduces to (Buzacott, 1967):

$$e_i = \frac{r_i}{r_i + p_i} \tag{4.4}$$

Let us introduce to the possible states in which a generic machine M_i of the flow line can be. Generally we say a machine is operational if it is processing a part. The machine M_i is *up* at time t if:

- it was not failed at time $t-1$ and no failure occurs at time t. In this case a piece is processed by the machine in the time unit
- it was failed at time $t-1$ and gets repaired at time t. In this case a piece is processed by the machine in the time unit since the machine is repaired at the beginning of the time unit.

The same machine is *down* at time t if:

- it was operational at time $t-1$ and a failure occurs at time t
- it was down at time $t-1$ and does not get repaired at time t.

In both down cases the machine does not produce any piece. Machine M_i is *starved* if it was operational at time $t-1$ and at time t is up and the upstream buffer B_{i-1} is empty. In this case M_i cannot work because it has no part in its input buffer.
Machine M_i is *blocked* at time t if it was operational at time $t-1$ and at time t is up and the downstream buffer B_i is full. In this case M_i does not work because it cannot unload the part in the buffer B_i.
Another assumption is that the first machine is never starved and the last machine is never blocked. This corresponds to assume there are always raw parts available in front of the line and an infinite buffer at the end of the system in which to store finished parts.

5.3 Analysis of two-machine line

In decomposition methods the line L is decomposed in different two-machine lines. The generic two-machine line $\ell(i)$ (with $i = 1, ..., K-1$) is composed of an upstream pseudo-machine $M^u(i)$, a downstream pseudo-machine $M^d(i)$ and a buffer $B(i)$ of size $N(i)$ equal to N_i (i.e. the same size of the corresponding buffer of the original line). The upstream pseudo-machine $M^u(i)$ represents the portion of the system between machines M_1 and M_i in the original line while the downstream pseudo-machine $M^d(i)$ represents the portion of the system between machines M_{i+1} and M_K in the original line. In more detail, machines $M^u(i)$ and $M^d(i)$ model the part flow upstream and downstream the buffer B_i in the original line respectively. See again the example of decomposition in Figure 4.8.
Each pseudo-machine can fail in several modes. We denote with $F^u(i)$ and $F^d(i)$ the total number of failure modes of upstream and downstream machines respectively. Also we distinct the failure modes of pseudo-machines between *local* failures and additional or *remote* failures. In particular pseudo-machine $M^u(i)$ can fail in local mode j, correspondingly with one out of the F_i real failures of the original M_i machine, with probability $p^u_j(i)$ (with $j = 1, ..., F_i$) and can get repaired from failure of type j with probability $r^u_j(i)$ (with $j = 1, ..., F_i$) at each time unit. The rest of the failure modes of the upstream machine are related to the interruption of flow. More in particular, the number of remote failure modes of $M^u(i)$ is equal to the number of different sources of interruption of flow due to starvation upstream buffer B_i. Similarly for

the downstream pseudo-machine $M^d(i)$. Pseudo-machine $M^d(i)$ can fail in local mode j, correspondingly with one out of the F_{i+1} real failures of the original M_{i+1} machine, with probability $p_j^d(i)$ (with $j = 1, ..., F_{i+1}$) and can get repaired from failure of type j with probability $r_j^d(i)$ (with $j = 1, ..., F_{i+1}$) at each time unit. The rest of failure modes of the downstream machine is related to the interruption of flow. More in particular, the number of remote failure modes of $M^d(i)$ is equal to the number of different sources of interruption of flow due to blocking downstream buffer B_i.

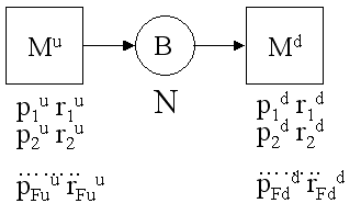


Figure 4.10. Two-machine line.

Given a two-machine line, $E(i)$ is the average production rate of the line, $Ps_j(i)$ is the probability of $M^d(i)$ being starved due to the failure mode j of the upstream pseudo-machine (with $j = 1, ..., F^u(i)$) and $Pb_j(i)$ is the probability of $M^u(i)$ being blocked due to the failure mode j of the downstream pseudo-machine (with $j = 1, ..., F^d(i)$). The total number of the failure modes of $M^u(i)$ and $M^d(i)$ is given (see also Figure 4.10) by:

$$F^u(i) = \sum_{k=1}^{i} F_k \quad F^d(i) = \sum_{k=i+1}^{K} F_k \tag{4.5}$$

Moreover we introduce:

- $E^u(i)$ and $E^d(i)$ the probabilities of $M^u(i)$ and $M^d(i)$ being operational. These values corresponds to the average throughput of upstream and downstream pseudo-machines respectively. It can be demonstrated (Tolio et al., 2002) that in a two-machine line $E^u(i)$ equals $E^d(i)$ because of the conservation of flow, that is $E^u(i) = E^d(i) = E(i)$.

- $D_j^u(i)$ and $D_j^d(i)$ the probabilities of $M^u(i)$ and $M^d(i)$ being failed in local mode j with $j = 1, ..., F_i$ and $j = 1, ..., F_{i+1}$ respectively

- $X_j^u(i)$ and $X_j^d(i)$ the probabilities of $M^u(i)$ and $M^d(i)$ being failed in remote mode j with $j = F_i + 1, ..., F^u(i)$ and $j = F_{i+1} + 1, ..., F^d(i)$ respectively.

Using the solution in (Tolio et al., 2002) we are able to calculate exactly the performance of two-machine lines. In the case of machines with one failure mode this solution coincides with that proposed by Gershwin (Gershwin, 1994). This solution, after several manipulations (for details refer to (Colledani. et al., 2004)), reduces to the exact closed relationship among average throughput and parameters of the two-machine line:

$$E = e^d + \frac{e^u - e^d}{1 - \frac{me^u}{le^d} X^{N-2}} \tag{4.6}$$

where the index i is omitted for simplicity of exposition, p^u, r^u, p^d, r^d are the failure and repair parameters of pseudo-machines, N is the buffer capacity (with $N \geq 4$), e^u and e^d are the efficiencies in isolation mode of upstream and downstream machines respectively of two-machine line $\ell(i)$, and X, l, m are parameters defined as follows:

$$\begin{aligned} X &= \frac{\frac{r^u+r^d-r^ur^d-p^ur^d}{p^u+p^d-p^up^d-r^up^d}}{\frac{r^u+r^d-r^ur^d-r^up^d}{p^u+p^d-p^up^d-p^ur^d}} \\ l &= \frac{r^u+r^d-r^ur^d-r^up^d}{r^up^d} \\ m &= \frac{r^u+r^d-r^ur^d-p^ur^d}{r^up^d} \end{aligned} \tag{4.7}$$

Relation (4.6) holds for X not equal to 1 and in all the other cases it reduces to:

$$E = \frac{Y\left[\left(\frac{1}{p^u} + \frac{1}{p^d} + 1\right) + (N-3)(Y+1)\right]}{Y\left(\frac{1}{p^u} + \frac{1}{p^d} + 2\right) + 2m + 2 + (N-3)(Y+1)^2} \tag{4.8}$$

where $Y = r^d/p^d = r^u/p^u$.

5.4 Decomposition equations

In order to approximate the behavior of the flow line with $K - 1$ different two-machine lines it is necessary to assign proper values to probabilities of local and remote failure of pseudo-machines. The purpose of this section is to calculate the unknown probabilities of failure and repair of pseudo-machines of each building block deriving from the decomposition of the original line L. These parameters are derived so that the material flow in the buffer of each two-machine line $\ell(i)$ closely matches the flow in the corresponding buffer of the original line.

In regards to local failure modes, the probabilities of failure and repair of the upstream machine $M^u(i)$ (i.e. $p_j^u(i)$ and $r_j^u(i)$ with $j = 1, ..., F_i$) are simply equal to the corresponding probabilities of the original machine (i.e. $p_{i,j}$ and $r_{i,j}$).
Let us go now to the remote failure modes of pseudo-machines. Remote failure modes model interruptions of flow in the system. In particular remote failure modes of the upstream pseudo-machine $M^u(i)$ model the interruption of flow at buffer B_i due to the starvation propagated from the upstream portion of the system; in this case the starvation is due to the failure of a machine upstream M_i. Similarly, remote failure modes of the downstream pseudo-machine $M^d(i)$ models the interruption of flow at buffer B_i due to the blocking propagated from the downstream portion of the system; in this case the blocking is due to the failure of a machine downstream M_{i+1}. The time to be repaired after a failure occurred with probability $p_j^u(i)$ with $j = F_i + 1, ..., F^u(i)$ (or $p_j^d(i)$, with $j = F_{i+1} + 1, ..., F^d(i)$) is approximately equal to the MTTR of the corresponding machine of the original line that has originated the interruption of flow plus the time necessary for pieces to reach the starved (or blocked) machine M_i (or M_{i+1}) modelled by $M^u(i)$ (or $M^d(i)$). Indeed, once the machine M_k, which has originated the starvation of machine M_i with $k < i$, is repaired it takes a certain amount of time for parts to reach the starved machine; this time depends on the portion of system between M_k and M_i that is empty because of starvation. As the flow of parts resumes and descends the line the starvation of machines ceases. Since processing times are scaled to units (see the assumptions of model), the time to resume the flow at machine M_i is equal to $i - k$. It is worthwhile to outline that this additional time has never been considered in previous decomposition models (Matta and Tolio, 1998). We use the notation $m(i, u, j)$ and $f(i, u, j)$ to indicate the machine and failure number of the original line that originates the starvation at building block i, modelled by the failure j of the pseudo-machine $M^u(i)$. In the same way $m(i, d, j)$ and $f(i, d, j)$ indicate the machine and failure number of the original line that originates the starvation at building block i, modelled by the failure j of the pseudo-machine $M^d(i)$. Thus we can write:

$$\begin{cases} \frac{1}{r_j^u(i)} = \frac{1}{r_{m(i,u,j),f(i,u,j)}} + i - m(i,u,j) & i = 2, \dots, K-1 \\ & m = 1, \dots, i-1 \\ & j = 1, \dots, F_m \\ \frac{1}{r_j^d(i)} = \frac{1}{r_{m(i,d,j),f(i,d,j)}} + i - m(i,d,j) & i = 1, \dots, K-2 \\ & m = i+2, \dots, K \\ & j = 1, \dots, F_m \end{cases} \tag{4.9}$$

Now the only unknown parameters of the two-machine $\ell(i)$ are $p_j^u(i)$ and $p_j^d(i)$ for every remote failure mode. Since every time machine $M^u(i)$ has a remote failure it eventually gets repaired, failure frequency must equal repair frequency for every failure mode; therefore we can write for every remote failure of the upstream and downstream pseudo-machines:

$$\begin{cases} p_j^u(i) = \frac{X_j^u(i)}{E^u(i)} \cdot r_j^u(i) & i = 2, \ldots, K-1 \\ & m = 1, \ldots, i-1 \\ & j = 1, \ldots, F_m \\ p_j^d(i) = \frac{X_j^d(i)}{E^d(i)} \cdot r_j^d(i) & i = 1, \ldots, K-2 \\ & m = i+2, \ldots, K \\ & j = 1, \ldots, F_m \end{cases} \tag{4.10}$$

where m is the machine that has originated the interruption of flow. In order to have coherence among the decomposed two-machine lines the conservation of flow must be imposed. Thus the following conditions must be respected:

$$E(i) = E(i-1) \quad i = 2, ..., K-1 \tag{4.11}$$

As said before, remote failures of $M^u(i)$ are introduced to mimic starvation, therefore the probabilities of remote failure states of $M^u(i)$ must equal the probabilities of starvation of $M^d(i-1)$ of the immediate preceding two-machine line for any remote failure mode:

$$X_j^u(i) = Ps_j(i-1) \quad i = 2, ..., K-1; j = 1, ..., F^u(i) \tag{4.12}$$

At the same way, remote failures of $M^d(i)$ are introduced to mimic blocking, therefore the probabilities of remote failure states of $M^d(i)$ must equal the probabilities of blocking of $M^u(i+1)$ of the immediate following two-machine line for any remote failure mode:

$$X_j^d(i) = Pb_j(i+1) \quad i = 1, ..., K-2 \tag{4.13}$$

Substituting (4.11), (4.12)(4.13) into equations (4.10) we obtain the final equations for remote failures:

$$\begin{cases} p_j^u(i) = \frac{Ps_j(i-1)}{E^u(i-1)} \cdot r_j^u(i) & i = 2, ..., K-1; j = 1, \ldots, F^u(i-1) \\ p_j^d(i) = \frac{Pb_j(i+1)}{E^d(i+1)} \cdot r_j^d(i) & i = 1, ..., K-2; j = 1, \ldots, F^d(i+1) \end{cases} \tag{4.14}$$

This form of decomposition equations helps us to apply the method in a simple way. Indeed, the unknown remote repair rates can be easily calculated by equations (4.9) while the unknown remote upstream (or downstream) failure rates can be calculated if production rate and starvation (or blocking) of the previous (or next) are known. This suggests to adopt an iterative algorithm that calculates the upstream (or downstream) unknowns on the basis of the solution of the previous (or next) two-machine line and assuming certain values for the downstream (or upstream) unknowns. Thus, equations (4.9) and (4.14) are used to write a very efficient recursive algorithm (Dallery et al., 1988) that allows the evaluation of the unknown parameters of the $K-1$ two-machine lines. In turn, each two-machine line can be evaluated with the approach proposed by Tolio et al. (Tolio et al., 2002) thus obtaining an estimate of the behavior of the original line. The following algorithm, known as the DDX algorithm (Dallery et al., 1988), allows the implementation of the described method:

1 **Initialization**.

- Decompose the original line in $K-1$ different two-machine lines.
- Set parameters of remote failures of pseudo-machines to some initial values between 0 and 1.
- Solve the first two-machine line using the technique in (Tolio et al., 2002).

2 **Update upstream pseudo-machines**.
For each $i = 2, \ldots, K-1$:

- Calculate upstream probabilities of remote failures and repairs using equations (4.9)(4.14).
- Solve the i-th two-machine line.

3 **Update downstream pseudo-machines**.
For each $i = K-2, \ldots, 1$:

- Calculate downstream probabilities of remote failures and repairs using equations (4.9)(4.14).
- Solve the i-th two-machine line.

4 **Convergence**. Goto step 2 until the values of probabilities of failure and repair of the last two iterations differ by a quantity lower than a specified ϵ.

Table 4.2. DMFL Real case: description of machines' failures.

Station 1	Breakdown of the boring tool
	Failure in the detection of a wrong position of the rotor
Station 2	Failure of the sensors
	Stop for jam of the rotor in the machine
Station 3	Stop for manual quality control of the rotor's diameter
	Wear of the disk that transmits the movement to the rotor during turning operation
Station 4	Wrong assembling of components
	Wrong position of the rotor in the transfer mechanism
Station 5	Part-box jam in the machine
	The rotor is not correctly positioned in the part-box.

5.5 Real case

The proposed method produces very accurate estimates of production rate of flow lines thus allowing its use as precise and fast method for evaluating the performance of Dedicated Manufacturing Flow Lines. The accurateness of the method has been estimated on several hundreds of cases in which the percentage error (estimated by comparing results with simulation) on production rate is always lower than 2% and very often below 1%. For an exhaustive set of numerical experiments refer to (Matta and Tolio, 1998) and (Matta and Bianchetti, 1997). In this section we only prove the applicability of the method to a real flow line that produces armature spiders for electrical engines. This line, composed by 5 stations and 4 buffers, performs both production and assembly operations. The raw parts, rotors, are in a large buffer in front of machine 1. Machine 1, a boring machine, machines the central hole of the part where afterwards at machine 2 a shaft is inserted; the third machine turns the external diameter of the rotor while in the fourth machine washers and sleeves are inserted in the shaft. The last machine unloads the parts and puts them in boxes. Since the operation times in each station are deterministic and very similar the line has a good balancing. The behavior of the line was observed for 3 weeks and data on occurrence of failures and repairs were collected. Different causes of machines breakdowns were identified, in particular two failure modes for each machine can occur during the functioning of the line (see Table 4.2).

Although the MTTFs and MTTRs of the real line are not all geometrically distributed the percentage errors over the results obtained by means of a detailed simulation are in the order of 1.3 % for the production rate and 3.6 % for the average buffer level thus proving the applicability of the method on real DMFL.

5.6 Some considerations

The decomposition technique described in this section has been applied also to assembly/disassembly systems, closed flow lines, and two-product flow lines with very accurate results. Actually, other decomposition techniques are the first one developed by Gershwin (Gershwin, 1994) and improved in its solution by Dallery et al. (Dallery et al., 1988), and the technique proposed by Dallery and Le Bihan (Le Bihan and Dallery, 1997). The former considers blocking and starvation phenomena in a first order approximation by modelling with one average failure both local failures and interruptions of flow. This method uses an analytic solution to calculate the performance of building blocks. The main advantage of this technique is in its simplicity of implementation. The second technique considers blocking and starvation phenomena in a second order approximation by modelling with two average failures both local failures and interruptions of flow. This method uses the same analytic solution of Gershwin to calculate the performance of building blocks. The main advantage of this technique is again the simplicity of implementation and provides generally more accurate results than the first order approximation.

The proposed technique considers blocking and starvation phenomena with an n-th order approximation and uses a numerical solution to calculate the performance of building blocks. In particular a polynomial of degree $F^u + F^d$ is solved with an efficient method which fully exploits the structure of the polynomial itself. This technique provides generally more accurate results than the other techniques. The solution of the building block is more complex to implement and it is motivated when failures have significantly different MTTRs. We think that the main advantage of the proposed technique is the simplicity and precision in considering the effect of blocking and starvation. This simplicity is due to the fact that the building block is more refined and provides much more information. Indeed this technique can be easily extended to analyze several different systems as the numerous works in this field demonstrate.

6. Performance evaluation of Flexible Manufacturing Systems

6.1 Introduction

The main peculiarities of FMSs are that processing times can be considered deterministic and the number of different products is generally large. The deterministic assumption on processing times is quite difficult to model and it has been faced by working on queuing networks with general processing times. Several works are present in literature. First Marie (Marie, 1979) analyzes the servers in isolation mode of a queuing network as $\lambda(n)/C_k/1$ queues (i.e. state dependent Poisson arrivals and a k-stage Coxian) and the calculation of the queue length distribution is used to fit an equivalent load-dependent exponential server in the approximate network. The paper of Marie is the starting point of several other related works. Yao and Buzacott propose the exponentialization approach for analyzing single-class Closed Queuing Networks (CQN) with general service times (Yao and Buzacott, 1985). According to this approach, the original network is represented by an approximate equivalent network composed of exponential servers with state dependent rates. Servers are analyzed in isolation mode as a $M/G/1$ queue taking into account the real processing time distribution. The results from the analysis in isolation mode is captured in the whole network by appropriately selecting the state dependent service rates of exponential servers in the approximate equivalent network. The analysis of the server in isolation mode with their general service times allows the consideration of the non-exponentiality assumption in the analysis. Baynat and Dallery apply the exponentialization approach to analyze multi-class CQN proposing an aggregation technique to study servers in isolation mode (Baynat and Dallery, 1993a) and unifying in a framework the existing exponentialization techniques (Baynat and Dallery, 1993b). Refer to (Papadopoulos et al., 1993) and (de Almeyda, 1998) for a complete view of works in this field.

The proposed method analyzes the modelled queuing network by means of approximate techniques of decomposition. The aim of the method is to reduce the complexity of the problem by analyzing several simple systems instead of a complex one. The multiple-class closed queuing network is decomposed in different single-class closed queuing sub-networks that are solved with classical product form solutions. In order to capture the behavior of the whole system in the decomposed single-class sub-networks, servers are analyzed in isolation by considering all the classes thus following the methodology of Baynat and Dallery. In particular each server of the decomposed network is analyzed with an

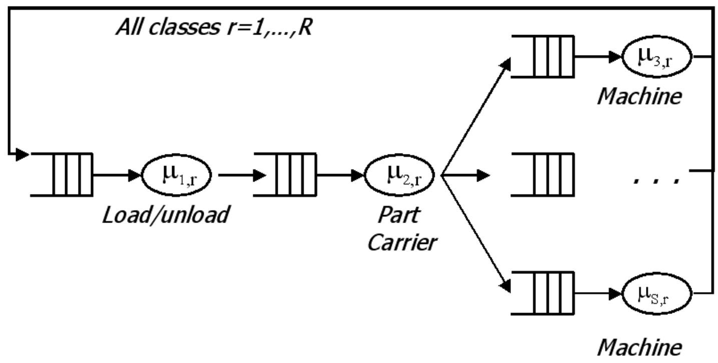


Figure 4.11. Queueing network of modelled FMS.

aggregation technique that allows calculating the throughput of single server systems with multiple classes. Dealing with multiple part types and deterministic processing times represent the main improvements of the proposed method over the existing techniques.

6.2 Analytical model: assumptions and notation

An FMS is modelled by means of a multiple-class closed queueing network in which clients represent pallets, and servers represent machining centers, load/unload stations and part carriers (see also Figure 4.11). We indicate with R and S respectively the number of classes and servers in the network.

Each class in the queueing network corresponds to a certain pallet type in the real system and we denote with N_r the number of customers of type r (with $r = 1, ..., R$) circulating in the system. Also we assume that customers cannot change the class type, that is a pallet cannot change its own fixture or more simply pallets are dedicated to part types.

Each server has a queue with infinite capacity immediately upstream to temporally store parts that wait to be served. The first server models one or more load/unload stations, the second one represents the part carrier while the other servers model machines of the system. The structure of the described network is known in literature as central server model.

Service rates in the network are assumed exponential and are state dependent, that is they depend on the number of customers in their queue. Service rates are denoted with $\mu_{i,r}(n_{i,r})$ where i indicates the server, r the class and $n_{i,r}$ represents the number of clients of class r that stay at server i. Regardless the assumption on service rates, the proposed method described in next sections perform well when applied to FMSs

with deterministic processing times because it fully captures the fact that distinct pallets require different processing times.
The path of customers in the network reflects that of pallets in the real system. Indeed, all customers always visit first the load/unload server and then the part carrier server before going to servers that models machining centers. After customers have completed their process cycle, they start again from the first server in the network; an implicit assumption here is that the number of customers is constant and equal to N_r (with $r = 1, ..., R$). The fact that in the real system the pallets need of the transporter for moving from machines to the load/unload station is modelled by considering two times the travel time in the service rates of the transporter node in the queuing network.
The routing of customers to servers is probabilistic and the visit ratios $V_{i,r}$ depend on the process cycle of parts clamped on pallets.
The tool flow in the FMS system is not modelled assuming implicitly that tools are always available when machines require them.
The queueing network has infinite buffers at each server and as a consequence machines are never blocked by the filling of downstream buffers. This assumption is reasonable because the supervisor that manages flows in the FMS avoids idle times of machines due to blocking by means of sophisticated dispatching rules.

6.3 Outline of the method

An FMS is modelled by a multi-class closed queueing network, denoted with L, in which each server represents a machine, or the part carrier or load/unload stations. In order to evaluate the performance of the network L, we use a general methodology based on a product-form approximation technique (Baynat and Dallery, 1993a) (Baynat and Dallery, 1993b). The idea of the method is to decompose the multiple-class system L into a set of single-class closed sub-networks ℓ_r (with $r = 1, ..., R$), each one related to a specific customer type r. The method associates R single class product-form sub-networks, each one modelling the behavior of a particular sub-network ℓ_r, to the original multi-class queuing system. Each sub-network is visited by only one class of customers and we denote by S_r the set of servers visited by class r. The total number of single-class queuing sub-networks is equal to the number of different types of classes in the original network L, i.e. it is equal to the number of different pallet types present in the FMS. Sub-networks are closed, that means the number of customers flowing in the sub-network ℓ_r is constant and equal to N_r.
Since each single-class is a Gordon-Newell network, the stationary solu-

tion of the sub-network ℓ_r is known and has the following form (Gordon and Newell, 1967):

$$P(\mathbf{n}) = \frac{1}{G_i} \prod_{i \in S_r} \left[\prod_{1}^{N_r} \frac{V_{i,r}}{\mu_{i,r}} \right] \tag{4.15}$$

where $\mathbf{n}$ is a vector representing the global state of the system, i.e. the number of customers at each server, $V_{i,r}$ is the visit ratio of class r at server i and G_i is a normalization constant.

Therefore, the performance of the multiple-class in the original network is approximated by the performance of a set of single-class sub-networks ℓ_r. The methodology is approximate because it assumes independence among the different classes. However, in order to capture the behavior of the whole system in the decomposed single-class sub-networks, servers are analyzed in isolation by considering all the classes. In particular each server of the decomposed network is analyzed with a multiple class aggregation technique that allows the calculation of the throughput of single server systems with multiple classes.

Each single-class sub-network has stations with load-dependent service rates; this allows the possibility of considering in sub-networks ℓ_r the effect of the rest of the system by selecting ad hoc values of load dependent service rates $\mu_{i,r}(n_{i,r})$. In particular, proper service rates are calculated by analyzing the servers of the network in isolation mode. The analysis of servers in isolation mode, i.e. the stations functioning independently from the rest of the system, provides the throughput of stations considering more than one class. Since the throughput of servers in isolation represents a good approximation of conditional throughput, we analyze the servers in isolation to approximate the service rates of load dependent servers. In particular service rates $\mu_{i,r}(n_{i,r})$ of servers in the original network are imposed equal to the conditional throughput $\nu_{i,r}(n_{i,r})$ of class r at the corresponding server i analyzed in isolation. Since each server is analyzed by considering all the classes that circulate in the system with their specific processing times, the method properly captures the dependencies among different classes providing very accurate results.

The method moves iteratively from the analysis of single-class queueing sub-networks to multiple-class servers in isolation mode until the set of decomposed systems fully represents the original multiple-class queueing network L.

After a set-up of network parameters, the first step of the method is to calculate the average arrival rate of customers entering in each node of the network. This task is rather straightforward because it is necessary

to solve single-class closed sub-networks by using the product form solution of equation (4.15).
After the average arrival rates of single servers are known, servers are analyzed in isolation mode so that their steady-state probabilities can be calculated as also the conditional throughput of the station. The conditional throughput of single servers analyzed in isolation mode is assumed (notice that an approximation is introduced in this step) to be the service rate of load dependent servers of the original network. Again arrival rates at servers are calculated solving the new single-class closed sub-queueing networks, and so on. The method continues until the unknown parameters $\mu_{i,r}(n_{i,r})$ converge to stationary values and performance indicators of the original network can easily be evaluated. It is possible to demonstrate that the method always converges because the service rates $\mu_{i,r}(n_{i,r})$ are the unknowns of a fixed point problem.

Summarizing, the described method can be implemented in an efficient algorithm with the following steps:

1 **Initialization**. The unknowns of the problem, i.e. service rates, are put equal to some initial values. To speed the convergence of the algorithm service rates are put equal to service rates of pallets in the real system.

2 **Calculation of arrival rates**. Average arrival rates of customers at each server of the network are calculated in this step. The solution of single-class closed queueing networks provides the values of $\lambda_{i,r}(n_{i,r})$ for each server and each class.

3 **Analysis of servers in isolation**. The solution of servers analyzed in isolation mode provides the conditional throughputs $\nu_{i,r}(n_{i,r})$ of the servers for each class. Then the following equation is imposed: $\nu_{i,r}(n_{i,r}) = \mu_{i,r}(n_{i,r})$ for each $i = 1, ..., S$ and $r = 1, ..., R$.

4 **Convergence condition**. The algorithm exits if service rates of the last two iterations differ by a quantity that is lower than a pre-defined small value ϵ_1, otherwise the algorithm moves again to step 2.

The analysis of servers in isolation mode requires the solving of a multiple-class load dependent server. The analytical model for studying this type of system is described in the next section.

6.4 Analysis of servers in isolation mode

In the overall algorithm it is necessary to use a performance evaluation tool for the calculation of the throughput in isolation. For classical queu-

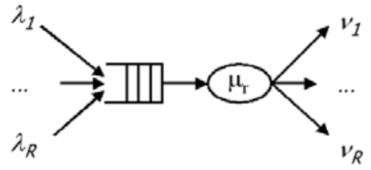


Figure 4.12. Multiple-class server in isolation.

ing servers (e.g. M/M/1, M/M/m, M/Er/1, etc.) solutions are already well known in technical literature and we do not deal with them. For multiple-class queueing servers exact methods are not available and existing approximate techniques perform poorly in practical applications.

For simplicity of exposition we assume parallel machines FMS, that is all the machines are equal and pallets can be routed indifferently to any machine of the system and, therefore, each server is visited by all the R classes circulating in the system. Also we drop for simplicity index i in this section referring equations and notations to the generic analyzed server i of the network. The arrival rates of customers at server i are known and denoted with $\lambda_r(n_r)$ (see also Figure 4.12); we remark that these values are known because they have been calculated at step 2 of the main algorithm. The service rate of the server depends on the type of customer the machine is working. The ranking rule of customers in the queue is the random rule, i.e. the next customer to be served is chosen randomly among the customers present in the queue, that approximates well the traditional FCFS (First Come First Served).

The system is completely defined by the vector state $(n_1, ..., n_R, s)$ where the first R components represent the number of customers in the analyzed server, and the last component represents the state of the server: $s = 0$ if the station is empty and $s = r$ (with $r = 1, ..., R$) if the station is serving customers of class r. The system behavior can be modelled as a Markovian process where the above vector represents its possible states. In order to evaluate the throughput of the analyzed server, it is necessary to solve such a Markov chain and therefore to calculate the stationary probabilities for each state of the system. An exact solution can be found by solving the Markov chain with traditional numerical methods. However, this approach is practicable only for very small systems. For large systems the total number of states of the Markov chain becomes too big and the approach can be unfeasible for practical applications. In the remainder of this section the approximate method to analyze the performance of multiple-class servers is described in detail.

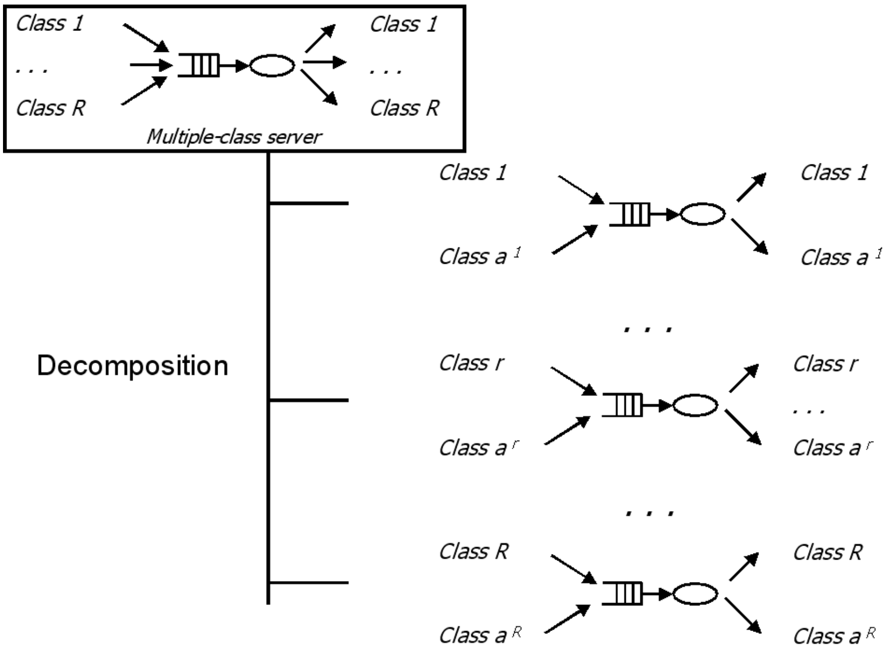


Figure 4.13. Decomposition of a multiple-class server in isolation mode.

The objective is to calculate the stationary probabilities $P(n_1, ..., n_R)$, on the basis of which the throughput of the server can be evaluated for all the classes. To do this we decompose (see Figure 4.13) the system in R sub-systems or building blocks that reproduce, in an approximate way, the behavior of the original server. Building blocks are composed of a server visited by two types of customers: the original class r and an aggregated class a^r that represents the behavior of all classes in the original system except class type r (see also Figure 4.14). We indicate with n_r^r and n_a^r the number of customers in the building block of classes r and a^r respectively, with $\lambda_r^r(n_r^r, n_a^r)$, $\lambda_a^r(n_r^r, n_a^r)$ their corresponding arrival rates and with $\mu_r^r(n_r^r)$, $\mu_a^r(n_a^r)$ their corresponding service rates. The generic state is represented by the vector n_r^r, n_a^r, s where s is the state of the server (i.e. $s = 0, r, a$). If we know the arrival and service rates of both classes it is possible to calculate the throughput of building blocks by solving numerically the related Markov chain. Therefore, for each building block deriving from the decomposition of the server in isolation mode it is necessary to find proper values of arrival and service rates for the two classes of customers (i.e. the classes r and a^r). These parameters must be appropriately selected so that the throughput of

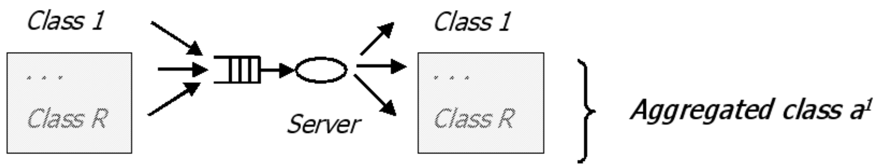


Figure 4.14. Aggregation of customers in one class.

parts that exit from the server is equal for all the building blocks. A set of equations concerning the equality of throughput among the building blocks allows the evaluation of building block parameters.

The decomposition reduces significantly the complexity of the analysis. Indeed, the analysis reduces to solve small two-class Markovian servers rather than a complex R-class server. The method is known as the aggregation technique and it has been first proposed by Baynat and Dallery in (Baynat and Dallery, 1996). However, the method proposed in the paper differs from that of Baynat and Dallery both in the underlying equations and in its application.
The differentiation among service rates in the decomposed building blocks allows a distinction to be made among the different classes of the network. This distinction is very important in cases in which service rates are very different and each one has low variability. Indeed, if service times are random variables, the advantages of analyzing them separately decreases as their variability increases because they tend to mix themselves into an average common distribution. On the contrary, if services times are deterministic, it is important to analyze them separately because they cannot be properly modelled by means of an average distribution.

Let us now analyze the building block parameters: number of customers, arrival and service rates. The number of customers for the classes r and a^r are equal respectively to:

$$\begin{cases} N_r^r = N_r & r = 1, \ldots, R \\ N_a^r = \sum\limits_{\forall k, k \neq r} N_k & r = 1, \ldots, R \end{cases} \tag{4.16}$$

The arrival rates of customers are state dependent type, that is they are function of the number of clients of both classes present in the system:

$$\lambda_r^r(n_r^r, n_a^r) = \left\{ \begin{array}{llll|l} n_a^r = 0 & n_a^r = 1 & n_a^r = ... & n_a^r = N_a^r & \\ \hline \lambda_r^r(0,0) & \lambda_r^r(0,1) & ... & \lambda_r^r(0) & n_r^r = 0 \\ \lambda_r^r(1,0) & \lambda_r^r(1,1) & ... & \lambda_r^r(1) & n_r^r = 1 \\ ... & ... & ... & ... & n_r^r = ... \\ 0 & 0 & 0 & 0 & n_r^r = N_r \end{array} \right. \tag{4.17}$$

If all clients of class a^r are in the analyzed system, i.e. $n_a^r = N_a^r$, parameter λ_r^r depends only on the number of clients of the class r. If all the class r customers are in the analyzed system, it is not possible a new arrival of class type r and therefore the arrival rate must be null. Given a certain value of n_r^r, the arrival rate λ_r^r increases as n_a^r increases because the customers of class r in the rest of the FMS have a lower competition to access to resources. Similarly, given a certain value of n_a^r, the arrival rate λ_r^r decreases as n_r^r increases. Similar equations and considerations hold for λ_a^r.

The state dependent service rates of servers in the queuing network take into account, even if not in an explicit way, the correlation among the different classes in the system. As a consequence, also the arrival rates, evaluated by solving the R single-class closed queueing subnetworks, consider the behavior of the other classes. Therefore, it seems reasonable to introduce another approximation by assuming that arrival rates at the building blocks, derived by the decomposition of the system, are function only of the number of customers of the same class type that stay at the server:

$$\left\{ \begin{array}{ll} \lambda_r^r(n_r^r, n_a^r) \simeq \lambda_r^r(n_r^r) & n_r^r = 0, 1, ..., N_r^r; r = 1, ..., R \\ \lambda_a^r(n_r^r, n_a^r) \simeq \lambda_a^r(n_a^r) & n_a^r = 0, 1, ..., N_a^r; r = 1, ..., R \end{array} \right. \tag{4.18}$$

Then we can say that the arrival rates of customers of class type r are approximately equal to those of the original system calculated at step 2 of the main algorithm:

$$\lambda_r^r(n_r^r) \simeq \lambda_r(n_r^r) \quad n_r^r = 0, 1, ..., N_r^r; r = 1, ..., R \tag{4.19}$$

The arrival rates of customers of the aggregated class type have to be calculated as a weighted sum of arrival rates of all the classes except type r.

We define $E^r(n_a^r)$ as the set containing all the possible combinations of numbers of customers that are not of class type r and that can be present at the analyzed server at the same time:

$$E^r(n_a^r) = \left\{ (n_1, ..., n_{r-1}, n_{r+1}, ..., n_R) : n_a^r = \sum_{\forall s: s \neq r} n_s \right\} \tag{4.20}$$

and with $P^r(n_a^r)$ the probability of being in that set of states. The arrival rate of the aggregated class is calculated as the weighted sum of arrival rates in the original system where the weights correspond to the probability of having n_a^r customers in the analyzed server:

$$\begin{cases} \lambda_a^r(n_a^r) = \dfrac{\sum\limits_{E^r(n_a^r)} \lambda_a^r(n_a^r) \cdot P^r(n_a^r)}{\sum\limits_{E^r(n_a^r)} P^r(n_a^r)}, & n_a^r = 0, ..., N_a^r - 1; r = 1, ..., R \\ \lambda_a^r(n_a^r) = 0, & n_a^r = N_a^r; r = 1, ..., R \end{cases} \tag{4.21}$$

Approximating the probability $P^r(n_a^r)$ by the product of the marginal probabilities $P_s(n_s)$ of each class s, other than class r, that visit the server, it is possible to write the following equations for any $r = 1, \ldots, R$:

$$\begin{cases} \lambda_a^r(n_a^r) = \dfrac{\sum\limits_{E^r(n_a^r)} \sum\limits_{\forall s: s \neq r}^{R} \lambda_s(n_s) \cdot \prod\limits_{\forall s: s \neq r} P_s(n_s)}{\prod\limits_{\forall s: s \neq r} P_s(n_s)}, & n_a^r = 0, ..., N_a^r - 1 \\ \lambda_a^r(n_a^r) = 0, & n_a^r = N_a^r \end{cases} \tag{4.22}$$

where $P_s(n_s)$ is the marginal probability that n_s customers of class s are present at the server. A good estimate for the values of marginal probabilities is taken from the last iteration. Similarly, service rates of the aggregated class are calculated as a weighted average of service rates in the original system for any $r = 1, \ldots, R$:

$$\mu_a^r(n_a^r) = \frac{\sum\limits_{E^r(n_a^r)} \sum\limits_{\forall s: s \neq r}^{R} \frac{\mu_s(n_s) n_s}{\sum\limits_{\forall s: s \neq r} n_s} \cdot \prod\limits_{\forall s: s \neq r} P_s(n_s)}{\prod\limits_{\forall s: s \neq r} P_s(n_s)}, \quad n_a^r = 1, ..., N_a^r - 1 \tag{4.23}$$

Service rates of class r in the building blocks correspond to the real service rates of that class. The equations (4.22) and (4.23) together with original arrival and service rates of class r allow the evaluation of the parameters of each building block. A different way of evaluating service rates for the aggregated class has been proposed by Baynat and Dallery in (Baynat and Dallery, 1996). However, in (Baynat and Dallery, 1996), as also in our case, the evaluation of marginal probabilities may be characterized by large errors when the number of customers is small. To deal with this problem other equations are used instead of (4.23).

If we denote with X_r and X_a^r respectively the throughput of class type r and the related aggregated class, it is possible to write the following equations of conservation of flow:

$$X_{r-1} + X_a^{r-1} = X_r + X_a^r, \quad r = 2, ..., R \tag{4.24}$$

The equations (4.22) and (4.24) together with original arrival and service rates of class r allow evaluating the parameters of each building block under the assumption that the aggregated class has state independent service rates: $\mu_a^r(n_a^r) = \mu_a^r$ for $n_a^r = 1, ..., N_a^r$. In practice these equations form a fixed point problem.
The described method can be implemented by the following algorithm:

1 **Initialization**. Calculate N_r^r and N_a^r. Calculate $\lambda_r^r(n_r^r)$ from relations (4.19). Assign initial values to the following coefficients:
$k = 0$, with k representing the number of iterations
$C(k) = 1$, with $C(k)$ representing a convergence coefficient at the iteration k.

2 **Analysis of building blocks**.
$k = k + 1$. For $r = 1, ..., R$:

- Calculate $\lambda_a^r(n_a^r)$ from the relation (4.22).
If $k = 1$ than calculate μ_a^r from equation (4.23), otherwise use the following equation:

$$\mu_a^r(k) = \frac{1}{C(k-1)} \cdot \frac{A_r(k-1))}{B(k-1)} + \left(1 - \frac{1}{C(k-1)}\right) \mu_a^r(k-1) \frac{A(k-1)}{D(k-1)} \tag{4.25}$$

with $A_r(k-1)$ representing the total weighted throughput out from building block r, $B(k-1)$ representing the total throughput out from the isolated server taking into account all the building blocks $r = 1, \ldots, R$, and $D(k-1)$ representing the average total weighted throughput averaged overall building blocks. These quantities are calculated in the previous iteration $k-1$ using equations (4.26)
- Calculate the marginal probabilities $P_r(n_r)$ for $n_r = 0, ..., N_r$.
- Solve the Markov chain of the r-th building block.
- Calculate the throughput values of the class r $X_r(k)$ and the aggregated class $X_a^r(k)$.

3 **Calculate throughput**. Calculate the average equivalent throughput referred to units of the first part type with the following quan-

tities:

$$\begin{aligned} A_r(k) &= \sum_{r=1}^{R} \left(X_r \cdot \tfrac{\mu_1}{\mu_r} + X_a^r \cdot \tfrac{\mu_1}{\mu_a^r} \right) \\ B(k) &= \sum_{r=1}^{R} X_r \cdot \tfrac{\mu_1}{\mu_r} \\ D(k-1) &= \tfrac{1}{R} \sum_{r=1}^{R} A_r(k) \end{aligned} \tag{4.26}$$

4 **Convergence condition**. Calculate the precision coefficient as:

$$precision(k) = \frac{|A(k) - B(k)|}{B(k)} \tag{4.27}$$

If $|precision(k) - precision(k-1)| > \epsilon_2$ than $C(k) = C(k-1) + 1$ and go to step 2, otherwise exits from the algorithm.

6.5 Numerical results

The purpose of this paragraph is to evaluate the accurateness of the proposed method in estimating the average throughput of FMS with deterministic processing times. To do this, the experimentation has been carried out by testing the method on several test cases and one real case. Numerical results from the proposed method are compared with those obtained by a discrete event simulation model of the analyzed FMS and by the application of the Mean Value Analysis algorithm, originally proposed in (Reiser and Levemberg, 1980), extended to the multiple-class case (Papadopoulos et al., 1993). In detail, the proposed algorithm has been implemented with values of parameters $\epsilon_1 = \epsilon_2$ equal to 10^{-4}. Results from simulation are considered the reference values to which the proposed method and the MVA are compared. The errors committed by the proposed method and the MVA algorithm are calculated as follows:

$$\%Err_{Analytical\ Value} = \frac{Analytical\ Value - Simulation\ Value}{Simulation\ Value} \cdot 100 \tag{4.28}$$

Errors are calculated both on the single class performance measures, e.g. average throughput per class, and on aggregated indexes, e.g. total equivalent throughput corresponding to the total number of pallets machined by the system in a certain period calculated by weighting the single class values with their processing times.

Test cases. The test cases have been defined by varying the number of pallets and machining centers for different product mix situations. Three levels of machining centers are considered from 2 up to 6 machines.

Table 4.3. Test case: Product mix 1 with long processing times [s].

Pallet type	*Loading/unloading time*	*Transport time*	*Processing time*
1	60	90	from 500 to 3500 with steps of 250
2	60	90	500
3	60	90	2000

The number of pallets is varied from unity to a limiting value defined by the system (i.e. the system has reached the saturation level) or by the limits of the proposed method (e.g. over a certain level of complexity the proposed method is too slow for practical applications).

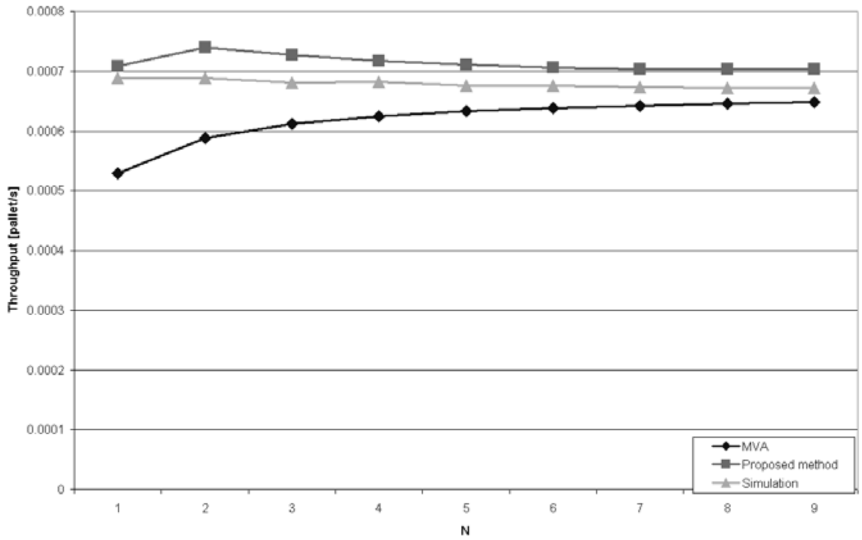
Let us consider the first product mix (data are reported in Table 4.3) characterized by having long processing times and consisting of three different types of pallets. The processing time of pallet type 1, denoted with t_1, is varied as shown in the table. For simplicity's sake, loading and travelling times have been considered to be the same for each pallet type; this assumption does not affect significantly the analysis because these times are much shorter than processing times. The number of machines in the FMS is 2 for a total number of servers in the modelled queuing network of 4. Finally the number of pallets per type N_r is equal for all types, i.e. $N = N_1 = N_2 = N_3$.

Figures 4.15-4.17 show the class average throughput for three different values of t_1; average values from simulation have a 95% half confidence interval around 10^{-6}. Notice that when the first pallet type assumes long processing times an increase of N, which corresponds to add in the system three pallets in total, penalizes the pallet with shortest processing time. This effect is shown in Figure 4.15a for classes 1 and 2 and is clearly visible in Figure 4.17b where the throughput of class 2 decreases as N increases. Notice that MVA never captures this phenomena.

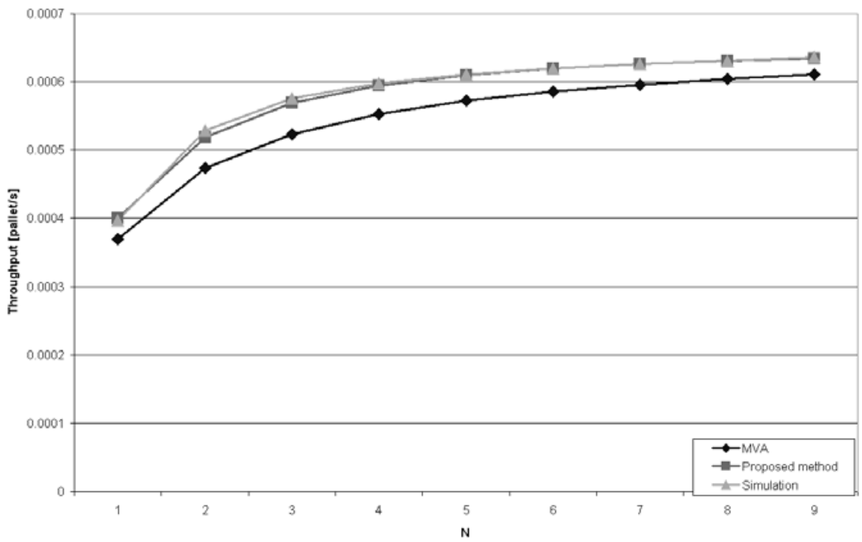
For each method (i.e. Mean Value Analysis (MVA), Proposed Method (PM) and Simulation (SIM)) the equivalent average throughput of the system is calculated taking the second pallet type as the reference for weighting throughput of single classes with the equation:

$$Th^{eq} = \sum_{r=1}^{R} Th_r \cdot \frac{t_r}{t_2} \tag{4.29}$$

where Th_r and t_r are the average throughput and processing time of pallet type r respectively. The equivalent throughput is shown in Figure 4.18 for three different values of t_1. Equivalent throughput estimated by



(a) Classes 1,2



(b) Class 3

Figure 4.15. Product mix 1: average value of class throughput as a function of N with $t_1 = 500s$.

simulation has a half confidence interval lower than $5 \cdot 10^{-5}$. It appears from the graphs that the average equivalent throughput estimated by

Table 4.4. Test case: Product mix 2 with brief processing times [s].

Pallet type	*Loading/unloading time*	*Transport time*	*Processing time*
1	60	90	2 machines: 10–2000 in steps of 10 4-6 machines: 100–2000 in steps of 100
2	60	90	300
3	60	90	900

the proposed method is quite accurate. Percentage errors of MVA and the proposed method on equivalent throughput are shown in Figure 4.19 for all the analyzed cases. Errors of MVA closely depend on the number of pallets in the system and they are never below the 3% in the most optimistic case. It seems qualitatively also that accurateness of MVA improves as t_1 increases. Errors of the proposed method seem to be more or less invariant after $N = 3$ and they are generally below 2 %.

The data of the second analyzed product mix is reported in Table 4.4. Again three types of pallets are considered. In this case processing times at the machining center are not as long as in the first test case and loading and travelling times are the same for each pallet type. Processing time of pallet type 1 is varied as shown in the table. The number of machines in the FMS is 2-4-6 for a total number of nodes in the modelled queuing network of 4-6-8. Finally the number of pallets per type N_r is equal for all types i.e. $N = N_1 = N_2 = N_3$.
The equivalent throughput and corresponding errors are shown in Figures 4.20–4.24. Equivalent throughput estimated by simulation has a 95 % half confidence interval lower than $5 \cdot 10^{-5}$. Accurateness of analytical methods obviously decreases as the number of machines in the system increases. The same considerations of product mix 1 holds also in these cases. A particular case is for six machines and $N = 10$ where MVA performs very well compared with simulation.

Real case. We consider a real part mix composed of five different part types of the automotive sector. Processing times of the part mix can be assumed deterministic and are reported in Table 4.5. The variability among processing times of different pallets is quite large and it could affect the analysis if it is not properly considered; indeed pallet types 1 and 3 require long processing times that are at least three times longer than those of pallet types 4 and 5. The analyzed FMS is composed of CNC machining centers characterized by tool changing time (from chip to chip) equal to 4 s and a pallet changing time equal to 15 s. Pallets

Table 4.5. Real case: part mix data [min].

Pallet type	*Processing time* [min]	*Loading/unloading time* [min]
1	24.0	0.7
2	17.1	2.0
3	24.2	1.0
4	6.6	2.0
5	9.7	0.7

are moved by a carrier at an average speed of 60 m/min and the pallet transport time is 90 s.

The experimentation has been carried out by varying two factors: the number of machines and the number of pallets circulating in the system. Since part types have different processing times, incrementing by a unit the number of pallets of a certain part type instead of another one has a different impact on the system performance. Therefore, we define an indicator ΔW that synthesizes the workload in time deriving from the introduction of a new pallet in the system. Starting from a minimal combination of pallets in which every part type has only one pallet that circulates in the FMS, the number of pallets is increased and this increment is measured by the equivalent workload calculated as follows:

$$\Delta W = \sum_{r=1}^{R} t_r \cdot \Delta N_r \tag{4.30}$$

where t_r and ΔN_r are respectively the processing time and the pallet variation of each part type in respect to the minimal combination. Table 4.6 reports the combination of pallets analyzed in numerical experiments.

Results are shown in the following graphs reported from Figure 4.26 to Figure 4.28. The graphs show the equivalent throughput, i.e. the throughput averaged taking into account the fact that pallets have different processing times, and the relative errors of the proposed method and MVA. It can be noticed in all graphs the proposed method is more accurate than MVA because it fully captures the different processing times of pallets. Also the error of the proposed method seems to be independent from the number of pallets in the system. On the contrary, MVA results depend on the additional workload confirming the several numerical results appeared in literature: the accuracy of the MVA improves as the population of the network increases. The error of the

Table 4.6. Real case: pallet combinations.

Pallet type 1	*Pallet type 2*	*Pallet type 3*	*Pallet type 4*	*Pallet type 5*	*Total pallets*	ΔW [%]
1	1	1	1	1	5	-
1	1	1	1	2	6	8
1	2	1	1	1	6	22
1	2	1	2	1	7	30
1	1	1	5	2	10	42
1	1	1	5	3	11	50
1	3	1	3	1	9	60
1	1	3	2	1	8	70
1	2	1	5	4	13	80
1	4	1	4	1	11	90
2	2	2	2	2	10	100
1	1	4	1	3	10	109
3	1	1	4	5	14	120
1	4	2	2	4	13	130
1	1	5	3	1	11	140
3	2	1	5	5	16	150
2	5	1	2	5	15	160
2	2	4	3	2	13	171
4	2	1	5	5	17	181
3	1	3	5	5	17	190
3	3	3	3	3	15	200

proposed method on the equivalent throughput is always lower than 3% in the two-machine case against an error of the MVA that is never lower than 6%. With regard to the 3-machine case the error of the proposed method is always lower than 3% against an error of the MVA that is never lower than 9%. In the 4-machine case the accuracy of both methods decreases even if results of the proposed method seem to be still acceptable for practical applications.

In all the other cases tested by the authors and not reported in this paper the proposed method always provides more accurate results than MVA. The algorithm converges in reasonable times for practical applications and obviously the computational effort depends on the number of classes and servers in the network.

7. Conclusions

In this chapter we have proposed a methodology for the detailed design of AMSs. In the methodology approximate analytical methods are used to calculate the performance of Dedicated Manufacturing Systems and

Flexible Manufacturing Systems. The proposed methods provide very accurate results and can be used instead of simulation in the first stage of the capacity planning process.

References

Baynat, B. and Dallery, Y. (1993a). Approximate techniques for general closed queueing networks with subnetworks having population constraints. *European Journal of Operational Research*, 18(1):250–264.

Baynat, B. and Dallery, Y. (1993b). A unified view of product-form approximation techniques for general closed queueing networks. *Performance Evaluation*, 18(1):205–224.

Baynat, B. and Dallery, Y. (1996). A product–form approximation method for general closed queueing networks with several classes of customers. *Performance Evaluation*, 24:165–188.

Buzacott, J. (1967). *Markov chain analysis of automatic transfer line with buffer stock*. PhD thesis, Univeristy of Birmingham.

Buzacott, J. and Shantikumar, J. (1993). *Stochastic models of Manufacturing Systems*. PTR Prentice Hall.

Buzacott, J. and Yao, D. (1986). Flexible manufacturing systems: A review of analytical models. *Management Science*, 32:890–905.

Colledani, M., Matta, A., and Tolio, T. (2003). Performance evaluation of production lines with finite buffer capacity producing two different products. *Proceedings of the 4th Aegean International Conference on Design and Analysis of Manufacturing Systems*.

Colledani., M., Matta, A., and Tolio, T. (2004). Analytical methods for optimal buffer space allocation. Technical Report 01-04, Politecnico di Milano-Dipartimento di Meccanica.

Dallery, Y., David, R., and Xie, X. (1988). An efficient algorithm for analysis of transfer lines with unreliable machines and finite. *IIE Transactions*, 20(3):280–283.

Dallery, Y. and Gershwin, S. (1992). Manufacturing flow line systems: A review of models and analytical results. *Queueing Systems*, 12.

de Almeyda, D. (1998). Agrgation des donnes pour l'valuation des performances de systmes flexibles de production par modles rseaux de files d'attente. *Operations research*, 32.

Dekleva, J. and Gaberc, M. (1994). Two phase method for designing flexible manufacturing systems. *International Journal of Production Research*, 32:683–691.

Frein, Y., Commault, C., and Dallery, Y. (1996). Modeling and analysis of closed-loop production lines with unreliable machines and finite buffers. *IIE Transactions*, 28:545–554.

Gershwin, S. (1987). An efficient decomposition algorithm for the approximate evaluation of tandem queue with finite storage space and blocking. *Operations Research*, 35(2):291–305.

Gershwin, S. (1994). *Manufacturing Systems Engineering*. PTR Prentice Hall.

Gershwin, S., Maggio, N., Matta, A., Tolio, T., and Werner, L. (2001). Analysis of loop networks by decomposition. *Third Aegean International Conference on Analysis and Modeling of Manufacturing Systems*, pages 239–248.

Gershwin, S. and Schor, J. (2000). Efficient algorithms for buffer space allocation. *Annals of Operations Research*, 93:91–116.

Gordon, W. and Newell, G. (1967). Closed queueing networks with exponential servers. *Operations Research*, 15:252–267.

Helber, S. (1999). *Performance analysis of flow lines with non-linear flow of material*. Springer.

Levantesi, R. (2002). *Analysis of multiple loop assembly/disassembly networks*. PhD thesis, Politecnico di Milano.

Levantesi, R., Matta, A., and Tolio, T. (2000). Performance evaluation of assembly/disassembly systems with deterministic processing times and multiple failure modes. *Special International Conference on Production Research*.

Levantesi, R., Matta, A., and Tolio, T. (2003a). *Analysis and Modeling of Manufacturing Systems*, chapter 9 - Performance evaluation of production lines with random processing times, multiple failure modes and finite buffer capacity - Part II: the decomposition, pages 201–219. Kluwer Academic Publishers.

Levantesi, R., Matta, A., and Tolio, T. (2003b). Performance evaluation of continuous production lines with machines having different processing times and multiple failure modes. *Performance Evaluation*, 51(2–4):247–268.

Marie, R. (1979). An approximate analytical method for general queuing netorks. *IEEE Transactions on Sofwtare Engineering*, 5:530–538.

Matta, A. and Bianchetti, D. (1997). Proposta di metodi analitici approssimati per lo studio delle prestazioni di linee automatizzate.

Matta, A. and Tolio, T. (1998). A method for performance evaluation of automated flow lines. *Annals of CIRP*, 47(1):373–376.

Matta, A., Tolio, T., Karaesmen, F., and Dallery, Y. (2001). An integrated approach for the configuration of automated manufacturing systems. *Robotics and CIM*, 17(1-2):19–26.

Le Bihan, H. and Dallery, Y. (1997). Homogenization techniques for the analysis of production lines with unreliable machines having different speeds. *Eur. J. Contr*, 3(3):200–215.

Papadopoulos, H., C., C. H., and Browne, J. (1993). *Queuing Theory in Manufacturing Systems Analysis and Design*. London: Chapman and Hall.

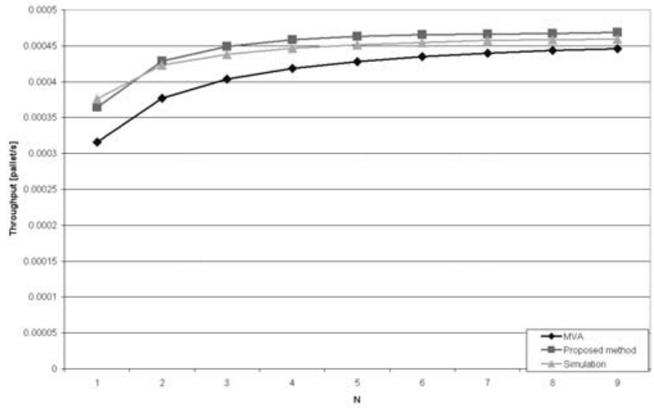
Reiser, M. and Levemberg, S. (1980). Mean value analysis of closed multiclass chain networks. *Journal of Associated Computer Machining*, 27:313–322.

Solot, P. and van Vliet, M. (1994). Analytical models for fms design optimization: a survey. *International Journal of Flexible Manufacturing Systems*, 6:209–233.

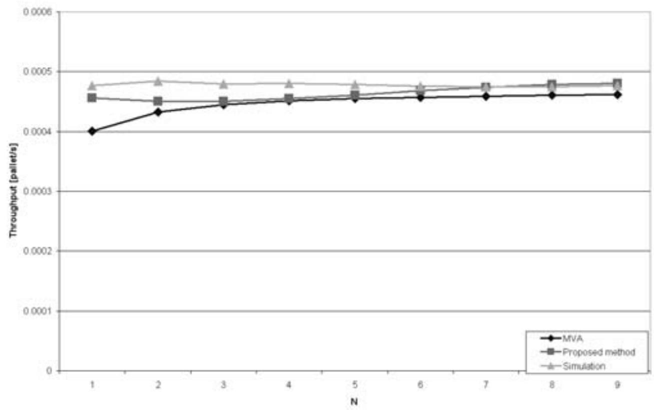
Starr, P. (1991). Integration of simulation and analytical submodels for supporting manufacturing decisions. *International Journal of Production Research*, 29:1733–1746.

Tolio, T., Gershwin, S., and Matta, A. (2002). Analysis of two-machine lines with multiple failure modes. *IIE Transactions*, 34(1):51–62.

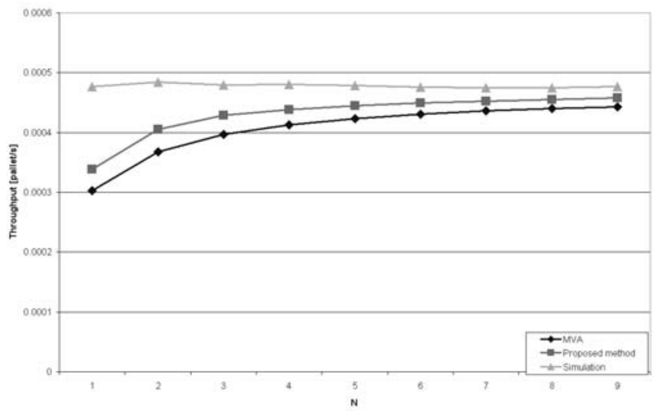
Yao, D. and Buzacott, J. (1985). The exponentialization approach to flexible manufacturing system models. *European Journal of Operational Research*, 24:410–416.



(a) Class 1

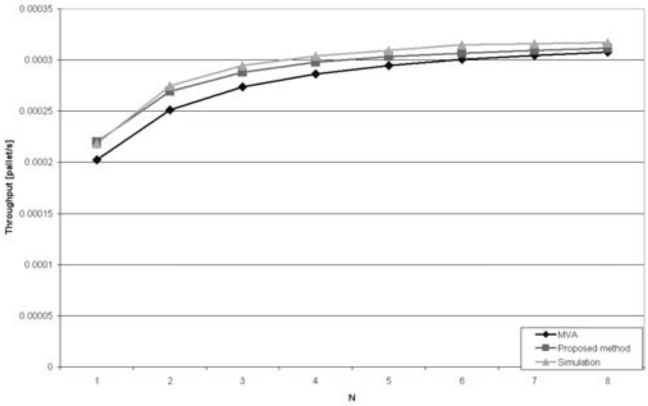


(b) Class 2

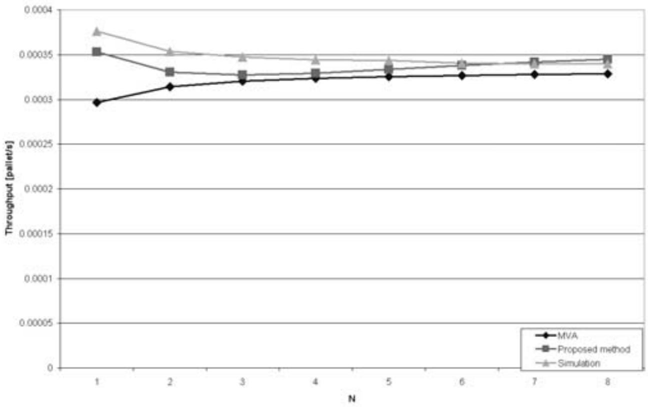


(c) Class 3

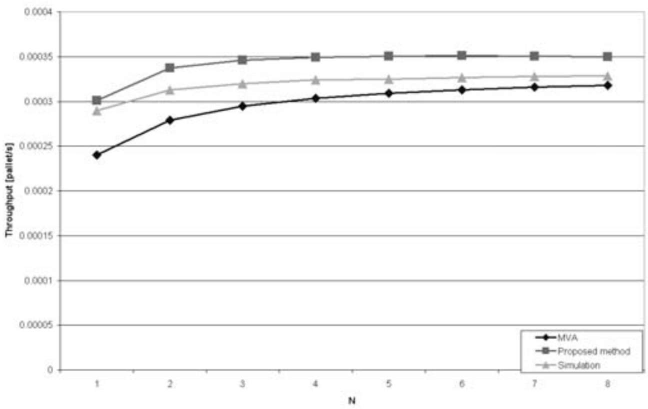
Figure 4.16. Product mix 1: average value of class throughput as a function of N with $t_1 = 1750s$.



(a) Class 1

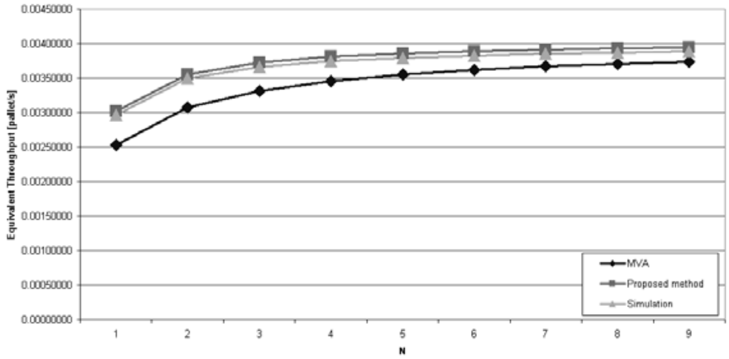


(b) Class 2

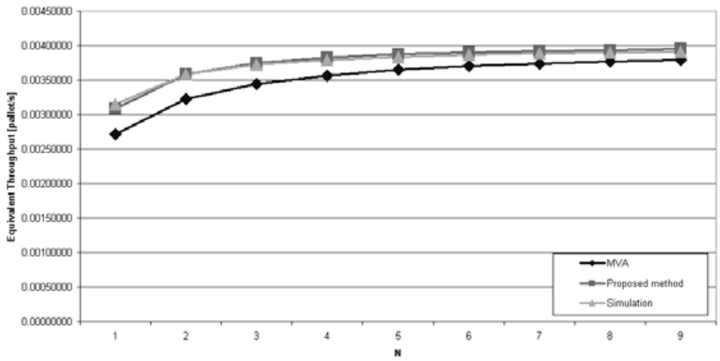


(c) Class 3

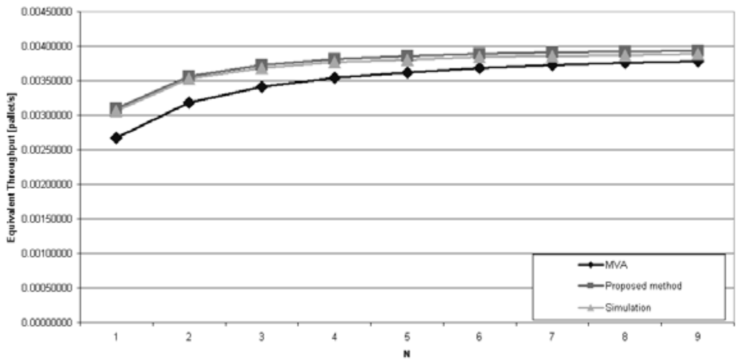
Figure 4.17. Product mix 1: average value of class throughput as a function of N with $t_1 = 3500s$.



(a) $t_1 = 500s$

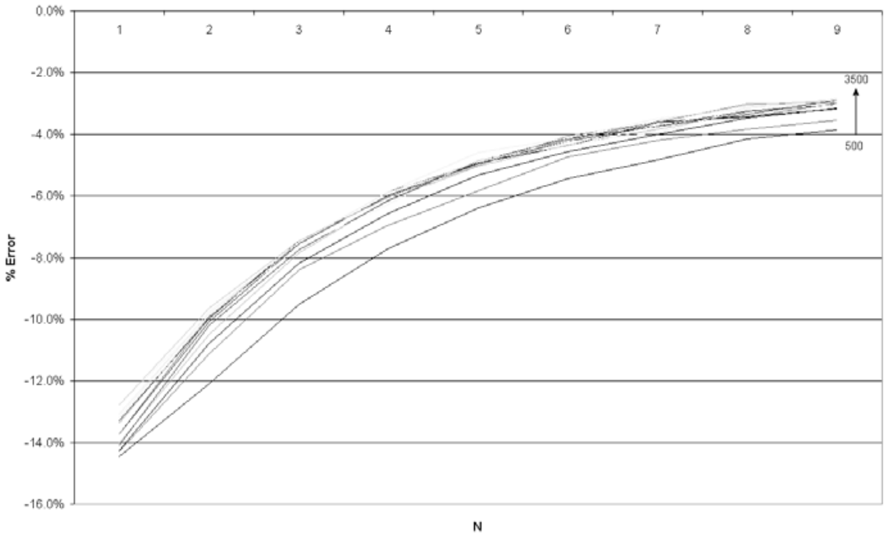


(b) $t_1 = 1750s$

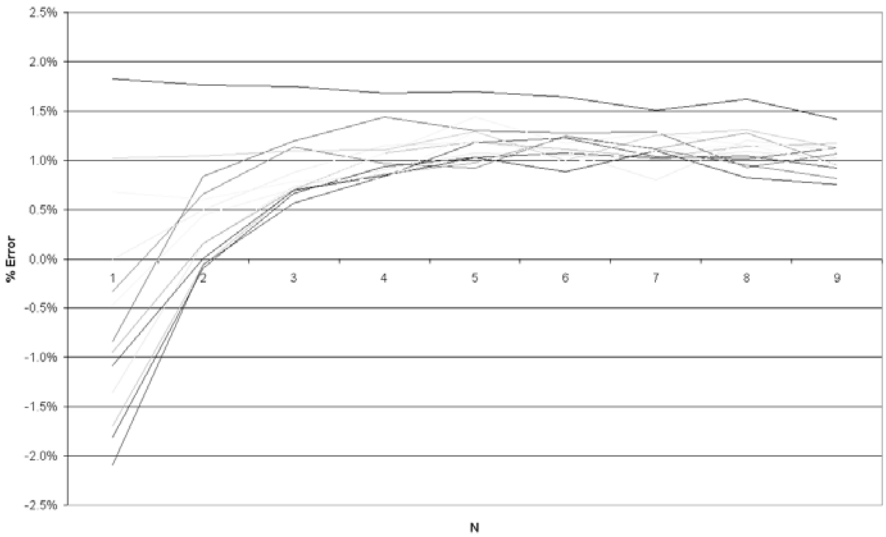


(c) $t_1 = 3500s$

Figure 4.18. Product mix 1: average value of equivalent throughput as a function of N for different values of t_1.

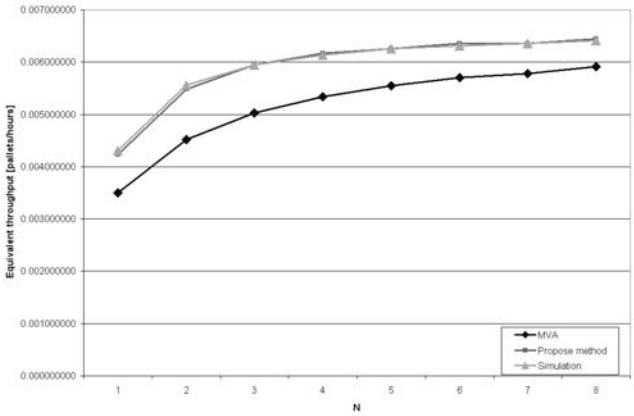


(a) MVA

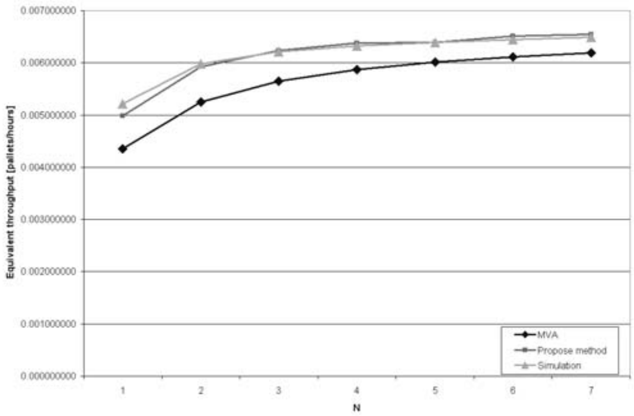


(b) Proposed method

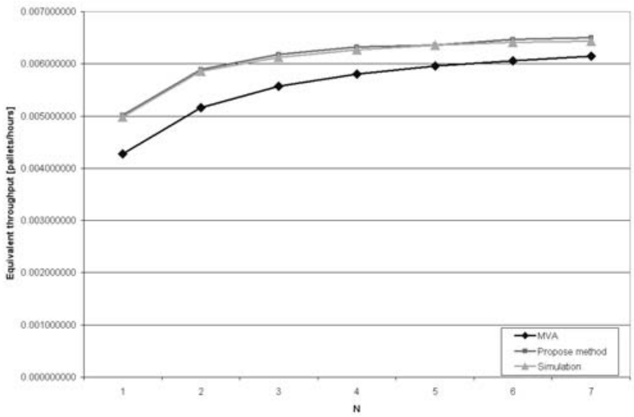
Figure 4.19. Product mix 1: percentage errors on equivalent throughput as a function of N for several values of t_1.



(a) $t_1 = 10s$

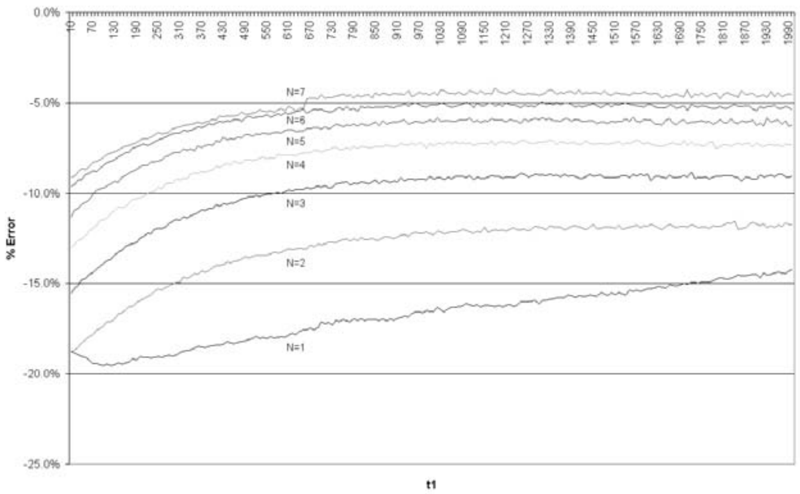


(b) $t_1 = 1000s$

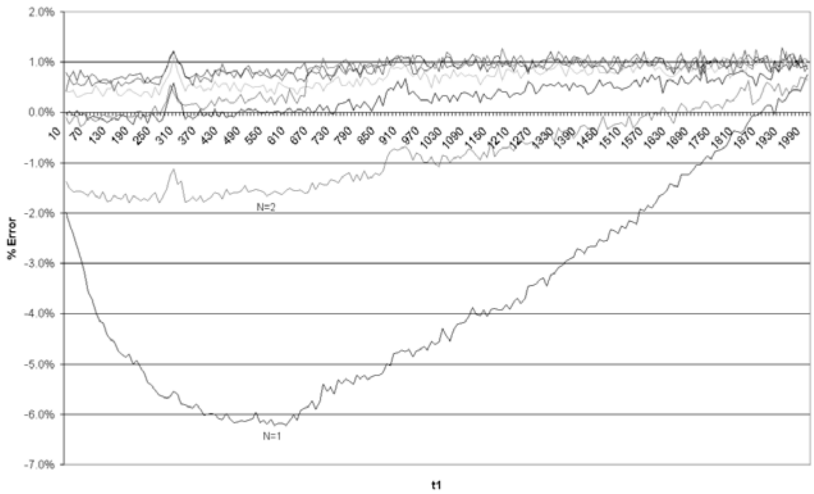


(c) $t_1 = 2000s$

Figure 4.20. Product mix 2: average PIPPO value of equivalent throughput with 2 machines as a function of N for different values of t_1.

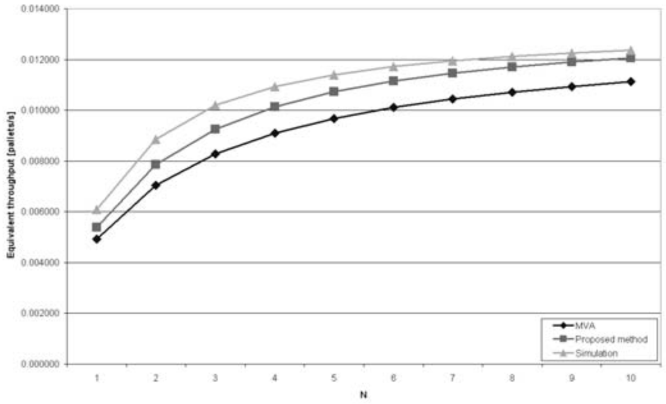


(a) MVA

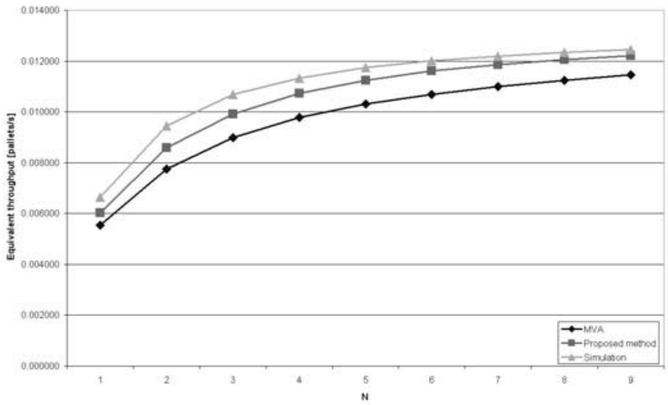


(b) Proposed method

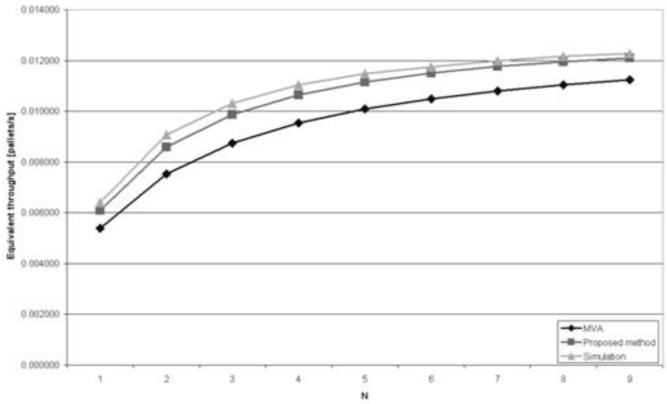
Figure 4.21. Product mix 2: percentage errors on equivalent throughput with 2 machines as a function of t_1 for different values of N.



(a) $t_1 = 10s$

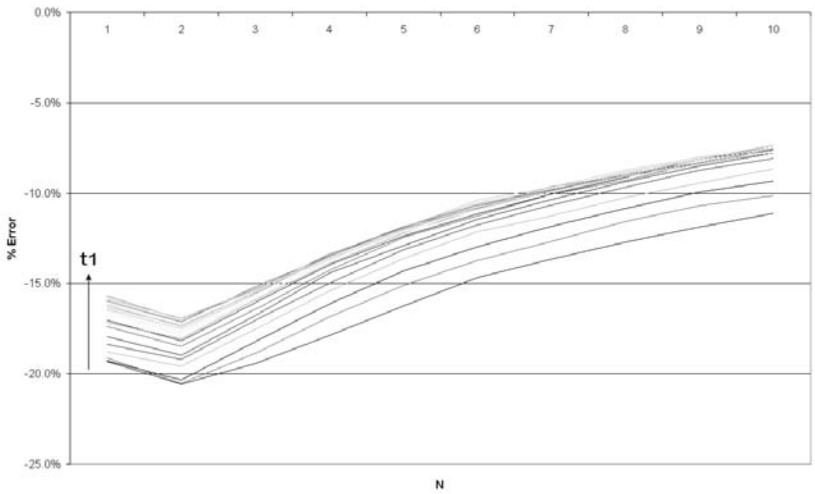


(b) $t_1 = 1000s$

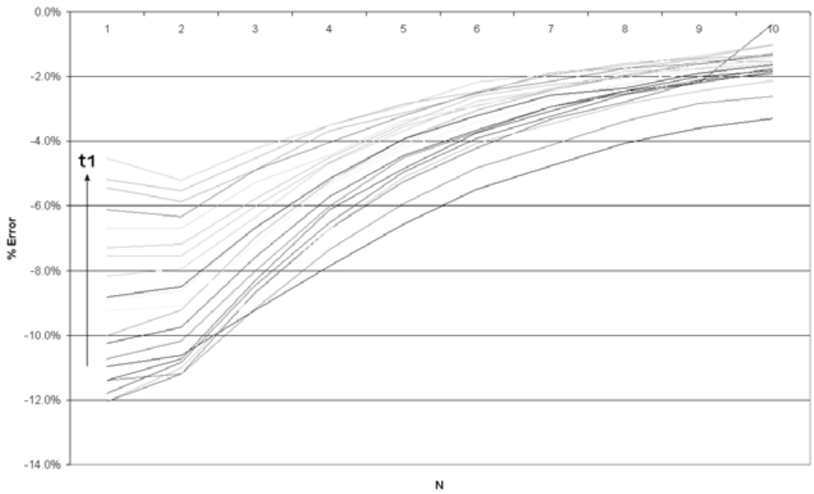


(c) $t_1 = 2000s$

Figure 4.22. Product mix 2: average value of equivalent throughput with 4 machines as a function of N for different values of t_1.

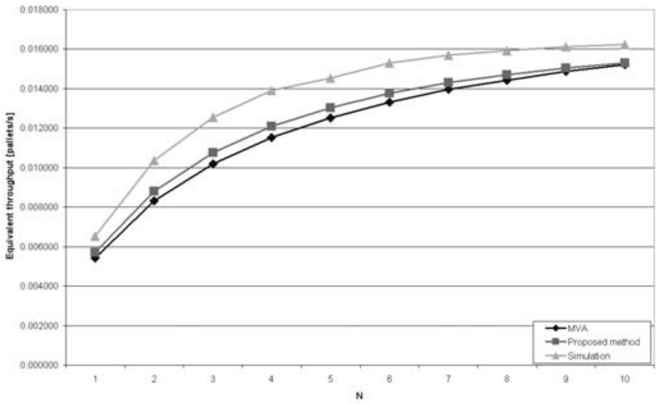


(a) MVA

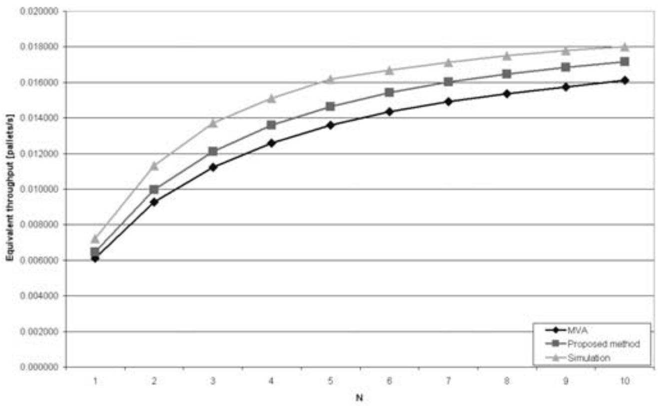


(b) Proposed method

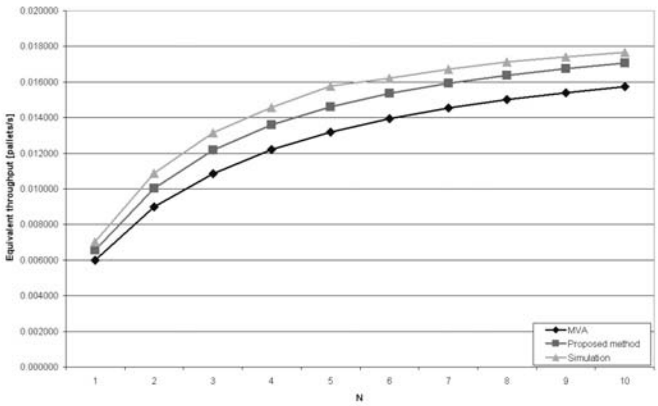
Figure 4.23. Product mix 2: percentage errors on equivalent throughput with 4 machines as a function of N for different values of t_1.



(a) $t_1 = 200s$

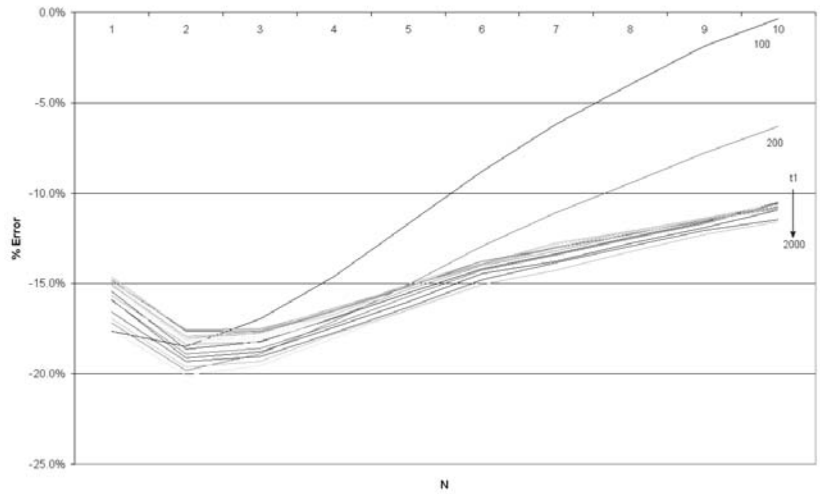


(b) $t_1 = 1000s$

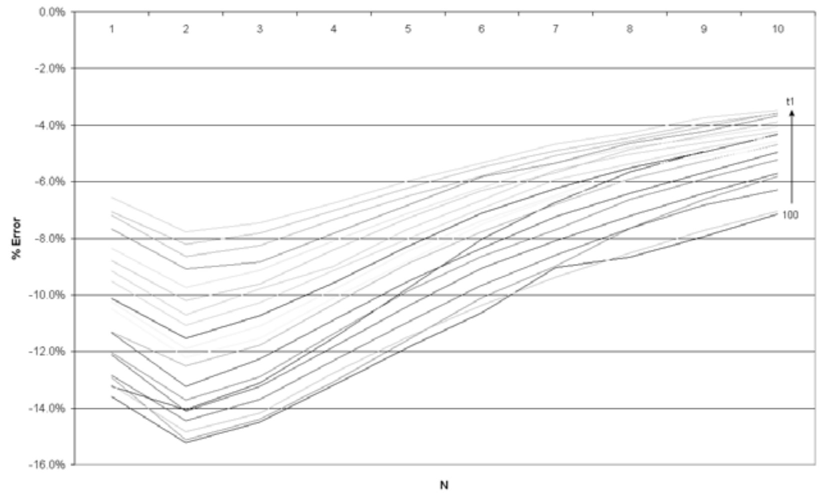


(c) $t_1 = 2000s$

Figure 4.24. Product mix 2: average value of equivalent throughput with 6 machines as a function of N for different values of t_1.

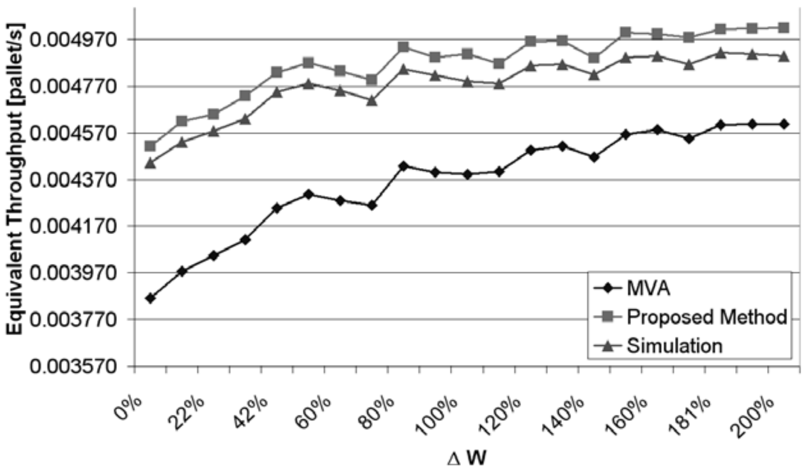


(a) MVA

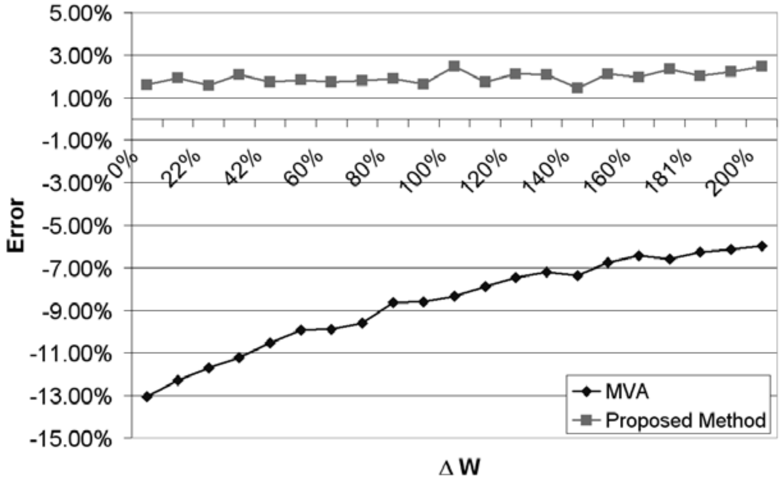


(b) Proposed method

Figure 4.25. Product mix 2: percentage errors on equivalent throughput with 6 machines as a function of N for different values of t_1.

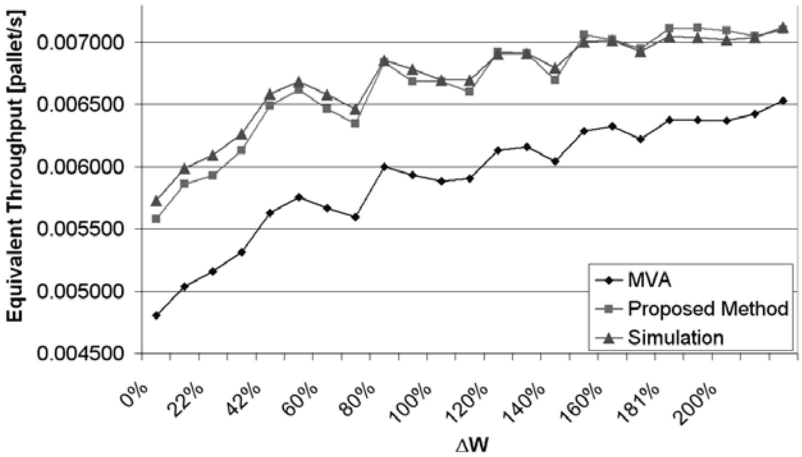


(a) Average equivalent throughput

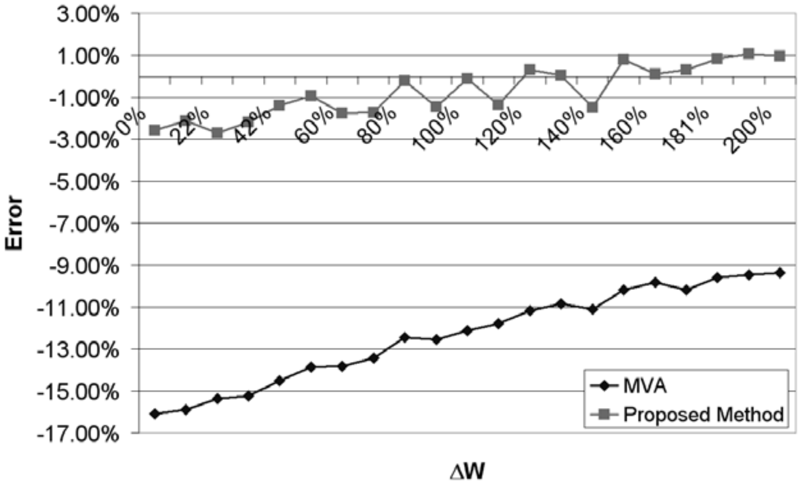


(b) Percentage errors

Figure 4.26. Real case with two-machines: average value of equivalent throughput and relative error as a function of additional workload.

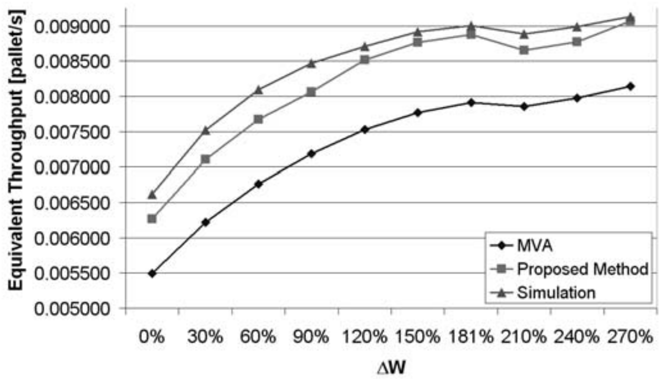


(a) Average equivalent throughput

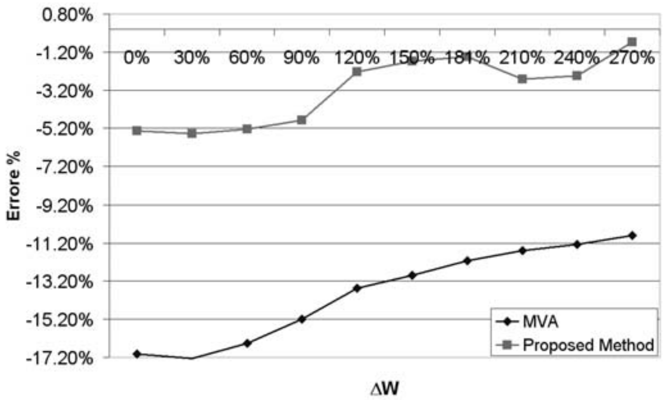


(b) Percentage errors

Figure 4.27. Real case with three-machines: average value of equivalent throughput and relative error as a function of additional workload.



(a) Average equivalent throughput



(b) Percentage errors

Figure 4.28. Real case with four-machines: average value of equivalent throughput and relative error as a function of additional workload.

Chapter 5

SELECTING CAPACITY PLAN

Dealing with non-stochastic uncertainty in the capacity planning process

A. Anglani, P. Caricato, A. Grieco, F. Nucci

Dipartimento Ingegneria dell'Innovazione

Università degli Studi di Lecce

Via per Monteroni, 73100 Lecce, Italy

{alfredo.anglani,pierpaolo.caricato,antonio.grieco,francesco.nucci}@unile.it

Abstract Nowadays, the frequency of decisions related with the configuration and capacity evaluation of manufacturing production system is increasing in more and more industrial sectors. In such a context, decision-makers have to take their actions in shorter times than they ever did in the past. This problem has increased in complexity because of the necessity to take into account all the sources of variability and each related level of uncertainty in the available data definition. For such a reason the capacity plan selection process is still an open question. In this chapter, a Decision Support System has been developed to help decision-makers to take productive capacity planning decisions according to uncertain characterization of the market evolution. The proposed methodology can be used to take strategic decisions over a long term programming horizon, allowing an effective comparison of user-defined strategies according to user-defined efficiency parameters. In the proposed approach, the strategy expansion evaluation concerns the designing of the state evolution, the representation of the system dynamic and the research of the suitable capacity plan. Finally, our approach has been validated by means of the reference case study.

Keywords: long-term capacity planning; Fuzzy-DEVS.

1. Introduction

The intense competition that characterizes present global markets forces companies to give quick responses to changing trends. For example, the introduction of new technologies, changes in customer demands or fluctuations in the cost of raw materials can force the necessity of modification in the current products or the conceive of new ones. To take advantage of these imbalances in the market, companies have to conceive, design and manufacture new products or to introduce modification in the current production plan quickly and inexpensively. Once the decision to issue a new product or to modify an existing one has been made, it may be necessary to increase the actual production capacity via the acquisition of new technologies or production resources. The problem is already known in literature and classified as 'production capacity planning problem'. Moreover, the frequency of decisions related with the configuration or expansion capacity evaluation of manufacturing production systems is increasing in many industrial sectors. This is due to different factors, such as the reduction of product life cycle, the increasing competition, etc. In such a context, decision-makers have to take their actions in shorter times than they have ever done in the past, taking, for example, quick decisions about different production system alternatives. Capacity planning is the process of determining the most cost-effective productive environment that meets the current and future demands of a manufacturing system.

Capacity planning estimates the amount of hardware resources needed to satisfy a set of performance requirement under a changing workload for a certain customer demand. In the proposed approach, the capacity planning problem is solved under the hypothesis that the initial configuration of the production system is supposed assigned in order to predict the impact of changes in the resource configuration. Such projections are necessary to determine the most suitable way to acquire resource for an already existing plan. The acquisition operation is carried out in different time periods in order to minimize enterprise costs and to satisfy customer demand.

The planning horizon, in which solutions to the capacity-planning is found, is a medium-long period. In order to manage the overall problem, it has supposed to decompose the planning horizon into sub-periods and to determine, among a set of promising configurations, for each subperiod the most promising one. In other term, the overall problem may be seen has the search of the minimum cost path on a oriented graph with arcs characterized by fuzzy costs. Efficient long-term capacity management is vital to any manufacturing firm, (Bretthauer, 1995). It has implications on competitive performance in terms of cost, delivery speed, dependability and flexibility. Moreover, in a manufacturing strategy, capacity is a structural decision category, dealing with dynamic capacity expansion and reduction relative to the long-term changes in demand levels.

Alternative system configurations are available for each period of time. Such configurations can be obtained by means of different acquisition strategies. In order to manage the overall problem, it has supposed to decompose the planning horizon into sub-periods. Each strategy leads to a specific configuration at the end of a given period by passing through a series of capacity acquisition. Our goal consists in determining, among a set of promising configurations (for each sub-period) and strategies the most promising one.

A production system is expanded because of different reasons such as the increasing of the volumes requested by the market and the arrival of a new product to be manufactured. In order to simplify the design phase of the system, most of the configuration parameters (demand, products, costs, etc.) can be assumed to be constant or, at the most, variable in some defined ways (statistical distributions), (Katok et al., 2003). However, in order to obtain a suitable modelling of such a changing environment, it may be necessary to exploit the available information expressed not in terms of stochastic uncertainty. A possible solution consists in using fuzzy set theory to handle such a kind of uncertainty.

The work described in this chapter is organized as follows. In Section 2, we proposed a comprehensive characterization of the reference problem. In particular, a literature analysis is reported. Section 3 describes the proposed approach. First, the strategy expansion evaluation is described referring to fuzzy set theory in order to manage the non-stochastic uncertainty. Then, the designing of a suitable expansion policy is considered. An example of the proposed approach application is reported. In Section 4, the case study is analyzed and the experimental results are reported. Finally, conclusions and further research directions are given in Section 6.

2. Problem statement

The considered problem is to determine the best expansion capacity planning among a set of promising solutions in the overall considered horizon by taking into account different level of uncertainty in the knowledge of market demand. The final goal of the proposed methodology is the minimization of the overall enterprise costs (externalization, production and investment costs) in the definition of the capacity acquisition plan according to the following basic hypothesis: (i) the production demand can be satisfied both internally and externally by acquiring suppliers production capacity and (ii) the demand level uncertainty in the planning horizon can not be modelled by probabilistic distributions for the lack of information related with the introduction in the production of new products; in this case only possibilistic assumptions may be made and they will be treated by fuzzy sets theory.

As previously reported, the goal of proposed method is to evaluate the performance of the configurations operating within a market characterized by a high level of uncertainty in order to determine the best resource acquisition plan in an production expansion hypothesis. The module will be developed, taking into account the output produced by the activity A3 'identification of capacity alternatives' and starting from the information and the results obtained in activity A1 'planning capacity at strategic level' activity. Refer to the IDEF0 A0 level diagram in Chapter 1.

The planning horizon, in which solutions to the capacity-planning will be found, is a medium-long period. The validity of the proposed approach has to be evaluated also taking into account the difficulties related within the search of a solution in a medium-long horizon, both for the high level of variability and for the difficulties in the estimation, for example, of the parameters of new markets in which the firm may be interested to operate in. In this perspective it is possible to decompose the overall planning horizon into sub-periods and, for each sub-period, to design a set of promising configurations. The research of a solution in a long period view imposes the necessity to represent the uncertainty of the decisional variables at different levels. In this view the representation by fuzzy set theory may be useful to represent the variability and the relative level of uncertainty. This is possible both in a long period view, in which it is very difficult to represent the probabilistic values and, in particular, in a short period view where, for example, it is necessary to define the demand of a new product. In this case, in fact, no historical data are available to infer correctly the probability distributions.

The objective is to develop a general approach in order to model the dynamics of the discrete fuzzy systems based on the solution of the minimum path problem within a oriented graph, in which at each arc are associated transactions whose occurrences are described by fuzzy variables. In this context the objective functions are aimed to measure the performance of the system from the economic point of view. The graph will be defined in terms of nodes and arcs in a such way that is possible to take into account all the alternatives within a decisional process, in order to provide a useful tool for the expansion capacity planning problem in advanced manufacturing systems. The innovative element is to solve the decisional problem under uncertainty conditions, described by means of fuzzy set theory. In particular, the representation in terms of possibility will be relative to the external variables describing the market in which the firm will operate.

In the following a short description of the model with the relative inputs and outputs is reported.

The input set to the module is made up of the following data. In the graph representation, the initial node represent the actual configuration of the production system under consideration. These following nodes represent the set of the promising configurations for each sub-period provided by the activity A3 'identification of capacity alternatives' (see Chapter 4). The operating parameters for each configuration may be defined in terms of ranges of possible values whereas it is supposed to have an uncertain knowledge about the market demand. Moreover, the module 'identification of capacity alternatives' provides in input to the developed model:

- the rules to change the configuration and thus the available production capacity;
- the list of the feasible transactions between the configurations of each successive sub-periods.

For each configuration and, eventually, for each one comprised between the configurations of two successive sub-periods (obtained by applying one of the strategies of capacity increasing), the production performance measured in terms of number of parts produced per type are supposed available and fixed. The last necessary input to the performance evaluation tool is the fuzzy representation of the final market in which the firm will operate. The considered parameters will be the production volume and the variety of different products required by the final market. The parameters must be defined in fuzzy terms for each elementary period. In the following it is reported the model to represent production system dynamics under a fuzzy market representation and the relative inference model, meanwhile the values of the support ranges of the fuzzy variables will be the output of the strategic DSS module. In particular, it is supposed will provide data about the production typology, the relative volumes, etc.

The evaluation model developed is able to represent the dynamics of fuzzy discrete systems and include an innovative algorithm to solve the TSP problem on fuzzy networks. The reference model must be able to represent and to simulate the performance of assigned production system configurations taking into considerations a fuzzy variability in the final market and the relative time dynamics. The model takes into consider not only the uncertainty in the input parameters but also simulate all the different evolutions consequence of the different capacity expansion adopted strategies. To reach these goals it has been necessary, in a first step, to deduct feasible capacity expansion strategies from the operative conditions in which each configuration is considered profitable. The different strategies must respect the constraints on the possible transitions between feasible configurations belonging to different periods.

The core of the performance evaluation tool consists of an innovative fuzzy simulator. The basic feature of the innovative approach is the possibility to supply information on the performance parameters in a unique simulation run. Furthermore, a heuristic search method, able to find the best route in fuzzy networks, has been developed within this module. To reach this goal, the approach presented in (Kwon et al., 1996) will be significantly enhanced by introducing the following additional features:

- a bi-directional communication between the model and the entity that generates external events
- the external transitions will be able to affect only some branches of the possible evolutions
- the model will keep track of any possible evolution, without trimming the evolution tree at any step through the “max-min rule”.

The final result will be a tool able to give usable results about system performance. The soundness of the obtained results will be demonstrated with respect to the ability of the tool in discarding fuzzy unfeasible solutions. Furthermore, a heuristic search method, able to find the best route in fuzzy networks, has been developed. In order to execute the experimental plan a demonstrative software module has been developed. As mentioned before, the model will simulate different evolutions, both in performance and economic terms, that may occur for different possible system configurations as result of uncertain market changes. For example, considering a set of promising configurations, it will be possible to obtain the value of the fuzzy variable representing the cost of these configurations, taking into considerations all the possible market evolutions. Once these values are known, it will be possible, by means of a well-designed ranking operator, to hit the ultimate target of the research plan, i.e. to select the set of promising configurations, feasible according to the budget limitations provided.

In order to reach the final goal the following activities have been carried out, see Figure 5.1. In Section 3 a description for such activities is reported.

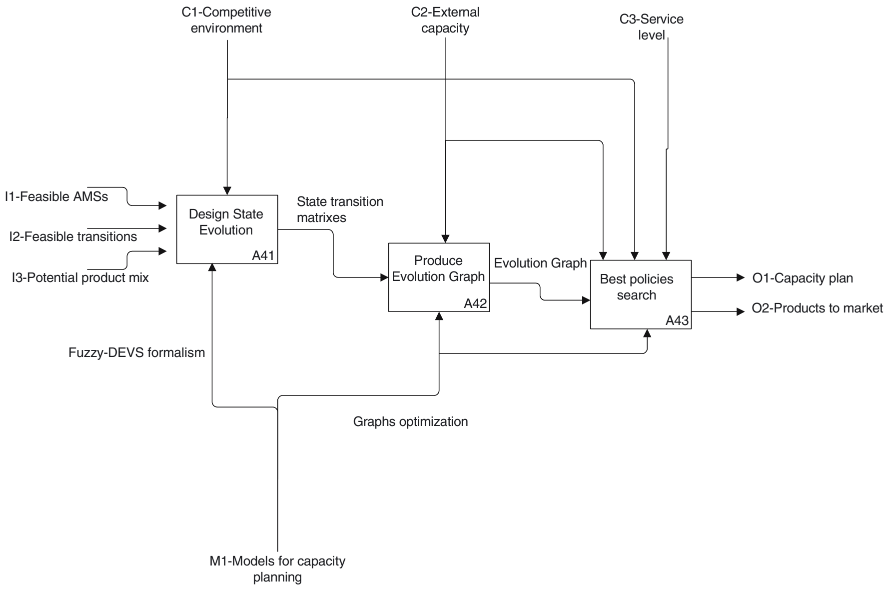


Figure 5.1. A4 level diagram.

2.1 The expansion capacity problem in literature

The problem of capacity expansion has been already faced in (Bhattacharyya, 1998) through a new methodology based on fuzzy Markovian chains. In the paper, the author models the high uncertainty of the market behavior and the high number of states in which the market can be in the future by means of fuzzy Markovian chains. Another work (Yager, 2002) considers the decision making under uncertainty problem by adopting fuzzy matrixes. However, no tool is available to face such a question within industrial environment, using information that are available for the firm. For this reason, we dealt with this problem by considering analytical tools for such decisional process in the industrial environment.

Making decisions under uncertainty is a persistent task faced by many decision-makers, as pointed out in a comprehensive historical review in (Bernstein, 1998). Whenever selections must be made among alternatives in which choices do not lead to well determined payoffs, the decision maker has to address the problem of comparing complex items whose difficulty often exceeds human ability to compare. One approach to address this problem, presented in (Yager, 2002) is to use valuation functions. These valuation functions convert the multifaceted uncertain outcome associated with an alternative into a single scalar value. This value provides a characterization of the decision makers perception of the worth the uncertain alternative being evaluated. The focus in (Yager, 2002) is on the construction of valuation functions. No software tool that applies the proposed methodology is implemented.

In (Herrera and Herrera-Viedma, 2000) a linguistic decision analysis is presented, focusing on the study of the steps to follow in the context of multi-criteria/multi-person decision making. Three steps are typically needed for solving a multi-criteria decision making problem under linguistic information: (i) the choice of the linguistic term set with its semantic in order to express the linguistic performance values according to all the criteria, (ii) the choice of the aggregation operator of linguistic information in order to aggregate the linguistic performance values, and (iii) the choice of the best alternatives. The paper proposes a deep insight into the methods to be used in each phase and exemplifies its approach through a case study in the automotive sector. The approach is highly beneficial when the performance values cannot be expressed by means of numerical values. On the other hand, it is not pretty efficient when dealing with complex data involving a system evolution throughout a wide time horizon.

In recent years, many firms have found it increasingly important to invest in technology to maintain a competitive edge. Technological improvements often require superior production methods, and some firms find themselves frequently evaluating opportunities for investments in new production resources. These decisions can easily become crucial to survival in a competitive market place.

Production investment decisions are extremely difficult because they involve planning under uncertainty. For example, when a new production resource provides manufacturing flexibility, the benefit of this flexibility can be easily underestimated.

As reported in (Jordan and Graves, 1995), while in capacity and flexibility planning, investment costs for flexible operations are typically quantified, it is less common to quantify the benefits because demand uncertainty is not explicitly considered by the planners. Since flexibility is expensive, this typically results in decisions not to invest in it.

The benefits of a new production resource are obtained by means of three factors: Lower cost due to superior performance, Increased capacity and Increased decision flexibility.

Indeed, cost savings may result if a new resource provides a more efficient production process or introduces a new dedicated process. If, at a particular stage, a new resource is added to the current production system, capacity at that stage may increase. If that stage previously formed a bottleneck, the throughput of the entire system increases yielding cost savings. The third source of benefit comes from increased decision flexibility (Benjaafar et al., 1995). Decision flexibility is the ability to postpone decisions until more information is obtained. When a new production resource is added to the current system, it can increase decision flexibility by either providing additional capacity where it is needed. To correctly estimate the impact of a flexible resource, a methodology must include all three sources of benefit.

In (Katok et al., 2003) several methods for evaluating resource acquisition decisions under uncertainty are examined. Since traditional methods may underestimate equipment benefit when part of this benefit comes from decision flexibility, a method for resource planning under uncertainty is developed. In particular in (Katok et al., 2003) a new model is developed to better represent advantages due to flexibility increase. Moreover a stochastic method is applied to evaluate the impact of new resources. This paper shows that this approach is more accurate than several commonly used methods: traditional approaches underestimate - in terms of flexibility - advantages due to new resources acquisition.

Flexibility planning has been studied extensively during the last decade. For a summary of flexibility categories and measures see (Sethi and Sethi, 1990) or (Gupta and Goyal, 1989).

In (Euwe and Wortmann, 1997) common problems concerned state-of-the-art planning support system as described. New issues address the developing of innovative systems to support strategic decision in the medium- and long-term planning levels. Current solutions do not consider several aspects, such as uncertainty influence, process flexibility and suppliers information

Long-term capacity management has implications on competitive performance in terms of cost and flexibility. In a manufacturing strategy, capacity is a structural decision category, dealing with dynamic capacity expansion and reduction relative to the long-term changes in demand levels (Olhager et al., 2001).

Three variables are commonly used to describe a capacity strategy (Hayes and Wheelwright, 1984): the type of capacity needed, the amount of capacity that should be added (or reduced), and the timing of capacity changes. Since the type of capacity strongly influences the amount that is to be added or reduced, the first two are normally discussed together in the so-called sizing problem. The timing variable in a capacity strategy is concerned with the balance between the (forecasted) demand for capacity and the supply of capacity. If there is a capacity demand surplus the utilization is high (low cost profile). A capacity supply surplus on the other hand creates a higher cost profile but due to the surplus capacity it is easier to maintain high delivery reliability and flexibility. The capacity strategy can thus be expressed as a trade-off between high utilization (low cost profile) and maintaining a capacity cushion (flexibility). Based such this, two capacity strategies can be identified, referred to as leading demand (capacity supply surplus) and lagging demand (capacity demand surplus). A good solution aims at finding an efficient trade-off between the two strategies in order to track the demand.

The presence of uncertainty in the considered question leads to carefully consider such an aspect in the methodology adopted in the support strategic decision system. Usually there are two directions used to handle uncertainties: Monte Carlo methods and stochastic programming methods. The latter approach can been used to model the capacity expansion problem, see (Wang and Sparrow, 1999). In this work a capacity expansion model is used in uncertainty demand condition. The impact of such uncertainty on costs is analyzed. The mixed integer non-linear programming (MINLP) model keeps into account uncertainty in order to optimize the profit. The model we presented describes the uncertainty as a discrete distribution and with only three possibilities. Nevertheless, it is demonstrated that neglecting uncertainty leads to an objective value which is far away from the correct answer.

3. The proposed methodology

The proposed methodology follows a stepwise approach, in order to achieve the best expansion strategy according to the information available about the possible behavior of the system's environment within the programming period.

First, a fuzzy-simulation based analysis is conducted in order to compare different, generally adoptable expansion strategies. Obviously, such policies are not optimal, but they allow obtaining a first description of the range of possible final states reachable by the system. These results are expressed in form of intervals in which the system variables can assume values. On the basis of the feasible configurations of the system through time, the second step exploits such information in order to design, through optimization methods, the optimal expansion policy the firm should adopt.

3.1 Strategy Expansion Evaluation

This work does not deal with the from scratch design issues, so the production system is assumed to be already known in terms of number of machines, part carriers and tools. The environment parameters that are considered in the system re-design activity involve the market, its demand values and the products requested by the customers.

The use of fuzzy simulation leads to the production of a scenario tree representing all feasible system evolutions. Indeed, starting from the initial state, each possible evolution based on specific feasible transitions is reproduced along time.

A set of corrective actions is available in order to change the system behavior in response to the market demand changes. Once a system state variation triggers a corrective action, the system state is altered in order to set a new state. By reproducing the system state variation along time in such a way, relevant information can be collected concerning the application of a management policy for particular demand behavior.

As an instance, the system may have been configured to produce a certain number of parts within a particular environment, but some corrections to its initial design could be necessary to react to future market evolutions (e.g. introduction of a new product into the part mix of the system). At a certain period t, the firm may evaluate, according to new market conditions, whether the existing system is able to operate in the modified environment or not. If a corrective action is to be taken, the method allows a quantitative characterization of the possible actions in terms of their results over the firm placement within the market, regardless of the operative actions that need to be taken in order to obtain this result.

The proposed fuzzy simulation tool evaluates the advantages and drawbacks of applying different alternative strategies to the analyzed system. For instance, acquiring new production capacity each time an even small increase of the market demand takes place may be a good strategy in fast growing markets, such as those involved in the new economy field, while waiting for larger demand volumes before taking any action would be a better strategy in less rapidly changing markets, such as the ones that operate within the automotive field. The firm should be able to calculate a series of performance parameters for each strategy, in order to select the best one.

On the basis of the performance parameters calculated applying the methodology proposed in the paper, the decisions about whether and when to expand the system can be taken, according to formally defined strategies. The fuzzy mathematics is used to model the uncertainty of the market behavior by means of the DEVS formalism.

A transition matrix describes the possible ways the system state may vary throughout the programming horizon. In particular, for each system state, the transition matrix provides the possibility of a particular transition toward another state. For simplicity's sake, let us consider a system state represented by the market demand for a certain product. A possible transition matrix between two consecutive periods is the one reported in Table 5.1, where three possible states are considered for a generic time t, while four different states are available for the following period. The table provides the possibility values for each available transition.

Table 5.1. Transition Matrix.

Demand	*200*	*300*	*350*	*400*
100	0.2	0.3	0.3	0.2
200	0.2	0.2	0.4	0.2
300	0.1	0.1	0.2	0.6

The first step of the proposed methodology is based on a modeling formalism called DEVS (Discrete Event System Specification). This was introduced in (Zeigler and Vahie, 1993). The DEVS formalism is a universal and general-purpose methodology to model systems characterized by an events-based behavior. Since the DEVS formalism is strictly deterministic, the system to be modeled (e.g. the market) is required to evolve according to deterministic rules. Therefore, DEVS formalism as is cannot be used to model systems with high variability as modern markets are. Fuzzy-DEVS approach was presented in (Kwon et al., 1996) as an enhancement to the DEVS formalism. By means of Fuzzy-DEVS, it is possible to specify, for each state in which the analyzed system can be, a set of possible evolutions of the system. Each of the possible paths is characterized by a possibility measure that synthesizes the fuzzy behavior of the system. In general, such a model keeps track of any possible evolution, without trimming the evolution tree at any step through the "max-min rule" (see (Kwon et al., 1996) for details).

In this work, a new formalism is introduced, based on the original Fuzzy-DEVS enhanced with new features:

- the possibility to have a bi-directional communication between the system and the entity that generates the external events;
- external transitions may affect only some branches of the possible evolutions of the system.

In the capacity expansion problem, it is necessary to identify the system, i.e. the internal and the external transitions of the new formalism. The system to be modeled is the market the firm operates in. Changes in the market can be considered as internal transitions, so internal is referred to the market. Each internal transition generates an output event that can be effectively used by the firm to take decisions about its behavior in the following periods. These modifications on the existing production system can be treated as events or external transitions, accomplished by the firm to adjust its strategies. The mapping of Fuzzy-DEVS elements within the modeled system is summarized in Table 5.2.

Table 5.2. Mapping of Fuzzy-DEVS elements within the modeled system

Fuzzy-DEVS element	*Modeled system element*
System state	n-dimensional vector (satisfied demand percentage, produced part type, ...)
Internal Transition	Market fuzzy evolution
External Transition	Decision maker actions
Output function	Performance parameters
Time advance function	Periods within the planning horizon

In order to implement the new formalism, a protocol has been used to enable communications between the modeled system and the decision-maker; through this protocol, the decision-maker gets feedback from the system (the outputs generated by the internal transitions) in response to actions previously taken or to the natural evolution of the system with the time.

The definition of the states through which the system can evolve has to keep track of parameters characterizing both the market and the firm. The state is defined as an n dimensional vector, representing the levels of the n parameters that the decision-maker has to control. An example of possible state evolutions with a mono-dimensional state vector is represented in Table 5.3; the parameter in the vector is the level of demand the firm has to satisfy at a certain time $t = 0, 1, \ldots, T$. Suppose that, at time 0, the market requests volumes with a rate of d_0 units per day and the firm has a capacity of $c_0 = d_0$ parts per day. Over the next period, the market demand assumes the new value d_1 that may be smaller, larger or equal to the previous one. On the basis of the comparison with the firm capacity, different situations can occur (see Table 5.3).

In order to describe transitions between possible states, market experts provide transition matrices, mostly based on their personal experience and knowledge of the target market. Obviously, the definition of the possible external transitions is strictly tied to the definition of the specific strategy to be evaluated.

Table 5.3. Example of state evolution.

State	*Description*
$d_1 = c_0$	The market demand d_1 does not change. Since the firm has enough capacity, the production system is perfectly designed on the market needs.
$d_1 > c_0$	The market demand d_1 increases. Since the firm does not produce the requested quantity to meet the market demand, a good solution could be acquiring new machines.
$d_1 < c_0$	The market demand d_1 decreases. Since the capacity of the system is c_0 units per day, the firm has an overcapacity that is not exploited to produce parts. No specific action can be taken in this particular case.

The result of each specific action is evaluated by considering the market behavior in the considered period in its fuzzy evolution. The cash flow related to the acquisition of production capacity is analyzed and the amount of sold products are computed.

3.2 Expansion policy design

The second step of the methodology consists in inferring the best expansion policy on the basis of the system analysis conducted in the first step. Once the available capacity expansion strategies have been examined by means of the previous tool, they need to be converted into production capacity expansion policies: i.e. operative actions to be taken by the firm. In fact, there may exist several ways to pursue the same strategy. Acquiring additional production capacity may be achieved buying new machining centers, increasing productivity by means of efficient conveyor system or even re-engineering the already available resources.

A software tool has been developed that allows the decision maker to design a production capacity expansion policy, according to the results obtained from the previous tool and exploiting the already available information about the uncertain evolution of the market. In particular, the strategy evaluation tool provides a path to be followed by the system throughout the planning horizon. On the basis of such temporal information, data concerning the configurations in different periods are supplied. Moreover, information regarding the feasible transitions from a fixed period to the next one are available. In order to obtain, for each period, detailed data concerning such configurations, it is necessary to decompose the “period” time unit in multiple “sub-periods”. For example, if the considered time unit at the strategic level is six-month length, it is possible to adopt two-month or three-month length sub-periods at the operative level.

For each period, the system can be in one of the feasible states: a sequence along the time periods of such states stands for a strategy to deal with the demand change. Once the best combination of system states through the planning horizon is found, the corresponding capacity expansion policy needs to be supplied to the firm. Note that starting from a state, only specific target states can be selected in order to perform a transition; this is because changes in the system are subjected to consistency constraints.

For each period t, it is assumed to be $s_1, s_2, \ldots, s_t$ the set of possible states of the systems. Each state s is characterized by specific architectural choices, production programs and conveying systems.

For the configuration analysis, the considered information, related to each feasible state, are reported below:

- fixed costs
- productive capacity for a given part type
- variable costs for the externalization of each part type
- variable costs for the non-utilization of an industrial facility for processing of a specific part type
- set of possible states reachable in the next period, together with the transition costs.

Since the basic temporal unit is the sub-period, while system state information are known on the basis of the time unit "period", it is necessary to decompose data concerning each period into information regarding the "sub-periods". For example, assuming that a system configuration s in a period t is partitioned in three sub-periods $(\lambda_1, \lambda_2, \lambda_3)$, the following parameter has to be defined.

- $C_{st} = (C_{st}^{\lambda_1}, C_{st}^{\lambda_2}, C_{st}^{\lambda_3})$: fixed costs;

moreover , for each part type p, the following further information is needed:

- $K_{spt} = (K_{spt}^{\lambda_1}, K_{spt}^{\lambda_2}, K_{spt}^{\lambda_3})$: production capacity for each part type;
- $V_{spt} = (V_{spt}^{\lambda_1}, V_{spt}^{\lambda_2}, V_{spt}^{\lambda_3})$: production variable costs for each part type;

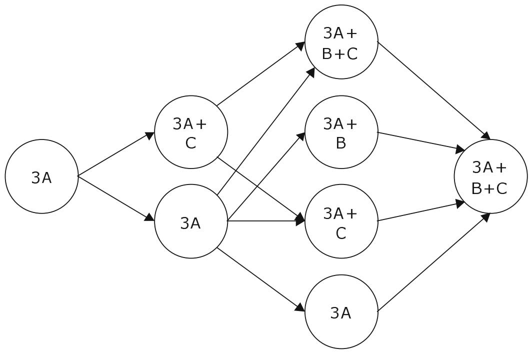


Figure 5.2. Feasible transitions.

- $E_{spt} = (E^{\lambda_1}_{spt}, E^{\lambda_2}_{spt}, E^{\lambda_3}_{spt})$: variable costs for the externalization for each part type;
- $I_{spt} = (I^{\lambda_1}_{spt}, I^{\lambda_2}_{spt}, I^{\lambda_3}_{spt})$: variable costs for the non-utilization of an industrial facility for processing of a specific part type;

However, when the system changes its configuration from a period to the next one, it is necessary to take into account the particular transition that takes place. For example, suppose the system is in state S_1 and three type A machines are present; whereas in the next state S_2, the system is made up of three type A machines, one type B machine and one conveyor apparatus C. The acquisition time for a type B machine is 4 months, while conveyor C is acquirable in two months. The transition may take place in different ways (see Figure 5.2).

Consequently, the previously considered parameters are different according to the way the transition is performed. Moreover, an additional parameter has to be considered in such a case: the cost to carry out the transition, divided among the sub-periods.

- $T_{s_1,s_2,t} = (T^{\lambda_1}_{s_1,s_2,t}, T^{\lambda_2}_{s_1,s_2,t}, T^{\lambda_3}_{s_1,s_2,t})$: cost to carry out a transition

The Decision Support System user supplies an uncertain characterization of the market evolution. In particular, on the basis of such user data, it is possible to estimate, in fuzzy terms, the correspondent market demand for each part type in each period. Different levels of uncertainty affect the demand: this can be represented by means of fuzzy set theory. At each period t and for each part type p, a fuzzy demand is hence determined. Such a quantity is decomposed along the various sub-periods. The outcome of such operation is a generic n-tuple of fuzzy numbers associated with the required quantities of part type p in the time period t.

- $\tilde{Q}_{pt} = (\tilde{Q}_{p\lambda_1}, \tilde{Q}_{p\lambda_2}, \tilde{Q}_{p\lambda_3})$: quantity of market demand for part type p

Similarly, on the basis of the information provided by the user, it is possible to forecast the selling price of various part types in each sub-period.

- $\tilde{P}_{pt} = (\tilde{P}_{p\lambda_1}, \tilde{P}_{p\lambda_2}, \tilde{P}_{p\lambda_3})$: forecasted selling price for part type p

Evolution graph. The set of the feasible evolutions the system may follow can be effectively represented by an oriented graph, in which each node is associated with a possible system state. Nodes represent the states in which the system can be in each period; while the arc weight linking connected nodes indicate the profits the firm obtains if such a transition is performed. Moreover, for each arc, the possibility associated with the transition occurring is available. For each sub-period, a set of nodes can be defined, each described by the parameters that consider the data previously discussed. In particular, for each node, the following parameters are assigned:

- C_n: fixed costs for the system configuration
- $K_n = (K_{p_1n}, K_{p_2n}, \ldots, K_{p_Pn})$: productive capacity for each part type
- $V_n = (V_{p_1n}, V_{p_2n}, \ldots, V_{p_Pn})$: production variable costs for each part type
- $E_n = (E_{p_1n}, E_{p_2n}, \ldots, E_{p_Pn})$: variable costs for the externalization for each part type
- $I_n = (I_{p_1n}, I_{p_2n}, \ldots, I_{p_Pn})$: variable costs for the non-utilization of an industrial facility for each part type
- $\tilde{Q}_n = (\tilde{Q}_{p_1n}, \tilde{Q}_{p_2n}, \ldots, \tilde{Q}_{p_Pn})$: quantity of market demand for part type p
- $\tilde{P}_n = (\tilde{P}_{p_1n}, \tilde{P}_{p_2n}, \ldots, \tilde{P}_{p_Pn})$: forecasted selling price for part type p

On the basis of the externalization rules adopted by the firm, it is possible to obtain further information concerning a generic node n, such as:

- $\tilde{X}_n = (\tilde{X}_{p_1n}, \tilde{X}_{p_2n}, \ldots, \tilde{X}_{p_Pn})$: part type quantity internally produced for part type p

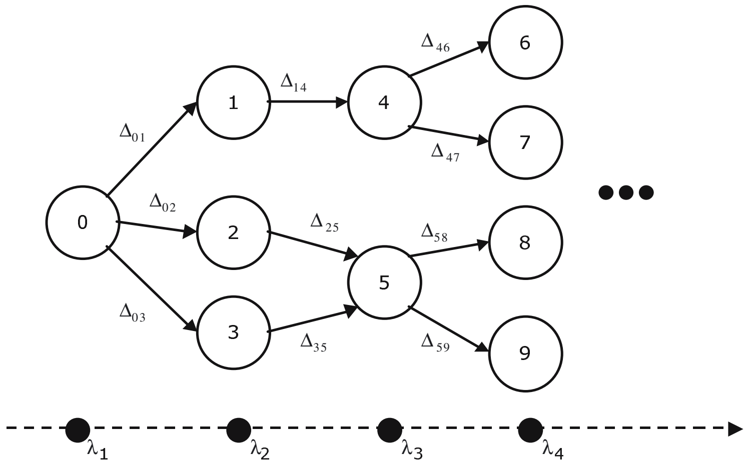


Figure 5.3. System evolution graph in sub-periods.

- $\tilde{Y}_n = (\tilde{Y}_{p_1n}, \tilde{Y}_{p_2n}, \ldots, \tilde{Y}_{p_Pn})$: part type quantity externally acquired for part type p
- $\tilde{W}_n = (\tilde{W}_{p_1n}, \tilde{W}_{p_2n}, \ldots, \tilde{W}_{p_Pn})$: residual quantity for part type p

By combining the available information about a node, it is possible to associate to each node n a synthetic economic fuzzy parameter $\tilde{Z}_n$ expressing the performance of the node in terms of node profit. The expression of such a parameter is the following:

- $\tilde{Z}_n = (\tilde{P}_n - V_n) \times \tilde{X}_n + (\tilde{P}_n - E_n) \times \tilde{Y}_n + I_n \times \tilde{W}_n - C_n$: node profit

For each arc linking two nodes, it is possible to associate a comprehensive performance parameter defined on the basis of the target node parameter and the transition costs.

- $(n_1, n_2) \rightarrow \Delta_{n_1n_2} = \tilde{Z}_{n_2} - T_{n_1n_2}$: arc profit

For each path ending in a node in the last sub-period, it is possible to assign a synthetic performance indicator of the considered evolution that is equal to the sum of the arch weights of the selected path. For example a possible evolution graph is represented in Figure 5.3. The selection of the best system configuration implies the determination of the optimal path between the source node and one of the nodes in the last sub-period.

Best policy search. In order to find the best sequence of feasible states, and hence to determine the best expansion policy, a two-step methodology is defined, as suggested in (Okada and Soper, 2000):

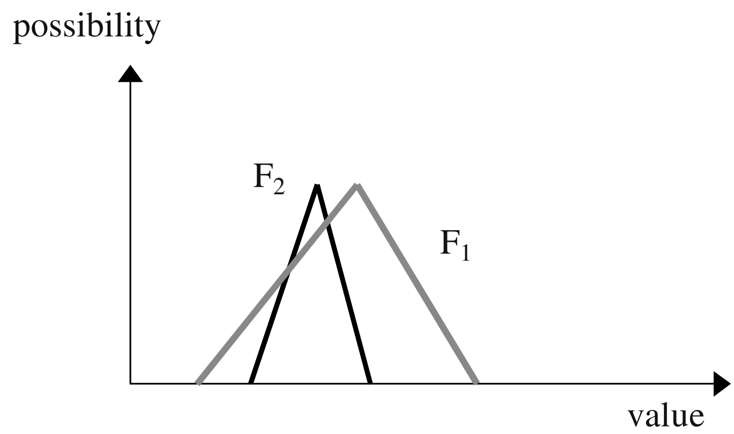


Figure 5.4. Fuzzy profits comparing.

1 Inferring the best k solution paths;

2 Comparing the k solution paths.

Firstly, the best k solutions are computed from the graph analysis. Then the k solutions are compared to select the best one. A possible alternative consists in the application of the Dijkstra method to find the best solution in a single step. This can lead to a final solution having a very low occurring possibility (the solution is optimal in economical terms but its possibility is very low). The possibility associated to a path is calculated as the minimum possibility in the path arcs. Moreover, since profits are fuzzy quantities, a two-step method leads to a deeper analysis of the graph. For example, if the two fuzzy numbers, representing the profit parameter, are those in Figure 5.4, the optimal solution can be both $F2$ (high average value) and $F1$ (lower average value having low dispersion).

Consequently, it is possible to infer:

- the correct choice among multiple fuzzy profits can not be deterministically performed
- the occurring possibilities have to be kept into account

In order to consider such constraints, the determination of the best k paths is considered. Then, a decision tool has to be used in order to select the optimal path among the best k ones.

For each path from the initial node to the final, leaf node, a "profit value" is associated with the corresponding production capacity expansion policy. Each arc has a possibility value associated with the transition occurrence. For each path, the parameter "occurrence possibility" is calculated as the minimum possibility encountered along the path.

The combination of the two parameters above leads to the definition of a "fuzzy profit" parameter. In order to compare the different values of such parameter as obtained for the different paths, it is necessary to consider a ranking based on two factors:

- the modal value of the fuzzy number (to be maximized);
- the dispersion of the fuzzy number (to be minimized).

The dispersion of the fuzzy number is the difference between maximum and minimum values the fuzzy number can assumes, divided by average value, that is:

$$DispersionIndex = \frac{Max - Min}{Average}$$

For example, the triangular fuzzy number $\pi = (1, 3, 5)$ has dispersion equal to: $(5 - 1)/3 = 4/3$

A "dominant solution" is a fuzzy number such that no other fuzzy number having both greater modal value and lower dispersion index exists (Pareto's dominance). If, comparing two fuzzy profits, it is not possible to have both the conditions true, then the two profits are not comparable in terms of Pareto's dominance. Dominated solutions are discarded, while dominant solutions are considered in the next step.

The algorithm to determine the best k solutions considers the information related to the graph that models the future evolutions of the system.

In order to rank the dominant fuzzy number, it is necessary to establish a user-defined ranking algorithm. If it is not possible to establish a strict ranking such as Pareto's dominance, an heuristic method has to be be used. For example, a possible algorithm used to compare two fuzzy numbers is described in (Anglani et al., 2000): such a method is based on the use of two-parameters. Among every possible system evolution, the best k solutions having the highest fuzzy profits are calculated. Since the ranking algorithm between fuzzy numbers is parametric, by varying the two parameters of such ranking method, different results can be obtained.

The decision tool does not rank the k best solutions at once; rather, their characteristics are showed. In this way the user can select the appropriate solution. For each of the k solutions, a circle is placed in a Cartesian chart (Figure 5.5), representing a fuzzy number that indicates a particular value of the considered parameter. The x-axis coordinate stands for the average of the fuzzy number, while the y-axis is

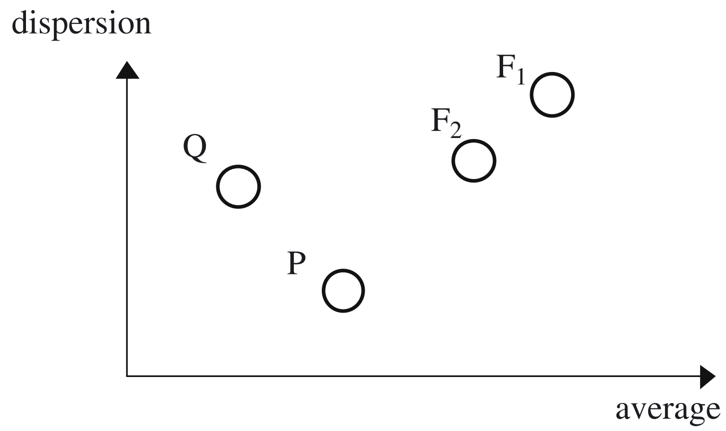


Figure 5.5. Pareto's dominance.

its dispersion index. Such an index is proportional to the width of the fuzzy number. The radius of the circle is proportional to the number of paths having the same couple of the $(profit, possibility)$ values. In other words, the radius corresponds to the number of hits during the graph analysis process - various evolutions can be represented by varying the two parameters of the ranking algorithm. For example, if the two solutions in Figure 5.4 have to be compared, on the basis of the particular firm policy, a high profit solution having low occurring possibility or a lower profit with higher possibility can be preferred. Such a decisional process is named "multi-objective programming" because more than a single parameter has to be considered (e.g. "profit" and "dispersion"). The k solutions are compared in order to maximize the fuzzy profit average and minimize the fuzzy profit dispersion. In order to select the best solution, the concept of "Pareto's dominance" is introduced (Okada and Soper, 2000). In Figure 5.5, each fuzzy number (each circle) represents a profit value; hence, a generic point P dominates another one Q, placed in the Northwest side: point P has better profit average with lower dispersion than Q. Therefore, Q is not considered as optimal candidate because the best solution belongs to the non-dominated solution set. Once the dominant solutions are calculated, the user has to perform the final selection.

3.3 Example

The proposed formalism can be effectively used to model real cases for obtaining an evaluation of the different strategies that can be pursued in the different cases. For such a reason, an application software has

Table 5.4. Example: economical and technical parameters.

Parameter	*EUR*
Price per unit	3
Internal cost per unit	1
External cost per unit	2
Initial production level	48000
Initial structural cost	0
Production capacity expansion cost	5000
Production capacity expansion unit	4800

been built in order to process the available information as described in the previous section. In order to demonstrate the way such a tool processes the information concerning the market behavior, a numerical example is presented. The case study refers to a firm that operates in the automotive sector. Economical and technical parameters, such as machining center costs and capabilities, operating costs and prices have been obtained through the analysis of a real case and interviews with experts and can be summarized as reported in Table 5.4.

A software implementation of the formalism has been developed in order to prove its applicability. The software tool works on the input information used to model both the market and the firm, such as:

- transition matrixes characterizing the market disposition towards changes;
- technical parameters describing the possible decisions the firm can take;
- economic parameters related to the costs of the firm decisions.

In order to characterize the market behavior, only the market demand parameter of the state has been considered, but other important parameters such as the "number of products requested by the market" can be taken into account. The variable x represents the state of the system, i.e. the percentage of the market demand that the current system configuration cannot satisfy. Negative values of x mean that the production system is over-sized, while null values stand for a perfect balancing between system capacity and market demand. Positive values imply lack of resources in order to meet the demand. The different strategies that have been evaluated are detailed in Table 5.5. Strategy $S1$ increases the capacity production when x is 20% for 3 consecutive states or x is 10% for 4 consecutive states. Whereas, strategy $S2$ triggers if x is 20% for 2 consecutive periods or x is 10% for 3 consecutive periods.

Table 5.5. Example: expansion capacity strategies.

Strategy	*Description*
S_0 (No reaction)	No capacity will be acquired whatever the level of the demand
S_1 (Conservative)	If $(X(t-2) = X(t-1) = X(t) = +20\%)$ Or $(X(t-3) = X(t-2) = X(t-1) = X(t) = +10\%))$ Then $X(t+1) \leftarrow 0\%$
S_2 (Very reactive)	If $(X(t-1) = X(t) = +20\%)$ Or $(X(t-2) = X(t-1) = X(t) = +10\%))$ Then $X(t+1) \leftarrow 0\%$

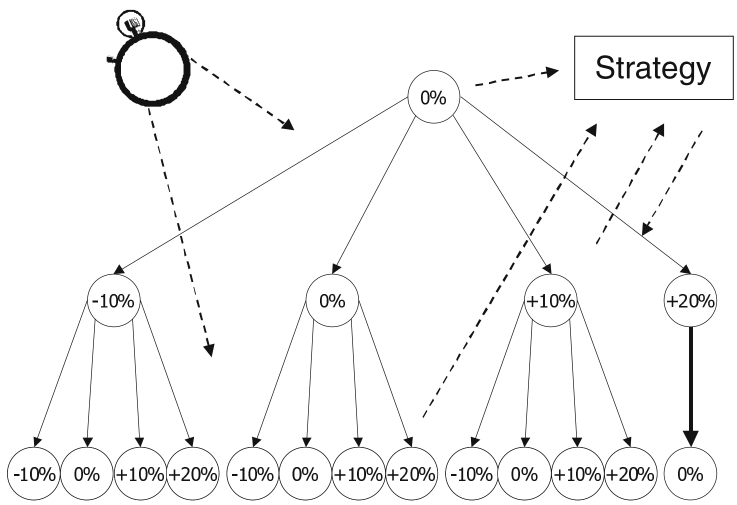


Figure 5.6. Example: system evolution as modeled by the proposed Fuzzy-DEVS enhancement.

Using the provided information, such a supporting tool calculates any possible evolution the system can follow according to the selected strategy-to-be-evaluated (see Figure 5.6).

A first, raw output of the tool consists of a list of all the possible paths the system can follow. Every step in a path carries information concerning:

- the possibility such a step has to happen;
- the costs associated with the step;
- the incomes associated with the step.

A comparison of the defined strategies is achieved analyzing such raw data obtained through the evolution path lists.

Qualitative analysis. A first processing of the tool produces, for each strategy, a "gathering chart" like the ones in Figure 5.7 and Figure 5.8. A gathering chart shows, in a qualitative way, the value of a performance parameter (e.g. variable costs, structural costs, income and profits) associated to the specific strategy that has to be evaluated. The horizontal axis represents the selected performance measure (it is the gain in Figure 5.7 and Figure 5.8). While the vertical axis represents the possibility level of having that value of performance. Each circle represents one or more paths that lead to the same profit value with the same possibility. The radius of the circle is proportional to the number of paths that lead to the same point: the larger the radius, the larger the multiplicity of the point.

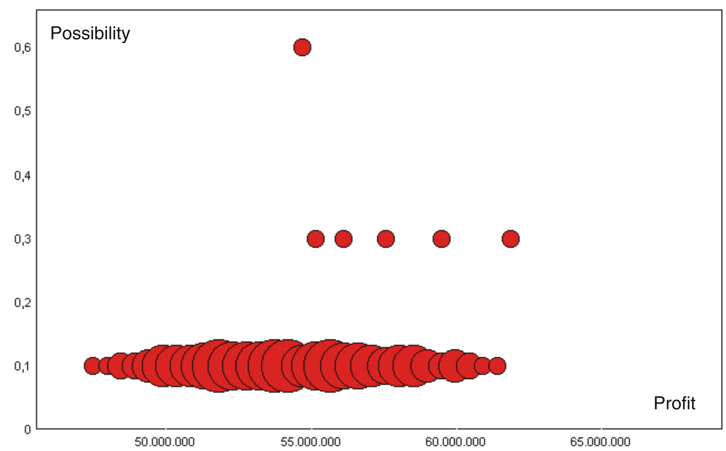


Figure 5.7. Example: "No-reaction" gathering charts.

Figure 5.7 shows the estimated profits and their corresponding possibility levels for the firm operating in a dynamic market and that adopts the no-reaction strategy. In this particular case, circles gather in three zones that correspond to low/average profits with different levels of possibility. Figure 5.8 shows the estimated profits and their corresponding possibility levels for the firm operating in a dynamic market and that adopts the very-reactive strategy. In this particular case the zones of the circles are five and compared to the previous ones, these zones are shifted towards areas in which the profits are higher and the possibility levels are equal. The gathering charts provide qualitative information

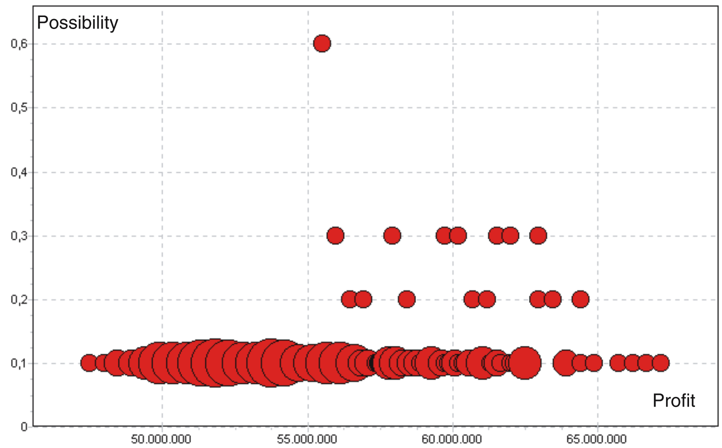


Figure 5.8. Example: "Very reactive" gathering charts.

on the decisions that can be taken by the firm. In the analyzed case, the circles gather in different zones according to the selected strategy. Thus, the qualitative analysis that can be led through the gathering charts provides a first evaluation of the different strategies, evaluating as a better strategy, in this case, the reactive one. It still lacks, anyway, a quantitative evaluation of the advantages and constraints deriving from adopting a certain strategy rather than the other one.

Quantitative analysis. A more detailed and quantitative analysis can be made using the EEM charts provided by the software. These charts are based on the Expected-Existence-Measure as defined in (Nguyen and Le, 1997), but adapted to the discrete case. A normalized fuzzy number is obtained from each gathering chart using the following equation:

$$(g_j, p_j, m_j) \rightarrow \left(g_j, \frac{m_j p_j}{\sum_j m_j p_j} \right) = (g_j, P_j)$$

Where:

- (g_j, p_j, m_j) is a circle in the gathering chart
- g_j is the parameter on the horizontal axis of the corresponding gathering chart
- p_j is the possibility level
- m_j is the multiplicity of the point

- (g_j, P_j) is a point in the fuzzy number corresponding to the considered chart

Using the so defined fuzzy numbers, it is possible to work out the charts shown in Figure 5.9 and Figure 5.10. The software can calculate EEM charts for any parameter used in the evolution paths. For instance, it is useful to have an EEM chart of the overall costs associated with the considered strategies. In the case shown in the Figure 5.9 and Figure 5.10, relative to a firm operating within a rapidly changing market, the three strategies defined above (no reaction, conservative and very reactive) have been considered. The firm maximum bearable cost can be fixed (horizontal axis) according to budget bounds. In correspondence of the given horizontal value, a possibility level is then fixed for each strategy. Considering the lowest and the highest values it is then possible to obtain an interval representing the confidence range of the evaluations provided by means of the software. In the analyzed case, the firm has a budget less than 4M EUR, so the confidence range is then [0.57;0.62]. Once this range is obtained, it is possible to get important information from the charts relative to other parameters such as costs (Figure 5.9) and profits (Figure 5.10). Considering the EEM chart for the profits and zooming it to the confidence interval, the chart shown in Figure 5.10 is obtained. In this interval, the very reactive guarantees profits between 5.68 M EUR and 5.77 M EUR, while no reaction strategy guarantee profits of 5.5M EUR. Thus, in the considered environment, the very reactive strategy is preferred and the benefits deriving from its choice are exactly quantified.

To get a more detailed analysis it is possible to study the charts relative to other parameters, such as structural costs, variable costs and income. In the considered case, such an analysis shows the gap between the very reactive strategy and the most conservative one. In particular, it is possible to observe:

- increasing of the total income, due to a better meeting of market needs;

- invariance of overall costs;

- increasing of structural costs;

- decreasing of variable costs.

The analysis can be carried out in even deeper details as long as more accurate parameters are used while creating the evolution paths.

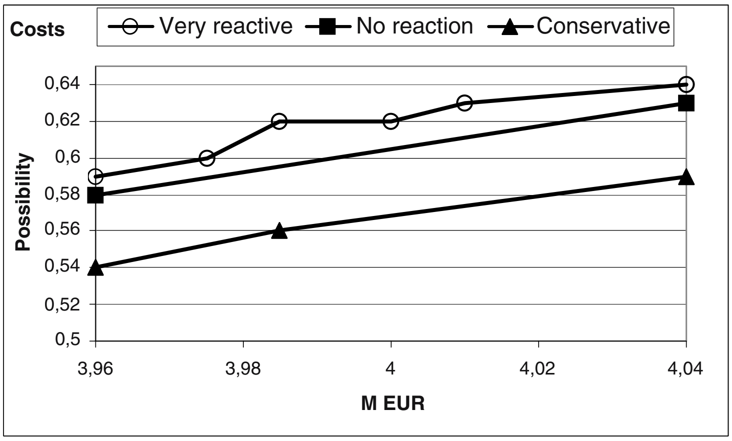


Figure 5.9. Example: cost EEM charts detail.

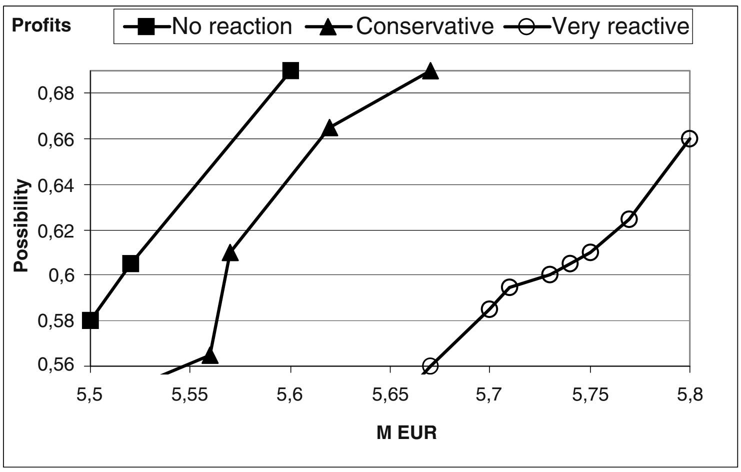


Figure 5.10. Example: profit EEM charts detail.

Table 5.6. Example: expansion actions.

Action	*Production capacity*	*Amortization share*	*Amortization periods*
A	2500	4000	3
B	3000	3000	4
C	2000	3500	2
D	2300	2000	3

Strategy design. The main previous step results consist in supplying the best strategy in terms of capacity expansion directions. For example, using such a tool it is possible to determine the best strategy is "very reactive". Moreover, the best feasible expansion alternatives are the non-dominated solutions in Figure 5.8 in terms of possibility-profit. On this gathering graph, Northeast solutions dominate Southwest ones - because solutions having high profit and high possibility values are preferred.

Four expansion evolutions are the results of this analysis. In order to establish the actual best strategy a further effort has to be performed. First, the actual expansion actions should be defined (in terms of possible machining type or conveyor system to be acquired). Then, the best policy has to be found on the basis of the performance indicators defined on the previously defined. In order to get the four expansion evolutions it is possible to select among different actual expansion actions (described in Table 5.6). By using these actions, the demand expansions can be pursued combining such actions.

The way such actions have to be combined, along time, in order to follow the strategy "very reactive" directives, is obtained by using the second step of our methodology. In such an example, each period is divided in two sub-periods. In particular considering the system state expansion as a 4-tuple:

$$
\begin{aligned}
&(no.A - typeactions, no.B - typeactions,\\
&no.C - typeactions, no.D - typeactions)
\end{aligned}
$$

The possible capacity expansion behaviors, along time, are represented in Table 9. Note that unfeasible system state transitions exist - e.g. in the evolution from period 4 to 5, it is not possible to pass from $(7, 4, 9, 5)$ to $(5, 7, 9, 5)$ because the assumption "no resource can be sold" is made. For this reason, a feasible transition matrix has to be defined. The system state $(5, 4, 6, 5)$ corresponds to the initial state. The 4-tuple values in Table 5.7 are added to $(5, 4, 6, 5)$ in order to obtain the system state 4 tuple, at a given period.

Table 5.7. Example: system state evolution.

Period	*States*
0	(5,4,6,5)
1	(5,4,6,5)
2	(5,4,6,5)
3	(5,4,6,5)
4	(5,4,6,5) (7,4,8,5) (7,4,9,5) (7,4,6,7) (5,6,8,5)
5	(8,4,9,5) (8,4,6,8) (5,7,9,5) (7,4,8,5) (7,4,9,5) (7,4,6,7) (5,6,8,5)
6	(8,4,9,5) (8,4,6,8) (5,7,9,5) (7,4,8,5) (7,4,9,5) (7,4,6,7) (5,6,8,5)
7	(8,4,9,5) (8,4,6,8) (5,7,9,5) (7,4,8,5) (7,4,9,5) (7,4,6,7) (5,6,8,5) (11,4,13,5) (11,4,14,5) (11,4,6,11) (5,10,11,5)
8	(7,4,8,5) (7,4,9,5) (7,4,6,7) (5,6,8,5) (11,4,13,5) (11,4,14,5) (11,4,6,11) (5,10,11,5)

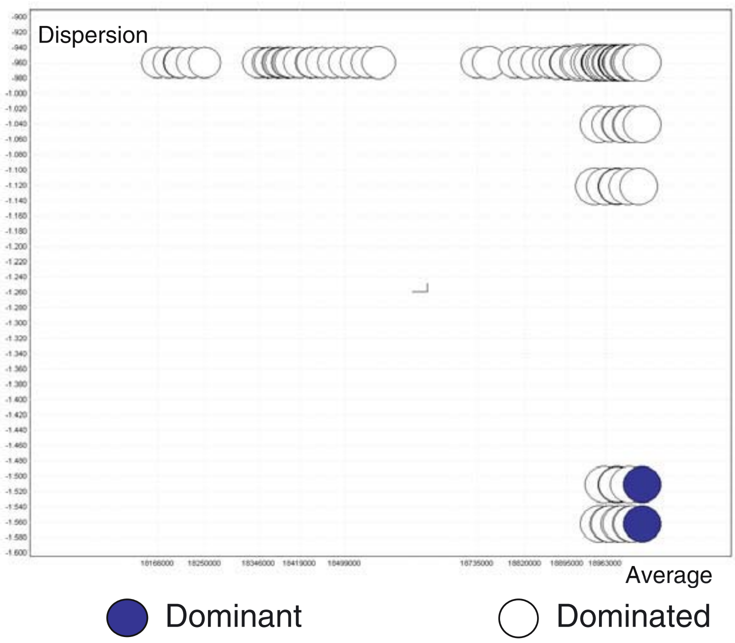


Figure 5.11. Dominance analysis diagram.

The application of the second step of our methodology leads to an optimal path list in which the feasible system evolutions are. Then, the dominance relation has to be analyzed in order to detect the best ones. The average-dispersion diagram is showed in Figure 5.11.

Table 5.8. Example: dominant solutions.

Period	*Dominant solution 1*	*Dominant solution 2*
0	(5,4,6,5)	(5,4,6,5)
1	(5,4,6,5)	(5,4,6,5)
2	(5,4,6,5)	(5,4,6,5)
3	(5,4,6,5)	(5,4,6,5)
4	(7,4,9,5)	(7,4,6,7)
5	(8,4,9,5)	(8,4,6,8)
6	(8,4,9,5)	(8,4,6,8)
7	(11,4,13,5)	(11,4,6,11)
8	(11,4,14,5)	(11,4,6,11)

Since the final choice has to be performed by the end-user, the two dominant solutions have to be analyzed. The solution having lower dispersion index is named "dominant solution 1", the other "dominant solution 2". In Table 5.8, the 4-tuple values of the state variables are supplied along with the various periods. As it is possible to note, in the interval period 0-3 the two solutions are equal. Then, the first solution increases the number of "type C actions", whereas the second one makes the quantity "type D actions" increases.

The analysis of the state variable values determines the correspondent optimal strategy-to-be-adopted. Finally, the best strategy to be pursued should be chosen according to subjective and not well-defined criteria. In such test case, since the two dominant solution have the same average profit, the "dominant solution 1" can be selected because of the lower dispersion index (see Figure 5.11).

4. Case study

4.1 Description

The proposed formalism can be effectively used to model real cases in order to obtain an evaluation of the different strategies that can be pursued in the different cases. For such a reason, an application software has been built in order to process the available information as described in the previous sections. In order to demonstrate the way such a tool can be used, a numerical example is presented.

The case study refers to a firm that operates in the automotive sector. Economical and technical parameters, such as machining center costs and capabilities, operating costs and prices have been obtained through the analysis of a real case and interviews with experts and can be summarized as described in the other unit reports.

From the architectural point of view, two different approaches exist to produce such goods. The former uses the *dedicated manufacturing flow line* architecture (DMFL), which can reach high production rates for a unique part type. The latter is based on the *Flexible Machining System* architecture (FMS) which can produce any part type with low production rates.

In the DMFL paradigm, the only variable to be set is the number of machines to be used, whereas in the FMS it is necessary to determine the number of machines (operating in parallel mode) along with the number of pallets for each part type.

Three classes of both DMFL and FMS system are available (each one having a specific machining center). For any FMS class, a load/unload station and a part carrier is needed. The pallet number parameter is variable and can vary both overall system production capacity and costs.

We consider five part types, with only three that can be obtained by using the DMFL. Any part type can be processed by FMS.

The state is defined as an n-tuple representing the number of machines for each DMFL system, the number of machines for each FMS system and the pallet number for each part type (they do not depend on the FMS class). The state definition is summarized in (5.1).

$$State = (L_1, L_2, L_3, F_1, F_2, F_3, P_1, P_2, P_3, P_4, P_5) \tag{5.1}$$

Where:

L_i is the number of machines for the DMFL class i ($i = 1, 2, 3$)

F_j is the number of machines for the FMS class j ($j = 1, 2, 3$)

P_k is the number of pallets for part type k ($k = 1, 2, 3, 4, 5$) in selected FMS class

In the test case, 12 periods (semesters) are considered: in each period, the analysis of the production capacity expansion is performed, so a specific number of feasible states is available. In particular, feasible transitions are described for each state of the current period toward the ones of the next period. The overall number of nodes is about 200, whereas the number of arcs is about 500. The description of the state number for each period is reported in Table 5.9. Input data on feasible configurations and transitions have been obtained by applying the models described in Chapter 4.

Table 5.9. Number of states for each period.

Period	*Number of states*
0	1
1	9
2	9
3	9
4	9
5	18
6	20
7	20
8	20
9	20
10	22
11	22
12	21

The hypotheses on the basis of such a distribution are:

- No resource can be dismissed (reduction of production capacity). For this reason each node at period t has a number of resources for each kind greater or equal to the corresponding one in the period $t-1$.
- Not all transitions from period $t-1$ to t having an increasing of capacity production are allowed.
- For each node in each period, at least a feasible transition toward the following period has to be provided.
- Any node in the period t can be reached from one node of the period $t-1$.
- All financial flows are discounted back (discounting back rate $r =$ 12%).

Input data are:

- architectures;
- resource classes (e.g. machining center, pallet) for each architecture;
- production capacity for each architecture and configuration;

Table 5.10. Dominant solutions.

Average[EUR]	*Dispersion[%]*
1072847	18.92%
1071928	11.72%
1071835	10.91%
1071235	9.843%
1070947	9.791%
1070659	9.744%

- market demand for each part type, in terms of minimum, maximum and average value (a triangular fuzzy number is defined by using such information);
- production costs for each part type;
- externalization costs;
- selling prices.

Costs for the non-utilization of an industrial facility for processing of a specific part type are supposed to be null in the case study.

4.2 Results

Computational test have been conducted on a 2GHz Pentium IV based PC system, equipped with 512MB RAM, MS Windows 2000 OS. The time needed to solve the presented problem is about four hours.

Optimal solutions and Pareto's dominance. The number of the obtained solutions is 258, each represented by a profit value expressed in the form of a triangular fuzzy number.

The graph (in which x-axis is the value in EUR currency of the average profit and y-axis is the fuzzy number dispersion index expressed in percentage)is reported in Figure 5.12.

Because of the necessity to both maximize the average profit and minimize the dispersion index, the dominant solutions are in the southeast region of the graph.

The list reporting the six dominant solutions is represented in Table 5.10, whereas the southeast region of the graph is in Figure 5.12, which highlights such dominant solutions is showed in Figure 5.13.

Configuration analysis. Starting from the dominance graph, it is important to analyze the meaning of the obtained solution in terms

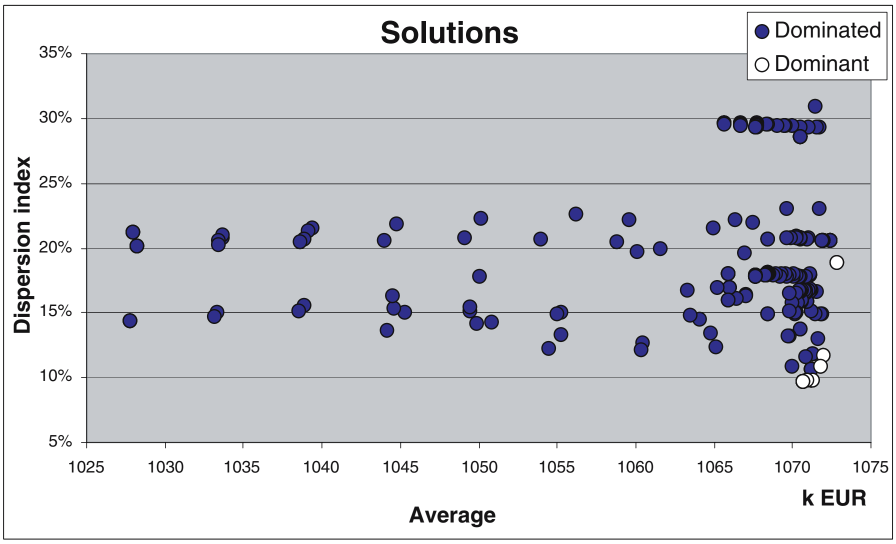


Figure 5.12. Complete set of the solution graph.

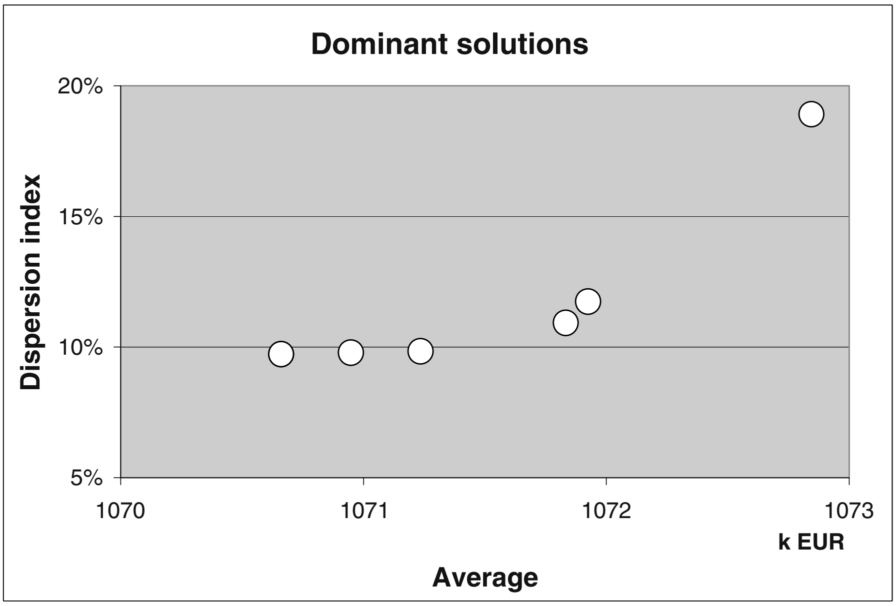


Figure 5.13. Dominant solutions.

Table 5.11. Performance of dominant solutions.

Solution Id	*Average profit*	*Dispersion index*
A	1072847	0.1892
B	1071928	0.1173
C	1071835	0.1091
D	1071235	0.0984
E	1070947	0.0979
F	1070659	0.0974

Table 5.12. Final states for the dominant solutions.

Solution Id	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*
A	1	1	1	0	0	4	16	8	1	5	2
B	2	1	1	5	0	0	15	8	2	4	3
C	1	1	1	5	0	0	16	8	2	4	3
D	1	1	1	4	0	0	16	8	2	4	3
E	1	1	1	5	0	0	16	8	2	4	3
F	1	1	1	0	4	0	16	7	1	5	2

of production system configuration and expansion capacity policy. In the following, the representation of the final states corresponding to the dominant solutions is given. Solutions are ordered in terms of average profit, in order to make it easier to spot them on the dominant graph. The greater the average profit is, the greater is the dispersion index: this implies multiple dominant solutions in the graph.

In Table 5.12, three classes of solutions can be identified on the basis of the class of adopted FMS.

With reference to Tables 5.11 and 5.12, we can note that:

- solution A is the only one using FMS class 3;
- solution B uses 2 machines of DMS class 1, but the same FMS class of solutions C, D and E;
- solutions C and E have the same final configuration, but having dominance coordinates different, the acquisition process is not the same in time;
- solution D uses fewer machining centers in the FMS;
- solution F adopts FMS class 2.

Table 5.13. Part type characteristics.

Part type	*1*	*2*	*3*	*4*	*5*
Profit Gi per piece if internally produced (EUR)	0.28	0.58	0.67	0.56	0.09
Profit Ge per piece if externally produced (EUR)	0.14	0.38	0.5	0.48	0.07
Profit increase $(Gi - Ge)/Ge$ for internal production compared to the external one (%).	100	53	34	17	29
Absolute difference between the profit values in internal and external production Gi-Ge (EUR)	0.14	0.2	0.17	0.08	0.02

The analysis of the 5 part types is described in Table 5.13. As it is possible to note, part type 1 and 2 are the most profitable to be produced. Therefore, it is possible to predict that such part types will be internally produced. On the contrary, the margin of profit decreases for part types 3, 4 and 5. Part type 3 has an average profitability level, whereas, part types 4 and 5 have a low profitability level, because of the low values in relative and absolute profit.

Acquisition operation analysis. In order to focus on the method used to acquire the resources, the system states evolution through time is displayed together with part types externally produced.

The acquisition mechanism for solution A is reported in Table 5.14. Initially, all the resources necessary to start the production are acquired. The only part types externally produced are 4 and 5. At period 5, an increase of the production capacity is performed for such part types. Only part type 3 is externally produced in minimal part (for such a reason no further resources are acquired).

The acquisition mechanism for solution B is reported in Table 5.15. Initially, the three less convenient part types are externally produced in part, then resources are acquired at period 5 and externalization is no more used. The following demand increase leads part type 3 to be produced externally in minimal part.

Acquisition mechanism for solution C is reported in Table 5.16. The only change consists in having less machining center for DMS class 1.

For solution D, the situation is identical (see Table 5.17). Indeed the only difference is the non-acquisition for an FMS machining center, compared to solution C. This leads to achieve lower profit and dispersion.

Table 5.14. Production capacity acquisition for Solution A.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	0	4	14	3	1	2	1	P4 P5
3	1	0	0	0	0	4	14	3	1	2	1	P4 P5
4	1	1	0	0	0	4	14	3	1	2	1	P4 P5
5	1	1	0	0	0	4	14	8	1	4	2	P3
6	1	1	1	0	0	4	14	8	1	4	2	P3
7	1	1	1	0	0	4	15	8	1	4	2	P3
8	1	1	1	0	0	4	16	8	1	5	2	P3
9	1	1	1	0	0	4	16	8	1	5	2	P3
10	1	1	1	0	0	4	16	8	1	5	2	P3
11	1	1	1	0	0	4	16	8	1	5	2	P3
12	1	1	1	0	0	4	16	8	1	5	2	P3

Table 5.15. Production capacity acquisition for Solution B.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
3	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
4	1	1	0	4	0	0	14	4	1	2	2	P3 P4 P5
5	1	1	0	4	0	0	14	8	2	4	3	
6	2	1	1	4	0	0	14	8	2	4	3	P3
7	2	1	1	5	0	0	15	8	2	4	3	P3
8	2	1	1	5	0	0	15	8	2	4	3	P3
9	2	1	1	5	0	0	15	8	2	4	3	P3
10	2	1	1	5	0	0	15	8	2	4	3	P3
11	2	1	1	5	0	0	15	8	2	4	3	P3
12	2	1	1	5	0	0	15	8	2	4	3	P3

Table 5.16. Production capacity acquisition for Solution C.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
3	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
4	1	1	0	4	0	0	14	4	1	2	2	P3 P4 P5
5	1	1	0	4	0	0	14	8	2	4	3	
6	1	1	1	4	0	0	14	8	2	4	3	P3
7	1	1	1	4	0	0	15	8	2	4	3	P3
8	1	1	1	4	0	0	15	8	2	4	3	P3
9	1	1	1	4	0	0	15	8	2	4	3	P3
10	1	1	1	5	0	0	15	8	2	4	3	P3
11	1	1	1	5	0	0	16	8	2	4	3	P3
12	1	1	1	5	0	0	16	8	2	4	3	P3

Table 5.17. Production capacity acquisition for Solution D.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
3	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
4	1	1	0	4	0	0	14	4	1	2	2	P3 P4 P5
5	1	1	0	4	0	0	14	8	2	4	3	
6	1	1	1	4	0	0	14	8	2	4	3	P3
7	1	1	1	4	0	0	15	8	2	4	3	P3
8	1	1	1	4	0	0	15	8	2	4	3	P3
9	1	1	1	4	0	0	15	8	2	4	3	P3
10	1	1	1	4	0	0	16	8	2	4	3	P3
11	1	1	1	4	0	0	16	8	2	4	3	P3
12	1	1	1	4	0	0	16	8	2	4	3	P3

Table 5.18. Production capacity acquisition for Solution E.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
3	1	0	0	3	0	0	14	4	1	2	2	P3 P4 P5
4	1	1	0	4	0	0	14	4	1	2	2	P3 P4 P5
5	1	1	0	4	0	0	14	8	2	4	3	
6	1	1	1	4	0	0	14	8	2	4	3	P3
7	1	1	1	4	0	0	15	8	2	4	3	P3
8	1	1	1	4	0	0	15	8	2	4	3	P3
9	1	1	1	4	0	0	15	8	2	4	3	P3
10	1	1	1	4	0	0	16	8	2	4	3	P3
11	1	1	1	5	0	0	16	8	2	4	3	P3
12	1	1	1	5	0	0	16	8	2	4	3	P3

Table 5.19. Production capacity acquisition for Solution F.

Period	*L1*	*L2*	*L3*	*F1*	*F2*	*F3*	*P1*	*P2*	*P3*	*P4*	*P5*	*Ext*
1	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	3	0	14	3	1	2	1	P3 P4 P5
3	1	0	0	0	3	0	14	3	1	2	1	P3 P4 P5
4	1	1	0	0	4	0	14	3	1	2	1	P3 P4 P5
5	1	1	0	0	4	0	15	7	1	3	2	P3
6	1	1	1	0	4	0	15	7	1	3	2	P3
7	1	1	1	0	4	0	16	7	1	4	2	P3
8	1	1	1	0	4	0	16	7	1	5	2	P3
9	1	1	1	0	4	0	16	7	1	5	2	P3
10	1	1	1	0	4	0	16	7	1	5	2	P3
11	1	1	1	0	4	0	16	7	1	5	2	P3
12	1	1	1	0	4	0	16	7	1	5	2	P3

Solution E is described in Table 5.18: 5 machining centers are acquired for the FMS as solution C does, leading to the same final state. The delayed acquisition of the machine at period 11, instead of 10 (solution C) provides a different profit.

Finally, solution F (reported in Table 5.19), follows the same logic as the previous ones. The only two differences are the different class of FMS and the constant use of external resource to produce part type 3, although in minimal part.

5. Conclusions

The formalism introduced consists in a Decision Support System able to model the firm behavior within uncertain environments. The DSS application that implements the proposed methodology can be used to take strategic decisions over a long term programming horizon, allowing an effective comparison of user-defined strategies according to user-defined efficiency parameters. Furthermore, the application can also be used to translate strategic decisions into concrete capacity expansion policies that behave according to the long term programmed strategy selected at the first step. Moreover, once a specific capacity expansion policy has been decided, and depending on problem specific parameters, quantitative indicators provide quality indices for the selected solutions, allowing the decision maker to specify the best actions the firm has to take in order to achieve the best performance.

References

Anglani, A., Grieco, A., Nucci, F., Semeraro, Q., and Tolio, T. (2000). A new algorithm to rank temporal fuzzy sets in fuzzy discrete event simulation. In *FUZZ-IEEE 2000 The 9th IEEE International Conference on Fuzzy Systems*, San Antonio, Texas, USA.

Benjaafar, S., Morin, T., and Talavage, J. (1995). The strategic value of flexibility in sequential decision making. *Eur J Oper Res*, 3(82).

Bernstein, P. (1998). *Against the Gods: The Remarkable Story of Risk.* John Wiley & Sons, New York.

Bhattacharyya, M. (1998). Fuzzy markovian decision process. *Fuzzy Sets and Systems*, (99):273–282.

Bretthauer, K. M. (1995). Capacity planning in networks of queues with manufacturing applications. *Mathl. Comput. Modellin*, 21(12):35–46.

Euwe, M. J. and Wortmann, H. (1997). Planning systems in the next century. *Computers in Industry*, (34):233–237.

Gupta, Y. and Goyal, S. (1989). Flexibility of manufacturing systems: Concepts and measurements. *Eur J Oper Res*, (43):119–135.

Hayes, R. and Wheelwright, S. (1984). *Restoring our Competitive Edge - Competing Through Manufacturing.* Wiley - New York.

Herrera, F. and Herrera-Viedma, E. (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, (115):67–82.

Jordan, W. and Graves, S. (1995). Principles on the benefits of manufacturing process flexibility. *Manage Sci*, 4(41).

Katok, E., Tarantino, W., and Harrison, T. P. (2003). Investment in production resource flexibility: An empirical investigation of methods for planning under uncertainty. *Naval Research Logistics*, 50:105–129.

Kwon, Y. W., Park, H., Jung, S., and Kim, T. (1996). Fuzzy-devs formalism: concepts, realization and applications. In *Proceedings Of The 1996 Conference On AI, Simulation and Planning In High Autonomy Systems*, pages 227–234.

Nguyen, Q. and Le, T. (1997). A fuzzy discrete-event simulation model. In *Proceeding of Australia-Pacific Forum on Intelligent Processing and manufacturing of Materials, Gold Coast, Queenland, Australia, July 14-17.*

Okada, S. and Soper, T. (2000). A shortest path problem on a network with fuzzy arc lengths. *Fuzzy Sets and Systems*, (109):129–140.

Olhager, J., Rudberg, M., and Wikner, J. (2001). Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning. *Int. J. Production Economics*, (69):215–225.

Sethi, K. and Sethi, S. (1990). Flexibility in manufacturing: A survey. *Int J Flexible Manuf Syst*, (2):289–328.

Wang, J. and Sparrow, F. T. (1999). The cost of uncertainty in capacity expansion problems. *Int. J. Energy Res.*, (23):1187–1198.

Yager, R. (2002). On the evaluation of uncertain courses of action. *Fuzzy Optimization and Decision Making*, (1):13–41.

Zeigler, B. and Vahie, S. (1993). Devs formalism and methodology: unity of conception/diversity of application. In *Proceedings of the 1993 Winter Simulation Conference.*

Chapter 6

FUZZY PERFORMANCE EVALUATOR OF AMSS

F. Caiazzo, R. Pasquino, V. Sergi, B. Spiezio
Dipartimento di Ingegneria Meccanica
Università degli Studi di Salerno
Via Ponte Don Melillo 1, 84084 Fisciano (Salerno), Italy
f.caiazzo@unisa.it,r.pasquino@unisa.it,v.sergi@unisa.it,b.spiezio@unisa.it

1. Introduction

In modern manufacturing environments, production system analysis is becoming more and more complex in consequence of an increasing use of integration and automation processes which call for far-reaching adjustments in the cost structures of firms: cuts on variable costs and corresponding increases in fixed costs with concomitant greater investment risks.

Due to the sheer complexity of both their production systems and associated investment risks, firms tend to earmark sizable proportions of their investment resources for the system design and management phases.

The most widely used analysis techniques are based on computer-aided simulation models, i.e. tools which are able to simulate any, even the most complex, aspects of a production system and which guarantee highly accurate results close to real-case scenarios.

Simulation techniques are used to obtain information on the behavior of a system by performing experiments on a representation of the system to be analyzed, called the model. These experiments are usually performed either on an existing system or on a system in the process of being designed. In the former case, simulation is preferred to other analysis methods in the following situations: when simulation offers cost-effective solutions guaranteeing less complex analysis procedures, when, in the absence of other exact or approximate analysis techniques, it is the only efficient method or, conversely, in the event such techniques would be available, when the experimentation requirements are not applicable to the production system to be analyzed. With respect to the latter case, examples in point are either nonexistent systems, i.e. systems still to be designed, or system upgrades which are still at the planning stage and have thus not yet been completed. In these situations, simulations can be used to predict and/or improve the performance characteristics of the system concerned.

In this connection, we also wish to mention that simulation techniques can only be used to evaluate a system, not reach an optimum solution. The relevant solutions must be worked out with the help of less complex techniques or by integrating the simulation model within an optimization routine capable of determining the optimal configuration of the system.

Static allocation is a generative technique which was widely used when less efficient computing equipment was available and is still adopted today for preliminary analysis. It is the simplest method of modelling a client-server type process. On the one hand, the process is assumed to be unrelated to time; on the other, reasonable cross-resource interactions are not considered. When this technique is used, the performance characteristics of the manufacturing system are determined at a stage when only the process plans of the tasks to be carried out within a given timeframe are known. Consequently, major shortcomings of the static allocation method stem from the fact that the effects of possible interactions between different resources will not be considered and that satisfactory results will only be obtained when large volumes of semifinished products are involved. Today, as a result of the increasing adoption of just-in-time, lean production and other similar techniques, large volumes of work-in process are the exception and this cannot fail to impact production system design and analysis methods.

Systems which require due attention to cross-resource interactions are generally analyzed based on queuing models, in which the individual input items of a production system are represented as queues which are then combined into a network.

As these analytical models are based on simplified assumptions, their accuracy levels are not the highest, though the results are certainly both more accurate and more significant than those obtainable using the static allocation method.

The models just mentioned (simulation, static allocation and queuing models) are based on the assumption that characteristic parameters such as interarrival times, service times and routing coefficients are either deterministic (static allocation) or probabilistic (queueing network and simulation). These hypotheses are certainly valid when historical data sets are available to describe the way these parameters are distributed, but they become less and less significant when it comes to developing a new system or when no or insufficient information is available on an existing system. In the latter case, experts on the system analyzed are asked to provide useful indications concerning the variables involved. Such information as is made available in this way is generally represented by linguistic phrases which can be properly represented using fuzzy numbers.

The use of fuzzy numbers to estimate the performance characteristics of production systems is associated with a number of difficulties which stem in part from the fact that the methods used to execute mathematical operations with fuzzy numbers are as yet not generally accepted and in part from the explosion of the support of the fuzzy sets fired by the considerable number of computations to be performed.

In this chapter we will analyse the problems associated with the description of vaguely known characteristic variables in a production system and the procedures to evaluate performance indices using queueing network models with fuzzy parameters.

2. Fuzzy sets and fuzzy numbers

Fuzzy sets theory was introduced by Lofti Zadeh at the end of the 1960s with the aim of providing a tool capable of describing problems in which vagueness stems from the absence of a criterion to distinguish clearly between different categories, rather than the absence of random variables.

Following a few years of mainly theoretical work, the earliest applications of fuzzy concepts began to appear mainly in fields such as controls, the representation of information and decision-making strategies. In the early 1980s, japanese researchers first explored the possibility of developing industrial applications using fuzzy logic. In a few years' time, fuzzy logic controls became an integral part of everyday life (Zimmermann, 1991).

The sector in which the use of fuzzy sets has failed to make headway is modelling, i.e. the development of mathematical models embodying fuzzy variables, because the relevant applications were found to be hardly significant as compared to the considerable computational effort involved.

Although a detailed description of fuzzy sets falls outside the scope of this chapter, for purposes of greater clarity it will be convenient to report a few definitions of fuzzy sets and fuzzy numbers and to illustrate the methods whereby arithmetic operations involving fuzzy numbers are executed.

2.1 Preliminary outline and definitions

A fuzzy set is just an extension of the mathematical notion of set (Klir and Yuan, 1995). A crisp set is defined by a boolean membership function μ which, when applied to any x element of the universe of discourse U, will return the true value if $x \in X$, and the false value in the opposite case. The membership function of a fuzzy set is not boolean. It is continuous, and its image is usually subsumed within $[0, 1]$; the truth values that can be determined with it are not only two, i.e. $x \in X$ and $x \notin X$, but an endless set of truth values comprised within $[0, 1]$. This means that to a certain extent any given element may be a member of more than one set at the same time. The membership degree is determined based on a membership function which is categorized as such only if it meets given requirements. In a large number of applications, these membership functions are not generic, but triangular or trapezoidal, which means that they can be defined by means of 3 or 4 parameters and that the membership degree of any component of the universe of discourse U can be rapidly calculated.

Definition - In a given set X, a fuzzy set of A in X is defined as a μ_A function:

$$\mu_A : X \to [0, 1]$$

μ_A is called the membership function of A and $\mu_A(x)$ is known as the membership degree of x within the given set A.

Definition - The support of a fuzzy set μ_A is the crisp set in which the value of the membership function is greater than zero:

$$supp\,(A) = \{x \in X : \mu_A(x) > 0\}$$

Definition - The elements of a fuzzy set μ_A whose membership degree is not lower than α are termed the *α-cut* of A:

$$^{\alpha}A = \{x \in X \ : \ \mu_A(x) \geq \alpha\}$$

Similarly, those with a membership degree higher than α are described as a strong *α-cut* of A:

$$^{\alpha+}A = \{x \in X \ : \ \mu_A(x) > \alpha\}$$

Each α-cut can be turned into a *special fuzzy set*:

$$_{\alpha}A(x) = \alpha \cdot {}^{\alpha}A(x)$$

From the above definitions it is possible to infer that a fuzzy set can be resolved into its constituent α-cuts and then restored by recourse to the following operation:

$$A = \bigcup_{\alpha \in [0,1]} {}_{\alpha}A(x)$$

Definition - The height of a fuzzy set A is defined as the highest membership degree of one or more elements subsumed within the set:

$$h(A) = \sup_{x \in X} (\mu_A(x))$$

Definition - A fuzzy set is defined normal if:

$$h(A) = \sup_{x \in X} (\mu_A(x)) = 1$$

Definition - In a given fuzzy set μ_A (defined by its set of real numbers $\Re$), μ_A is termed convex if:

$$\mu_A(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \geq \min[\mu_A(x_1), \mu_A(x_2)]$$

for each x_1, $x_2 \in \Re$ and with $\lambda \in [0,1]$
Definition - A fuzzy number $\tilde{A}$ is a fuzzy set with the following characteristics:

- it is normal then $h(A) = \sup(\mu_A(x)) = 1$;
- it has minimal support;
- ${}^{\alpha}A$ is a closed set for each $\alpha \in (0,1]$.

A triangular fuzzy number $\tilde{A}$ is a fuzzy set with a triangular membership function. It can be defined by the triple (a, b, c) such that:

$$\begin{cases} \mu_A(x) = 0 & x < a \\ \mu_A(x) = \frac{x-a}{b-a} & a \leq x \leq b \\ \mu_A(x) = \frac{c-x}{c-b} & b \leq x \leq c \\ \mu_A(x) = 0 & x > c \end{cases}$$

Triangular fuzzy numbers are widely used in fuzzy modelling because they can be represented without difficulty and handled using simple methods, for instance interval arithmetic.

2.2 Fuzzy arithmetic

Numerous potential approaches to the development of fuzzy arithmetic were proposed several years ago, but the only ones that have so far been accepted by most researchers and are widely used for arithmetic operations on fuzzy numbers are the fuzzy extension principle and interval arithmetic. This is also due to the fact that later research has not resulted in any generalized innovations.

The principle governing the fuzzification of crisp functions is known as the fuzzy extension principle. Introduced by Zadeh in 1975 , it entails that in a given function $\Phi : X^n \to Y$, the extension of Φ is defined as:

$$\Phi^* : F(X)^n \to F(Y)$$

$$\Phi^*(\mu_1, \ldots, \mu_n)(y) = \sup_{y=\Phi(x_1,\ldots,x_n)} \left(\min\left\{\mu(x_1), \ldots, \mu(x_n)\right\}\right)$$

The extension principle can be directly used to carry out elementary arithmetic operations $\{+, -, *, /\}$ on two fuzzy numbers $\tilde{A}$ and $\tilde{B}$.

Summation:

$$\mu_{A+B}(z) = \sup_{z=x+y} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

Subtraction:

$$\mu_{A-B}(z) = \sup_{z=x-y} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

Multiplication:

$$\mu_{A\cdot B}(z) = \sup_{z=x\cdot y} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

Division:

$$\mu_{A/B}(z) = \sup_{z=x/y} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

Maximum:

$$\mu_{\max[A,B]}(z) = \sup_{z=\max[x,y]} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

Minimum:

$$\mu_{\min[A,B]}(z) = \sup_{z=\min[x,y]} \left(\min\left\{\mu_A(x), \mu_B(y)\right\}\right)$$

An alternative approach (Kaufmann and Gupta, 1985) which further advances fuzzy arithmetic is based on interval arithmetic and, therefore, on the backing out and restoration of the fuzzy sets.

If A and B are two intervals of real numbers $A = [a_1, a_2]$, $B = [b_1, b_2]$, the four elementary operations are defined as follows:

$$\begin{aligned}
[a_1,a_2]+[b_1,b_2] = &\ [a_1+b_1, a_2+b_2]\\
[a_1,a_2]-[b_1,b_2] = &\ [a_1-b_2, a_2-b_1]\\
[a_1,a_2]\cdot[b_1,b_2] = &\ [\min(a_1\cdot b_1, a_1\cdot b_2, a_2\cdot b_1, a_2\cdot b_2)\ ,\\
&\ \max(a_1\cdot b_1, a_1\cdot b_2, a_2\cdot b_1, a_2\cdot b_2)]\\
[a_1,a_2]\,/\,[b_1,b_2] = &\ [\min(a_1/b_1, a_1/b_2, a_2/b_1, a_2/b_2)\ ,\\
&\ \max(a_1/b_1, a_1/b_2, a_2/b_1, a_2/b_2)]
\end{aligned}$$

In the fourth of these operations it is assumed that $0 \notin [b_1, b_2]$.

Moreover, if the extremes of intervals A and B are positive, the multiplication and division operations are reduced to:

$$[a_1,a_2]\cdot[b_1,b_2] = [a_1\cdot b_1, a_2\cdot b_2]$$

$$[a_1,a_2]\,/\,[b_1,b_2] = [a_1/b_2, a_2/b_1]$$

Due to its comparatively low complexity, interval arithmetic is the method of choice for operations on fuzzy numbers. The considerable computational effort entailed in solving even comparatively small problems is generally overlooked because of the high processing speed of modern computers. However, according as the variables involved increase in number, computing times are also seen to increase dramatically. Suffice it to think that a 1.19 GHz processing unit takes 0.01 s to add up two fuzzy numbers divided into 101 α-cuts in a Matlab environment.

To overcome the severe limitations of interval arithmetic, researchers have mainly concentrated on methods capable of reducing both the computational complexities of operations on fuzzy numbers and the explosion of the support which occurs when numerous or recursive computations are to be performed.

A review of the literature shows a tendency to improve on widely accepted techniques, rather than working out innovative computation procedures.

A case in point is George J. Klir, in (Klir, 1997). While confirming the paramount role of interval arithmetic, he suggests limiting the number of values that a fuzzy variable can assume when it is iteratively subjected to the same operation.

Examples are subtractions and divisions involving the same interval of two real number $A = [\underline{a}, \overline{a}]$ and performed in line with the canons of interval arithmetic:

$$[\underline{a}, \overline{a}] - [\underline{a}, \overline{a}] \qquad [\underline{a}, \overline{a}]/[\underline{a}, \overline{a}]$$

with the following result:

$$[\underline{a} - \overline{a}, \overline{a} - \underline{a}] \qquad [\underline{a}/\overline{a}, \overline{a}/\underline{a}]$$

Using the method suggested by Klir, the result is:

$$[\underline{a} - \underline{a}, \overline{a} - \overline{a}] = [0, 0] \qquad [\underline{a}/\underline{a}, \overline{a}/\overline{a}] = [1, 1]$$

because a variable, even if fuzzy, cannot simultaneously assume different values.

The effect of this constraint is to enable a crisp result even setting out from fuzzy input data. This solution seems to be appropriate especially in the light of the fact that the vagueness of a variable depends on the kind of information available, not on the range of values that this can assume. In a situation with numerous vague variables, each of these adds to the total level of uncertainty, which means that arithmetic computations will have to be carried out in respect of all possible combinations.

Operations on intervals performed in line with Klir's constraints produce the effect of somewhat reducing the support of the fuzzy sets which constitute the results of the problem analyzed. To account for the fact that his proposal is not particularly innovative, Klir argues that fuzzy numerical arithmetic is still at its initial stage and that many mathematical and computational issues need to be solved before it can become fully operational.

Other researchers have attempted to reduce the computational complexity of interval arithmetic by introducing alternative representations of fuzzy numbers describing shape based on just a few significant parameters useful in implementing arithmetic computations. A particularly interesting study on this subject is reported in Albrecht Irion (Irion, 1998). In his work, Irion introduced a new kind representation pattern of triangular fuzzy number consisting of a quintuple of values and used this coding method to devise an arithmetic calculation method alternative to interval arithmetic. Irion has introduced the following characteristic values of the quintuple: the element with the highest membership degree (m), the lower extreme (l), the upper extreme (r), the height of the triangle (h), and the upper limit (b). The meanings of these parameters are illustrated in Figure 6.1.

Based on this representation, Irion has worked out a method that can be used to perform both arithmetic and logical operations involving fuzzy sets.

Let $\tilde{A}$ and $\tilde{B}$ be two triangular fuzzy numbers that can be summarized in the two quintuples: $(m_A, r_A, l_A, h_A, b_A)$ and $(m_B, r_B, l_B, h_B, b_B)$.

Arithmetic operations on these two numbers can be performed based on the indications provided in the table below.

The shortcoming inherent in this method is the fact that only triangular numbers can be represented and, consequently, subjected to logic-

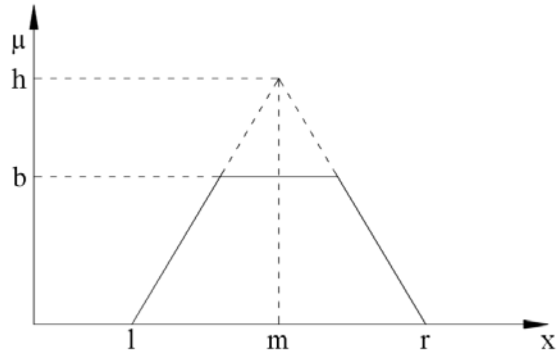


Figure 6.1. Representation of a fuzzy number in line with A. Irion's proposal.

Table 6.1. Arithmetic operations with fuzzy numbers in line with Irion's representation method.

Operation	m_C	l_C	r_C	h_C	b_C
$-a_F$	$-m_A$	$-r_A$	$-l_A$	$-h_A$	$-b_A$
$1/a_F$	$1/m_A$	$1/r_A$	$1/l_A$	h_A	b_A
$a_F + b_F$	$m_A + m_B$	$l_A + l_B$	$r_A + r_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F - b_F$	$m_A - m_B$	$l_A - r_B$	$r_A - l_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F \cdot b_F$ $a_F, b_F > 0$	$m_A \cdot m_B$	$l_A \cdot l_B$	$r_A \cdot r_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F \cdot b_F$ $a_F, b_F < 0$	$m_A \cdot m_B$	$r_A \cdot r_B$	$l_A \cdot l_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F \cdot b_F$ $a_F > 0, b_F < 0$	$m_A \cdot m_B$	$r_A \cdot l_B$	$l_A \cdot r_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F \cdot b_F$ $a_F < 0, b_F > 0$	$m_A \cdot m_B$	$l_A \cdot r_B$	$r_A \cdot l_B$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F^{b_F}$ $a_F > 1, b_F > 0$	$m_A^{m_B}$	$l_A^{l_B}$	$r_A^{r_B}$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F^{b_F}$ $0 < a_F < 1$ $b_F < 0$	$m_A^{m_B}$	$r_A^{r_B}$	$l_A^{l_B}$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F^{b_F}$ $a_F > 1, b_F < 0$	$m_A^{m_B}$	$r_A^{l_B}$	$l_A^{r_B}$	$\min(h_A, h_B)$	$\min(b_A, b_B)$
$a_F^{b_F}$ $0 < a_F < 1$ $b_F > 0$	$m_A^{m_B}$	$l_A^{r_B}$	$r_A^{l_B}$	$\min(h_A, h_B)$	$\min(b_A, b_B)$

mathematical operations. The two methods just described are far from universally valid tools, in terms that they are not applicable in every situation.

3. Describing uncertainty

In engineering, uncertainty is often modelled by recourse to descriptive statistics, i.e. using probability distribution to model the parameters involved.

This approach is certainly appropriate in situations characterized by the availability of a large quantity of data, but is less effective when the smaller number of input data available is not enough to describe a random variable. In this case, it is necessary to make hypotheses concerning possible probability distributions which may even be in conflict with the actual process flow.

Very often, details concerning a process, e.g. processing time, the time needed to carry out a given maintenance action, etc., can only be obtained by interviewing an expert and it may happen that this expert even lacks the competency needed to provide a mathematical description of the trend of a variable. More often than not, the expert will cast this information in a type of language which is ill suited to be modelled into a probability distribution.

An alternative tool to describe uncertainty is fuzzy sets theory, based on which system variables can be represented by formulating linguistic variables.

A linguistic variable is characterized by a quintuple:

$$(x, T, U, g, m)$$

where x is the name of the variable, $T(x)$ is the set of linguistic terms of x which are related to the variable whose range of values is comprised within the universe U, g is a syntactical rule applicable to the generation of linguistic terms, and m is a semantic rule which assigns a given meaning to each linguistic term.

Figure 6.2 illustrates a linguistic variable in which " $T(x)$" is represented by the set [very low,..., very high], " U" is the set of possible values that the variable " x" can assume, " g" is the syntactical rule which combines each value of the set T [very low,, very high] with the corresponding fuzzy set, " m" is a semantic rule used to assign the correct meaning to each component of the set " $T(x)$".

Based on linguistic variables, the indications provided by a process expert can be quickly and simply coded using mathematical magnitudes, i.e. fuzzy sets, which describe the vagueness of such data and can be handled by recourse to appropriate mathematical tools.

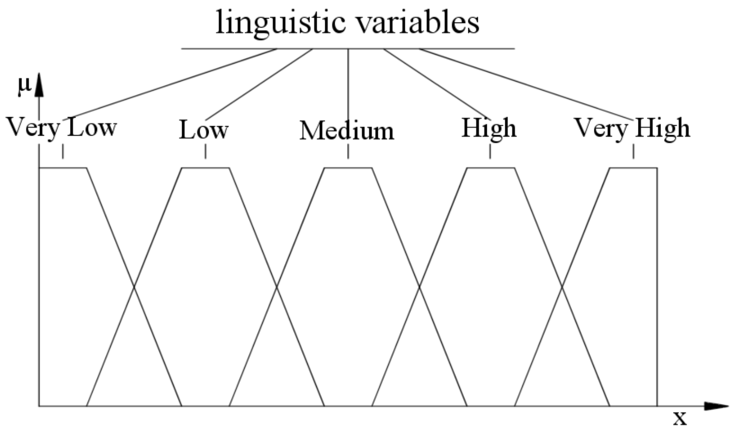


Figure 6.2. Structure of a linguistic variable.

The fuzzy sets generally used to represent such information are characterized by triangular or trapezoidal membership functions. This does not automatically entail that other categories of functions are not suited to describe the relevant uncertainty. It only involves that at least in comparatively low-detail studies these functions cover a wide range of situations typical of systems design.

In manufacturing systems design, a typical piece of "linguistic information" provided by an expert might be the phrase "processing takes about 3 minutes". This phrase can be turned into a triangular fuzzy number in which the component with the maximum membership degree is 3 (Figure 6.3 shows the fuzzy set which represents the linguistic variable concerned).

Similarly, the phrase "the interarrival time may range between 3 and 5 minutes" can be turned into a trapezoidal fuzzy set in which the components with the highest membership degrees are between 3 and 5 (Figure 6.4).

The indications concerning the support of the fuzzy set (uncertainty of the information) will also be provided by the expert. As will be shown in the paragraph on fuzzy sets, focused interviews can be used to reconstruct the shape of the investigated set in lesser or greater detail, according to need, using interpolation techniques.

4. Linguistic modifiers

Hedges are special linguistic terms which may change the truth value of a given proposition. Words such as very, much, somewhat, extremely,

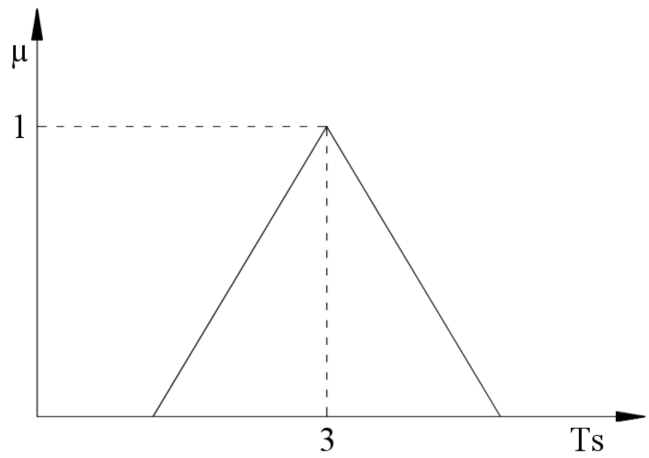


Figure 6.3. The triangular fuzzy number for the phrase "processing takes about 3 minutes".

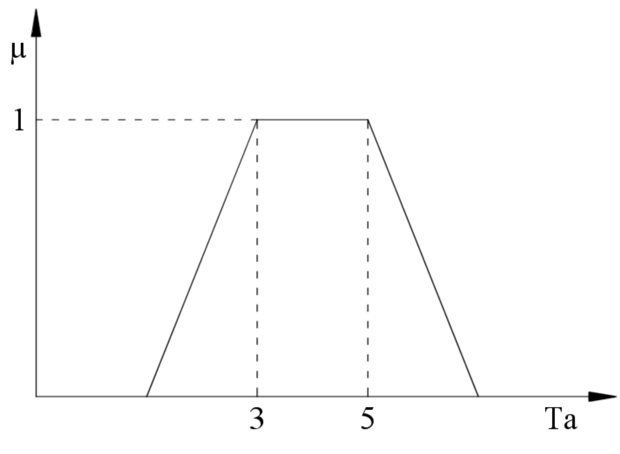


Figure 6.4. The trapezoidal fuzzy number for the phrase "interarrival time takes between 3 and 5 minutes".

etc., are examples of linguistic modifiers. Literally speaking, a hedge is a barrier and a constraint. In the case of a fuzzy set, it denotes a modifier which acts on the shape of the set. The proposition "*x is young*", for example, which may assume the truth value "*x is young, this is true*", can be modified by the hedge "quite" as follows:

- x is quite young, this is true;
- x is young, this is quite true;
- x is quite young, this is quite true.

In general, when a fuzzy proposition such as:

$$p : x \;\; is \;\; F$$

is imposed along with a linguistic modifier H, the proposition can be changed as follows:

$$Hp : x \;\; is \;\; HF$$

where HF reflects the attribute fuzzy obtained by applying the modifier to the attribute F. Other changes can be made when modifiers are applied to the truth value.

It is worth specifying that no modifiers can be applied to crisp attributes; e.g. the phrases "quite horizontal" or "quite parallel" or "quite rectangular" make no sense in our common language and are consequently out of place in classical logic.

The linguistic modifier H can be interpreted as a unary operation h performed on the interval $[0, 1]$; e.g. the modifier "quite" is often interpreted using the unary operation $h(a) = a^2$, while the modifier "enough" can be interpreted using the unary operation $h(a) = \sqrt{a}$ with $a \in [0, 1]$ (Figure 6.5).

When a fuzzy attribute is combined with a modifier such as H which is represented by the unary operation h, the modified fuzzy attribute is determined for each $x \in X$ by equation:

$$HF(x) = h(F(x))$$

If $h(a) < a$ with $a \in [0, 1]$, the modifier is termed strong; if $h(a) > a$ with $a \in [0, 1]$, it is termed weak. In the particular case in which $h(a) = a$ the operator is termed an identity modifier.

A strong modifier increases the truth value of the fuzzy set to which it is applied and thus reduces the truth value of the proposition to which

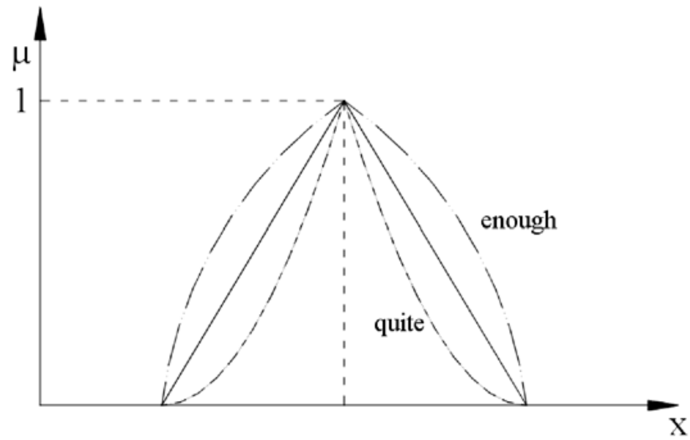


Figure 6.5. Applying the modifiers "quite" and "enough" to a fuzzy set.

it is applied. Conversely, a weak modifier weakens the attribute, and thus the truth value associated with it, and increases the truth value associated with it. Let us consider, for example, three propositions:

- p_1: the processing time is short;
- p_2: the processing time is quite short;
- p_3: the processing time is short enough .

The variation in the truth value produced by the two modifiers can be immediately calculated considering that where the membership degree of the processing time 13.5 in the fuzzy set "short processing times" is 0.5, then the membership degree of the same processing time in the fuzzy set "very short processing times" will be $0.25 = 0.5^2$ and that of the phrase "the processing time is quite short" will be $0.707 = \sqrt{0.5}$ (Figure 6.6).

These values are fully consistent with a pragmatic intuition of some sort: the stronger the assertion, the weaker its truth value and vice versa.

Moreover, it is convenient to point out that a modifier such as h will satisfy the following conditions:

- $h\,(0) = 0 \qquad h\,(1) = 1$;
- h is a continuous function;
- if h is a strong modifier, then h^{-1} is a weak modifier;

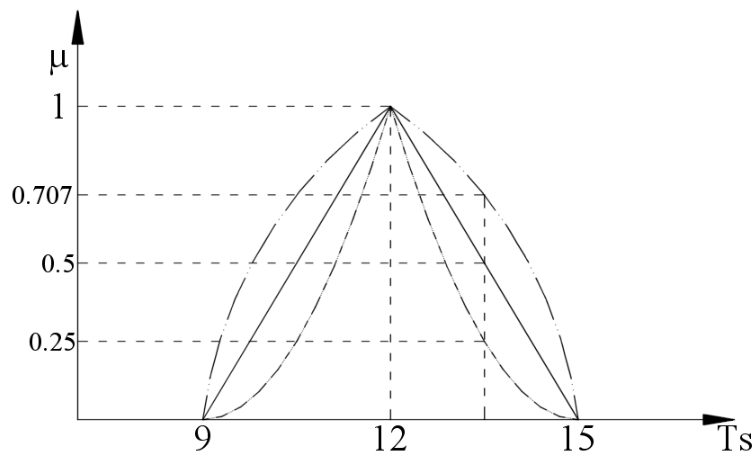


Figure 6.6. The variation in truth value produced by applying the modifier "quite" and enough to the fuzzy set "short processing time".

- in a given modifier such as g, the functions $g \circ h$ and $h \circ g$ are likewise modifiers, moreover, if g and h are both strong (or weak) modifiers, when used in combination they will produce a strong (or weak) modifier.

A family of functions which satisfy the above-mentioned conditions is:

$$h_\alpha(a) = a^\alpha$$

if $\alpha < 1$, then h_α is a weak modifier; if $\alpha > 1$, h_α is a strong modifier. Using this family of modifiers, it is possible to fully represent all the possible linguistic variations of a fuzzy variable.

5. Constructing fuzzy sets

As mentioned repeatedly in the foregoing paragraphs, fuzzy sets are used as "formal describers" of linguistic variables in a given context.

Numerous methods have been adopted to construct membership functions, but the most important of these – both because they secure satisfactory results and because they are by now generally accepted – are the direct and indirect methods. Analytical and computational techniques have mainly been used over these past years either to construct fuzzy sets from small historical sets or to "interpret" data provided by experts in the manners pointed out above.

Regardless of the method used, direct or indirect, membership functions are reconstructed based on the information provided by one or more experts. The difference between these two methods lies basically in the fact that the former uses explicit information which can be directly turned into membership functions, while the latter uses expert information of a generic kind, so as to avert the risk that the structure of the fuzzy set may be affected by personal opinions.

When the direct method is used with a given universe of discourse X or a set which is assumed to include a definition of the fuzzy variable whose shape is to be reconstructed, an expert is asked to assign to each $x \in X$ element a membership degree $\mu_A(x)$ capable of fairly representing the bearing that the element concerned has on the fuzzy set A. This can be done by either providing a membership function already expressed in mathematical terms, or by imposing membership degrees for some elements of the set X.

When more experts are interviewed, the simplest method for reconstructing the fuzzy membership function is to determine the mean value of the membership degrees $a_i(x)$ that have been assigned to the element x. However, at times it may be convenient to consider the degrees of competence of the individual experts asked to state their opinions on the problem concerned. In this case a competency degree c_i with $i = 1...N$ being N the total number of experts interviewed, and it is assumed that the membership degree of one element within the set reflects the weighted average of the membership degrees with the competency degrees:

$$\mu_A(x) = \sum_{i=1}^{N} c_i \cdot a_i(x) \qquad \forall x \in X$$

and that the competency degree c_i is a normalized parameter, i.e. a parameter which satisfies the condition:

$$\sum_{i=1}^{N} c_i = 1$$

As mentioned above, the drawback of the direct method is that the expert must be asked to provide a comparatively exact answer which in some cases may result in arbitrary results. The indirect method aims to reduce the risk of arbitrary results by substituting direct membership degree estimates for the results of comparisons between pairs of elements subsumed within the set X. The expert is asked to compare pairs of elements selected from those subsumed within the set $[x_1, x_2, \ldots, x_n]$ and to assign a relative membership degree to the set A. The results of multiple comparisons give rise to a square matrix whose p_{ij} elements represent the relationship (a_i/a_j) between the membership degrees of the elements x_i and x_j within the fuzzy set. The values of individual membership degrees a_i can be determined by solving a simple system of linear equations. When the direct method is used and more than one experts is involved, the values stated by these can be weighted, as in the situation described above for the direct method, based on the different competency degrees of the individual experts.

For the formulations of fuzzy sets to be really useful, the results must be interpolated in such a way as to obtain mathematical-computational expressions capable of providing a membership degree value for each of the elements subsumed within the universe of discourse. Regardless of the technique used, a basic requirement is to obtain a convex set which always satisfies the condition that:

$$\mu\left(\lambda \cdot x_1 + (1-\lambda) \cdot x_2\right) \geq \min\left[\mu\left(x_1\right), \mu\left(x_2\right)\right]$$

both with x_1, $x_2 \in \Re$ and with $\lambda \in [0,1]$.

Apart from the least squares method, the most widely used interpolation technique seems to be Lagrange interpolation. The relevant interpolation function is:

$$f(x) = a_1 L_1(x) + a_2 L_2(x) + \ldots + a_n L_n(x)$$

where:

$$L_i(x) = \frac{(x-a_1)\ldots(x-a_{i-1})(x-a_{i+1})\ldots(x-a_n)}{(x_i-a_1)\ldots(x_i-a_{i-1})(x_i-a_{i+1})\ldots(x_i-a_n)}$$

while the pairs $(x_1, a_1), (x_2, a_2), \ldots, (x_n, a_n)$ are experimental data or indications provided by experts.

An alternative method for constructing fuzzy sets is to have the experimental points learnt by a neural network. The network acts as a universal interpolator in that it accurately follows the trends of these points. The problem with neural networks is they do not guarantee that the fuzzy sets will be convex because data overfitting might result in generating fuzzy sets with shapes other than convex. Moreover, as neural networks make it impossible to generate a mathematical expression of the interpolation function, they can only be used when suitable hardware/software aids are available.

6. Queuing systems

A queuing system can be described as a system in which tasks are seen to arrive in order to obtain a given service and one or more servers

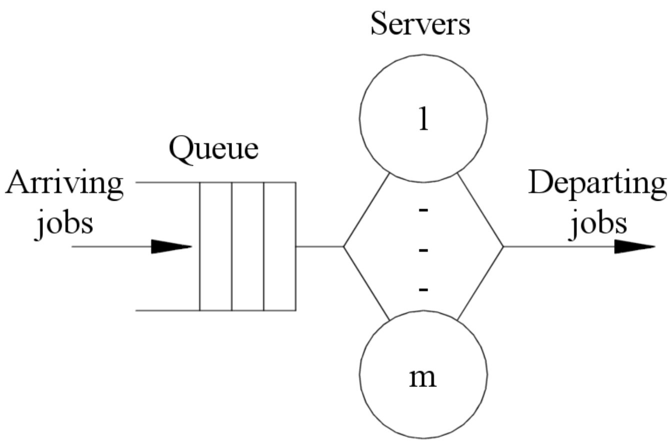


Figure 6.7. Configuration of a client-server system.

are waiting to provide the relevant service. The first researcher to address this subject was Erlang. In the early years of the 20th century, Erlang published a number of studies reporting applications of probability theory on telephone conversations. Until the 1950s, Erlang's theories were only used in the telecommunications sector; later on, numerous innovations led to applications in the aeronautics and warehouse management sectors and, more recently, in the analysis of processing units and computer networks.

The elements to be identified and characterized in a system that has to be modelled according to the queuing system format include the probability distribution of the time interval at which the tasks concerned are expected to reach the server station, the capacity of the queue constituted by the tasks waiting to gain access to the server, the probability distribution of the times when service is provided, the number of servers and the applicable queue discipline (Gross and Harris, 1985).

The five characteristic parameters just mentioned are usually summarized by means of the Kendall notation A/B/x/y/Z (A is the arrival pattern, B service's scheme, x the number of servers, y the system capacity and Z the queue discipline).

Figure 6.7 illustrates the configuration of a queuing system.

The following are the main efficiency parameters of a queuing system:

- average number of items within the system L and average queue length L_q;
- average time a items stays within the system W and average queuing time W_q;

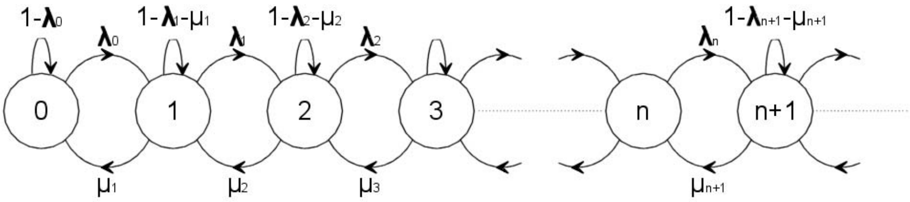


Figure 6.8. The birth/death process.

- utilization of the resources ρ;
- throughput value or system productivity X.

The problem will only be solved if the analysis is conducted in a stationary condition and provided simplified hypotheses are made concerning both service system configuration (exponential distribution of service times) and arrival rates (Poissonian process). In all other cases it is possible to obtain approximate solutions or bounds which subsume the real solution.

In an M/M/1 system, i.e. a boundless capacity system with a single server and exponential interarrival and service times, the trend of the system can be modelled using a simple continuous-time Markov chain (Bolch et al., 1998): an endless birth/death process (Figure 6.8).

If we write the balancing equations for a generic state bearing in mind that no death is possible when no items are circulating within the system, we can determine the probability of a steady state condition and, consequently, the performance characteristics of the system.

The literature on this subject (Li and Lee, 1989; Negi and Lee, 1992) reports a number of queuing system models based on fuzzy parameters in which traditional techniques can be used to work out the relevant solutions.

In (Li and Lee, 1989), Li and Lee address a single server system FM/FM/1 in which arrival and service rates are Poissonian processes with fuzzy parameters. Based on the concept of the α-cut, the FM/FM/1 system is turned into an M/M/1 queue family. As soon as the solution for a non-fuzzy system is known, the relevant crisp solution can be calculated and the fuzzy solution is reconstructed.

Besides studying the FM/FM/1 system, Li and Lee also used α-cuts to work out solutions for F/M/1, M/F/1 and F/F/1 systems.

As will be shown below, the importance of FM/FM/1 systems, whose analysis can be easily extended to FM/FM/s systems, lies in the fact that they provide a method to solve queue networks with fuzzy parameters.

7. Open queuing network models

Very often, a manufacturing system can be modelled as a queuing network, i.e. a set of elementary systems composed of a number of variously interconnected queues. The method for working out solutions with systems that can be modelled as open queuing networks is owed to Jackson (Gross and Harris, 1985). Having made a number of suitable hypotheses, Jackson formulated the following theorem. If, in a queuing system, the ergodicity hypothesis is satisfied for each node, then the probability of a steady state condition of the network can be expressed as the product of the probability levels of the stationary conditions of each node:

$$\pi\left(L_1, L_2, \ldots, L_n\right) = \pi_1\left(L_1\right) \cdot \pi_2\left(L_2\right) \cdot \ldots \cdot \pi_n\left(L_n\right)$$

A direct effect of this theorem is that each node of this network can be considered as independent of the others, so that the performance characteristics of the whole network will be the sum of the performance characteristics of the individual nodes.

To work out a solution for the network, the net rate of the arrivals at each node must be calculated based on the following relationship:

$$\lambda_i = \lambda_{0i} + \sum_{j=1}^{N} \lambda_j \cdot p_{ji}$$

where λ_i is the net arrivals rate at the i-th node, λ_{0i} the outer arrivals rate at the i-th node, p_{ji} the probability that an item will migrate from the j node to the i node and N the number of machines comprised within the network.

The mathematical models that have been developed so far in order to work out solutions for open queuing networks with fuzzy parameters were invariably based either on the extension principle or on the application of interval arithmetic to the relationships needed to determine the performance characteristics of crisp queuing systems.

The problems associated with the explosion of the support of the performance characteristics are now experienced at the stage when it comes to solving the linear equation system reported above, where fuzzy sets are found in place of crisp arrivals rates. Research studies reported in the literature address this problem by substituting fuzzy sets for crisp arrival rate values, but this procedure has not led to satisfactory solutions. A far more interesting case is that in which a solution is worked out by means of optimization techniques, although these techniques require constraints which are not easily identifiable when it comes to determining net arrival rates.

8. Closed queuing network models

The characteristic of a closed queuing network is the fact that the number of tasks circulating within it remains constant in time.

In analytical terms, the product form solutions for these networks can be expressed in very simple terms, but state probabilities computing requires considerable computational effort.

This is why numerous algorithms have been developed to determine the performance characteristics of closed queuing network models. The most versatile and efficient of these are the convolution algorithm and mean value analysis. The former has been little used in determining the performance characteristics of manufacturing systems with fuzzy parameters. The reason for this is the greater computational effort entailed and, as mentioned before, the fact that an increasing number of operations may unduly extend the support. Conversely, mean value analysis has been fairly extensively addressed by a number of authors and, as will be shown below, its results have so far proved fairly satisfactory.

8.1 Mean Value Analysis

The mean value analysis algorithm (MVA) was purposely developed by Reiser and Levemberg (Reiser and Levemberg, 1980) for use in analyzing closed queuing networks. It is based on two fundamental equations and can be used to determine the mean value of important indices such as average queuing time, throughput and the average number of parts in each node.

A closed queuing system (Figure 6.9) is made up of a number N of items or parts which are circulating in it and remain constant in time. The system also comprises M nodes, each of which is characterized by a service parameter μ_i and by an average number of visits ν_i.

The two laws based on which performance can be iteratively determined are:

- *Little's law*, which defines the relation between the average number of items at each node (L) and the average queuing time (W) by

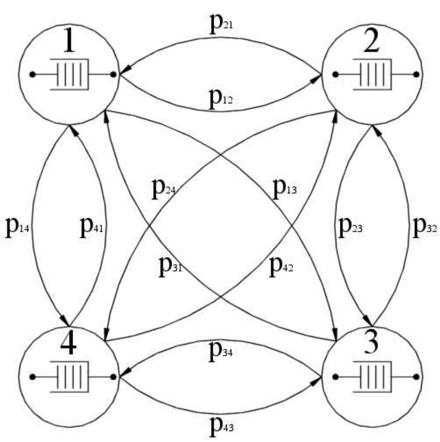


Figure 6.9. Model of a typical totally interconnected closed queuing network with four nodes.

means of the arrivals rate (λ):

$$L = \lambda \cdot W$$

- *Theorem of the distribution at arrival time.* Based on the arrival theorem, in a closed product-form queuing network, the pdf of the number of jobs seen at the time of arrival at a node i when k tasks are circulating in the network is equal to the pdf of the number of jobs at this node with one less job in the network.

These two laws are both intuitive and mathematical laws. Indeed, a quite banal observation concerning Little's law is the following: when an item spends in the system an average number of x units of time and y additional items are simultaneously introduced into the system during each unit of time, the average number of items will be the product $x \cdot y$. An intuitive justification of the second law is the fact that an item which is just arriving at a given node is as yet not comprised within the queue, and that consequently the system cannot contain more than $k-1$ items in a position to interfere with the new entrant.

The MVA algorithm can be implemented both for single-class systems, i.e. systems comprising a single type of items (e.g. when aggregate analysis are performed), and for multi-class systems, i.e. systems in which distinct categories of information are to be obtained for each type of items.

Both in single and multi class systems, the equation used to determine performance is based on the distribution of arrivals theorem and links the average queuing time to the number of items present at a node:

$$W_i^n = ts_i \cdot \left(1 + L_i^{n-1}\right)$$

When a only one station is present at a single node, a fairly simple observation is that an item's average queuing time at the i-th node is the sum of the time needed to process the items already comprised within the queue and the time needed to process those arriving.

The next step is to determine the throughput of the system as a whole:

$$X^n = \frac{n}{\sum_{i=1}^{M} W_i^n \cdot \nu_i}$$

that of a single node:

$$X_i^n = \nu_i \cdot X^n$$

and, lastly, the average number of items queuing at one node:

$$L_i^n = X^n \cdot \nu_i \cdot W_i^n$$

If the process illustrated above is reiterated N times (number of items circulating within the system), all of the performance characteristics can be determined using $n = 1, \ldots, N$.

8.2 Background on MVA fuzzy solutions

In the past few years, the extension principle and interval arithmetic were the main tools used to develop fuzzy queuing models. Although the relevant solutions are not directly usable due to computational and, to a certain extent, conceptual complexities, they are useful as benchmarks to conduct comparisons with those obtained through the adoption of alternative methods.

The researchers who seem to have most effectively advanced the "fuzzification" of the MVA algorithm are Jo, Tsujimura, Gen, Yamazaki and Lee (Jo et al. 1997) and Perrone and Noto La Diega (Perrone and Noto La Diega, 1998).

The study done by Eastern researches is simply a fuzzification of the MVA algorithm. They suggested a rather uncritical approach, consisting of the simple substitution of fuzzy numbers for the characteristic equations solved within MVA. Although this approach may appear formally correct, in practice the relevant solutions are quite insignificant, except for the fact that the performance values obtained are those with the highest membership degree with respect to the fuzzy set considered.

Evidence of the scant significance of the results obtained by merely embodying fuzzy numbers in the MVA algorithm is provided by a very

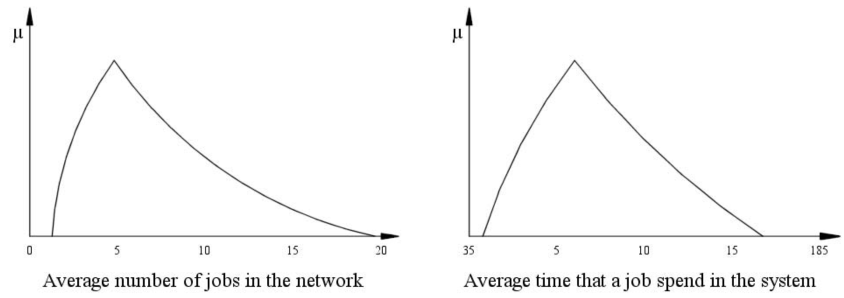


Figure 6.10. Fuzzy sets reflecting lead time and average number of tasks in the system.

simple example. Let us imagine that five same-type items are circulating in a closed queuing network and that each of these needs to be processed at two work stations. Assuming that processing times are represented by triangular fuzzy numbers whose characteristic values are $\tilde{t}s_1 = (14,\ 15.5, 16)$ and $\tilde{t}s_2 = (5.5,\ 7, 8)$, the values reported in Figure 6.10 will be obtained for the average number of items in the system and average lead time parameters.

Lead times range between a minimum of 40.8 and a maximum of 160.3 units of time, while the number of items present in the system ranges between a minimum value of 1.27 and a maximum value of 19.64.

The first observation to be made is that the domain of the fuzzy set which reflects lead times does not afford any inferences because the extreme values of the support do correspond to possible events, although they are associated with a minimum membership degree. One further remark concerns the work in process as a whole. One of the values to be determined is the average number of items within the system: this must be 5 and should not change throughout the evolution of the algorithm. Variability is linked to the fact that the average number of items within the system has been computed as the sum of the average number of items at each node. This relation is valid when the algorithm is iterated using crisp parameters, but it is either not valid or contradictory as soon as fuzzy numbers are used.

Compared to the method illustrated above, Perrone and Noto La Diega's approach shows innovative elements stemming from the introduction of a relation between fuzzy variables which reduces the support of the fuzzy sets which is obtained when the algebraic operations are reiterated. With slight changes, this relation makes use of Klir's constraint (Klir, 1997), i.e. the concept that a relation which embodies fuzzy numbers cannot be independent of the value of the variable.

To shed light on this relation, let us consider two triangular fuzzy numbers, $\tilde{A} = [a_1, a_2, a_3]$ and $\tilde{C} = [c_1, c_2, c_3]$, and execute the operation $\tilde{B} = \tilde{A} - \left(\tilde{A} - \tilde{C}\right)$ under due consideration of all possible combinations. The result is:

$$\begin{matrix} a_1 - (a_1 - c_3) & a_1 - (a_2 - c_2) & a_1 - (a_3 - c_1) \\ a_2 - (a_1 - c_3) & a_2 - (a_2 - c_2) & a_2 - (a_3 - c_1) \\ a_3 - (a_1 - c_3) & a_3 - (a_2 - c_2) & a_3 - (a_3 - c_1) \end{matrix}$$

However, as we examine the matrix we note that only the events on the diagonal are possible; indeed, once the a_1 has occurred, the event at the $(1, 2)$ location within the matrix can never take place because it is the result of the combination of two different values of the same variable. It is worth noting that where a subtraction had been performed in line with interval arithmetic, the result would have been much wider than that obtained using the relationship $[a_1 - (a_3 - c_1), a_3 - (a_1 - c_3)]$ reported above.

Based on this relation, Perrone and Noto La Diega developed their mean value analysis algorithm and the results they obtained proved very interesting above all in terms of the reduction of the support of fuzzy sets of the performance characteristics of the system. Their algorithm provides for a number of constraints associated with the average number of items circulating within the system, and these constraints are imposed at each iteration when the system's average lead time is computed.

The result we obtain using the same data as in the previous example is much more acceptable, in terms of significance, because the domain of the support is much smaller.

Although, in strictly 'engineering' terms, the results obtainable based on this algorithm are more satisfactory than those afforded by the other system, the structure of the algorithm itself is somewhat contradictory. Indeed, when we examine the values obtained for the total lead time [64.12 78.31 88.07] and throughput [0.0595 0.0638 0.0747] (Figure 6.11), only for the α-cuts with $\alpha = 0$ obtained after 5 iterations we note that the total lead time (W_{tot}) is the sum of the individual lead times (W_1, W_2), but that Little's law $L_{tot} = X \cdot W_{tot}$ is no longer applicable, in terms that the total lead time computed in line with Little's law $W_{tot} = L_{tot}/X$ is no longer the sum of the individual lead time. Another shortcoming of this algorithm emerges as soon as we increase the number of machines involved: in a system composed of no more than

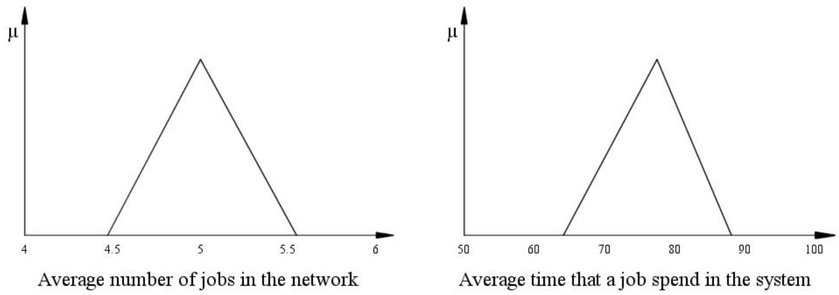


Figure 6.11. Fuzzy sets reflecting average number and lead time of tasks within the system.

three machines the number of combinations of possible events is seen to increase excessively, so that the algorithm can no longer be handled either in logical terms (selections needed to evaluate possible events) or in computational terms (all of the possible combinations of events to be considered).

Despite its limitations, the algorithm has the merit of reusing the notion of constraining the number of possible events and applying it to the case under review. This notion is not only useful in solving analytical models with fuzzy parameters; provided it is properly interpreted, it can also be extended to other problems involving uncertain variables.

9. The method proposed: single-class case

Compared to the techniques illustrated so far, the method proposed is based on the braking down of the fuzzy parameter MVA problem into sub-problems which can both be more easily solved and then reassembled to reconstruct the fuzzy sets related to performance characteristics.

The parameters needed to implement the MVA algorithm for a single-class system – both when crisp and fuzzy numbers are used – are average service time at each node Ts_i and the average number of visits ν_i. In this paper, we only use this method to analyze the case in which the average service times at each node are fuzzy variables; the case in which fuzzy numbers are used to determine the average number of visits will be dealt with in a subsequent paper.

The calculation method concerned is developed based on a very simple and intuitive assumption which can be mathematically demonstrated and is reported below.

In a closed queuing network system made up of M machines and N items and with fuzzy service time distribution parameters, let us consider the corresponding α-cuts with $\alpha \in [0, 1]$. If, from among all possible combinations of the processing time interval extremes, we select those which subsume the upper (lower) extreme of the processing time of the i-th node and the lower (upper) extremes of the remaining $M - 1$ nodes, by reiterating the MVA algorithm for this combination we will obtain the upper (lower) extremes of the α-cuts related to the average number of items and the average queuing time at the i-th node.

On closer analysis, this thesis is a highly intuitive one. Considering the combination formed by the upper extreme of the i-th node and the lower extreme of the remaining nodes, the i-th node can be assumed to act as a sort of bottleneck in respect of the processing times range because there is no combination of values at which, when the α-cut is assumed to remain constant, the queue present at the node considered will be greater than the one reflected in the above-said combination.

The same can be said of the combination which makes a pair with the one reported above, i.e. that in which the extreme of the i-th node processing times to be combined with the upper extremes of the remaining nodes is the lower one.

As mentioned above, the same MVA algorithm can be used to resolve the problem into sub-problems which can be more easily solved because they involve computational complexities only.

The first step is discretising the processing time fuzzy sets into the corresponding α-cuts:

$$^{\alpha}Ts_i = \{ts_i : Ts \geq \alpha\}$$

For purposes of greater clarity, each interval will be identified by the following notation:

$$^{\alpha}Ts_i = [\underline{^{\alpha}ts_i}, \overline{^{\alpha}ts_i}] \qquad \alpha \in [0, 1]$$

From among the α-cuts thus obtained, we only select those which in line with the thesis posited are related to the performance characteristics we wish to determine:

$$\underline{^{\alpha}P_1} = [\underline{^{\alpha}ts_1}, \overline{^{\alpha}ts_2}, \ldots, \overline{^{\alpha}ts_M}], \ldots, \underline{^{\alpha}P_M} = [\overline{^{\alpha}ts_1}, \overline{^{\alpha}ts_2}, \ldots, \underline{^{\alpha}ts_M}]$$

will be considered for all lower extremes, and

$$\overline{{}^{\alpha}P_1} = [\overline{{}^{\alpha}ts_1}, \underline{{}^{\alpha}ts_2}, \ldots, \underline{{}^{\alpha}ts_M}], \ldots, \overline{{}^{\alpha}P_M} = [\underline{{}^{\alpha}ts_1}, \underline{{}^{\alpha}ts_2}, \ldots, \overline{{}^{\alpha}ts_M}]$$

for all upper extremes.

For instance, by solving $\underline{{}^{\alpha}P_i}$ we determine the average number of tasks and the average lead time at the i-th node, while the upper extremes of these characteristics are determined by solving $\overline{{}^{\alpha}P_i}$.

It is clear that the performance characteristics of individual nodes can only be determined by solving $2 \cdot M \cdot h$, where M is the number of machines and h the number of α-cuts into which the fuzzy sets are resolved.

The remaining elements of the network are solved using interval arithmetic to determine the total lead time, the throughput and the average number of items circulating within the system:

$$\tilde{W}_{tot} = \sum_i \tilde{W}_i \qquad \tilde{L}_{tot} = \sum_i \tilde{L}_i \qquad \tilde{X} = \frac{L_{tot}}{\tilde{W}_{tot}}$$

A comparison with the algorithms illustrated above shows that:

- the support of the fuzzy sets that are obtained using this procedure are smaller than those obtained using the first algorithm illustrated above, but comparable to those obtained by recourse to Perrone and Noto La Diega's algorithm;
- the laws applicable to the evolution of the MVA algorithms are nowhere contradicted (no violation of Little's law);
- as compared to the great computational complexity associated with Perrone and Noto La Diega's algorithm, fewer and simpler operations are needed as the problem grows increasingly complex.

Despite its undoubted potential, this method too has a limitation, because the algorithm uses, not the fuzzy numbers themselves, but the crisp numbers generated by their braking out.

10. The method proposed: multi-class case

By means of simple integrations, the method just illustrated can be extended to a multi-class case, i.e. a system which subsumes R categories of items.

Let us assume that the system to be analyzed can be modelled as a closed queuing network system comprising M machines and R categories of items. Assuming that n_r items with $r = 1 \ldots R$ are comprised in the system for each category of items and that the machine processing time for each item category is a fuzzy number $\tilde{t}s_{mr}$ where the subscripts $m = 1 \ldots M$ and $r = 1 \ldots R$ respectively reflect the m-th machine and the r-th category. Under this hypothesis the performance characteristics of the whole network can be determined as reported below. The lower extreme of the fuzzy set which reflects the average number of tasks waiting to be performed or the average queuing time at the m-th node for each category of item can be computed as follows: first, the processing times are subdivided into α-cuts, then the multi-class MVA algorithm is solved based on the combination of the upper (lower) extremes of the processing times of all the items present at the m-th node and the lower (upper) extremes at the remaining $M - 1$ nodes. In this way, as already shown for the single-class case, the highest possible number of items obtainable based on all the possible combinations will be seen to be queuing up at the m-th machine.

11. The algorithm for the method proposed: single-class case

The algorithm is described below.

1 Data collection and construction of the fuzzy sets for the linguistic variables.

2 Computation of the average processing times of the individual items categories:

$$\tilde{T}s_i = \sum_{j=1}^{R} \tilde{T}s_{ij} \cdot r_j \quad i = 1, \ldots, M \quad \text{with } r_j = \frac{V_j}{\sum_{h=1}^{R} V_h} \quad j = 1, \ldots, R$$

where V_j represent the volume of production of the j-th class of part.

3 Computation of the average number of visits at each individual node:

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \cdots \\ \nu_M \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{M1} \\ p_{12} & p_{22} & \cdots & p_{M2} \\ \cdots & \cdots & \cdots & \cdots \\ p_{1M} & p_{2M} & \cdots & p_{MM} \end{bmatrix} \cdot \begin{bmatrix} \nu_1 \\ \nu_2 \\ \cdots \\ \nu_M \end{bmatrix}$$

the condition that $\nu_1 = 1$ is needed for the system of equations to have one and only one solution.

4 Discretization of the processing times fuzzy sets into the corresponding α-cuts:

$${}^{\alpha}\tilde{T}s_i = \left\{ ts_i : \tilde{T}s_i \geq \alpha \right\}$$

5 Selection of the processing time combinations needed to calculate the lower extremes of the numbers of queuing items and average queuing times at each node:

$$\underline{{}^0P_1} = [\underline{{}^0ts_1}, \overline{{}^0ts_2}, \dots, \overline{{}^0ts_M}], \dots, \underline{{}^0P_M} = [\overline{{}^0ts_1}, \overline{{}^0ts_2}, \dots, \underline{{}^0ts_M}]$$

....

$$\underline{{}^1P_1} = [\underline{{}^1ts_1}, \overline{{}^1ts_2}, \dots, \overline{{}^1ts_M}], \dots, \underline{{}^1P_M} = [\overline{{}^1ts_1}, \overline{{}^1ts_2}, \dots, \underline{{}^1ts_M}]$$

6 Selection of the processing time combinations needed to calculate the upper extremes of the numbers of queuing items and average queuing times at each node:

$$\overline{{}^0P_1} = [\overline{{}^0ts_1}, \underline{{}^0ts_2}, \dots, \underline{{}^0ts_M}], \dots, \overline{{}^0P_M} = [\underline{{}^0ts_1}, \underline{{}^0ts_2}, \dots, \overline{{}^0ts_M}]$$

....

$$\overline{{}^1P_1} = [\overline{{}^1ts_1}, \underline{{}^1ts_2}, \dots, \underline{{}^1ts_M}], \dots, \overline{{}^1P_M} = [\underline{{}^1ts_1}, \underline{{}^1ts_2}, \dots, \overline{{}^1ts_M}]$$

7 Computation of the average queuing time and the average number of parts queuing up at the i-th node. Iteration of the MVA algorithm for the processing time combinations: $\underline{{}^\alpha P_i}$, using $i = 1 \dots M$ and $\alpha \in [0, 1]$

$$\begin{array}{c}
{}^\alpha W_1^n = \overline{{}^\alpha ts_1} \cdot \left(1 +^\alpha L_1^{n-1}\right) \\
\dots \\
\underline{{}^\alpha W_i^n} = \underline{{}^\alpha ts_i} \cdot \left(1 + \underline{{}^\alpha L_i^{n-1}}\right) \\
\dots \\
{}^\alpha W_M^n = \overline{{}^\alpha ts_M} \cdot \left(1 +^\alpha L_M^{n-1}\right) \\
X^n = \dfrac{n}{\sum\limits_{j=1 \neq i}^{M} {}^\alpha W_j^n \cdot \nu_j + \underline{{}^\alpha W_i^n} \cdot \nu_i} \\
{}^\alpha L_1^n = X^n \cdot \nu_1 \cdot^\alpha W_1^n \\
\dots \\
\underline{{}^\alpha L_i^n} = X^n \cdot \nu_i \cdot \underline{{}^\alpha W_i^n} \\
\dots \\
{}^\alpha L_M^n = X^n \cdot \nu_M \cdot^\alpha W_M^n
\end{array} \tag{6.1}$$

and for the $\overline{{}^\alpha P_i}$ combinations with $i = 1, \dots, M$ and $\alpha \in [0, 1]$

$$
\begin{aligned}
{}^{\alpha}W_1^n &= \underline{{}^{\alpha}ts_1} \cdot \left(1 +^{\alpha} L_1^{n-1}\right) \\
&\dots \\
\overline{{}^{\alpha}W_i^n} &= \overline{{}^{\alpha}ts_i} \cdot \left(1 + \overline{{}^{\alpha}L_i^{n-1}}\right) \\
&\dots \\
{}^{\alpha}W_M^n &= \underline{{}^{\alpha}ts_M} \cdot \left(1 +^{\alpha} L_M^{n-1}\right) \\
X^n &= \frac{n}{\sum_{j=1\neq i}^{M} {}^{\alpha}W_j^n \cdot \nu_j + \overline{{}^{\alpha}W_i^n} \cdot \nu_i} \\
{}^{\alpha}L_1^n &= X^n \cdot \nu_1 \cdot^{\alpha} W_1^n \\
&\dots \\
\overline{{}^{\alpha}L_i^n} &= X^n \cdot \nu_i \cdot \overline{{}^{\alpha}W_i^n} \\
&\dots \\
{}^{\alpha}L_M^n &= X^n \cdot \nu_M \cdot^{\alpha} W_M^n
\end{aligned}
\tag{6.2}
$$

8 Reconstruction of the fuzzy sets which represent the average number of items and the average queuing time at each node:
Reconstruction of the intervals for each node:

$$
\begin{aligned}
{}^{\alpha}W_1 &= [\underline{{}^{\alpha}W_1^N}, \overline{{}^{\alpha}W_1^N}] \\
&\dots \\
{}^{\alpha}W_M &= [\underline{{}^{\alpha}W_M^N}, \overline{{}^{\alpha}W_M^N}] \\
{}^{\alpha}L_1 &= [\underline{{}^{\alpha}L_1^N}, \overline{{}^{\alpha}L_1^N}] \\
&\dots \\
{}^{\alpha}L_M &= [\underline{{}^{\alpha}L_M^N}, \overline{{}^{\alpha}L_M^N}]
\end{aligned}
\tag{6.3}
$$

Construction of the special set:

$$
\begin{aligned}
{}^{\alpha}W_1 &= \alpha \cdot^{\alpha} W_1 \\
&\dots \\
{}^{\alpha}W_M &= \alpha \cdot^{\alpha} W_M \\
{}^{\alpha}L_1 &= \alpha \cdot^{\alpha} L_1 \\
&\dots \\
{}^{\alpha}L_M &= \alpha \cdot^{\alpha} L_M \\
{}^{\alpha}L_1 &= \alpha \cdot^{\alpha} L_1 \\
&\dots \\
{}^{\alpha}L_M &= \alpha \cdot^{\alpha} L_M
\end{aligned}
\tag{6.4}
$$

Reconstruction of the fuzzy sets:

$$
\tilde{W}_1 = \bigcup_{\alpha\in[0,1]} {}_{\alpha}W_1 \qquad \dots \qquad \tilde{W}_M = \bigcup_{\alpha\in[0,1]} {}_{\alpha}W_M
$$

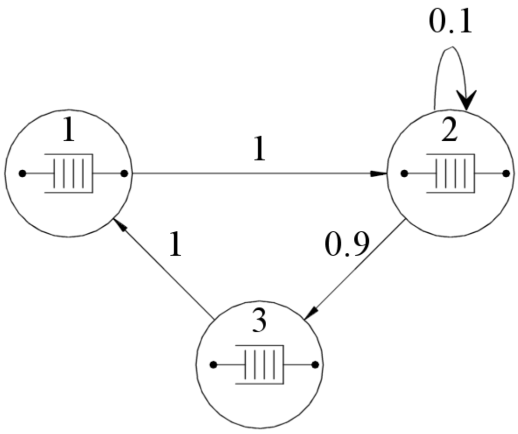


Figure 6.12. Configuration of a queuing system with one reiteration.

$$\tilde{L}_1 = \bigcup_{\alpha \in [0,1]} {}_{\alpha}L_1 \qquad \ldots \qquad \tilde{L}_M = \bigcup_{\alpha \in [0,1]} {}_{\alpha}L_M$$

9 Calculation of the average lead time, the fictitious average number of tasks within the system and the throughput value (interval arithmetic):

$$\tilde{W}_{tot} = \sum_i \tilde{W}_i \qquad \tilde{L}_{tot} = \sum_i \tilde{L}_i \qquad \tilde{X} = \frac{L_{tot}}{\tilde{W}_{tot}}$$

12. A sample application

To illustrate their method, the authors have solved a system of the kind reported in Figure 6.12.

Ten items are circulating within the system and the relevant machine processing time is:

$$\tilde{T}s_1 = (10, 11, 12) \qquad \tilde{T}s_2 = (8, 9, 10) \qquad \tilde{T}s_3 = (6, 7, 8)$$

The routing probability matrix is:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 1 & 0 & 0 \end{bmatrix}$$

The system to calculate the average number of visits is:

$$\begin{cases} \nu_1 = \nu_1 \cdot p_{11} + \nu_2 \cdot p_{21} + \nu_3 \cdot p_{31} = 1 \\ \nu_2 = \nu_1 \cdot p_{12} + \nu_2 \cdot p_{22} + \nu_3 \cdot p_{32} \\ \nu_3 = \nu_1 \cdot p_{13} + \nu_2 \cdot p_{23} + \nu_3 \cdot p_{33} \end{cases}$$

and this, once solved, leads to the following values:

$$\nu_1 = \nu_3 = 1 \qquad \nu_2 = 1.11$$

The fuzzy sets (only two α-cuts in the example) are solved and the processing time combinations needed to calculate the lower and upper extremes of the interval:

$$\underline{{}^0P_1} = [10, 10, 8], \underline{{}^0P_2} = [12, 8, 8], \underline{{}^0P_3} = [12, 10, 6],$$

$$\underline{{}^1P_1} = \underline{{}^1P_2} = \underline{{}^1P_3} = [11, 9, 7]$$

$$\overline{{}^0P_1} = [12, 8, 6], \overline{{}^0P_2} = [10, 10, 6], \overline{{}^0P_1} = [10, 8, 8]$$

$$\overline{{}^1P_1} = \overline{{}^1P_2} = \overline{{}^1P_3} = [11, 9, 7]$$

are selected.

It is worth mentioning that the combinations needed to calculate the lower and upper extremes with $\alpha = 1$ coincide with each other because the sets for the system's performance characteristics are fuzzy sets.

To determine the average number of tasks and average queuing time we reiterate the MVA algorithm:

$$\begin{aligned}
\tilde{L}_1 &= (3.36, 5.1, 6.74)\\
\tilde{W}_1 &= (40.47, 60.1, 82.66)\\
\tilde{L}_2 &= (2.2, 3.57, 5.36)\\
\tilde{W}_2 &= (24.54, 38.15, 57.23)\\
\tilde{L}_3 &= (0.85, 1.36, 2.27)\\
\tilde{W}_3 &= (10.96, 16.1, 28.84)
\end{aligned}$$

The lead time and the throughput value can be determined by adding up the results thus obtained:

$$\tilde{W}_{tot} = (75.97, 114.35, 168.73)$$

$$\tilde{X} = (0.059, 0.087, 0.132)$$

13. Conclusions

Compared to the case in which iterations are made in line with the extension principle or interval arithmetic procedures, the method for solving the fuzzy MVA algorithm proposed in this chapter does not unduly widen the range of the fuzzy system performance characteristics sets.

The computational complexity of this method, i.e. the number of operations that will have to be performed, is comparable to that of interval arithmetic. Moreover, the structure of the algorithm is less complex because the relevant operations can be executed using crisp numbers.

To establish the true potential of this method, fuzzy routing coefficients could be introduced without performing any operations on the (α-cuts) intervals. In this case the values of the fuzzy performance characteristics would again be determined by solving the crisp problem. Attempts in this direction have already been made with respect to flow-shop type processes and their earliest results are already available.

References

Bolch, G., Greiner,S., de Meer, H., and Trivedi S. K. (1998). *Queueing Networks and Markov Chain*. John Wiley & Sons, New York.

Brandimarte, P., Villa, A. (1995). *Advanced Models for Manufacturing Systems Managment*. CRC Press, New York.

Gross, D., and Harris, C.M. (1985). *Fundamentals of Queueing Theory*. John Wiley & Sons, New York.

Irion, A. (1998). Fuzzy rules and fuzzy functions: A combination of logic and arithmetic operation for fuzzy numbers. *Fuzzy Sets and System*, 99: 49-56.

Kaufmann, A. and Gupta, M. M. (1985). *Introduction to Fuzzy Arithmetic*. Van Nostrand Reinhold, New York.

Klir, G.J., and Yuan, Y. B. (1995). *Fuzzy sets and fuzzy logic*. Prentice-Hall, New Jersey.

Klir, G.J. (1997). Fuzzy arithmetic with requisite constraints. *Fuzzy Sets and System*, 91: 165-175.

Jo, J.B., Tsujimura, Y., Gen, M., Yamazaki, G., Lee, J.U. and Noto La Diega, S. (1997). Performance of Multiclass BCMP Model for Computer System based on Fuzzy Set Theory. *Computers Industrial Engineering*, 33: 557-560.

Li, R.J., and Lee, E.S. (1989). Analysis of fuzzy queues. *Computers Mathematical Applications*, 17: 1143-1147.

Negi, D.S. and Lee, E.S. (1992). Analysis and simulation of fuzzy queues. *Fuzzy Sets and System*, 46: 321-330.

Perrone, G. and Noto La Diega, S. (1998). Fuzzy methods for analysing fuzzy production environment. *Robotics and Computer-Integrated Manufacturing*, 9: 465-474.

Reiser, M. and Levemberg, S.S. (1980). Mean value analysis of closed multiclass chain networks. *Journal Associated Computer Machining*, 27: 313-322.

Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and System*: 1, 3-28.

Zimmermann. H.J. (1991). *Fuzzy Set Theory and Its Application.* Kluwer-Nijhoff, Boston.