Subcontracting as a capacity management tool in multi-project repair shops

Citation for published version (APA):
Keizers, J. M. (2000). *Subcontracting as a capacity management tool in multi-project repair shops*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Technische Universiteit Eindhoven. https://doi.org/10.6100/IR537276

DOI:
10.6100/IR537276

Document status and date:
Published: 01/01/2000

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 19. May. 2025

Subcontracting as a Capacity Management Tool in Multi-Project Repair Shops

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Keizers, Joris M.

Subcontracting as a Capacity Management Tool in Multi-Project Repair Shops / by Joris Maria Keizers. – Eindhoven : Technische Universiteit Eindhoven, 2000. – Proefschrift.
ISBN 90-386-0743-1
NUGI 684
Keywords: Subcontracting; Production control; Maintenance management

Druk: PlantijnCasparie Heerhugowaard bv

Subcontracting as a Capacity Management Tool in Multi-Project Repair Shops

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr. M. Rem, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 3 oktober 2000 om 16.00 uur

door

Joris Maria Keizers

geboren te Nijmegen

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. J.W.M. Bertrand
en
prof.dr. J. Wessels

Acknowledgements

This thesis is the result of the research that I conducted between May 1996 and July 2000 at the Technische Universiteit Eindhoven. Of course, this research could not have been completed without the support of others. First of all, I would like to thank my supervisors. The interesting and fruitful discussions with Will Bertrand have been a great inspiration to me. Jaap Wessels joint the team almost two years ago and played an important role in the validation phase of this research. Henk Zijm and Aart van Harten are acknowledged for their useful comments on the draft version of this thesis. Furthermore, I would also like to extend my gratitude to Ivo Adan and Jan van der Wal, who were of great importance for realizing Chapter 5.

I am indebted to the Royal Netherlands Navy. They have not only been the financier of this research, but, more importantly, they provided me with the access to all kinds of processes and information in their maintenance organization. There are many people who where willing to answer my many questions and critically comment on ideas. In particular, I would like to thank Peter Calbo, who has been of great help.

My colleagues at the university made my stay there a very pleasant one. Especially Kim van Oorschot (keep going during your last months!), Ebbe Negenman (didn't we have a great course on negotiating?), Willem de Paepe (one day I'll beat you!), Wenny Raaymakers (I had a great time in Budapest) and Jan-Peter Vos (no more 'trainpooling') are acknowledged.

I express my gratitude to my friends and family, for their interest and the necessary distraction from work. In particular, I would like to thank my parents for their support and confidence in me throughout the years. And finally Nienke, you are the living proof that my stay in Eindhoven has resulted in much more than just this thesis!

Den Haag, July 2000 Joris Keizers

Contents

1

Introduction

1.1 Maintenance

Maintenance has existed since time immemorial. It is not unlikely that the first cavemen were faced with the problem of keeping their spearheads sharp enough for the hunt. In later ages, the coaches had to be kept operational for transport and travelling. It is not surprising then, that most maintenance activities were only carried out on a trial-and-error base, resulting in corrective maintenance. It is not until the second part of the 20th century that maintenance received more attention, which is a consequence of World War II and the aerospace developments (Corder 1992). Ever since, industrial companies have become more and more aware that maintenance could reduce a part of the uncertainty of the production processes. Therefore, their belief that “maintenance only costs” has turned into the belief that “maintenance also pays”. Furthermore, in everyday life people are faced - whether aware or not - with maintenance: we give preventive maintenance to our cars and corrective maintenance to a flat tyre. Companies like Bic and Gillette make fortunes due to the fact that people apply repair-by-replacement activities for pencils and razor blades (Patton 1980).

For scientific researchers, maintenance is an area of growing interest, demonstrated by the increasing number of scientific publications about maintenance and maintenance management. Several reasons can be attributed for this phenomenon. Firstly, organizations realize that maintenance may contribute to the long-term objective of producing a high quantity of their products whilst maintaining a high quality. Secondly, the widespread mechanization and automation has increased the capital tied up in production equipment which makes the output of the organizations increasingly dependent on the reliability of the technical systems. Thirdly, the increasing complexity of the technical systems has resulted in a need for more sophisticated maintenance models and control policies. Another reason is that managers have become aware that the productivity may be increased by maintenance control (Paz & Leigh 1993, Dekker 1996 and Cho & Parlar 1991). And finally, many very interesting problems that ask for a scientific approach arise from the field of maintenance management.

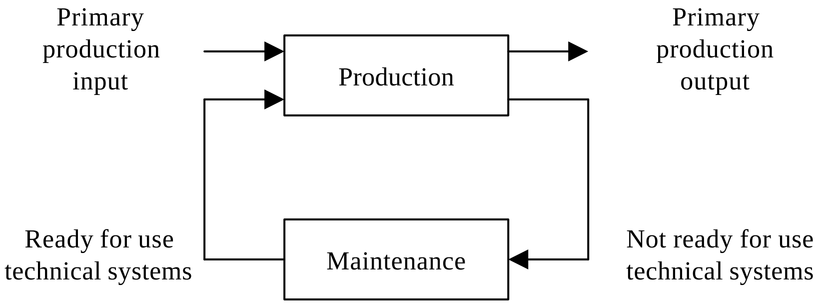


Figure 1.1. The relationship between production and maintenance (Gits 1994).

This thesis is about the planning and control of maintenance departments that are concerned with the objective to provide an operational availability of the technical systems that is in accordance with the production schemes of one or more production departments. According to Gits (1994), in a production department, the primary production input (material, energy and manpower) is transformed into the primary production output (see Figure 1.1). To facilitate this transformation, the maintenance department has to ensure that the production means are kept in a condition in which they are required to fulfil their requested functions. Therefore Gits (1992) defines maintenance as "*the total of activities required to retain the systems in, or restore them to the state necessary for fulfillment of the production function*". In this definition, “retaining in” corresponds with preventive maintenance and “restoring to” with corrective maintenance. Finally, we assume that this relationship between the maintenance and production department has been accomplished by long term contracts, which legally make the maintenance departments responsible for the prompt delivery of the maintenance activities. This timely delivery can be achieved by using either the internal or external resources (outsourcing, subcontracting).

In general, three types of resources are needed for maintenance activities: manpower, tools and materials. Particularly manpower is an important resource because it is needed during the entire processing time of a maintenance job. This simple observation has important consequences for the planning and control: the availability (or unavailability) of the maintenance operators determines the actual progress of the maintenance projects on hand. Moreover, our conjecture is that the planning and control of the maintenance resources and resource requirements is the key control decision in achieving a satisfactory performance. Tools generally do not appear to be bottleneck resources. For work on the role of materials in the maintenance process, especially in the type of maintenance organizations that are under scrutiny in this research, we refer to the work done by Rustenburg *et al.* (2000a, 2000b). Later in this chapter, we return to the topic of handling materials in maintenance.

In this thesis we will look at a special class of maintenance departments, namely the ones consisting of parallel *multi-project repair shops.* After having described their demand and capacity characteristics in the next section, a definition of these shops can be given.

1.2 Production control in multi-project repair shops

In the remainder of this thesis, we are solely interested in the maintenance function in Figure 1.1, assuming that appropriate agreements are made between the production and the maintenance department, resulting in a certain demand that cannot be influenced by the maintenance department. Therefore, in talking about *production control*, we consider the control of the ‘production’ process in the maintenance department: i.e. transforming the non-ready-for-use systems into ready-for-use systems. According to Bertrand *et al*. (1990), production control can be defined as *the co-ordination of supply and production activities in manufacturing systems to achieve a specific delivery flexibility and delivery reliability at minimum cost*. This clearly is also applicable to repair shops in case we regard a repair shop as a manufacturing system that is faced with (maintenance) demand and has to produce non-failed technical systems out of the failed technical systems. However, in order to judge or design a production control structure for the type of maintenance organizations under scrutiny, it is of utmost importance to investigate demand characteristics and capacity resource characteristics.

1.2.1 Demand characteristics

Different demand classes. Three different classes of demand can be distinguished for the maintenance organizations in this research. All these types of demand compete for the same capacity resources. At first, part of the capacity is consumed by large preventive overhauls of entire (or part of) production facilities. These are called *preventive maintenance projects.* During these planned projects, non-urgent failed components are repaired or replaced and other components or entire sub-systems are inspected (and repaired or replaced, if necessary) with the aim to reduce the probability of failures in the next operational period. These overhauls may be carried out either during night shifts or weekends, or during maintenance shut-downs. In both cases it is important to complete the overhaul within the agreed time slot. Delivering after the due date directly influences the production schedules of the production department. Secondly, unplanned breakdowns of operational technical systems create *corrective maintenance projects* for the maintenance department. Because the objective of the maintenance organization is to maximize the operational availability of the technical systems in the production department, capacity should be assigned immediately to such a failed system. Consequently they disturb the progress of the preventive projects. And finally, the *repairable maintenance projects* originate from the possibility to replace a failed technical (sub)system by a non-failed one which is kept in stock, instead of repairing it. After this exchange of components, the failed component asks for maintenance capacity in order to be available for future component exchanges.

No due date flexibility. The relationship between the production and maintenance department (see Figure 1.1) makes that all projects (with their requested due dates) need to be accepted. After all, if projects are postponed, the operational availability of the production department is reduced. Consequently, a powerful decision function like the assignment of due dates based on available capacity cannot be applied.

Uncertainty. Two types of uncertainty can be distinguished. At first, the processing times of the projects are uncertain. Upon arrival of a project, the exact condition of the technical systems (and thus the needed repair actions) cannot be established. Therefore, upon arrival the actual processing times can only be estimated, for example using a database with historical data. Secondly, the random breakdowns of technical systems make that at unknown points in time emergency projects arrive, which immediately require capacity.

Multi-project structure. Projects of all three demand classes dynamically arrive over time, all competing for capacity of various types of capacity resources. This means that choices have to be made about prioritizing and sequencing the projects of the different classes.

1.2.2 Capacity resource characteristics

Parallel multi-project repair shops. The maintenance organization consists of a certain number of parallel operating repair shops, each consisting of a (small) number of operators. A repair shop is responsible for a given set of technical systems. Because of the highly skilled human resources, the volume flexibility of the workforce is (especially on the short term) rather limited and only offered by working late. The repair shops have the predicate 'multi-project', because they are faced with the three classes of maintenance, all competing for their resources.

Subcontracting. Volume flexibility of the work force can be created by subcontracting. Because the lead times for subcontracting are generally long, this decision has to be made sufficiently early in time. Within this research we assume that the subcontracting lead time is fixed. Using this fixed lead time, the subcontracting capacity is assumed to be infinite: subcontractors deliver in time with a hundred percent reliability.

Summarizing, we define a *multi-project repair shop* as a cluster of a fixed number of parallel servers (i.e. maintenance operators), which are confronted with projects of three different classes: preventive, emergency and repairable projects.

1.3 Problem formulation

Feiler (1972) already stated that "*it has almost become axiomatic that planned project schedules will be optimistic. The only question is, By how much*?". It is not daring to say that this situation has not changed during the almost 30 years that have passed. Regardless the industry in which they are operating, organizations still have a lot of difficulties to complete their projects in time. Recently, Icmeli-Tukel & Rom (1998) conducted a survey about the characteristics of projects in different industries varying from manufacturing to software organizations. Using a questionnaire which has been returned by 91 practitioners (out of a total of 320), they conclude among others that 54% of the respondents face project delays often or all the time. Clearly, it still is a common thing in project management that the completion dates of the projects are uncontrolled. Of course this is a very undesirable situation. The delivery performance towards the customers is a crucial performance indicator in most of today's markets. Many different markets can be typified as buyer markets. Customers are becoming more and more demanding and require accurate deliveries from their suppliers. If an organization cannot meet these high demands, its continuity is in severe danger.

During the past decades lists have been developed in the literature with critical factors affecting the successful completion of projects, by e.g. Schonberger (1981), Baker *et al.* (1983), Morris & Hough (1987), Icmeli-Tukel & Rom (1998). Commonly reported factors can be divided into organizational oriented and production control oriented factors. The main organizational oriented factors are: the management support, the project manager's performance on the job, client consultation and predetermination of the success criteria. Regarding the production control oriented factors, the availability of resources and recognition of uncertainty in processing times are considered to be crucial. However, despite the huge body of knowledge available in the scientific literature about project planning and control, it still seems hard to meet the agreed due dates in real-life.

This thesis is about maintenance organizations satisfying the characteristics described in Section 1.2. These organizations definitely are project-driven. Moreover, they may be considered as a subset of all organizations that are faced with project management. To get more insight in the problems with which these organizations are faced, we base this research on a case study at SEWACO. This maintenance organization of the Royal Netherlands Navy (RNLN) is responsible for the maintenance, repair and modification of the sensor, weapon and command systems of the Navy and is a prototype of the class of maintenance organizations under scrutiny. They perform projects varying from small repairs to the executions of large overhauls. Moreover, multiple projects are in process at the same time. Given that projects may be divided into many sub-projects for different repair shops, the capacity complexity enormously increases the difficulty of planning and controlling the projects.

To check to what extent the SEWACO faces an uncontrolled delivery performance, we conducted a survey about the due date reliability of the 4773 projects which had their due dates in 1998. The results of this delivery performance audit are largely consistent with the earlier mentioned performances in other industries: less than 40% of the projects were indeed delivered to the fleet before or at the due date. It is noteworthy that that during the last years several reorganizations and projects have (unsuccessfully) been initiated with the aim to achieve a higher and satisfactory delivery performance.

From a control point of view, even more interesting than the answer to Feiler's question is the answer to Schonberger's (1981) question: *Why are projects late?* Our conjecture is that due to the capacity complexity and the uncertainty, the planning and control of the maintenance resources and resource requirements are the most important key control decisions in achieving a satisfactory delivery performance. Therefore, we take SEWACO as a typical maintenance organization and use this case to answer the following research question:

> Q_1: *Can the planning and control of the maintenance resources and resource requirements be improved given the demand characteristics of*

> *the maintenance organization under scrutiny? If so, what information about demand characteristics is crucial for performance improvement?*

Once this question has been answered, the following research question logically comes up:

> Q_2: *Is it possible - based on deduction and validation - to design a framework for control which does sufficiently pay attention to the identified critical demand characteristics such that the delivery performance will be under control? If so, what should this framework look like?*

According to Bertrand *et al.* (1990), a proper framework for control should identify the main decision functions, their goals, their (inter)relationships and the organizational positions to which these decision functions should be assigned. A decision function expresses the decision as a function of the conditions (i.e. decisions from the next higher organizational echelon) and the current state of the system. Usually, a decision function is represented by a small-scale mathematical model. These models may not contain all (detailed) process information, although all model characteristics can be found in real-life. Nowadays, it is widely agreed that these models do provide necessary knowledge. It is also known that just applying the results of the theoretical studies does not solve the real-life control problems. This is simply due to the fact that real-life control problems are much richer than theoretical models, both in terms of the complexity of the problems to be solved and the intelligence and control means available for the planners for solving the problems. As a consequence, it is impossible to mathematically model the processes at a detailed level. Consequently, transferring the knowledge that has been obtained from the small-scale models is crucial for a controlled performance. Therefore, a relevant third research question can be formulated:

> Q_3: *How can knowledge of theoretical models be applied to improve the performance of real-life processes, if it is not possible to mathematically model the processes at a detailed level?*

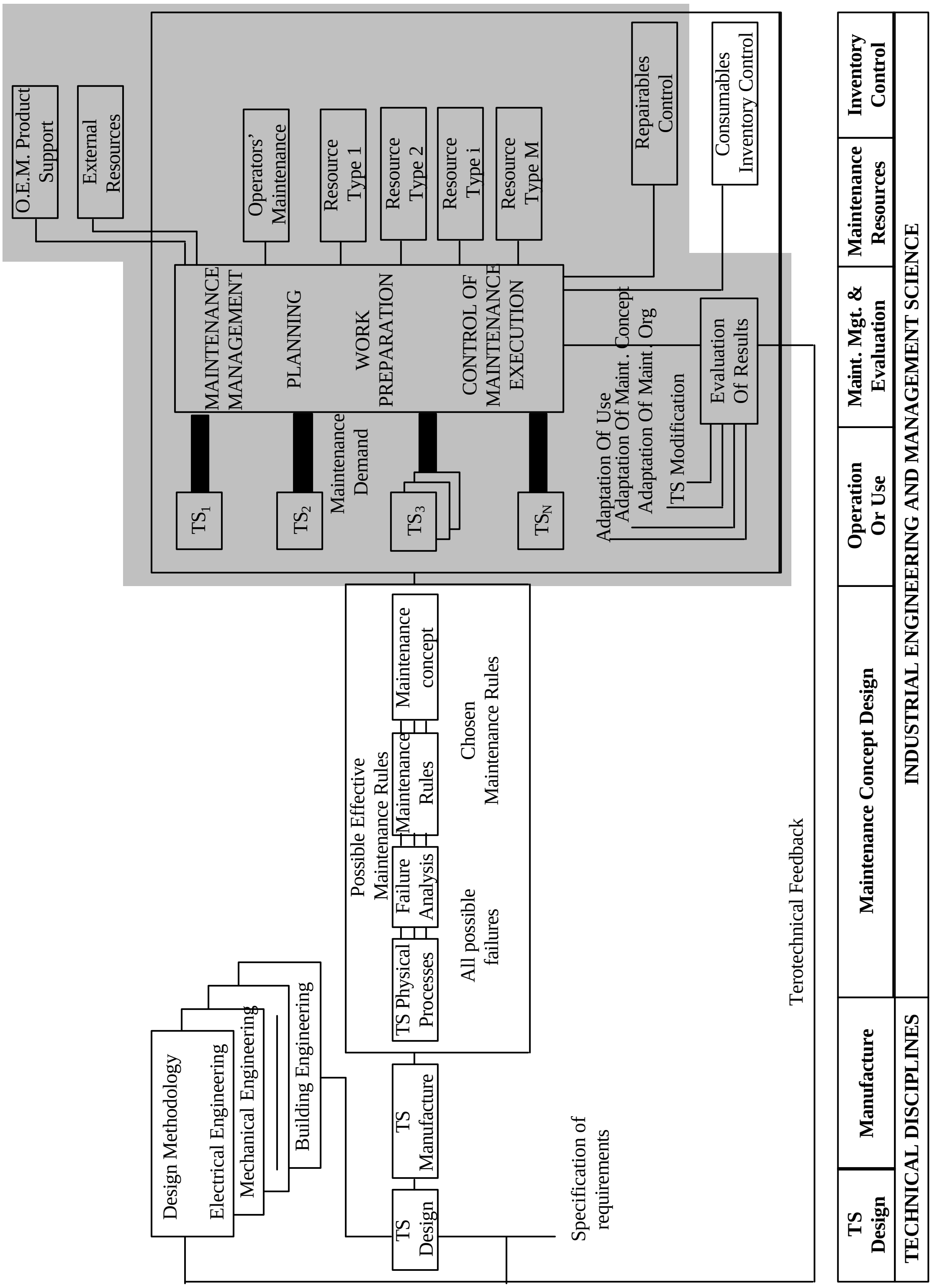


Figure 1.2. The boundaries of the research (shaded area) depicted within the EUT maintenance model.

These three research questions are to be answered in the remainder of this thesis. As a final note of this section we discuss the boundaries of this study. A lot of sub-functions and sub-processes influence the demand for capacity resources and their performance. Geraerds (1992) introduced the EUT maintenance model which specifies these sub-functions and sub-processes, connected by their interrelations (see Figure 1.2). It shows maintenance as it appears in an organization, seen from the point of view of maintenance management. Therefore, the model is suitable for defining our area of interest. The development of maintenance concepts has received a lot of scientific attention, by e.g. Cho & Parlar (1991), Gits (1992), Dekker (1995) and Silver & Fiechter (1995). Scientific knowledge about maintenance resource management is less amply available. Combining this statement with our conjecture that the planning and control of the maintenance resources and resource requirements are the most important key control decisions, we restrict ourselves in this thesis to the shaded area in Figure 1.2. The answers to the three research questions will be given in terms of variables that can be controlled by decisions in this area.

1.4 Literature review

We already noticed that the interest of scientists to maintenance has increased during the past three decades. Most of the topics that are distinguished in the EUT maintenance model have received considerable attention and have produced (besides a lot of new knowledge) new problems in the particular areas. However, hardly any literature is available about the control of multi-project repair shops. This control problem takes up a particular position in literature, as it stands in two worlds. Firstly, it is highly connected to the technical world of maintenance. In Section 1.2, we systematically described the characteristics of maintenance demand that may briefly be summarized by the predicates ‘dynamic’ and ‘highly stochastic’. These characteristics - together with the possibilities and restrictions of the capacity resources - need careful attention in the phase of designing the production control structure. Secondly, we can lean on the available knowledge about structuring complex control problems and the design of production control structures. In this section we merely look at the literature about similar and slightly different control problems in maintenance. The body of knowledge about the (design of) production control systems receives considerable attention at a later stage of this research when the second research question is to be answered. Now consecutively pay attention to the literature about project planning in maintenance, repair shop control and related research at the Royal Netherlands Navy.

Uncertainty is unmistakable connected to maintenance management. Another important characteristic is the multi-project structure: multiple projects compete for the capacity of various repair shops with limited capacity. Some authors focus more on the aspect of planning under fixed capacity, rather than on the uncertainty. As a result, scheduling tools have been developed for project planning in maintenance environments, assuming no uncertainty. Chen (1994), for example, presents a scheduling tool for multiple maintenance projects at a copper mine based on deterministic 0-1 programming. Kralj & Petrovic (1995) describe a branch and bound algorithm that determines the best maintenance schedule subject to some objective function. Taylor (1996) uses a linear programming model to assign maintenance tasks with deterministic processing times to time slots. And finally, De Boer (1998) presents a decision support system for ship maintenance planning. Although all of the above mentioned studies are based on methods for deterministic problems, De Boer (1998) emphasizes the need for putting some time allowances around the estimated processing times to cope with uncertainty.

Part of the maintenance activities may be carried out on a repair-by-replacement base: technical systems are disassembled into smaller pieces for various repair shops. Failed components - so-called repairables - are taken out and replaced by non-failed ones from the depot. The component that has been taken out is sent to the maintenance department. After having received the necessary maintenance (inspection and/or repair) it will be sent to the depot for future use. This operating procedure is particularly used in the aircraft industry, see e.g. Schneeweiß & Schröder (1992). Design decisions in such maintenance environments are, for example, the determination of the capacity levels and the total number of repairables of each type in the system. Control decisions mainly focus on keeping the inventory of non-failed parts in balance with the demand. Or, in other words: given the failed repairables, the current inventory levels and the demand rates for the different types of repairables, what is an optimal sequence to repair the failed repairables? Such problems are studied by e.g. Hausman & Scudder (1982), Albright (1988), Lee (1989) and Schneeweiss & Schröder (1992).

In relation to this research, several other researches have been conducted at the Royal Netherlands Navy, in the past few years. At the Royal Netherlands Naval Dockyard, a production re-engineering project has been carried out with the goal to improve the logistic performance of the dockyard (Zijm, 1996). As part of this project, De Waard (1999) designs an organizational structure for large scale maintenance organizations. He comes up with seven rules for which holds that the application should result into an improved controllability. Some of these rules are (i) integration of preparation and execution of the activities, (ii) using a hierarchical control structure with respect to capacity planning, (iii) constructing multi-disciplinary teams at the lowest organizational level and (iv) the necessity to use information technology to support the various decisions. As another part of the re-engineering project, De Boer (1998) has designed a decision support system. This hierarchical system contains decision levels for rough-cut capacity planning and resource-constrained project scheduling, and can be used for - among other things - lead time setting and crisis management. Finally, Rustenburg *et al.* (2000a, 2000b) study the initial purchase and resupply of repairables, given annual budget restrictions. Using a system-approach, they show that the capital tied up in the current inventory, can drastically be reduced. One of the conditions for this inventory reduction is a controlled throughput time of repairables in the repair shops, which is one of the objectives of research in this research.

To the best of our knowledge, no literature is available about the control problem in repair shops where the three different classes of maintenance projects as distinguished in Section 1.2, compete for the same capacity resources.

1.5 Research methodology

Research has to be carried out according to rigorous research designs in order to produce scientific knowledge. According to Reynolds (1971), scientific knowledge should aim at (i) organizing and categorizing ‘things’, (ii) predicting future events, (iii) explaining past events, (iv) understanding what causes events and (v) having potential for the control of events. Applied to the field of Operations Management, Simon (1967) states that it is our mission to conduct research that contributes to our field as a scientific discipline and to apply this knowledge in real-life processes.

Moreover, Simon notices that nowadays the interests of managers have shifted towards the understanding what factors affect successfully managed production processes. At the same time, operations management researchers have had a turn to a new approach: *theory-driven empirical research*. Melnyk & Handfield (1998) describe this as an "*approach to research that is driven to provide better insight and understanding into issues by using empirical data to build and develop better theories*". Starting with theory, the researcher uses the empirical data to further build, validate and modify the theory. According to Hunt (1991), these starting theories may be based on extensions of past research, they can be derived from logic or come from serendipity. This is exactly the approach we take in this thesis. Or, like Melnyk & Handfield (1998) state, quoting Barber from Michigan State University: "*Theory without data is a bother*; *data without theory is a nightmare*".

In this research we use a case study in a maintenance organization of the Royal Netherlands Navy to gain insight in the problems which are inherent to this type of organizations. Yin (1981) states that the distinguishing characteristic of a case study is that it "*attempts to examine: (a) a contemporary phenomenon in its real-life context, especially when (b) the boundaries between phenomenon and context is not clearly evident*." We are aware of the grievance of some researchers against case study (based) research. Miles (1979) states that, "*without renewed efforts at methodological inquiry, qualitative research on organizations cannot be expected to transcend story-telling*". We would rather weaken this statement by postulating that *without a theoretical framework*, a researcher is indeed in severe danger of providing description without wider meaning. To overcome these problems, we split up the case study in two parts. In the first part (Chapter 2), we objectively observe the processes and measure the performance without intervention with these processes. In the second part of the case study (Chapter 3), we use the empirical observations together with scientific literature on order planning and due date performance, to postulate our hypothesis about the causes of poor performance. At that stage, we have come to the approach of theory-based empirical research.

One final note on case study based research should be made. Hartley (1994) emphasizes that in a single case study the disentangling of what is unique to this organization from what is common to other organizations can be difficult. Figure 1.3

depicts the place of the SEWACO case study in this research. For all organizations satisfying the characteristics in Section 1.2, it holds that if no suitable control structure is used, the delivery performance will be uncontrolled and unsatisfactory. Moreover, all organizations in this class should have the same necessary (but at this stage not yet known) elements in the control structure in order to ensure a controlled and satisfactory delivery performance. However, in this research we observe only one instance of this class of problem situations. Their current control structure is only one way of dealing with the process characteristics, based on their perception and knowledge of the processes. As a consequence, it is very likely that part of their problems do not originate solely from the characteristics, but are due to their way of (perhaps inadequately) planning and controlling the processes. We have to be aware of this entanglement of problems throughout this research.

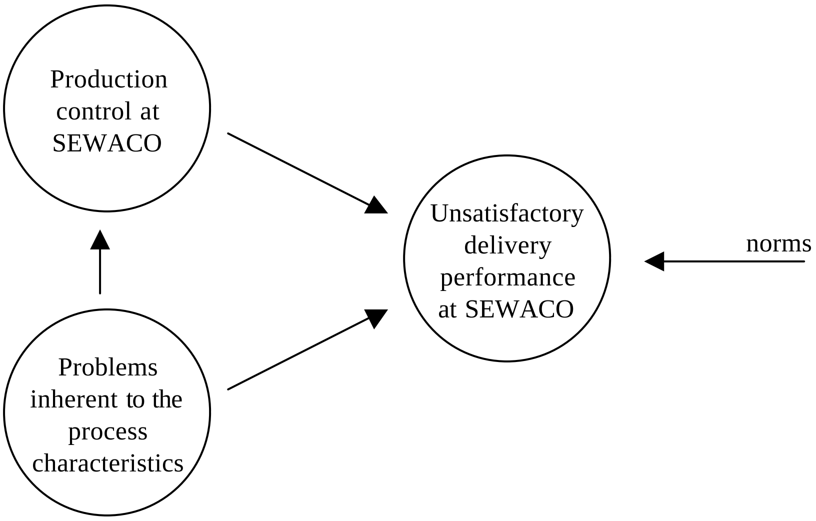


Figure 1.3. Relation between the process characteristic and the delivery performance in the SEWACO case.

Once the causes for poor performance are found, we use the theoretical results on hierarchical control structures to deduct and develop a suitable control structure (Chapter 4). We develop a small-scale model to fill the (possible) lack in theoretical knowledge about the functions of the designed control structure (Chapter 5). It is already stated that most models in operations research (and especially in the area of maintenance management) lack application in real-life situations. Scarf (1997) emphasizes that modelers should not only restrict themselves to complicated models which cannot be explained to practitioners. Instead, they should consider what might be gained (both for practitioners *and* academic people!) from restricting the attention to simple models, and approximate the solutions to problems of interest to decision makers. Endorsing this remark, we put the emphasis on applying the knowledge obtained from our research to improve the performance of real-life processes at SEWACO (Chapter 6, 7, 8). In Chapter 9, we discuss the domain to which the results of this thesis can be generalized. Finally, Chapter 10 presents the conclusions.

2

The Royal Netherlands Navy Case I - Introduction

2.1 Motivation

This research is inspired by a case carried out at SEWACO. The purpose of this chapter is to present the processes within a real-life maintenance organization and their control problems. We first explain the history and current tasks of the Royal Netherlands Navy (RNLN). We use these to derive the objectives for SEWACO, we describe the technical systems, the capacity resources, the organizational structure and the actual used planning and control procedures. We first give an objective overview of these topics, without immediately evaluating these principles. Subsequently, we discuss the resulting delivery performance. Finally, we put on our production control ‘glasses’ and use these to lay the finger on the sore spot and come to hypotheses about the causes of uncontrolled performance.

At the start of this research, the board of directors of the RNLN had a general complaint about the unsatisfactory delivery performance. In their opinion, the number of maintenance projects that was completed before the agreed due date was too low.

Several attempts had been made during the last years trying to increase this performance, but hardly any improvements were observed. These efforts gave rise to the thought that there was a lack of knowledge within the organization about the way in which this delivery performance could and should be controlled. Therefore the board of directors made out a case for a more structural analysis and approach, resulting in this research.

2.2 The Royal Netherlands Navy

This section is focused on the tasks of the RNLN and the role of SEWACO in adequately supporting these. We briefly describe the history of the RNLN and its current tasks. For more detailed information about these subjects, the interested reader is referred to Eekhout *et al*. (1988) and the Ministry of Defense (2000).

The formation of the RNLN goes back over a 500 years. At that time, the Dutch merchant navy and offshore were faced with a lot of piracy and plundering. Therefore, Maximiliaan van Habsburg decided to structure the control of the seagoing and called the Admiralty into being. This authority had to supervise the maritime jurisdiction and warfare. The date of this formation, that is somewhere in 1488, is generally regarded as the foundation of the RNLN. The following centuries, famous seamen like Piet Heyn, Maarten Hapertsz Tromp and Michiel Adriaensz de Ruyter ruled the seas. The present-day frigates still bear their names. Nowadays, more than 500 years after the formation of the Navy, the situation of international security has drastically changed: there are no longer serious threats and aggressions towards the territory of the Netherlands. However, ethnical contrasts, rising nationalism and fundamentalism have caused new regions of agitation in the world, that might threaten the international system of law and therefore also the interests of the Netherlands. As a consequence, the tasks of the Navy have changed: the strategy for security is now primarily focused on preventing large-scale conflicts. For these tasks, the Netherlands closely work together at an international level with the United Nations and the NATO. For actual interventions, permission is not only necessary at this international level but also at a national level. If indeed permission is given for intervention, the means of intervention range from only pressurizing by presence to economic blockades at sea and even actions with violence. For decisive interventions, the Navy can rely on operating strength in three different forms. They can not only operate at the surface of the water (frigates and supply ships), but also above it (helicopters, airplanes) and underneath it (submarines). Together, these form the ‘means of production’ of the Navy to ensure international security and fighting strength at sea. And consequently, the deterioration of these means makes that maintenance is necessary for an unhampered functioning of these systems.

Nowadays there are three different naval maintenance organizations. At first, the Royal Netherlands Navy Dockyard (“*Rijkswerf*”) is responsible for the maintenance, repair and modification of platform systems on board of all warships. These systems comprise the hull, the propulsion system and all supporting systems as energy supply systems, climate control, etc. Secondly, the Marine Electronic and Optical Company (“*MEOB*”) is responsible for all electronic and optical systems. And finally, SEWACO is responsible for the sensor, weapon, and command systems of the RNLN. The three maintenance organizations maintain a non-overlapping set of technical systems, whereas all naval technical systems are processed by one of these three organizations. Therefore, one can speak of a tommy or a monopolistic situation. However, once the technical systems are sent to one of the maintenance organizations, this organization is allowed to subcontract (part of) this maintenance project to external organizations. At this moment the three maintenance organizations are working on a merger to form one large naval maintenance organization that employs a about 2400 people. Although this should soon become effective, in this research we restrict ourselves to SEWACO.

2.3 The maintenance organization

SEWACO is located at the base of the RNLN in Den Helder, which is in the northern part of the Netherlands. As stated before, the objective of the RNLN is to provide international security and fighting strength at sea. In peacetime, this objective implies that the RNLN has to cooperate in missions, both individually as well as in cooperation with other countries (e.g. NATO allies). Schedules for these missions are made a few years in advance by the Dutch Commander of the Naval Forces. These schedules indicate when each ship (and the technical systems on it) is requested in an operational condition. Consequently, the objective of SEWACO is:

> *Providing an optimal maintenance of the technical sewaco systems against minimal cost, in order to bring (or keep) the operational availability of the RNLN into conformity with the demands in compliance with the demands on quality, occupational health and environment.*

Unlike in civil (profit) industries where performance indicators are used like due date reliability, throughput time and revenues, the performance of the ‘production department’ of the RNLN - which indeed is the fleet that has to provide international security and fighting strength at sea - is not that easily measurable. The performance of the maintenance department is easier to quantify, because they have to ensure that the operational availability of the technical sewaco systems is in accordance with the mission and maintenance schemes of the Navy. And as a result, SEWACO can steer on performance indicators like e.g. tardiness and due date reliability. These indicators are highly correlated with the operational availability of the RNLN. The lower (resp. higher) the tardiness (resp. due date reliability), the higher the operational availability. Accordingly, the relation between production and maintenance can be sketched by transforming Figure 1.1 into Figure 2.1 for this particular case. From this figure, it directly follows that the *raison d’être* of SEWACO is the wear of the technical systems.

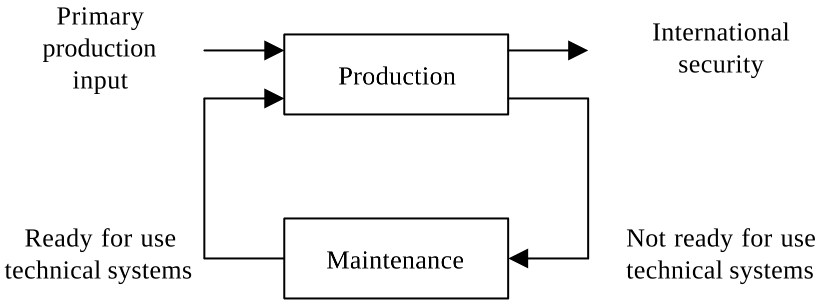


Figure 2.1. The relationship between the fleet and SEWACO.

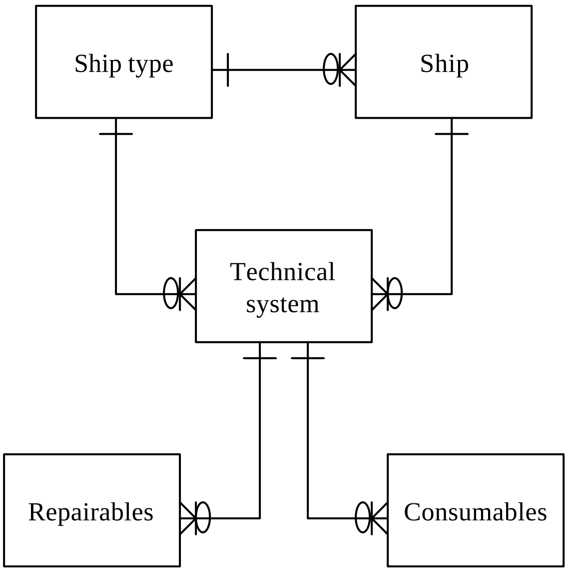


Figure 2.2. Product structure diagram.

2.3.1 Technical systems

SEWACO is responsible for the maintenance of the sensor, weapon, and command systems of the RNLN. These systems - and their availabilities - are of utmost importance: in wartime they can make the difference between life and death. A lot of these systems can be considered as high-tech systems: the most recent knowledge is used to make them as reliable, accurate and advanced as possible. One of the similarities between the different systems is their complexity: highly qualified operators are necessary to carry out the maintenance tasks. Figure 2.2 depicts the product structure diagram. Each ship (e.g. Tromp, De Ruyter) belongs to a certain ship type class (e.g. a frigate, a minesweeper). Each ship type has a number of prescribed technical systems (e.g. a goalkeeper) on board. If we descend in the product structure of these systems, we may distinguish consumables and repairables. Repairable items are items for which it is technically possible and economically beneficial to repair the item after a failure. After a failure, the failed repairable can be taken out of the system and replaced by a serviceable one. The failed repairable is sent to the maintenance organization in order to be restored into a non-failed condition. Since condemnation of repairables regularly occurs, it should be stated that resupply of repairables sometimes is needed in order to keep the number of repairables at a satisfactory level. Consumables are either components which lose their identities during usage and therefore cannot be re-used or for which repair is too expensive (repair cost exceeds replacement costs) or technically not possible.

2.3.2 Demand

Demands for the capacity resources of SEWACO arise for the larger part from the fleet of the RNLN. Besides, a small fraction of SEWACO capacity is being utilized for systems from other divisions of the Dutch armed forces (land forces and air force) and by foreign naval forces. Actual demand is determined by the breakdowns of the technical systems. At a more aggregated level we are able to pronounce upon demand characteristics, as there are (for example) arrival intensity, batch sizes, processing times. These characteristics are mainly determined by the following factors:

- The number of ships of the RNLN;
- Reliability of the technical systems;
- Operation intensity of the technical systems;
- Maintenance concepts of the technical systems.

SEWACO actually has only a little say in the factors, because these are for the larger part determined by the RNLN. Although characteristics of each of these four factors can be derived and used for forecasts in this study, the actual demand for SEWACO capacity is regarded as exogenous.

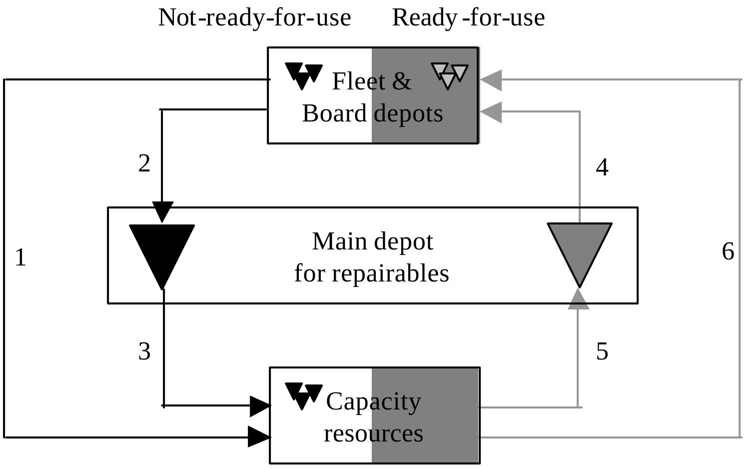


Figure 2.3. Rotation of technical systems in the closed system with the flows of not-ready-for-use (black) technical systems and ready-for-use (gray) technical systems.

How demand actually comes to SEWACO can best be explained on the basis of Figure 2.3. This figure represents the so-called “closed system” in which technical systems (as far as these are repairable sub-systems) are rotating. As can be seen, the systems can be either in a not-ready-for-use (NRFU) status or a ready-for-use (RFU) status. We say that a (sub-)system is NRFU if either it has failed or has come into a certain condition which requires preventive maintenance.

Figure 2.3 depicts the various locations in which the technical systems can be: the fleet and the board depots, the main depot for repairables and the repair shops (waiting for or being processed by the capacity resources). The fleet consists of the naval vessels. Each ship has its own board depot for the storage of spare parts and failed sub-systems. The RFU technical systems on board can be either operational or not, depending on the question whether a particular system is requested to be used or not. Besides, there are also RFU repairables which are not immediately requested to be in use, but used as spare parts and kept in the board depot. Upon failure of an operational technical system, the NRFU repairable is taken out (if possible), it is put in the board depot and it is replaced by a RFU one (if any available). After the mission of the ship, the NRFU repairables in the board depot are sent to the main depot (see flow (2) in Figure 2.3) and exchanged with RFU systems (4). In case the level of RFU repairables in the main depot drops under a predefined stock level, NRFU repairables are sent to the maintenance depot (3). After having received maintenance, these are sent back to the main depot in a RFU condition (5). Unlike the flow of repairables, corrective and preventive repairs result into the direct flow to the maintenance department (1). After maintenance, the technical systems are directly sent back to the fleet (6). Finally, it should be noted that the different flows of maintenance compete for the same capacity resources.

Based upon the closed system representation in Figure 2.3, different types of demand are distinguished at SEWACO, each having their own characteristics with different implications for the planning and control. Each maintenance assignment of the RNLN for SEWACO is called a *project*, regardless its size. A project may consist of only one *job*, but also of many jobs, all asking for capacity of different types of capacity resources. A job is carried out by only one operator. Based upon the consequences for planning and control, three types of projects are distinguished at SEWACO: *preventive*, *corrective* and *repairable* maintenance projects.

Firstly, preventive maintenance projects (about 57% of the 1998 workload expressed in hours) comprehend the overhaul of entire warships. These overhauls are planned years ahead and have a throughput time of about 10 months. Within this time period, all technical SEWACO systems need to be revised and repaired if needed. These projects explode into a lot of jobs, asking for capacity of various repair shops. Because no - or hardly any - precedence relations exist within a project, all jobs within a single project have identical external due dates. This due date mostly is the start date of the next mission. Consequently, delivering at least one job after this due date makes that the entire overhaul is completed late. Note that these projects are customer order driven.

Secondly, corrective maintenance projects (20% of the 1998 workload) are distinguished. Corrective projects arise from the breakdowns of operational systems. As soon as (a part of) an operational system fails, the operational availability of the fleet is reduced and immediate repair is needed to fix the problem. In general, these projects do not take that much time. However, their unexpected arrivals and the travelling time of an operator to the ship involved, makes that they disturb the progress of the preventive maintenance projects on hand. Note that these projects are customer order driven.

And finally, repairable maintenance projects (23% of the 1998 workload) constitute the third category. As shown in Figure 2.3, NRFU repairables are stocked in the main depot waiting for maintenance to be turned into a RFU condition. An individual repairable project mostly concerns only one (relatively) small job and requires only one type of resource. The necessity to process these jobs depends on the inventory levels of RFU repairables. Therefore, the due dates are set by SEWACO itself, based on these inventory levels. However, exceeding this due date does not have a direct negative result on the delivery performance. However, the further a repairable project is postponed, the smaller is the probability that demand from the fleet for their board depots can be filled from stock. In real-life, the time between failure and the next moment that this repairable is required to be RFU is relatively large. As opposed to the first two classes, note that these projects are inventory level driven.

2.3.3 SEWACO organization

In this section we analyze in more detail the SEWACO organization in a bottom-up approach. At the operational level, operators are carrying out maintenance jobs. For a particular job, one may need materials, tools, documentation and at least one operator. It is a characteristic of maintenance jobs that this operator is needed during the entire execution and therefore their availability is an important control variable. Their main tasks are to inspect, test, repair and modify the technical systems of the RNLN. Some jobs can be accelerated if two or more operators are simultaneously working at the same job, but this is not the case for most jobs. Because of the state-of-the-art technology of the technical systems, operators need to be well-trained and they periodically need new training.

Operators are - based upon their knowledge and skills - clustered into repair shops. People in one repair shop are more or less equally skilled and therefore mutually interchangeable. These operators work in parallel: after having repaired a technical sub-system, this component is finished and has not to be processed by another operator in the same repair shop. Each repair shop is responsible for the maintenance of a predefined set of technical systems, whereas each technical sub-system can be maintained by only one repair shop: there is a one-to-one relation between each technical sub-system and the repair shop which is responsible for its maintenance. The number of operators within a repair shop varies between 1 and 12, whereas on average there are about 4 operators per repair shop. Within SEWACO there are 75 repair shops.

Finally, the repair shops which are responsible for sub-systems that belong to the same technical system are (both organizational and physically) clustered into a repair centre. In this way an atmosphere is created where the commitment and involvement of the employees with the end products is maximized. An individual repair centre can be regarded as more or less self-contained: there are hardly any precedence relations for the projects between the different repair centres. Therefore, the integral responsibility for a timely and correct delivery is put mainly at the individual repair centres. As Figure 2.4 shows, SEWACO can be regarded as a network of repair centres with their own buffers with arrived projects and jobs on hand.

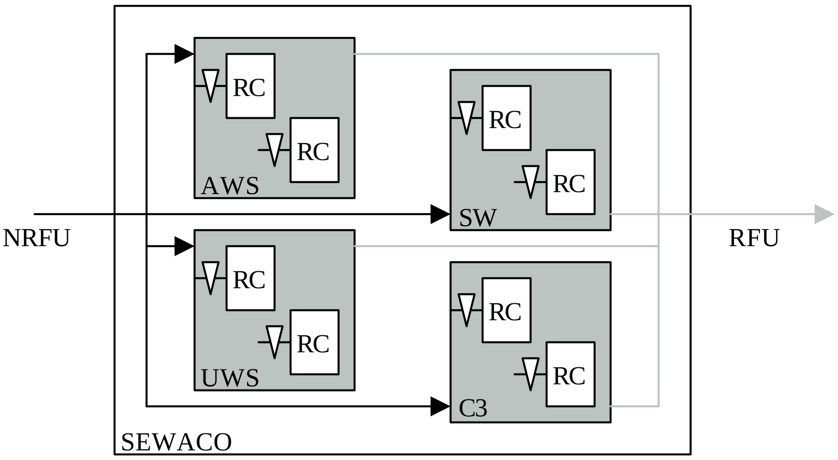


Figure 2.4. The SEWACO design: four SMOs and some of their repair centres (RC).

Based on the type of technical systems for which the repair centres are responsible, each repair centre belongs to one of the four Sub-Maintenance Organizations (SMOs). These SMOs, together with the Material Coordination, form the core of the organization. The SMOs are charged with the maintenance of technical systems and the supply of the non-failed repairables to the ships. Because ‘customer-orientation’ has been one of the key principles in the design of the current SEWACO organization, one created (physically) short lines of communication. Therefore, the choice of these four SMOs is based on a clustering of technical systems, resulting in these four SMOs:

- Above Water Systems (AWS);
- Under Water Systems (UWS);
- Command Control and Communication (C3);
- Specific Workshops (SW).

For AWS, UWS and C3 it may be more or less obvious which type of technical systems they maintain. The last one, SW, is responsible for the maintenance of the ammunition. Each SMO is supported by a Production Support and a Maintenance Engineering department (see Figure 2.5). Production Support is responsible for the capacity planning, detecting bottlenecks and preventing idle time. Maintenance Engineering has to support the operators with technical information and to train the operators on the job.

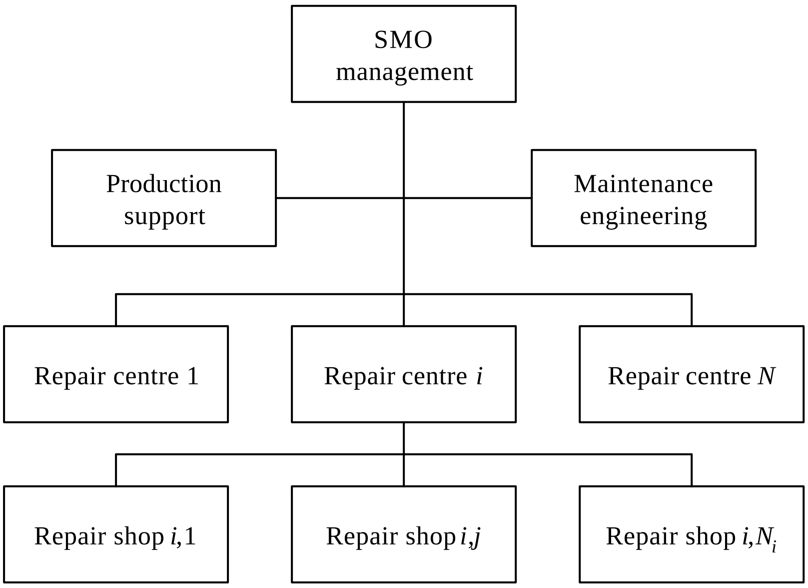


Figure 2.5. Organization chart of a SMO.

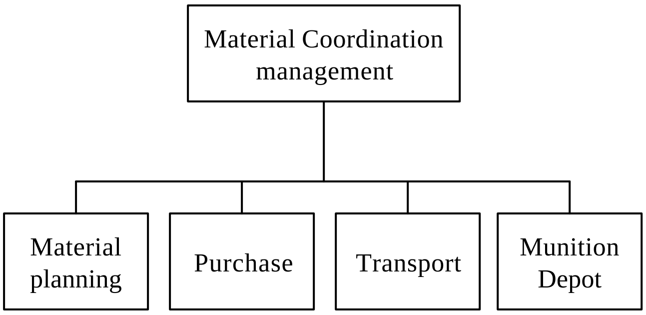


Figure 2.6. Organization chart of the Material Coordination

The main task of the Material Coordination (see Figure 2.6) can simply be summarized as ensuring that the right materials (consumables and repairables) are in the right quantity, at the right place, and at the right time. Their main control instruments are monitoring the stock levels of non-failed repairables at the main and local board depots and sending failed ones to the repair centres if one of the stock levels is not sufficiently high (see flow 3 in Figure 2.3). Besides, they have to make sure that the preventive and corrective projects are transported to SEWACO if (part of) these projects are carried out - in a decoupled way - inside the SEWACO building (see flow 1 in Figure 2.3).

2.4 Planning and control: the current situation

In this section we take a closer look at the currently used procedures for planning and control. We first analyze the strategic and tactic parameters. Then we describe the aggregate planning, the due date assignment and repair centre control. We conclude this section with an investigation of the performance, which is a result of the current way of planning and controlling.

2.4.1 Strategic and tactic parameters

The objective of SEWACO is discussed in Section 2.3.1 and can be summarized as “providing an optimal maintenance against satisfactory costs”. This mission is mainly availability oriented. In order to support the operational control decisions that should contribute to this mission, a set of strategic and tactic agreements (or parameters) is derived from it. These agreements are written down in a yearly covenant between the fleet and SEWACO. Surprisingly, these parameters are ‘capacity utilization oriented’, rather than ‘availability oriented’. These parameters include amongst others the percentage of hours actually spent on customer projects, and the ratio between direct and indirect personnel. Concerning the availability, only agreements are made about the speed of response in case of corrective maintenance projects.

The customers frequently complain about the delivery performance of SEWACO. The maintenance organization is aware of this grievance and periodically measures the delivery performance towards the customer. Surprisingly, at a strategic level no (or not yet) agreements are made about the delivery performance.

2.4.2 Aggregate planning

At an aggregate level, conditions need to be created for the repair centres to allow them to attain the stated goals. Roughly speaking, at this level they have to provide the needed capacity and materials for the individual repair centers.

Demand for maintenance is to a large extent determined by the Navy. The Commander of the Naval Forces is responsible for the Mission & Maintenance Scheme. This scheme roughly depicts for a period of ten years when and which Naval units are flying, sailing or receiving maintenance. In general, there are only a few short time mutations in this scheme. Mutations may occur (among other things) if the RNLN is requested to operate in NATO alliance (e.g. Gulf war, Yugoslavia). From the Mission & Maintenance Scheme the Sail, Fly, Maintenance & Exercise Schedule is derived. This schedule, with a horizon of only one year, is a more detailed projection of the Mission & Maintenance Scheme. SEWACO has little influence in this phase. This little influence can be used to more or less smooth the workload arising from large projects.

On a yearly base, the Commander of the Naval Forces and SEWACO management agree on the total budget for resource capacities (manpower) and materials which are needed to perform the maintenance activities. These budgets are based on the Sail, Fly, Maintenance & Exercise Schedule as well as on forecasts about corrective maintenance, repairable recoveries and appointments with other customers (e.g. Land Forces, Air Force, foreign military forces). These agreements are confirmed in a covenant. In this covenant the different types of human resources (irrespective of the skills) within the organization are aggregated into a single number of total hours of manpower per year and compared to the estimated demand. However, due to the fact that all employees of the organization are on the payroll of the Dutch government and thus laying off is not that easy, the total number of employees cannot fluctuate much over time. Therefore, maintenance periods are more or less matched with the availability of manpower, as explained in the previous paragraph.

2.4.3 Due date assignment

In most production environments, the due date assignment function is a powerful tool to control the work load and (indirectly) the delivery performance. For maintenance organizations (especially the ones which are part of a production facility) it holds that all maintenance jobs need to be accepted in order to achieve high availabilities in the production facilities. After all, the shorter the down time, the higher the availability. Moreover, in such a case the due dates are more or less dictated by the production facility. The SEWACO case is an example of such a situation, as the relation between the fleet and maintenance department implies (see Figure 2.1). Whereas SEWACO has a little influence in the design of the long term Mission & Maintenance Scheme, on the 'short term' (i.e. within a year) there is no influence, all projects with their requested due dates need to be accepted. These due dates are called the *original due dates*.

Original Due Date	Number	Percentage	Conditional adjustment (calendar days)
Advanced	19	0.4%	-77
Postponed	1961	41.1%	+151
Kept unchanged	2793	58.5%	0
Total	4773	100%	+62

Table 2.1. Due date adjustments for the projects with an original due date in 1998.

However, before and during the processing of the individual projects, it remains possible to adjust the due dates, resulting in an *adjusted due date.* This adjusted due date can either be *earlier* than the original due date (advancing) or *later* than the original due date (postponing). For projects in which more than one repair centre is involved, this adjustment has to be done at a central level, but in case only one repair centre is involved, this can be done at repair centre level. On top of that, in both cases the original due date can be adjusted without consulting the commander of the concerned ship. Over time, this has become a generally accepted procedure within SEWACO to control the delivery performance. After all: postponing a due date generally means increasing the delivery performance. However, this has serious negative consequences for the overall closed system performance: the availability of the fleet decreases! The due date adjusting behavior is shown in Table 2.1. For this table, we have constructed a database containing all projects having their *original due dates* in 1998. Of all these 4773 projects as much as 41.1% are postponed, with an average postponement of 151 calendar days. The other way around - advancing the due dates - only sparsely happened: 0.4% of the projects, with an average advancement of 77 calendar days. On average, the due dates are postponed with 62 calendar days.

To what extent the original and adjusted due dates represent any sense of reality (given the available capacity) can easily be shown. Therefore we constructed a database containing two types of projects. Firstly, this database contains all jobs which were on hand on January 1, 1998. Some of these may be in backlog (i.e. their original due date was in 1997), others may have due dates in 1998 or even later. Secondly, this database contains all projects which arrived during 1998. From the projects in the database we can extract some interesting views on the requirements for capacity resources and their availabilities in a graphical way, at any level in the organization, ranging from an individual repair shop with only one operator to the entire organization.

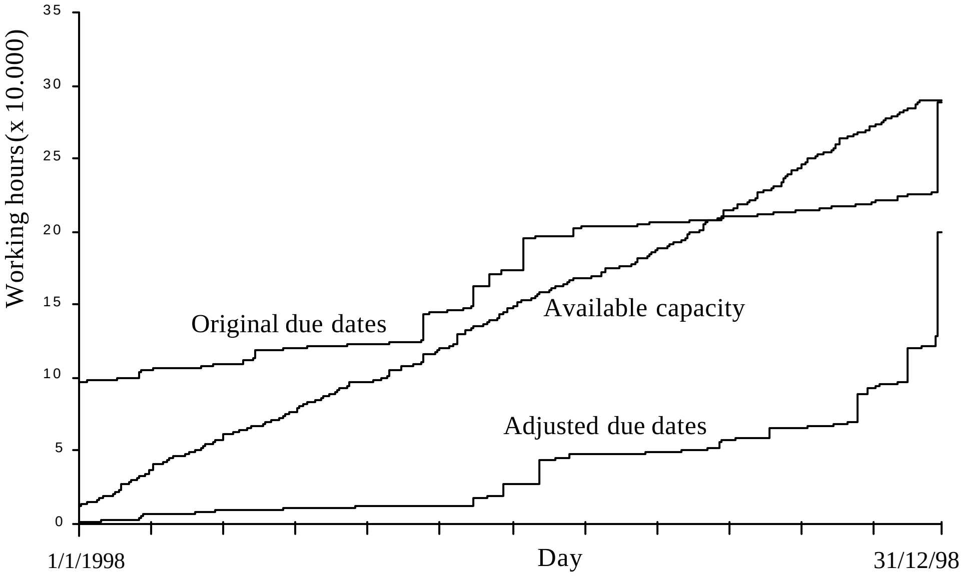


Figure 2.7. Available cumulative capacity, required cumulative capacity given the original due dates and required cumulative capacity (all in working hours) given the adjusted due dates for the entire SEWACO organization in 1998.

Based on this database, for each project we have projected its required demand for capacity at its *original due date*, for various levels in the organization. The accumulation of all projects in the database at organizational level is presented in Figure 2.7. Appendix A shows this figure for lower levels in the organization. A particular point on the requirement line (say at day t) represents the minimal needed cumulative capacity (expressed in working hours) between January 1st and day t, that is needed to process all the projects with an original due date before day t and that are not yet finished before January 1, 1998. Clearly, the intersection of this line with the y-axis depicts the backlog at January 1, 1998. In a similar way, another line represents the requirements in case all projects are projected at their *adjusted due date.* Finally, the availability line shows the used cumulative resource capacity between January 1st, 1998 and any particular day in 1998. This figure shows that original due dates are far from realistic when they are compared to the availability of resource capacity. On a yearly base, the cumulative required resource capacity is more or less equal to the cumulative available resource capacity, observing the intersection of the availability line and the original due date line at December 31th. However, it is not until the end of October that the resource availability exceeds the demand. Once the cumulative capacity requirements are measured against the adjusted due dates, it becomes clear that the cumulative capacity availability exceeds the requirements. This difference is about 100,000 hours. Or in other words, the due dates adjustments make that a necessary condition for a controlled and high performance - availability exceeds demand - is satisfied. Finally, this picture shows that there are two peaks in demand: the first peak occurs somewhat in the middle of the year, whereas the second peak appears at December 31, 1998. Appendix A shows that this mismatch is also common practice at lower levels of the organization.

Summarizing, this figure shows that indeed on a yearly base the availability of the capacity resources is in accordance with the requirements. However, the dynamics and uncertainty of the demand make this is an insufficient condition for control. Moreover, it can easily be seen that it is impossible to obtain a controlled process under the current way of planning and controlling. To fulfill the capacity requirements, it should be possible to flexibly vary the capacity level during the year: it seems to be impossible to control the processes if no use is made of subcontracting.

2.4.4 Repair centre control

Based on the belief that the best end-products arise when the commitment and involvement of the employees with the end-products is maximized, the integral responsibility for a timely and correct delivery of the products is being put on the individual repair centres as much as possible. The decentralized Production Support departments should assist in this task. The goals and means of both the repair centres and the Production Support are analyzed.

Job number	Estimated repair time	Number of historical events	Historical repair times
G08582	24.0	0	-
G08583	48.0	3	32.0; 23.5; 63.5
G08584	12.0	1	12.5
G08585	192.0	1	102.7
G08586	80.0	0	-
G08587	160.0	1	259.5
G08588	4.0	1	4.5
G08590	4.0	1	3.5

Table 2.2. Estimated workload in working hours for the preventive overhaul of a technical system on the frigate ‘Tromp’ and historical repair times in 1997 and 1998.

The primary goal of the Production Support is to make sure that capacity utilization targets are met. They should report to the SMO manager both the periods of expected capacity shortages as well as capacity surplus. For that purpose they have insight in the future capacity availability and already arrived projects demanding for capacity. The knowledge about arrived projects is used as a base to derive estimations for the future processing times at job level, using historical data. However, this may not be an easy task, for the following two reasons. Firstly, even for preventive maintenance of the same type of components, repair times may vary a lot from time to time. In some cases only inspection and some cleaning is necessary, whereas in other cases some hidden failures in the system come up and a lot of effort should be put in diagnosing, repairing and testing the components. Secondly, for every type of technical system it holds that there are only a few items within the RNLN. Considering that the reliability of the systems is generally high, the average time between two consecutive maintenance activities at a similar technical system is relatively long. Therefore, it needs no further comment that the record with historical repair times may be rather limited. This seriously influences the accuracy of the estimations of the repair times. This problem is illustrated on the basis of Table 2.2, which shows the jobs in a overhaul of a technical system at the frigate ‘Tromp’. From this table it is clear that estimating the repair times based on historical data may be a hard task. However, it can also be shown that one currently does a rather good job in estimating the actual repair times. An analysis of the ratios of on the one hand the difference between the actual and estimated repair time and on the other hand the estimated repair time for all jobs that are processed and finished in 1998 (N = 11353) has resulted in Table 2.3. We draw the following conclusions. Firstly, the values of the ratios strongly vary: the database produces ratios in a range between -99.8% and 3031.25%. Note that in the latter case, the actual repair time is equal to (about) 30 times the estimated processing time! Secondly, the mass of the distribution is at the right side of zero: almost 50% of the observations have ratios which are between 0% and 10%. Finally, looking at the *absolute* deviation, it has been derived that on average the actual repair time is about 0.7 working hours longer than the estimated repair time. The average repair time of the 11353 jobs is 26.8 hours. The standard deviation of these repair times is 22.24.

Ratio interval (%)	Number of observations	Percentage	Cumulative percentage
(-100;-50)	496	4.4%	4.4%
[-50;-40)	335	3.0%	7.3%
[-40;-30)	382	3.4%	10.7%
[-30;-20)	438	3.9%	14.5%
[-20;-10)	866	7.6%	22.2%
[-10;0)	1848	16.3%	38.5%
[0;10)	5664	49.9%	88.3%
[10;20)	295	2.6%	90.9%
[20;30)	185	1.6%	92.6%
[30;40)	152	1.3%	93.9%
[40;50)	70	0.6%	94.5%
[50; →)	622	5.5%	100.0%

Table 2.3. Distribution of the actual repair times minus the estimated repair times, divided by the estimated repair times, for the jobs that are processed in 1998 (N = 11353).

The Business Planning System (BPS) is the main tool available for the Production Support to make the (future) capacity requirements visible. However, observing everyday practice has learnt us that this module is hardly used for this purpose. Looking more carefully at the basic algorithms in the software package, the following drawbacks and shortcomings can be distinguished. The planning and control algorithms:

- treat each project in the same way, irrespective its characteristics;
- assign priorities to projects in a way that has no relation to the operational availability of the technical systems;
- do not reserve capacity for unplanned corrective jobs.
- intend to plan high priority jobs at the latest possible start date.
- have an inclination to minimize throughput time, rather than minimizing due tardiness;
- generate new schedules only on a weekly base.

The purpose of the repair centres is to ensure the timely delivery of the projects, which indeed is perpendicular to the objective of the Production Support. The latter has to ensure a high utilization rate. It is known from theory that high utilization rates go with high throughput times. Work order release, sequencing rules and ordering materials and manuals are the main decision functions to achieve a satisfactory performance. However, the actual procedures for using these decision functions vary from shop to shop. The repair shop manager has a lot of freedom in establishing these rules.

2.4.5 Performance

Now that the processes of planning and control have been explained in more detail, a closer look is taken at the performance of the organization. As we are mainly interested in the operational availability, attention is especially paid to the delivery performance of SEWACO. For an objective representation we constructed a database containing all projects with an original due date between January 1, 1998 and December 31, 1998 (which is exactly the same database which has formed the basis for the construction of Table 2.1). Table 2.4 shows the 1998 delivery performance at the company level. It turns out that indeed a very low percentage of the projects has indeed been completed before their original due dates. Even measuring with the adjusted due dates, it still cannot be said that the performance is satisfactory. Note that this observation is completely in accordance with the literature in Chapter 1: "*projects are always late*". In Appendix B, the 1998 performances for the different repair shops is shown.

Projects	Completed before original due date	Completed before adjusted due date
Preventive	47.0%	70.8%
Corrective	34.7%	54.4%
Repairables	39.0%	71.0%
Total	38.7%	65.3%

Table 2.4. Percentage of all SEWACO projects with original due date in 1998 and completed in time.

Projects	Original due date			Adjusted due date		
	Lateness	Conditional tardiness	Conditional earliness	Lateness	Conditional tardiness	Conditional earliness
Preventive	-3	65	-81	-48	28	-80
Corrective	26	53	-24	-4	31	-33
Repairables	33	97	-66	-41	53	-80

Table 2.5. Due date deviations measured in calendar days for projects with original due dates in 1998.

However, not only the percentage of projects completed in time is of interest, but also the due date deviations. Therefore we have measured for the three distinguished types of projects the lateness, the conditional tardiness and the conditional earliness (see Table 2.5). The lateness is defined as the average deviation (measured in calendar days) between the project due dates and the project completion dates, where a negative (resp. positive) lateness means that on average the projects are delivered before (resp. after) the agreed due dates. The conditional tardiness (resp. earliness) is defined as the average lateness measured for all the projects that have been delivered after (resp. before) the agreed due dates. Note that both the lateness and conditional tardiness are censored variables: for the projects which were not yet completed at the end of 1998, it holds that the completion date is not yet known and therefore set at

January 1, 1999. As a consequence, the real values are higher than the ones depicted in Table 2.5. Based upon Table 2.5, we may draw a few conclusions. Firstly - as may be expected from the results shown in Table 2.1 and Figure 2.7 - large differences are found between the original and adjusted due dates. Re-planning indeed is a common phenomenon and therefore processes seem to be uncontrolled. Besides, it turns out that adjusting the due date indeed means postponing (rather than advancing) the due date, because the conditional earliness for the original due dates is (more or less) equal to the conditional earliness for the adjusted due dates. Secondly, the due date adherence is poor. Projects that are late, tremendously exceed their due dates. Thirdly, because the delivery performances for preventive projects outperform the performances of the other types of projects, it is suggested that preventive projects are better controlled. Finally, it can be concluded that (despite the fact that the majority of the projects is being delivered after the due date) a lot of projects are being delivered way to early, as the conditional earliness shows. This can be a direct cause of the mismatch between availability of and requirements for capacity: at the short term the repair shop manager tries to deliver at least some projects in time using a sequencing rule which gives priority to those jobs which are available and still (far) before their due dates.

2.5 Local versus process inherent problems

In the previous section we have looked in more detail at the processes within SEWACO and the resulting performance. It may be obvious that the performance is unsatisfactory and that several attempts have unsuccessfully been made to improve this performance: no serious improvements have been obtained. Earlier, we suggested that two types of causes should be distinguished for this poor performance. On the one hand there are local causes which only play a role in the SEWACO case, and on the other hand there are causes which are inherent to this type of industry and thus are assumed to bother also other maintenance organizations of this type. We try to isolate this last type of problems from the SEWACO situation and use them to refine the problem formulation stated in Chapter 1.

2.5.1 SEWACO related problems

The case study at SEWACO has learnt us some very interesting topics which are not assumed to bother the delivery performance of other maintenance organizations, satisfying the conditions stated in Chapter 1, but are specific for this organization. In the remainder of this subsection these topics are presented.

- The strategic parameters that are currently in use are determined at the navy level. These parameters are mainly oriented towards the prevention of idle time in the shops, rather than fleet-availability oriented. As a result, a satisfactory performance - in terms of the *current* strategic parameters - can be attained by keeping the operators busy. However, the commanders of the fleet do not consider this as a satisfactory performance. This mainly follows from their grievance about the current operational availability of the various systems and the delivery performance. So, a first step towards a controlled delivery performance should be the choice of operational availability oriented strategic parameters.
- A lot of responsibility is put at the different repair centres. The management delegates the integral responsibility for a timely and correct delivery of the products as much as possible to the individual repair centres. Besides the fact that no threshold values for this ‘timely and correct delivery’ are given and that delegation of responsibilities mainly is assumed to be reprehensible, this also creates a lot of freedom within the individual repair shops. Clearly, the freedom exceeds the knowledge and perception of the decision makers at repair shop level.
- In the current control structure, the Production Support and the repair centres are assumed to co-operate in achieving satisfactory due date performance and capacity utilization. Because the objective of Production Support is to achieve a satisfactory utilization of. the capacity resources and the repair centres face the problem to achieve a satisfactory delivery performance, both goals are conflicting.
- It turns out that the available tools for supporting the control decisions are not geared to the processes and therefore not supportive. One of the results of this phenomenon is that the people in a repair centre face problems both a the short and medium term. This leads them into a vicious circle: the short term disturbances do not allow them to think about the processes and how to prevent or reduce these problems by an adequate medium term planning, which therefore

does not reduce the short term problems. This is confirmed by the values for the conditional earliness and conditional tardiness: besides delivering many projects too late, one is trying to compensate this on the short run by delivering other projects way too early.

2.5.2 Problems inherent to processes

Solely looking at the processes and their characteristics, we come to the following observation. We have seen that the maintenance organizations under scrutiny in this thesis, have tight relations with the production departments. These relations make that basically all projects need to be accepted and thus the influence of the maintenance organization in setting the due dates is very restricted. Consequently, we can classify the demand pattern as *lumpy*. This means that it is of utmost importance to match demand and resource availability at the medium term: postponing control decisions is very undesirable because short term capacity adjustment are impossible. Furthermore, we have seen that just having fixed internal capacity levels is not sufficient. Capacity should be flexible to meet the customer’s demand.

Two conditions need to be met in order to solve this problem. Firstly, the flexible part of the capacity should be used. This can be achieved by subcontracting. Secondly, the planners need to have clear insights in the requirements for capacity in the near future. Nowadays, they find this hard due to the uncertainty and the capacity complexity. To first solve this second problem, in the next chapter we try to model the behavior of the current planners. This behavior needs to be understood in order to explain what they do wrong and to come to performance improvement. Furthermore, factors are derived which currently are not adequately being used during planning and control.

3

The Royal Netherlands Navy Case II - Diagnosing Project Planning Performance

3.1 Introduction

It is obvious that the performance of SEWACO can be improved: this is motivated by the actual performance (which is considered as unsatisfactory), the demand characteristics (which show that uncertainty exists but it does not seem to be insurmountable) and the available control decisions (advancing and postponing projects within their allowed time slots and the possibility to subcontract).

We have seen that the processes in the different repair centres are complex and stochastic. The planners in the Production Support departments work together with the people in the repair centres. Their main task is (based on the structure and work contents of the projects) to make sufficient internal and external capacity available to perform the projects in time. Checking whether their behavior can be adjusted in order to achieve performance improvement is difficult. Like in many other real-life situations, the exact way in which this capacity and demand matching process is carried out is difficult to establish, mainly because various situational factors play a role, such as the level of experience of the planners, individual (assumed) knowledge about characteristics of the projects and the repair shop status, expectations regarding capacity available in the various repair shops, etcetera. In this chapter, we therefore first derive from scientific theory the characteristics that should be taken into account when controlling the project due date performance. Subsequently, we apply a diagnosis tool that has been developed by Keizers *et al*. (2000) to 12 months of real-life data and check whether these characteristics are adequately being used in the current planning and control processes. This knowledge can be used for performance improvement and should therefore be incorporated in the design of the framework for control.

The decision freedom of the planners at SEWACO is rather limited: the due dates are set by the Navy. As a consequence, the planners are faced with *time driven planning*: the project due dates are fixed and the aim is to minimize the costs of external capacity. This is opposite to *resource driven planning*, where resource availability levels are fixed, and the goal is to meet due dates as much as possible (see Giebels *et al.*, 2000). Both problems have many similarities. At least, in both problems the same factors regarding structure and work contents of the projects on the one hand, and the characteristics of the repair shops on the other hand, need to be considered. Therefore, as a basis for this chapter, we use the theoretical research on (*resource driven*) order planning and due date performance, which has produced much knowledge about how to assign attainable due dates to orders. For some theoretical models optimal due date assignment procedures have been obtained with respect to a given objective; for other models, the theoretical results indicate important factors that determine due date performance. These results focus more on due date performance than the results on time driven planning do.

The results on resource driven order planning have been generated within the context of theoretical models, that is, small scale models assuming idealized orders, order arrival patterns, order processing times, resource structures, etcetera. We may assume that these results are of use for improving due date performance in real-life situations. However, real-life situations differ in many, often unknown, ways from the theoretical models, although many of the characteristics of the theoretical models can be recognized in real-life situations. Wiers (1996), for instance, studied real-life planning and scheduling systems and concluded that these are much richer than theoretical models, both in terms of the complexity of the problems to be solved and the intelligence and control means available for the planners solving the problems. It is now widely agreed that just applying the results of theoretical research does not solve the real-life planning problem. However, human planners and schedulers who are integral parts of the planning and scheduling functions in a manufacturing system can use the theoretical results as a guideline for their practice, but they still are fully aware of the complexity of the real-life problems and the various means available for tackling the problems (McKay *et al.* 1995a, 1995b). Recent research by Crawford *et al.* (1999) has shown that planners in manufacturing systems have many different roles, ranging from mediator and problem solver, to information channel and problem escalator. Planners therefore seem to do more than just taking decisions; they seem to influence the process in various ways, and seem to take into account all kinds of information that is not represented in the theoretical models that are being used for studying the planning and control processes. These human planning and control processes are difficult, if not impossible, to model. Therefore, we do not attempt to model the short term scheduling and sequencing decisions. Instead, we focus on the actual decisions of the planners who have to create solvable sequencing and scheduling problems for the detailed level. A relevant question therefore is in what way use can be made of the theoretical results to adapt the planning behavior of the SEWACO planners in a right direction.

This question could be answered by comparing the current planning policies with the available prescriptive scientific research and then identifying potential for improvement. However, in view of the huge complexity of the real-life situation at SEWACO, it will be clear that it makes no sense to try to develop optimal order planning and capacity planning policies and compare them with the actual order and capacity planning decisions taken in order to search for ways to improve the planning performance. After all, developing a detailed description of the processes is impossible. However, we can derive from theory the project and shop characteristics that are assumed to influence the due date performance and then check whether these characteristics have been sufficiently taken into account at SEWACO, without developing optimal order and capacity planning policies. For this purpose we applied logistic regression to 12 months of real-life data of the projects and capacity resources of the maintenance organization. The regression results reveal the characteristics which are not adequately being used in the current planning processes. Once these characteristics have been identified, in a next step, the regression results can be used for performance improvement and the design of the control structure.

The remainder of this chapter is organized as follows. In Section 3.2 a short literature overview is given of order planning and due date assignment research. In Section 3.3 this knowledge is applied to the SEWACO case. Section 3.4 discusses the theoretical factors that are assumed to determine order planning and due date performance. Section 3.5 presents the logistic regression model and its application to the case data. And finally, Section 3.6 presents the conclusions of this chapter.

3.2 Literature review

Scientific literature on dynamic stochastic order planning and due date performance is mainly concerned with so-called single component orders, that is orders that consist of the manufacturing of one part, according to some specific linear routing of operations in the shop. Research into due date assignment aiming at a high due date performance has among others been done by Conway *et al*. (1969), Eilon & Chowdhury (1976), Kanet & Hayya (1982), Bertrand (1983), Baker (1984) and Enns (1994). The major findings of this research are that the due date performance, measured by the order tardiness and variance in order lateness, can be improved by:

- assigning internal due dates which are proportional to the number of operations in the order and/or the total workload in the order (Conway *et al*. 1969);
- assigning internal due dates which reflect the workload in the shop (Eilon & Chowdhury 1976, Bertrand 1983);
- assigning external due dates as the sum of the internal due dates plus an allowance that is proportional to the standard deviation of the internal order lateness (Bertrand 1983, Enns 1994);

- using internal due date oriented priority rules for sequencing the projects in the shop (Conway *et al.* 1969, Kanet & Hayya 1982, Baker 1984).

Scientific literature on due date performance for orders that consist of more components which have to be completed at a common due date is rather scarce. In the literature this type of order is referred to as an assembly order. Maxwell & Mehra (1968) were the first to study the effect of various priority rules on the throughput time and due date performance of assembly orders. Their main findings were that for these orders it is important that the priority rule synchronizes the progress of the component orders that belong to the same assembly orders. Siegel (1971) extended this research by investigating a wide variety of due date assignment and priority rules for assembly orders. He identified three factors which influence the assembly order throughput time:

- Pacing; that is synchronizing the progress of orders belonging to the same assembly order;
- Acceleration; that is giving priority to assembly orders with a relatively low number of operations remaining;
- Determining due dates based on structural dependencies: that is considering the number of components per order when assigning a due date.

Fry *et al.* (1989a, 1989b) studied an enhancement of the due date assignment rules studied by Siegel (1971) by including workload data in calculating the coefficients in the due date assignment rule. They observed a substantial improvement in due date performance. Adam *et al.* (1993) developed a due date assignment system which dynamically updates the coefficients in the due date rule depending on both the order mix on the shop floor and the workload. Adam *et al.* (1987) investigated priority rules for multi level assembly orders which take into account work remaining and order due dates. Philipoom *et al.* (1991) investigated priority rules which take into account work remaining and relative position of an operation in its routing. All these studies showed some performance improvements under certain circumstances.

The main general results of order planning and due date performance research in assembly job shops are consistent with the results for component job shops. Important factors are the order work content and workload in the shops, distinguishing external due dates from internal due dates and basing job priority in the shop on the order due dates. However, for assembly orders it is a general finding that the external order due date should increase with the number of components of the order (although not proportionally). Moreover, for assembly shops, the order sequencing problem is much more complex than for component shops, due to the conflicting priorities that arise from the various perspectives (number of components remaining, slack time remaining, processing times) from which the sequencing problem can be considered. Currently, there is no priority rule available which takes into account all these perspectives and which has shown to perform well under a wide variety of shop and job conditions.

As opposite to the literature on resource driven planning, the main body of research on time driven planning is capacity utilization oriented, rather than due date performance oriented. However, still some relevant comments should and could be made about this literature. Giebels (2000) describes the EtoPlan concept. The EtoPlan order planning methodology supports the integration of the decision function 'order acceptance' and the higher level planning tasks in both process planning (i.e. rough technological planning of aggregate production activities) and production planning (i.e. resource loading). The resource loading function depends on the processing times, number of processing steps and the uncertainty of the processing times.

3.3 Order planning at SEWACO

From a queuing theory perspective, a repair shop at SEWACO can be regarded as a multi-server work centre. The SEWACO consists of 75 repair shops. The number of operators within a shop varies between 1 and 12, with an average of 4 operators. Each repair shop can be characterized by its *demand characteristics*, expressed in terms of the *mixture of demand classes* and the *work content predictability.* The latter characteristic reflects to what extent the planners are capable of estimating the repair times. This predictability varies for the different repair shops, because each technical system has its own complexities. Therefore the uncertainty of the repair times varies from shop to shop. For a certain repair shop, the work content predictability is defined as the correlation between the ex-ante estimations and the ex-post realizations of the repair times. The closer this value is to one, the less uncertainty in the repair shops.

The demand characteristics measured at SMO level are listed in Table 3.1. Because each SMO is supported by its own Production Support department, the planning procedures are at least consistent within each SMO.

SMO	Fraction preventive (in hours)	Fraction corrective (in hours)	Fraction repairable (in hours)	Work content predictability
AWS	0.60	0.21	0.19	0.89
UWS	0.51	0.27	0.22	0.94
C3	0.49	0.30	0.21	0.87
SW	0.61	0.12	0.27	0.97

Table 3.1. Process characteristics of the various SMOs based on all projects with due dates in 1998.

Referring to the earlier mentioned work by Crawford *et al.* (1999), the role of the planners in these four Production Support departments can mainly be typified as an information channel and problem escalator. The due dates are primarily determined by the customer (i.e. the Navy). The primary function of the planners is to advance and postpone projects between their earliest start date and due date to identify future capacity problems. The capacity problems may either concern capacity shortages or capacity surpluses. In case a period of capacity shortage is detected, the planners should escalate the problem by notifying the concerned SMO manager. Subsequently, it is this manager’s task to decide whether or not to hire extra personnel or subcontract part of the current workload. In case a period of capacity surplus is detected, the planner’s task is to advance the maintenance of repairables, if possible. If a project is internally processed, the actual sequencing decisions of the different projects and jobs within one repair shop are left to the repair shop level, without interference of the planners.

Process analyses and interviews did reveal some interesting facts about the operating procedures used by the planners. The planners indeed understand that they should use some kind of workload information for their practice. Moreover, in their opinion they do a good job in creating an apparently balanced workload. At the medium term a (at least in their eyes) reassuring workload is created: the requirements for capacity do not exceed the availabilities and enough projects are available to keep the operators busy in the near future. As a result, none or hardly any projects are recommended for subcontracting. However, on the short term a totally different picture exists. The capacity loading has become highly irregular and tardy projects are the rule rather than the exception. At this stage of the planning process, because of the short term volume inflexibility, no control decisions are available for the planners in order to solve capacity shortages. It turned out that the planners are not aware that this short term situation is (partly) based on their medium term behavior. An often heard, although not justified, sigh from the planners: “Fortunately, within a few months, it will be less hectic in the shops.” This phenomenon is shown in Figure 3.1. It should be noted that this contrast between short and medium term is not exclusive for this real-life situation. It is common practice in industry.

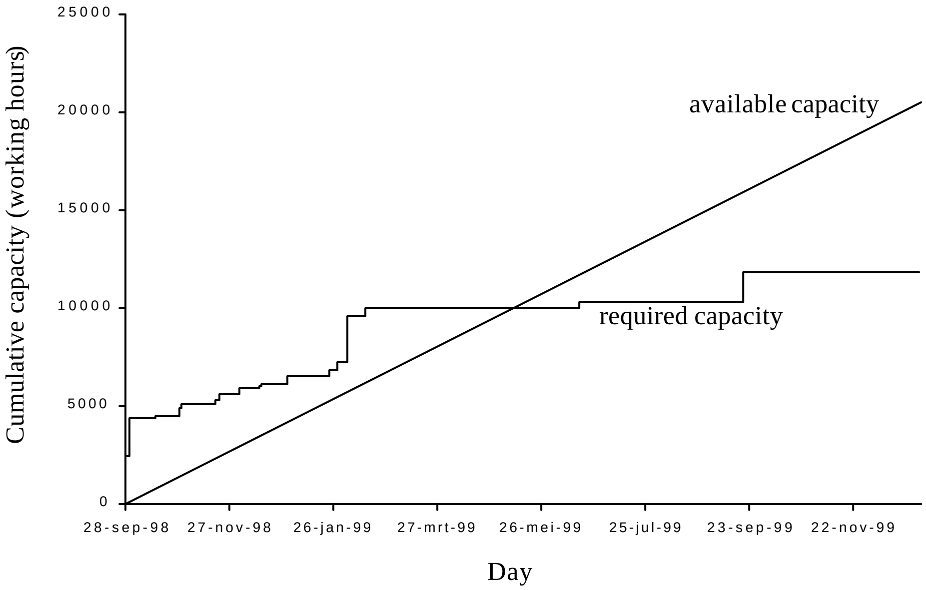


Figure 3.1. Expected available capacity and expected required capacity due to arrived projects for repair shop 640E, seen at September 28, 1998.

Furthermore, it followed from the database analyses that subcontracting is hardly used in the current control system. Conversely, if this subcontracting possibility would have been used properly, (i.e. at the right time warnings had been generated and SMO managers did subcontract part of the workload), the number of short term problems could have been reduced, the performance could have been improved and the processes would have been controlled.

Summarizing, we conclude that apparently at the medium term inadequate process models are used. Otherwise, there would not have been that many short term problems. Clearly, the perception of the planners does not coincide with the real-life processes. In order to achieve performance improvement, the perception of the planners and the planning policies - which depend upon the perception - have to be changed. Relevant information on project and shop characteristics need to be included to give subcontracting notifications at the right moment.

3.4 Selection of case relevant characteristics

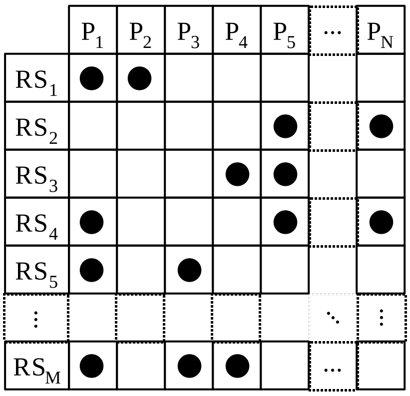
We already mentioned that a typical maintenance project at SEWACO consists of a number of operations where each operation has to be performed at one and only one of the 75 repair shops. Therefore a repair project can be considered as an assembly order with components that require only one operation. Each operation has to take place in a predetermined repair shop. The assembly nature is induced by the requirement that all maintenance jobs of a project have to be completed before the scheduled departure of the ship. Maintenance projects may differ in terms of the number of shops involved, the number of jobs to be performed, lead times, and in terms of the job processing times per shop. Therefore, referring to the discussion of the relevant literature in Section 3.2 - it seems obvious that the following factors should be taken into account for managing the due date performance in the repair shops:

- number of parallel repair shops involved in a project;
- number of jobs per project;
- lead time;
- processing times estimations;
- workloads in the shops.

Remark that due to the typical structure of the projects (i.e. a project is divided over the various repair shops with no or hardly any precedence relations both *between* and *within* the repair shops), we do not have to consider these relations.

Regarding the latter factor, workloads in the shops, the results in literature have been obtained for simulated shop models, consisting of a number of single machine work centres with equal job arrival rate and job processing characteristics. This is quite different from the structure of the maintenance organization where the number of operators per repair shop differs from 1 to 12, the average processing times per shop vary from 9.1 to 102.5 working hours and the mixture of the three demand classes also varies (see Table 3.1). It is known from queuing theory that shop waiting times depend on the number of servers, processing times and utilization rates for the different demand classes. We therefore may expect that, apart from variation in the overall shop workload over time, also differences in shop characteristics will be of importance for managing the due date performance. Thus we postulate that, for managing the due date performance of a maintenance project, the workload and shop characteristics must be taken into account.

Figure 3.2 depicts a project-process matrix. This shows for a given set of projects $P_1,\ldots,P_N$ which of the repair shops $RS_1,\ldots,RS_K$ are involved in the execution of the projects. From this figure we see that a project competes for capacity of one or more repair shops and conversely, each repair shop has several projects competing for its capacity. Consequently, we distinguish *project* and *process* characteristics. Project characteristics are project-related (e.g. number of jobs, processing times), whereas process characteristics are repair shop related (mixture of demand classes, work content predictability). To investigate the remaining relationship of the process and project characteristics on the delivery performance we have constructed a large database containing the 549 preventive maintenance projects which have been completed during the year 1998. We only consider these projects for the following reason. Corrective maintenance projects immediately ask for capacity and therefore can seriously disturb the progress of preventive maintenance projects. To ensure a timely delivery of the preventive maintenance projects, sufficient slack should be assigned to them. If the slack is not sufficiently large, this can be extended by for example subcontracting. Repairables are not planned in advance; these can be used to smooth the workload. This is possible because of their (relatively) large waiting time allowance.



	P_1	P_2	P_3	P_4	P_5	⋯	P_N
RS_1	●	●					
RS_2					●		●
RS_3				●	●		
RS_4	●				●		●
RS_5	●		●				
⋮						⋱	⋮
RS_M	●		●	●		⋯	

Figure 3.2. An example of a project-process matrix.

Based on the scientific literature we have selected the following variables as relevant project characteristics:

$X_{1,i}$ estimated processing time for the entire project i (in working hours);

$X_{2,i}$ total number of jobs exploding from project i;

$X_{3,i}$ total number of repair shops involved in project i;

$X_{4,i}$ allowed lead time for project i (in calendar days);

where $i = 1,...,549$.

Characteristic	Minimum value	Maximum value	Mean	Standard deviation
X_1	1	9321	179.7	637.2
X_2	1	146	4.9	13.7
X_3	1	15	1.6	1.3
X_4	1	641	154.0	129.9
X_5	0	1	0.28	0.4
X_6	0	1	0.31	0.5
X_7	0	1	0.15	0.4
X_8	0	1	0.35	0.5

Table 3.2. Characteristics of the 549 projects.

Regarding the process variables, we have observed that these are somewhat perpendicular to the project characteristics: each project consumes resource capacity from one or more repair shops (see Figure 3.2). Still we want to know what their contributions are to the delivery performance. Instead of assigning a variable number (depending of the number of repair shops involved in a particular project) of explanatory process variables to the project, we describe the process characteristics at project level by introducing dummy variables. Because each SMO is supported by its own group of planners, it is likely that for all repair shops in one SMO, the same process knowledge and perception are used for coping with the shop characteristics. As a result, we could introduce the following dummy variables:

$X_{5,i}$ dummy variable which takes the value 1 if AWS is involved in project i and 0 otherwise;

$X_{6,i}$ dummy variable which takes the value 1 if UWS is involved in project i and 0 otherwise;

$X_{7,i}$ dummy variable which takes the value 1 if C3 is involved in project i and 0 otherwise;

$X_{8,i}$ dummy variable which takes the value 1 if SW is involved in project i and 0 otherwise.

Note that all the values of the process and project characteristics can be observed a priori, that is before the project has even been released to the shop. And of course, finally we have the response variable Y_i which takes the value 1 if project i has been finished in time and 0 otherwise. Obviously, this response variable cannot be measured until project i is completed or its due date has been exceeded. Information of the values of the various project and process characteristics is summarized in Table 3.2.

It should be noted that the selection of including these response and explanatory variables in the model is not a unique one. For example, to include some information about the due date exceeding the response variable Y_i could be set to one in case the exceeding is less than 10% of the lead time. Furthermore, to investigate to what extent the variance in the performance can be addressed to only the project or only the process characteristics, additional logistic regression models could be estimated which include only part of the project and process characteristics. As an illustration we use the above defined response and explanatory variables.

3.5 Diagnosing the causes of poor performance

In this section we show the method we have used for finding the causes of poor performance and redefine our conjecture regarding the poor performance of the current control situation in terms of the model that we are going to use. The basic principle is to impose as few assumptions as possible on the type of the relation between the (dependent) delivery performance and the independent characteristics: we want to let the data speak for themselves. The maintenance organization is regarded as a black box: the values of the input and output variables are known. However, we cannot exactly know how the performance is related to the input variables, given the current way of controlling the processes. In such a case logistic regression is an excellent method to carefully and objectively investigate this relationship. The most important benefit of this method is that logistic regression models do not intend to represent the exact relationship between the explanatory and response variables. Instead, it focusses on estimating the probability of the occurrence of a certain outcome (e.g. "having success" or "being delivered in time") as a function of the known explanatory variables, $X_1, \ldots, X_8$, in the following way:

$$P\left(success \middle| X_1, \ldots, X_8\right) = \frac{e^{\hat{b}_0 + \hat{b}_1 X_1 + \ldots + \hat{b}_8 X_8}}{1 + e^{\hat{b}_0 + \hat{b}_1 X_1 + \ldots + \hat{b}_8 X_8}}, \qquad (3.1)$$

where the coefficients $\hat{b}_0, \ldots, \hat{b}_8$ are to be estimated and thus depend upon all the entries in the database. The parameter $\hat{b}_0$ plays the role of an intercept, whereas the values of $\hat{b}_1, \ldots, \hat{b}_8$ are associated with the project and process characteristics.

The purpose of logistic regression is as any other model-building technique in statistics: "*to find the best fitting and most parsimoniously, yet biologically reasonable model to describe the relationship between an outcome and a set of independent variables*" (Hosmer and Lemeshow, 1989). The major difference between ordinary linear regression and logistic regression is that in the latter case, the response variable should be dichotomous: it can only take two different values. Hence, because the response variable Y_i is a binary variable, logistic regression can now be applied to our data set to derive the coefficients $\hat{b}_0,\ldots,\hat{b}_8$. Note that the relation (3.1) always produces estimated probabilities that are between 0 and 1 and which thus are meaningful and interpretable.

During the past decades, logistic regression has become a valuable and standard method in many scientific fields, like social and health sciences. Its use in Operations Management is still limited. However, researchers start to see that it can indeed be a valuable tool and logistic regression is also gaining ground in operations management. For example, Tansey *et al*. (1996) demonstrate the use of it in management research.

Before actually applying the logistic regression analysis, we first transform our stated conjecture about the inadequate use of process and project characteristics into relevant terms of the regression analysis. Consider a production situation which is under control. In such a situation, the availability of resource capacity will be adjusted to the project and process characteristics in such a way that the probability of completing the projects before the due date is in accordance with some predefined and satisfactory norm. The latter should hold regardless the values of the individual project and process characteristics, whether a project is complex or not. When we apply logistic regression to our data set, for j = 1,...,8 the coefficient $\hat{b}_j$ describes the influence of the j-th associated project or process characteristic on the delivery performance. In case coefficient $\hat{b}_j$ is significantly larger (resp. smaller) than zero, it follows from equation (3.1) that under the current control system, the j-th associated project or process characteristic of a project is positively (resp. negatively) related to the delivery performance. In general terms: in case the value of $\hat{b}_j$ significantly differs from zero (regardless its sign), the current control system poorly performs with respect to the j-th process or project characteristic. Note that in a controlled system in which the $\hat{b}_j$'s do not significantly differ from zero for j = 1,...,8, the a-priori estimated probability of completing a project in time equals

$$P\left(success \middle| X_1,\ldots,X_8\right)=\frac{e^{\hat{b}_0}}{1+e^{\hat{b}_0}}, \tag{3.2}$$

which clearly does not depend on the project and process characteristics.

Applying logistic regression to the 549 cases results into the set of estimated coefficients $\hat{b}_0,\ldots,\hat{b}_8$, that are shown in Table 3.3. We will not go into detail about the way in which the parameters $\hat{b}_0,\ldots,\hat{b}_8$ are estimated. We confine ourselves to the statement that the fitting procedure is based on a maximum likelihood method. In a very general sense it can be stated that this method yields values for the unknown parameters which maximize the probability of obtaining the observed set of output data. For more technical information on this topic, the reader is referred to e.g. Agresti (1990) or Hosmer and Lemeshow (1989).

Characteristic	$\hat{b}_j$	Standard error	Significance
Intercept	-0.2111	0.4329	0.6259
X_1	-0.0007	0.0003	0.0357
X_2	0.0034	0.0147	0.8192
X_3	-0.0402	0.1212	0.7405
X_4	0.0084	0.0010	0.0000
X_5	-1.2262	0.4779	0.0103
X_6	-0.9131	0.4773	0.0557
X_7	-1.0274	0.5060	0.0423
X_8	0.2710	0.5505	0.6225

Table 3.3. Estimates for the coefficients of the project and process characteristics

Before interpreting the coefficient estimates, a goodness-of-fit measure is needed to judge the quality of our predicted model. Only in case the model is well fitting, the obtained coefficients can meaningfully be interpreted. Whereas in linear regression analysis the (adjusted) R^2 is an adequate test statistic (i.e. the fraction of variance explained by the model) for goodness of fit, literature on logistic regression analysis hardly pays any attention to this statistic. The R^2 statistic turns out not to be widely useful and not that powerful and interpretable as it is in linear regression analysis. Moreover, Hagle and Mitchell (1992) remark that logistic regression analysis relatively lacks of diagnostics. However, Agresti (1990) proposes various other test statistics, but no generally accepted and generally used test statistic is available yet. Therefore, we have constructed a goodness-of-fit statistic especially for the case at hand.

We base the construction of a useful test statistic for this case on the specific characteristics of this type of control problems. Most logistic regression models estimate the probability of the presence (or absence) of a certain feature as a function of other characteristics, e.g. the presence of coronary heart disease as a function of age (Hosmer and Lemeshow, 1989). In these studies both the explanatory and the response variables are measured at the same time. In our model this is not the case. The explanatory characteristics are measured upon arrival of the project at the organization. At that stage, the dependent variable is not yet known. The completion date of a project depends upon a lot of factors and events, some of which are random: work in process, absence of the operators by illness, deviations of the estimated processing times, the possible arrival of rush orders, the quality of the planning and control, etc. The response variable (i.e. early or late delivery) can only be measured *after* the completion of a project. At the start of a project, we can only give an estimation of the *probability* of completing it in time. We use this observation to construct a suitable test statistic.

Based upon the coefficients in Table 3.3 and equation (3.1), we produce for each of the 549 projects ex-post a prediction whether or not the project will be completed in time. Let this prediction be denoted by $\hat{Y}_i$ with i = 1,..,549. It is obvious that $\hat{Y}_i$ can take any value between 0 and 1. For each project, we know whether it was completed in time. This is denoted by the binary variable Y_i. Subsequently, the obtained doublets $(Y_i, \hat{Y}_i)$ have been plotted in a *classification plot*. For this purpose, we divided the interval [0;1) into T disjoint sub-intervals with equal lengths such that the union of all these sub-intervals forms the interval [0;1), that is: $[0;T^{-1})$, $[T^{-1};2T^{-1})$, … , $[(T-1)T^{-1};1)$. Each project can be assigned to exactly one of the T intervals such that its prediction $\hat{Y}_i$ falls into that particular interval. Now, we are ready to construct our goodness of fit measure. For t = 1,...,T, let f_t be the number of projects with its prediction in the t-th interval and let f_t^0 (resp. f_t^1) be the number of projects with its prediction in this t-th interval and which have been delivered too late (resp. in time). Clearly

$$\sum_{t=1}^{T} f_t^0 + \sum_{t=1}^{T} f_t^1 = \sum_{t=1}^{T} f_t = 549 \tag{3.3}$$

Consider the t-th interval, that is $[(t-1)T^{-1}; tT^{-1})$. In case the estimated model fits well, it is to be expected that a fraction q_t of all these f_t projects (which have a predicted probability $\hat{Y}_i$ of being delivered in time that is between $(t-1)T^{-1}$ and tT^{-1}), indeed is delivered in time. Of course, this fraction q_t is an increasing function of t. After all, in case the estimated model is well fitting, a high (resp. low) fraction of the projects with a high (resp. low) predicted probability of success should be delivered in time. Therefore, we assume:

$$q_t = \frac{(t-1)T^{-1} + tT^{-1}}{2} \tag{3.4}$$

Furthermore, we denote the actual fraction of projects with a prediction in the t-th interval and that indeed are on time, by $\hat{q}_t$, thus:

$$\hat{q}_t = \frac{f_t^1}{f_t} \tag{3.5}$$

The model is well fitting in case the deviations between the $\hat{q}_t$'s and the q_t's are sufficiently small. The following test-statistic, say γ, can be created, that represents the deviation between the $\hat{q}_t$'s and the q_t's:

$$g = \sum_{\substack{t=1 \\ f_t > 0}}^{T} \left(q_t - \hat{q}_t\right)^2 \tag{3.6}$$

The actual distribution of the test-statistic (3.6) depends on T and the values of the different f_t's. The estimated model is said to be well fitting in case the test-statistic γ is smaller than $g_{.95}$, with $g_{.95}$ the 95*th* percentile of the random variable γ in (3.6).

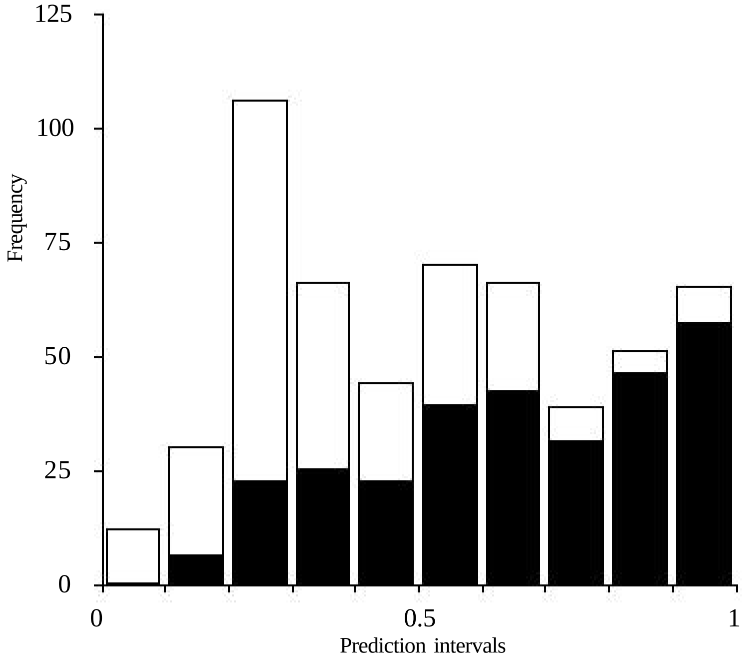


Figure 3.3. Classification plot for the logistic regression with f_t^0 (resp. f_t^1) in white (resp. black).

To actually apply the goodness-of-fit test in this case, first an appropriate value for T should be chosen. In this analysis we use the “power of 2 rule”: in case of N observations, it should hold that $2^{T-1} < N \leq 2^T$ (Ryan, 1989). Substituting N = 549 gives T = 10 intervals. The resulting classification plot for the logistic regression is shown in Figure 3.3. As we may expect for a well fitting model, at the left and right hand side of this figure only a small fraction has wrongly been predicted by the model, whereas in the middle of the plot about half of the projects have correctly been predicted. More formally, if we test the hypothesis that the estimated model fits the relation between the dependent and the independent variables at the 0.05 level of significance, we have to reject it in case the test statistic (3.6), that is γ = 0.020, exceeds the value of $g_{.95}$. Because the derivation of the exact distribution function of γ is intractable, we derive this distribution function by simulation: 1,000,000 drawings of γ are conducted in the following way. Because we know the values of the different

f_t's and their associated probabilities q_t, for each of the 1,000,000 drawings, we draw the T values for the f_t^1's to construct 1,000,000 values of γ. Note that f_t^1 has a binomial distribution function with parameters f_t and q_t. The resulting estimated probability density function is depicted in Figure 3.4. Furthermore, the value of $g_{.95}$ can be estimated from these data. It follows that $\hat{g}_{.95} = 0.066$ which is larger than 0.020. Therefore, our estimated regression model passes the goodness-of-fit test. Clearly, the fitted model is an acceptable (black box) representation of the effect of the current planning and control procedures in the maintenance organization.

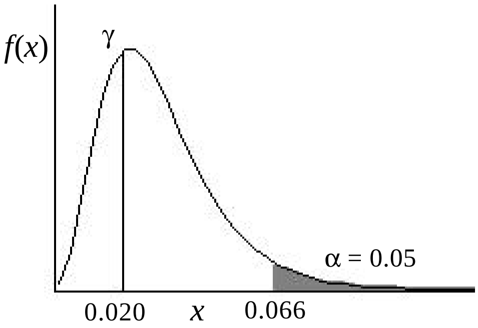


Figure 3.4. Estimated probability density function of the γ-statistic, applied to the project and process information.

To show that the choice of the value T does not have a large impact on the fitting results, the goodness-of-fit test has also be conducted for different values of T, that is $T = 5, 15, 20$. The results, which are summarized in Table 3.4, show that also for these values of T, the goodness-of-test also successfully is passed.

T	g	$\hat{g}_{.95}$
5	0.004	0.019
10	0.020	0.066
15	0.063	0.141
20	0.086	0.330

Table 3.4. Goodness-of-fit statistics for various values of *T*.

3.6 Conclusions

As a final step, the emphasis shifts to the interpretation of the estimates of the coefficients in the logistic regression model. Now that we know that the model fits, the next question is: what do the values of the coefficient mean for the planners in the maintenance organization if they want to improve the delivery performance? Before answering this question, a closer look should be taken at the type of relationship between the explanatory and dependent variables. This relationship is not only *non-linear* (the change in the dependent variable associated with a one-unit change in an independent variable depends on the value of the independent variable) but also *non-additive* (the change in the dependent variable associated with a one-unit change in an independent variable depends on the value of the other independent variables). As a result the individual values of the estimates $\hat{b}_j$ are not that easy to interpret. However, referring to the beginning of this section, our main interest is whether the estimates significantly differ from zero or not. And if so, whether these are smaller or larger than zero. The relation (3.1) tells us that the probability of having success is an increasing function of the term $\sum_{j=1}^{8} \hat{b}_j X_j$. Therefore, if we consider a 0.05 level of significance, the following conclusions can be drawn (based on the coefficient estimates with their levels of significance displayed in Table 3.3).

- The estimated processing time is negatively related ($\hat{b}_1 = -0.0007$) to the probability of delivering in time: the larger the estimated processing time, the smaller the probability of completing the project within the allowed lead time.
- The lead time is positively related ($\hat{b}_4 = 0.0084$) to the probability of delivering in time: the larger the assigned lead time, the larger is this probability.
- In spite of the fact that the estimate $\hat{b}_2$ (associated with the number of jobs in a project) does not significantly differ from zero, this characteristic is highly correlated with the estimated processing time (correlation is 0.743). Due to this multi-collinearity this estimate becomes significantly smaller than zero in case the variable X_1 is omitted in the logistic regression.
- The number of involved repair shops seems to be used correctly ($\hat{b}_3 = -0.0402$) in the current planning and control. However, if we take a closer look at the

coefficient estimates for the process characteristics we see that the number of repair shop indeed is related to the dependent variable in the following manner.

- Two out of the four dummy process variables (concerning AWS, C3) are negatively related to delivery performance. Besides, the significance of the dummy variable concerning UWS is not that high that the conclusion of not significantly differing from zero is justified. This means that it seems that the planners from the Specific Workshops correctly use the process characteristics, unlike their colleagues at the other SMOs. This is in accordance with the expectations that can be derived from Table 3.1: SW has not only the highest values for the predictability of the repair times and the fractions of preventive maintenance and maintenance of repairables, but also the smallest value for the fraction of corrective maintenance projects. Clearly, the planning policies in use are valid for environments with little (or no) uncertainty (in this case SW), but lack applicability in the SMOs with more uncertainty and dynamics. Therefore, the control structure should include the information about uncertainty to get a controlled performance.
- Furthermore, observing the coefficients it turns out that the more SMOs are involved in a project (i.e. the more dummy process variables are equal to one), the smaller the probability of completing in time. This is due to the negative values of the coefficients of the process characteristics.

We now have carried out the analysis to what extent the relevant process and project characteristics are adequately being used by the planners. Their behavior is described by equation (3.1) together with the estimated coefficients in Table 3.3. From these coefficients we learn that a control policy should include knowledge about the (uncertainty of) processing times, lead times, the number of shops involved in the various projects on hand and the mixture of the demand classes in the repair shops.

4

Hierarchical Framework for Control

4.1 Introduction

In the previous chapters, we have described the environment in which multi-project repair shops are operating. We discussed the demand and resource characteristics as well as the processes in the shops. With regard to the SEWACO case, process analyses and interviews did reveal some interesting facts about the operating procedures used by the planners. The planners indeed understand that they should use some kind of workload information for their practice. Moreover, in their opinion they adequately use workload information. Clearly: at the medium term an (*apparently*) balanced workload is created. However, at the short term a totally different situation exists. The capacity load has become highly irregular and tardy projects are the rule rather than the exception. In the previous chapter we have explained this phenomenon by the fact that not all relevant project and process characteristics are adequately used. In this chapter we use the knowledge about these characteristics, the processes and the available literature on control structures to structure the production control problem and determine a suitable control structure.

Research on control structures goes back to the early days of Operations Research. From the late 1950s until the early 1970s, several attempts in Operations Research have been made to solve complex capacity planning problems for several types of production situations using one large mathematical model. Work on this type of models has been carried out by e.g., Manne (1958), Dzielinski *et al.* (1963), Dzielinski & Gomory (1965) and Lasdon & Terjung (1971). This often resulted in models that had overwhelming data requirements and - as these models are based on forecasts for every single item to produce - extremely large forecast errors. To overcome these problems, several researchers came up with hierarchical approaches of production and capacity planning. These are discussed in more detail in the next section.

4.2 Literature on hierarchical planning and control

After being confronted with the difficulties in controlling real-life situations by large-scale (deterministic) models, researchers introduced another approach of production planning: the hierarchical approach. In this hierarchical approach, two schools can be distinguished. The first generation of hierarchical models is based on *model decomposition* and introduced by Hax, Meal and Bitran (see the work of Hax & Meal 1975, Bitran & Hax 1977, Hax 1978, Bitran, Haas & Hax 1981). Their model decomposition approach is based on the idea that some models are too large to efficiently solve it in one iteration. To overcome this problem, the model is decomposed into different hierarchical levels. According to Bitran & Hax (1977), in a hierarchical framework "*higher level decisions impose restrictions on lower level actions, and lower level decisions provide the necessary feedback to reevaluate higher level actions*". In each iteration, the problem at each decomposition level is solved and gives restrictions for the next lower level. The three levels of product aggregation that are considered are (i) aggregate plans for product types for a number of time periods, (ii) disaggregation of aggregate plans into production quantities for each product family and (iii) disaggregation to production quantities per item per time period. Although the complexity has been reduced, still at the time of solving the production planning models, detailed information should be known. However, it may occur that this information is not available at the same time. As an answer to this inadequacy, the second generation of hierarchical control structures has been introduced and advocated by e.g. Holt *et al.* (1960), Bertrand & Wortmann (1981), Dempster *et al.* (1981), Meal (1984), Bertrand *et al.* (1990) and Schneeweiss (1999). Similar to the first school, these researchers also advocate the decomposition of control problems into smaller sub-problems to reduce complexity. However, they proceed with a fundamentally different view: their approach is based on a *problem decomposition*. Production planning and control is considered as a hierarchy of decision problems with decision variables that have different effectuation times (i.e. the time elapsing between the time at which the decision is made and the time at which the impact on the system becomes effective). Consequently, in case the different decisions are ranked in order of effectuation time, a decision level creates conditions for the lower levels. These conditions are usually not equal to the decision itself, since the outcome of a decision may differ due to the dynamics and uncertainty of the environment.

For several complex real-life production situations that have some similarities with the maintenance situation, hierarchical control structures have successfully been designed and discussed in literature. In these studies, the problem decomposition approach is used, based on the fact that different decisions have different effectuation times. For instance, Schneeweiss & Schröder (1992) discuss a hierarchical planning and scheduling framework for the control of the repair shops at the Deutsche Lufthansa. This company, which only carries out repair-by-replacement, aims at minimizing the amount of capital tied up in repairables, whilst maintaining a certain service level. Key decisions in this control problem are the purchase of repairables and the sequencing of the repairables in the different repair shops. Clearly, the first decision has a longer effectuation time than the second. Consequently, in this framework a medium and short term model are distinguished. The first (medium term) model determines the optimal provision of repairables that is necessary to achieve the target service level. This model does not use the current state of the repair shops as an input variable for the decisions, but is based on demand and repair time characteristics. The second (short term) model is a real time model for shop floor control that guarantees the overall requested service level by using priority rules. Bertrand & Muntslag (1993) describe the role of MRP II planning systems in engineer-to-order firms. They conclude that due to the important role of the customer order, the customer-specific product specifications and the uncertainty in both product and production, MRP II systems are not suitable for control. Instead, they propose a hierarchical production control framework, which better suits these specific characteristics. This framework is based on the finding that the decision functions (i) customer order acceptance and due date assignment, (ii) sub-order assignment and outsourcing, (iii) work order release and (iv) work sequencing have different effectuation times. Also Raaymakers (1999) discusses the topic of hierarchical production control. She studies the problem of production control in batch process industries and introduces a control structure and distinguishes three decision functions based on different effectuation times: (i) capacity adjustment, (ii) order acceptance and capacity loading and (iii) scheduling. The capacity adjustment function has the longest effectuation time, while at the medium time, orders arrive and are accepted (or not) and, if so, assigned to a specific planning period. The resource allocation has the shortest effectuation time and deals with the operational use of the resources. Subsequently, aggregate models are developed to support the order acceptance and capacity loading function.

Keeping this in mind and observing the complexities of the planning and control problems in the maintenance organization, we can conclude that a (problem decomposition approach based) hierarchical control structure may be a very suitable type of control structure for the maintenance organizations. From these researches we adopt the following principles that are assumed to be relevant for our control problem. First, because different decisions have different effectuation times, some decisions need to be taken based upon assumptions regarding the future (independent of the current state of the system), whereas other decisions may be based upon the current state of the system. Secondly, if we assume that in case of two decision levels, the one having the longest effectuation time is higher in the hierarchical control structure, then decisions taken at a certain level create conditions for the lower decision levels. Summarizing, we state that a control structure should identify the main decision functions, their goals, their interrelationships and the hierarchical positions to which these functions should be assigned in order to enable us to match the actual output with the required output. Now that we have chosen for a hierarchical production control structure and discussed the main body of relevant work on these types of structures, we have come to the stage of designing a framework for control for the maintenance organizations under scrutiny.

4.3 Objective

The objective of the maintenance organization is to ensure that the operational availability of the technical systems is (as good as possible) in accordance with the production schemes. This can be translated into claims for the performance for the three different demand classes. At first, corrective maintenance projects need to be completed as soon as possible after their arrival. Secondly, a sufficiently high fraction of the preventive maintenance projects must be completed before their due dates and thirdly, the out-of-stock probability of ready for use repairables in the depot may not exceed a certain prescribed level.

On the other hand, the organization is faced with budget restrictions, which should be satisfied while maximizing the operational availability of the technical systems. These budget restrictions concern upper limits on the costs of total internal and external capacity. Projects can be executed by internal or external capacity. However, the size of the internal workforce is fixed. To what degree external resources are utilized depends on the question whether or not subcontracting is used. Therefore, the costs of total capacity over a given time horizon is equal to the fixed (salary) costs of the internal capacity (regardless their degree of utilization) plus the costs of the actually used external capacity (i.e. the subcontracted projects or hiring of extra personnel).

This observation is used to design a four level production control structure. Figure 4.1 depicts these levels and their mutual interrelationships. The following six different decisions are divided over these four levels.

- capacity and repairables determination;
- class transition;
- job release;
- subcontracting;
- pulling repairables;
- job-operator allocation.

The explicit position of the decisions and their impact is shown in Figure 4.2. In the remainder of this chapter we discuss this structure in more detail and point out which topics need further research.

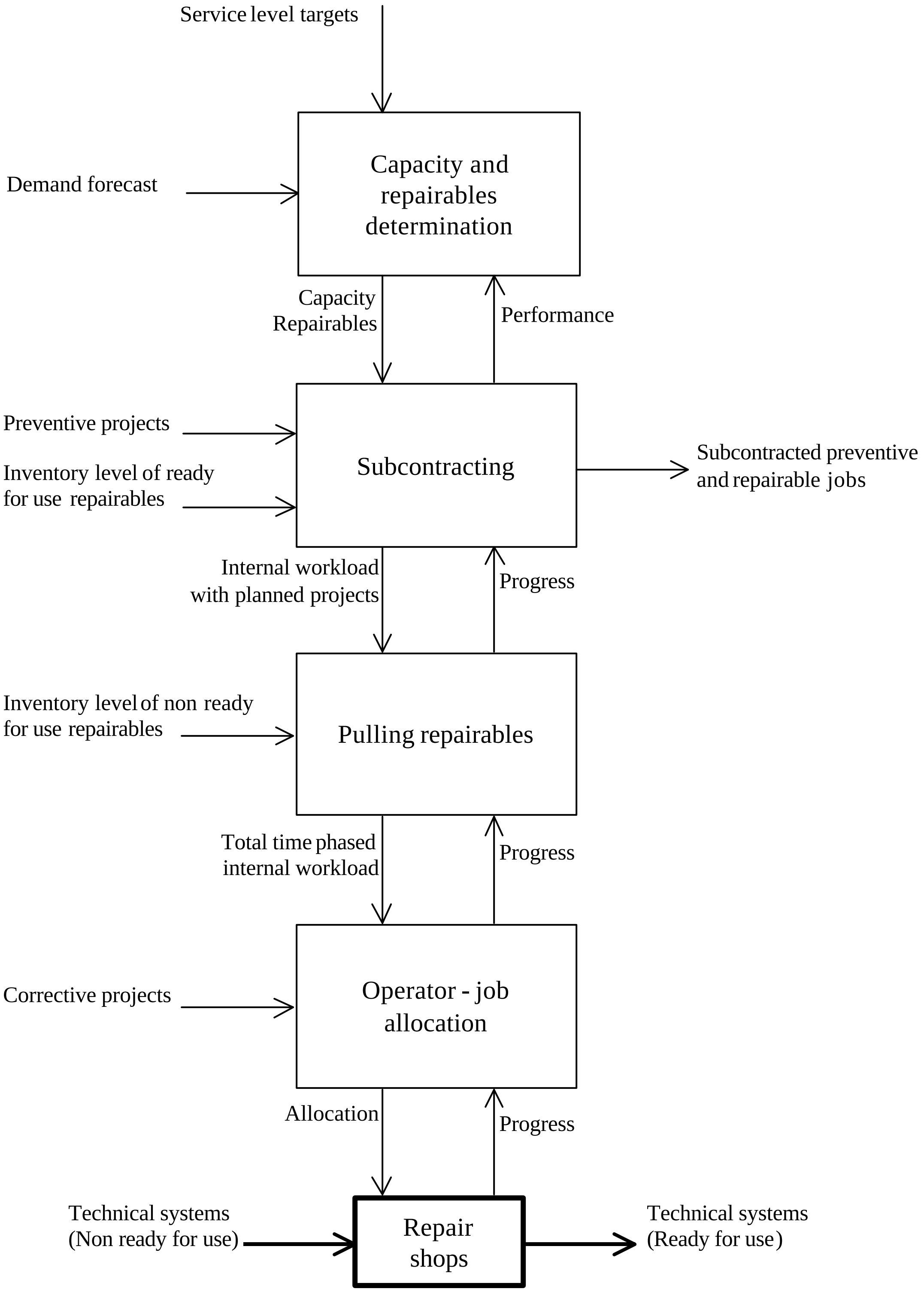


Figure 4.1. Hierarchical control structure for the maintenance organizations with the four different decision levels and their mutual relationships.

4.4 Capacity and repairables determination

The role of this control level is to create such conditions that one will not meet problems at medium term and short term level, whilst achieving the stated targets. Decisions at this level are independent of the maintenance jobs currently on hand. It concerns the long term matching of the required resource capacities with the available capacity to meet the stated goals. The following decisions can be distinguished at this level. At first, the maintenance and production schemes need to be constructed. These schemes depict when the technical systems need to be used for production and when preventive maintenance should be carried out. Obviously, this decision should be taken in cooperation with the production facility. Research on this subject has been carried out by e.g. Van Dijkhuizen (1998). Secondly, for each technical system it needs to be determined whether it should normally be internally or externally maintained. For example, because the organization does not have the knowledge in-house for certain activities, it may be cheaper to have some systems always maintained by other organizations than hiring and training extra personnel. The latter, so-called strategic outsourcing, has to be arranged in long term contracts. Research on this topic has been done by e.g., Venkatesan (1992), Gupta & Zhender (1994) and Embleton & Wright (1998). Thirdly, once it has been decided what technical systems are maintained internally, the capacity levels should be geared to the demand forecasts. We already postulated that in environments that face a lot of uncertainty and have no (or little) short term volume flexibility in capacity levels, the decision functions subcontracting and pulling repairables should be used to achieve both a satisfactory performance and a satisfactory high utilization rate of the internal capacity resources. Consequently, the internal capacity should be smaller than the total demand for maintenance arising from the non-strategic outsourced technical systems. De Boer (1998) pays attention to the subcontracting decision in maintenance project management. However, he assumes deterministic repair times. To the best of our knowledge, hardly any research is available about the balancing of internal capacity and possibilities for external capacities subcontracting on a job-by-job basis in a stochastic environment. Finally, this level is concerned with the determination of the total number of repairables in the entire system (fleet, depot and maintenance). Studies of Schrady (1967), Sherbrooke (1968) and Muckstadt (1973), have shown that the total number of repairables positively influences the time slack of recovering failed repairables. The required number of repairables depends among others upon target service levels, failure rates and repair times.

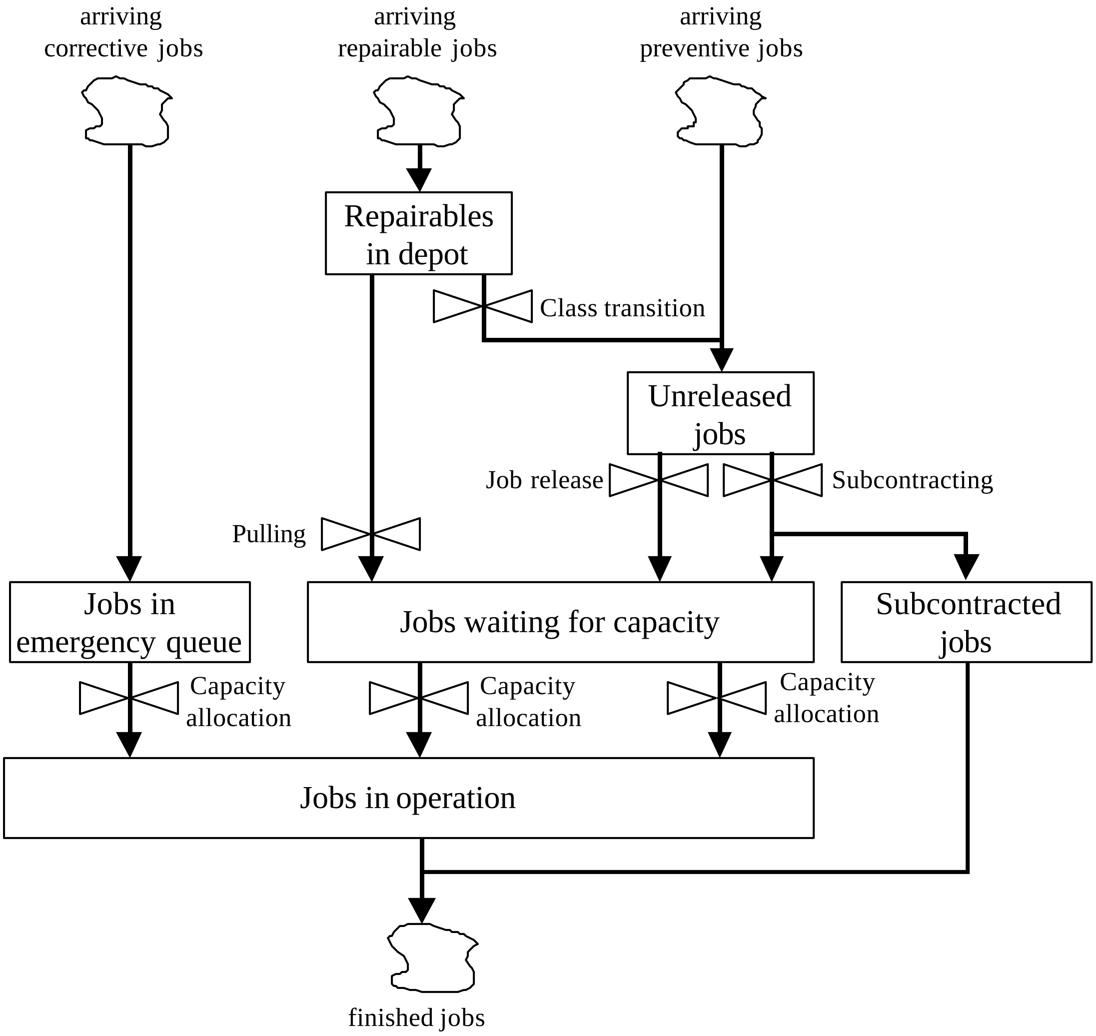


Figure 4.2. Basic control decision functions in relation to the job flow.

4.5 Subcontracting

The decisions taken at this level depend on the current state of the repair shops. This state can be described by the projects on hand and the available knowledge about planned, but not yet arrived, projects. This medium term decision level has a planning horizon of a few months, which is in the order of the subcontracting lead times of complex technical systems. In the SEWACO case, this decision level is the lowest decision level at which capacity volume flexibility can be created. Therefore it is the task of the planners at this decision level to match the demand for resources and the resource availability within the planning horizon, such that a satisfactory delivery performance can be achieved by the short term decisions (i.e. the job-operator allocation), taking into account the uncertainty about the processing times and arrival times of the projects. It should be emphasized that the subcontracting decision for a project should not necessarily has be taken *before* or *at* the start of the project. In case the subcontracting lead time is smaller than the time between the start date and due date of the project (which often is the case), this decision can be taken *during* the execution of a project, while part of its jobs are already released or even completed.

The subcontracting decision has to be made in three steps. The first step is material oriented. The inventory levels of the ready for use repairables need to be evaluated to decide whether or not these levels are sufficiently high to fulfill future demand in the planning horizon with a prescribed probability (e.g. 90%). This decision can for example be based on the work of Schneeweiss & Schröder (1992), in which they express for each repairable item the probability of being out-of-stock, as a function of, among others things, the current inventory levels and failure rates. With respect to our class of maintenance organizations, if this method foresees shortages, failed repairables need to receive a due date which is in the planning horizon: these have become ‘planned’. This is called the *class transition decision*: the repairable project can now be treated as a planned preventive project. So, material requirements are transformed into capacity requirements. The second step is to accumulate all capacity requirements arising from the preventive projects and planned repairable projects with due dates in the planning horizon and which currently are not yet *released* to the shops. One has to account for the fact that besides the already known preventive and repairable maintenance projects, their progress may be disturbed by the arrival of unplanned (and high priority) corrective maintenance projects: a certain time slack needs to be assigned to the projects. Furthermore, the uncertainty of the processing times of the projects should also be reflected in this slack. It should be noted that due to the project structure of the demand, monitoring the progress of a project, means monitoring the progress of the jobs of the project in all the repair shops involved in the project. From literature and also from the observations in the SEWACO case, we know that in case projects have an assembly structure, *completion delays* occur. A completion delay occurs whenever a completed sub-project in a repair shop is required to wait for the completion of a parallel sub-project in another repair shop and appears at the end of a project. A commonly used method for controlling the completion dates of projects that are split up over various parallel (repair) shops is to assign internal due dates to each sub-project, which lie before the external due date. It will be clear that the difference between an internal and external due dates increases if the number of parallel shops increases, i.e. more time slack is required. Finally, the third step (if necessary), is the creation of capacity slack. *Subcontracting* (or hiring extra personnel) can create this slack: it decreases the workload, which has to be internally processed within a fixed time period. As a result, the probability of delivering the project in time increases.

Very little research has been carried out on the integration of subcontracting at a tactical and operational level. Levy & Tayi (1989) study scheduling strategies in a client-contractor relation in a project environment where the client at each stage of the project is allowed to reject the current work. Furthermore, they investigate the role of subcontracting in project management and derive subcontracting policies. Kamien & Li (1990) study a production planning model that explicitly considers subcontracting as a planning tool, in a market at which two production facilities are operating. They conclude that both players can reduce the variability in both production and inventory. Furthermore, Bertrand & Sridharan (1993) evaluate different subcontracting policies for low volume component manufacturing. They conclude that policies which include the workload in the shop outperform (in terms of utilization rate and delivery performance) others, which do not seek to use this information. Therefore, the subcontracting decision should be based on the workload factors that are currently not adequately being used (see Chapter 3).

Besides the technical points about the functioning of subcontracting, some practical points of caution need to be discussed. At first, there are some circumstances in which subcontracting may not be possible at all. For example due to technological, financial, time or political reasons, some technical systems may not be candidates for subcontracting. This emphasizes the need for a clever *job release function*: before the latest subcontracting decision of a project, only jobs need to be released which can only be processed internally (if possible). Consequently, at the time of the subcontracting decisions, a relatively large fraction of jobs is available as subcontracting candidates. Secondly, once it has been decided that indeed jobs are subcontracted, there is the issue of selecting the subcontractors and managing the contract once it is signed. According to Shapiro (1986) and Kamien & Li (1990), the criteria here could be quality, on-time delivery and conformance to engineering specification. Thirdly, it should be noted that within the organization, resistance may be encountered to subcontracting for psychological reasons. According to Venkatesan (1992), manufacturing managers have a strong incentive to insource production, as more parts mean more responsibility, more authority and thus bigger salaries. Moreover, he stated that "*many managers seem to take it personally that a little garage machine shop could outperform them - though an impartial observer could see why it might*". Levy & Tayi (1989) state that managers use the subcontractors to blame them for the poor performance. However, blaming them does not mean that they indeed are responsible for delays: in a questionnaire among 91 project management consultants, Icmeli-Tukel & Rom (1998) conclude that subcontractors can only be blamed for 10% of the delays in project management. These psychological factors may not be neglected in attempt to implement the subcontracting decision. Moreover, it is not unlikely that these factors also play a role in the SEWACO case. Besides the fact that the project and process characteristics are not adequately being used, the above listed factors may also have contributed to the fact that the subcontracting valve is hardly used. Anyhow, it should be clear that attention needs to be paid to these factors when implementing the subcontracting decision function.

4.6 Pulling repairables

In the discussion about the use of subcontracting it appeared that time slack is needed in order to ensure a sufficiently high probability of delivering the projects in time. Consequently, the expected lateness of the projects (deviation between due dates and completion dates) is negative, meaning that on average a project is completed a certain time period before its due date. As a result, if no countermeasures are taken, the decision makers run the risk that operators are idle during a significant fraction of the time: there is no new work. After all, subcontracting causes a decreased demand for internal resources. In the light of this discussion, it should also be stated that the discussion of the subcontracting function has taught us that subcontracting is a so-called *one-directional* control mechanism: it can only temporarily increase the total available capacity (internal plus external). Therefore, the total available capacity is always at least equal to the internal capacity. In times of capacity surplus, the capacity levels can not be decreased. The resulting idle time can be prevented if there are projects available, which have start dates in the planning horizon and due dates beyond the planning horizon. If this is not the case, non-ready-for-use repairables can be taken from the depot to be processed, which is called *pulling repairables*. This pulling decision has a certain *pulling lead time* (i.e. the time elapsing between the moment at which the pulling decision is taken and the moment at which the repairables are available in the repair shops). This lead time is made up of the time to search the non-ready-for-use repairables, the transportation time between the depot and the repair shops, the time to collect the required documents or manuals and the time for work preparation and estimating the processing times.

4.7 Job - operator allocation

The lowest decision function, the short-term control, has the shortest effectuation time and is concerned with the actual allocation of the jobs on hand to the operators. This decision comprises both the scheduling of the preventive and planned repairable jobs on hand and - in case there are none - the scheduling of the non-urgent repairable jobs. Furthermore, jobs need to be interrupted as soon as corrective projects arrive. If the number of emergency jobs in the emergency queue exceeds the number of operators, also the sequencing of emergency jobs has become a necessary decision. Consequently, the capacity allocation concerns both *sequencing* and *interruption* decisions. Surprisingly, little research has been done on maintenance scheduling in stochastic environments. However, due to the dynamics and uncertainty of maintenance, the scheduling decision is a very complex task. Paz & Leigh (1993) discuss a list-scheduling approach, in which the jobs on hand are sorted according to some priority rule. As soon as an operator finishes a job, the job with the highest priority is assigned to him or her. Due to the dynamics and uncertainty, the sequence of this list varies over time. It should be stated that the success of this approach heavily depends on the chosen priority rule. Literature on priority rules is ample and it is reviewed in Chapter 3. Summarizing, it can be stated that - considering the assembly nature of the projects - the priority rule should be based on the number of operations left in the project (Siegel 1971, Philipoom 1991) and the progress of subprojects in other repair shops (Maxwell & Mehra 1968 and Siegel 1971). Furthermore, a typical characteristic of maintenance projects is that expert knowledge is crucial for generating good schedules (“Unlike for other systems, for this technical system it is better to perform job A before job B”). Therefore, this technical knowledge (which can hardly be formalized) should also be incorporated in the job-operator allocation function.

4.8 Summary

At this stage of the research, we have identified the problems that are encountered by the maintenance organizations. Furthermore, we have described the processes within the organization and have postulated a hierarchical control structure. A literature review on the main body of work about hierarchical control structures has been carried out, for the various decision functions. This resulted in the observation that - unlike for aggregate planning and short term allocation decisions - the use of subcontracting and pulling repairables has not received considerable attention in literature. Both at the level of determining internal and external capacity and at the subcontracting level a lack of scientific knowledge exists. Before the internal and external capacity determination can be carried out, more knowledge needs to be available of how to use the subcontracting decision at the medium term level. Therefore, in this thesis, we restrict ourselves to the design of a useful tool for the subcontracting level. Once this has been developed, directions for further research at the higher levels are given.

The main problem for the planners that are concerned with the subcontracting decision is to cope with the uncertainty over the subcontracting lead time. At the time the subcontracting decision is taken, the results become not effective until the projects are returned at the end of the subcontracting lead time. Clearly, the planners have to foresee future capacity problems. In the SEWACO case, we have seen that the current medium term perception of the planners does not coincide with reality. To investigate the role of the identified project and process characteristics, we develop in the next chapter a small-scale and simplified model to research this question.

Once this role is established, we devote another chapter to make this model operational for control in large-scale and complex situations. This large-scale model should give insight in the role of the different project and process characteristics on the completion date of the project. Subsequently, a subcontracting decision function can be derived.

5

A simplified Model for Integrating Subcontracting and Internal Capacity Planning

5.1 Introduction

The previous chapter has shown that the major lack of knowledge in the presented control structure concerns the integration of the internal capacity planning and the subcontracting decision at an operational level. Furthermore, we have noticed that this knowledge is crucial for the medium term control to match the capacity requirements and availabilities. In other words, at the medium term level, the planners need to have insight in the maximum workload levels of the different repair shops such that the planned projects are completed before their due dates with a sufficiently high probability even if corrective projects are immediately carried out. It may be obvious that these maximum workload levels vary for the different repair shops, depending on the characteristics of the processes.

Therefore, in this chapter we develop a small-scale and simplified model to describe the processes within a single repair shop and to identify the relation between on the one hand the project and process characteristics and on the other hand the makespan (and thus completion date) probability distribution function. This chapter produces the principles on which the subcontracting decision should be based. We restrict ourselves in this chapter to a single multi-server repair shop, which has to carry out only one project, consisting of a given number of jobs. In the next chapter, this model will be embedded in a more generic decision rule for multiple projects at the same time in a maintenance organization consisting of multiple repair shops.

A single preventive planned maintenance project can be seen as a set of jobs that has to be completed within a certain time period. To process these jobs, a number of parallel operators (i.e. servers) is available. The project progress may be disturbed by the arrival of corrective jobs. The high priority character of these latter jobs makes that we can consider them as *emergency* jobs. Literature on steady-state throughput time distributions in multi-customer class systems is ample. However, surprising little is known about the case of a fixed horizon, in which we are interested in the probability distribution function of the throughput time of a certain given set of jobs interrupted by higher priority jobs with multiple parallel servers and random repair times. Gaver (1962), Kim & Aden (1997) and Keizers & Bertrand (1998) present some results for the single-server shop. However, for the multi-server case, no literature is available. The approach sketched in this chapter is based on Keizers *et al.* (1998). We first consider a simplified situation, in which a project consists of jobs with identically and exponentially distributed repair times. Once we have found the makespan distribution, we consider the case of non-identically and non-exponentially distributed repair times. However, it should be noted that this model is still a simplified model. We come back to this later on in this chapter.

5.2 Problem formulation

In this section, we formulate the mathematical problem. Consider an idle repair shop with c parallel servers, as depicted in Figure 5.1. Regarding this shop and the processes in the shop, we make the following assumptions:

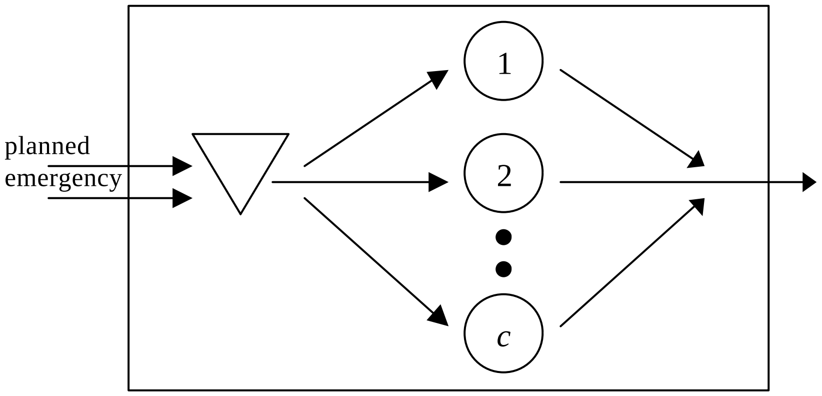


Figure 5.1. A c-server repair shop.

- at time $t = 0$, a project arrives containing n_0 planned jobs;
- because there are no precedence relations between the jobs, all n_0 jobs have identical due dates, say $t = T_D$;
- each jobs has to be processed by only one operator;
- the processing time of each planned job is exponentially distributed with mean $1/m_P$;
- emergency jobs arrive according to a Poisson process with rate λ and have exponentially distributed processing times with mean $1/m_E$;
- the emergency jobs have absolute *pre-emptive resume* priority over the planned jobs: upon arrival of an emergency job, one of the planned jobs in service is interrupted and the emergency job is immediately taken into service.
- for stability we have to require that $l < cm_E$.

Regarding the interruption process, the following comments have to be made. If the total number of emergency jobs in the system is already larger than or equal to c the emergency job is put in the emergency queue. If a server has finished an emergency job and the queue with emergency jobs is empty, a planned job in the queue is resumed from the point at which interruption took place.

If C_i is the completion time of planned job i ($i = 1,..., n_0$), the makespan of the entire project is equal to $\max\{C_i | i = 1,\ldots,n_0\}$. This makespan is determined by the processing times of the emergency and internally processed planned jobs. The state of the project can be described by the pair (n,m), with n the number of unfinished planned and m the number of unfinished emergency jobs in the shop. Let $T(n,m)$ denote the random variable 'remaining makespan' for state (n,m). An example of a realization of the makespan is shown in Figure 5.2. The problem we are facing is to calculate the probability of delivering the entire project in time, that is, with m_0 the number of emergency jobs at $t = 0$:

$$P(T(n_0,m_0)\leq T_D). \tag{5.1}$$

Once we are able to calculate this probability, we can use it to control the delivery performance of the repair shop according to the targets set by the organization. Note that at this stage, we neglect the repairables. However in case it turned out that the inventory level of repairables that are ready for use will become critical if there are not at least n_r repairables processed within the time interval $[0,T_D)$, the project with the n_0 jobs can simply be transformed into a project with $n_0' = n_0 + n_r$ jobs, without affecting the sketched approach.

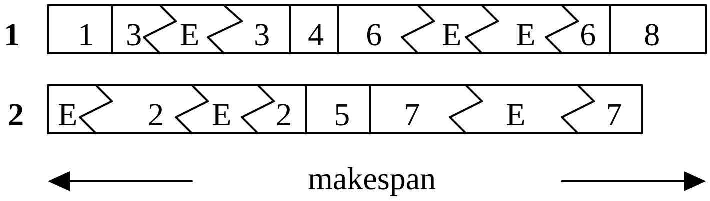


Figure 5.2. Realized makespan of a project with 8 planned jobs in a 2-server repair shop with initially 1 emergency job (E).

5.3 The distribution of the makespan

In this section we present an approach to approximate the distribution of the makespan. We first analytically derive an expression for all moments of $T(n_0,m_0)$. Then we approximate the distribution of the makespan by fitting an appropriate distribution on the first two moments.

5.3.1 The transition structure

While processing the project, the remaining number of planned jobs n decreases from n_0 to zero and the number of emergency jobs m fluctuates between zero and infinity. In state (n,m), one of the following three events may occur:

- with rate l an emergency job arrives yielding a transition towards state $(n,m+1)$;
- if there is at least one emergency job in the system, then with rate $\min(m,c)m_E$ one of the servers finishes an emergency job resulting in a transition towards state $(n,m-1)$.
- because of the pre-emptive resume interruption process by emergency orders, a planned job can only be finished if there are at c-1 emergency jobs in the system. With rate $\min(c-m,n)m_P$ one of the servers then finishes a planned job and a transition occurs towards state $(n-1,m)$.

The project is finished as soon as the system enters one of the states (0,0), (0,1),..., (0,c-1). The state-transition rate diagram is depicted in Figure 5.3.

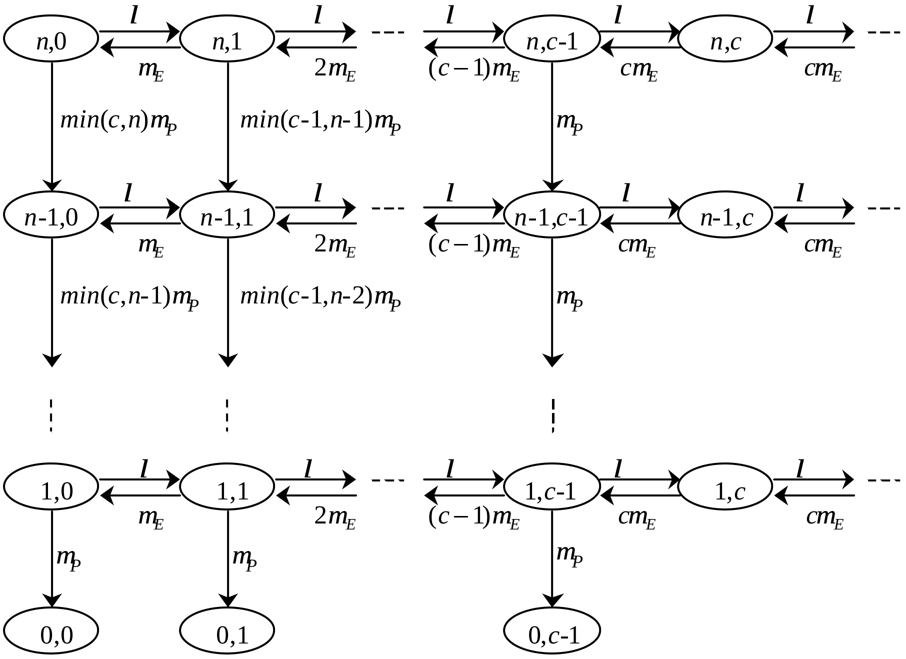


Figure 5.3. State-transition rate diagram for the c-server repair shop.

5.3.2 Recursive relations for *T*(*n*,*m*)

Let $T(n,m)$ be the project makespan if there are initially *n* planned and *m* emergency jobs in the shop. From the state-transition rate diagram, we can obtain relations between the random variables $T(n,m)$. First of all we note that the time it takes to reach state $(n,c+m-1)$ starting from state $(n,c+m)$ is distributed as a *busy period BP* in an *M*/*M*/1 queue with arrival rate l and service rate cm_E. Thus for $m \geq 0$ and $n > 0$ we have

$$T(n,c+m) = BP + T(n,c+m-1) \tag{5.2}$$

Next, let $Z(q)$ denote an exponentially distributed random variable with mean $1/q$. Then we have for $m < c$ and $n > 0$

$$T(n,m) = Z(q(n,m)) + \begin{cases} T(n,m+1) & w.p. \quad l\,/q(n,m) \\ T(n,m-1) & w.p. \quad mm_E\,/q(n,m) \\ T(n-1,m) & w.p. \quad \min(c-m,n)m_P\,/q(n,m) \end{cases} \tag{5.3}$$

with

$$q(n,m) = l + mm_E + \min(c-m,n)m_P$$

For $n = 0$ the project is finished, thus $T(0,m) = 0$ for $m = 0,...,c-1$.

5.3.3 The Laplace Stieltjes Transform of *T*(*n*,*m*)

From these relations we can obtain recursive relations for the Laplace Stieltjes Transform (LST) of $T(n,m)$. From this we could obtain the exact distribution by numerical inversion (see Abate & Whitt 1992). We follow another approach and use the LST to obtain the (first two) moments of $T(n,m)$. If $f(n,m)(s)$ is the LST of $T(n,m)$, that is $f(n,m)(s) = E\left(e^{-sT(n,m)}\right)$ for $s \geq 0$, then

$$E\left(T(n,m)^k\right) = (-1)^k f^{(k)}(n,m)(0), \qquad k = 1,2,3,\ldots \tag{5.4}$$

where the superscript (k) denotes the k-th derivative with respect to s. From these first two moments we derive an excellent (as we will see) approximation for the distribution of $T(n,m)$. From relation (5.2) we immediately obtain for $m \geq c$ and $n > 0$,

$$f(n,m)(s) = b(s) f(n,m-1) = \ldots = b(s)^{m-c+1} f(n,c-1)(s), \tag{5.5}$$

where $b(s)$ is the LST of the busy period *BP*, which is given by (see e.g. Kleinrock 1975)

$$b(s) = \frac{l + cm_E + s - \sqrt{(l + cm_E + s)^2 - 4\,l\,cm_E}}{2\,l}.$$

Relation (5.3) immediately yields for $m < c$ and $n > 0$ that

$$\begin{aligned}(q(n,m)+s) f(n,m)(s) = {} & l\, f(n,m+1)(s) + m m_E f(n,m-1)(s) \\ & + \min(c-m,n) m_P f(n-1,m)(s).\end{aligned} \tag{5.6}$$

Finally, we have

$$f(0,m)(s) = 1, \qquad m = 0,\ldots,c-1. \tag{5.7}$$

Note that the transform $f(n,c)(s)$ occurs in equation (5.6) for $m = c$-1. This transform can be eliminated by substituting relation (5.5) for $m = c$. This gives a set of linear equations for $m < c$ which in vector-matrix notation can be written as:

$$A_n(s) f_n(s) = D_{n-1} f_{n-1}(s), \qquad n = 1,2,3,\ldots \tag{5.8}$$

$$f_0(s) = (1,\ldots,1)^T, \tag{5.9}$$

where $f_n(s) = (f(n,0)(s),\dots, f(n,c-1)(s))^T$ and

$$A_n(s) = \begin{pmatrix} q(n,0)+s & -l & & & 0 \\ -m_E & q(n,1)+s & & & \\ & -2m_E & \ddots & & \\ & & & q(n,c-2)+s & -l \\ 0 & & & -(c-1)m_E & -lb(s)+q(n,c-1)+s \end{pmatrix},$$

$$D_{n-1} = \begin{pmatrix} \min(c,n)m_P & & & 0 \\ & \min(c-1,n)m_P & & \\ & & \ddots & \\ 0 & & & m_P \end{pmatrix}.$$

For each value of s the equations (5.8) and (5.9) can easily be solved recursively for $n = 1,2,3,\dots$. This concludes the determination of $f(n,m)(s)$.

5.3.4 The moments of *T*(*n*,*m*)

According to (5.4), the computation of the first and second moment of $T(n,m)$ requires the first and second derivative of $f(n,m)(s)$ at s = 0. Differentiating the system (5.8)-(5.9) once and twice with respect to s results in:

$$f_n^{(1)}(0) = [A_n(0)]^{-1}\left[D_{n-1} f_{n-1}^{(1)}(0) - A_n^{(1)}(0)\right], \qquad n = 1,2,3,\dots \tag{5.10}$$

$$f_0^{(1)}(0) = (0,\dots,0)^T \tag{5.11}$$

and

$$f_n^{(2)}(0) = [A_n(0)]^{-1}\left[D_{n-1} f_{n-1}^{(2)}(0) - A_n^{(2)}(0) - 2A_n^{(1)}(0) f_n^{(1)}(0)\right], \qquad n = 1,2,3,\dots \tag{5.12}$$

$$f_0^{(2)}(0) = (0,\dots,0)^T. \tag{5.13}$$

Note that $A_n(0)$ is a transient generator and thus $[A_n(0)]^{-1}$ exists. Furthermore, because $A_n(0)$ is a tridiagonal matrix, the equations (5.10) and (5.12) can be solved very efficiently for respectively $f_n^{(1)}(0)$ and $f_n^{(2)}(0)$ (see e.g. Atkinson 1989). Based on the formulas (5.5), (5.8)-(5.13) it is possible to recursively calculate the first two moments of $T(n,m)$ for any given value of n and m. An algorithm for this calculation is given in Appendix C.

5.3.5 Fitting

Finally, the first two moments are used to fit an appropriate distribution. The distribution we use is a mixture of two Erlang distributions, one with k-1 and one with k phases, having the same scale parameter m (cf. Tijms, 1994). For details see the Appendix D. This results in the approximation

$$P(T(n,m) \le T_D) \approx 1 - p\sum_{j=0}^{k-2} e^{-mT_D}\frac{(mT_D)^j}{j!} - (1-p)\sum_{j=0}^{k-1} e^{-mT_D}\frac{(mT_D)^j}{j!}\,. \tag{5.14}$$

If the standard deviation of $T(n,m)$ is larger than its mean then we need another type of distribution, such as a hyperexponential. However, observing the actual uncertainty in the different repair shops, this is not relevant in the present situation.

5.4 Discussion of the approximation

Though the first and second moment of $T(n,m)$ are theoretically and exactly determined, the fitting procedure may cause a difference between the approximated probability (5.14) and the exact probability. In this section we investigate the accuracy of this approximation. We generate eight projects, each with different repair and arrival rates, different starting conditions (n_0, m_0) and different number of parallel servers. The projects are listed in Table 5.1. Project 1 is used as the basic example: the others are derived from it by varying some characteristics.

Project	m_0	m_E	l	n_0	m_P	c	T_D
1	0	0.25	0.25	50	0.25	4	100
2	0	0.4	0.2	50	0.2	4	100
3	0	1	3	50	1	4	80
4	0	0.1	0.1	10	0.1	4	100
5	0	0.1	0.1	50	0.2	4	100
6	0	0.5	0.5	50	0.5	2	110
7	0	0.1	0.1	50	0.1	15	100
8	20	0.5	0.25	40	0.25	4	100

Table 5.1. Project characteristics of the 8 projects and process characteristics of the repair shops in which they are processed.

Project	$E(\cdot)$	$s(\cdot)$	70th perc.		80th perc.		90th perc.		95th perc.		99th perc.	
			S	A	S	A	S	A	S	A	S	A
1	69.4	12.6	75.2	75.3	79.6	79.6	86.0	85.7	91.6	90.6	103.2	101.3
2	76.1	12.0	81.8	82.0	85.9	86.2	91.8	92.5	97.0	97.0	107.4	106.6
3	49.6	18.5	56.2	57.4	63.1	64.6	74.1	73.9	84.4	83.7	107.4	99.7
4	40.3	15.9	46.3	46.7	52.2	53.5	61.5	61.2	70.2	70.9	89.2	92.8
5	85.6	17.7	93.1	94.0	99.4	100.4	109.1	108.7	117.8	115.8	136.4	133.3
6	100.2	24.5	110.7	110.8	119.5	119.7	132.8	132.3	144.8	145.7	169.5	160.6
7	57.6	13.3	62.6	63.6	67.3	68.3	74.9	75.9	82.1	81.2	98.3	92.8
8	60.9	9.2	65.3	65.5	68.4	68.6	73.1	72.9	77.0	76.0	85.1	82.3

Table 5.2. Expectations, standard deviations and approximated (A) and simulated (S) percentiles of the makespan distribution of the 8 projects.

The difference between the approximated and the exact value is investigated by simulation, because we do not have the exact value. For each project, the mean and standard deviation of $T(n,m)$ are exactly calculated, and the 70th, 80th, 90th, 95th and the 99th percentiles are both simulated and approximated. Each simulation is based on 10^6 replications. As we see, the results (listed in Table 5.2) are excellent for the percentiles 70 upto 95. For percentile 99 the differences however are not negligible. However, in real life the 99th percentile is usually not used: decision making in highly dynamic and stochastic environments is usually aimed at achieving a performance of at most 95 percent.

5.5 Subcontracting policy

In this section we discuss how the approach can be used to develop a subcontracting policy in the repair shop. We assume that the subcontracting decision has to be taken as soon as the project has started, that is at $t = 0$. However, in case the subcontracting lead time (i.e. the time between making the subcontracting decision and having the subcontracted jobs returned), denoted by L_0, is substantially smaller than T_D, the subcontracting decision can be postponed until the latest possible subcontracting time $t = T_D - L_0$, without affecting the way in which the computations are carried out. It should be noted that postponing the subcontracting decision means that more information about the progress of the project is available, at the time the subcontracting decision is made. It may not be surprising that as a result, on average less time slack has to be created.

Let a be the required delivery performance (e.g. $a = 0.95$). Define $\hat{n}(a, m)$ as the maximum number of planned jobs which can be completed before the due date T_D with a probability of at least a, given that at time $t = 0$ there are m_0 emergency jobs in the shop. So,

$$\hat{n}(a, m_0) = \max\{n | P(T(n, m_0) \le T_D) \ge a, \quad n = 0,1,2,\ldots\} \tag{5.15}$$

Then, the value of $\hat{n}(a, m_0)$ can simply be calculated (for instance with a bi-section method). By comparing $\hat{n}(0.95, m_0)$ (with m_0 from Table 5.1) listed in Table 5.3 to n_0 in Table 5.1, we see that only for the projects 3, 5 and 6 subcontracting is necessary if one has to complete these in time with a probability of at least 0.95.

Furthermore, the expected project makespan is shown in Table 5.3, where $\hat{T}$ is a short-hand notation for the makespan $T(\hat{n}(0.95, m_0), m_0)$.

Project	$\hat{n}(0.95, m_0)$	$P(\hat{T} \leq T_D)$	$E(\hat{T})$	$E(T_D - \hat{T})^+$
1	55	0.956	76.1	25.0
2	51	0.960	77.6	23.3
3	47	0.952	46.6	34.3
4	17	0.955	63.6	37.7
5	41	0.956	70.6	30.3
6	36	0.952	72.2	40.7
7	75	0.953	75.5	25.1
8	58	0.953	81.5	19.5

Table 5.3. Maximum number of jobs to be processed internally and resulting performance measures for the 8 projects.

In general, internally processing at most $\hat{n}(a, m)$ jobs will result in a low capacity utilization in the interval $[0, T_D)$. If no other activities are carried out in that interval besides the planned project and emergency jobs, Table 5.3 shows the expected remaining time that all servers are only used for emergency jobs (if any), i.e. $E(T_D - \hat{T})^+$. This illustrates the importance of the pulling decision at medium term level to ensure that enough jobs are available in the interval $[0, T_D)$, which have due dates after T_D.

5.6 Extension

The model assumes that all planned jobs have exponentially distributed processing times with the same mean. In real-life this will not be the case. A project may contain both small and large jobs and processing times may not be exponentially distributed. We present an extension of the model to be able to cope with more general and non-identical processing time distributions. In case the processing times have non-identical distributions, *sequencing* and *interruption rules* also effect the makespan distribution. An important question is to what extent this more generic model also gives acceptable approximations. In this section we discuss the generic model. Furthermore, we test the quality of the approximation for different sequencing rules.

Sequencing rules are rules to determine the sequence in which the jobs of a project are processed by the operators. In case all jobs have different expected repair times, literature suggests using sequencing rules which assign jobs to the operators in order of decreasing expected repair times (ELPT rule: Expected Longest Processing Time first). This has a positive effect on the makespan: the smaller jobs keep the (expected) difference between the first and last finished operator as small as possible. The smaller the makespan, the higher the probability of completing the project in time. For the case of exponential repair times without emergency jobs, Pinedo (1995) shows that the ELPT sequencing rule gives optimal results with regard to the delivery performance. However, in real-life precedence relations may exist between the jobs, so this rule is not always fully applicable. Interruption rules are rules to determine the planned job (and thus the operator) which is interrupted upon arrival of an emergency job. In case jobs have non-identical (residual) repair time distributions, the interruption rule affects the makespan distribution. In real-life this choice may depend on the difficulties of interrupting the various planned jobs currently on hand.

Without highly complicating it, the model cannot be extended with sequencing and interruption rules. Therefore, the model is not capable to exactly determine the moments of the makespan. However, the queuing model and fitting procedure can be extended to give at least approximations for these moments. Consider a project consisting of n_0 jobs. Let the random variables $X_1,\dots,X_{n_0}$ denote their processing times with means $E(X_i)$ and standard deviations $s(X_i)$. Further let Y be the total demand for capacity resources induced by the n_0 jobs, that is $Y = X_1 + \dots + X_{n_0}$. Assuming that the coefficient of variation of Y is smaller than 1, we can use the same fitting procedure as in Subsection 5.3.5 to fit a mixture of two Erlang distributions. If so, the problem is transformed from processing n_0 non identical jobs into a problem in which with probability p' (resp. $1-p'$) the project consists of $k'-1$ (resp. k')

exponential phases with identical mean $1/m'$ which have to be processed. Notice that no longer a physical relation exists between a certain job and a phase. Subsequently, let $E(S(m))$ (resp. $s(S(m))$) be the expectation (resp. standard deviation) of the makespan if the work load Y is represented as a mixture of Erlang distributions and there initially are m emergency jobs in the shop. This gives

$$E(S(m)) = p'E(T(k'-1,m)) + (1-p')E(T(k',m)) \tag{5.16}$$

and

$$E\left(S(m)^2\right) = p'E\left(T(k'-1,m)^2\right) + (1-p')E\left(T(k',m)^2\right) \tag{5.17}$$

and thus

$$s(S(m)) = \sqrt{E\left(S(m)^2\right) - E^2(S(m))}. \tag{5.18}$$

Finally, after again applying the fitting procedure based on $E(S(m))$ and $s(S(m))$, the random variable $S(m)$ is approximated by a mixture of two Erlang distributions.

Type	q	m	$E(X)$	$s(X)$
I	2	1	2	$\sqrt{2}$
II	2	0.5	4	$\sqrt{8}$
III	4	1	4	$\sqrt{4}$

Table 5.4. Erlang parameters (q,m) and resulting moments of the three different job types.

In real-life situations, technical considerations may prevent using a clever sequencing rule. As a result, under each sequencing rule the makespan distribution may differ. Therefore, we are interested in the difference between the simulated makespan distribution (in which each sequencing rule can be implemented) and the approximated makespan distribution by the model built (which neglects sequencing rules because the project has been transformed into a project with jobs having identically and exponentially distributed processing times) for both clever and unsophisticated sequencing rules. Therefore, we investigate the difference between the model results and the makespan in case simulations are carried with an ELPT and random sequencing rule. Furthermore, we assume that the interruption process can hardly be influenced due to technological constraints and therefore we use a random interruption rule.

Project	n_I	n_{II}	n_{III}
1	40	0	10
2	10	0	40
3	40	10	0
4	10	40	0
5	0	40	10
6	0	10	40

Table 5.5. The number of type I, II and III jobs in the 6 different projects.

We consider several projects, where each is assumed to consist of n_I (resp. n_{II} and n_{III}) type I (resp. II and III) planned jobs. The repair time of each job is assumed to be Erlang(q, m) distributed: the mean equals q/m and the variance equals q/m^2. The parameter values q and m depend on the job type. Table 5.4 displays the characteristics of each job type and Table 5.5 shows a list of the projects. The shop conditions are for each project kept at $m = 0$, $l = 0$, $m_E = 1$ and $c = 4$. Both for the ELPT and random sequencing rule, 10^6 replications are carried out. Results are listed in Table 5.6.

Again, it turns out that the difference between the theoretical makespan distribution and the simulated makespan distribution is again acceptably small. As a result, we conclude that regardless the sequencing and interruption rules used, the model gives acceptable predictions for the makespan distribution. However, it should be stated that as may be expected from the foregoing, the different percentiles are smallest in case the ELPT rule is used. After all, this rule is aimed at synchronizing the point in time at which the different operators finish their jobs.

Project	$E(S(0))$			$s(S(0))$			90th perc.			95th perc.		
	A	SR	SE	A	SR	SE	A	SR	SE	A	SR	SE
1	40.7	41.4	41.1	4.8	4.9	4.9	46.8	47.8	47.4	49.2	49.8	49.4
2	60.7	61.7	61.1	5.8	6.0	5.9	68.1	69.5	68.7	70.3	72.0	71.1
3	41.0	41.6	41.0	5.3	5.5	5.3	47.7	48.8	47.9	50.2	51.3	50.2
4	61.5	62.3	61.1	7.4	7.5	7.2	70.9	72.1	70.6	74.1	75.2	73.5
5	68.1	68.8	68.5	7.6	7.7	7.5	77.7	79.0	78.3	80.6	82.2	81.4
6	67.5	68.6	68.5	6.5	6.7	6.6	75.9	77.3	77.1	79.1	80.0	79.7

Table 5.6. Approximations (A) and simulated values under non-identical repair time distributions and random (SR) resp. ELPT (SE) sequencing rule for the expectation, standard deviation, 90th and 95th percentile of the makespan distribution.

It is to be expected that in case of more varying processing times of the planned jobs, the approximation will be less accurate, because now sequencing and interruption rules have more impact on the determination of the makespan. Furthermore, one should be careful in case the standard deviation of the processing times of the planned jobs is much smaller than their mean. Then the approximation in fact splits each planned job into a large number of small exponential subjobs, the makespan of which will be (much) smaller than the actual one.

In case of non-identically and non-exponentially distributed repair times, the subcontracting decision is not that trivial as solving problem (5.15). In case at the latest subcontracting time the probability of delivering in time is less than a, just applying formula (5.15) gives the maximum number of phases which can internally be processed while keeping the probability larger than a. Because there is no physical relation between a phase and an actual job, the subcontracting decision problem should be solved in another, for example “what-if”, approach. Also job parameters like subcontracting costs and subcontracting lead times play an important role now. We come back to this problem in the next chapters.

5.7 Conclusion

In this chapter we have shown that we are able to derive excellent approximations for the probability distribution of the makespan of a bunch of planned jobs which are disturbed by the arrivals of emergency jobs, in case all repair times have identically and exponentially distributed repair times. Moreover, we have shown that also in the case of non-identically and non-exponentially distributed repair times, the sketched approach generates satisfactory - although less accurate than in the first case - approximations. Now that we are able to describe the subcontracting and internal capacity planning for a single multi-server repair shop, the knowledge obtained from this model should be used to facilitate the medium term control decisions in more complex situations. This problem is subject of the next chapter.

6

Design of Subcontracting and Pulling Decision rules

6.1 Introduction

At this stage of the research, we have investigated which factors need to be considered whilst planning and controlling the projects and the processes in multi-server repair shops. Moreover, we have used the Navy case to establish which factors are currently not being used. Subsequently, in the previous chapter, we have taken a closer look at the role of each of these factors in the planning process. This has been done by developing a small-scale mathematical model which describes the probability of delivering a project in time as a function of the project and process characteristics. The results of the logistic regression and the analyses of the queuing model together can be regarded as the physics and mechanics of the maintenance processes: these make the behavior of the processes comprehensible for us. However, our objective is not only to describe the processes, but merely to *control* them. This chapter solely addresses the problem of transforming the descriptive knowledge into decision rules that are capable of controlling the processes.

The subcontracting decision function is concerned with the task to determine, based on the current state of the system, whether or not subcontracting is necessary and, if so, what part of the workload needs to be subcontracted. The aim of this decision function is to ensure a satisfactory delivery performance. The pulling decision is concerned with the decision whether failed repairables should be taken from stock and prepared for repair. Given the internal and subcontracted workload, this function should ensure a utilization of the internal capacity that is as high as possible. These days, managers have strong requirements regarding the outcomes of the control decisions. Managers cannot allow a period of poor performance whilst emphasizing that the long run performance is stable. Recent research (to which we will refer in the next section) has shown that the short term performance of production control systems (highly) improves in case information about the current state of the production system is included in the decision rule.

A complication in developing the control decisions may be that the repair shops encounter a demand pattern that has been classified as being dynamic and stochastic. As a result, the state of the production system, that is described by the projects on hand together with their due dates, (heavily) fluctuates in time. Furthermore, the process model that has been developed in Chapter 5 is only a simplified process model. This model differs from the real-life situation in (among other things) the number of repair shops involved in a project, the number of projects on hand and the processing time distributions. Besides, some events or phenomena are also not modeled in the process model (e.g. varying speed of work, influence of learning curves on processing times) which do play a role in real-life. Consequently, at the time the subcontracting and pulling decisions should be taken, there is only a rough idea about the future states, rather than detailed knowledge. It is unlikely that the future projects coincide with the expectations.

In the next section we show by a means of a literature review that principles and concepts from control theory can be of great help when applying theoretical models in real-life situations. Furthermore, these concepts are based on the use of current state information for control. Therefore, this makes these concepts - at least in theory - also capable of providing a satisfactory robustness of the short term delivery performance.

6.2 Control and control theory

Webster's New Collegiate Dictionary defines control as "*to exercise restraining of directing influence over, to have power over*". With respect to production control, Axsäter (1985) states that it is the purpose of a production control system to transform incomplete information about the market and available production resources into coordinated plans for production. Most research on production control aims at achieving satisfactory values for the performance measures (such as production rate, throughput and average WIP inventory levels) in the steady state (i.e. long run), while assuming a specific stable environment. Literature on this subject is ample and needs no further references. In our opinion, these studies indeed provide very valuable knowledge about the role of various factors on the performance. However, they are hardly applicable for short term control. As opposed to the steady state, only little research has been addressed to the transient state (i.e. short-term) behavior of production systems. Lin & Cochran (1990) state that production control based on mean value analysis and steady-state conditions is incapable of making on-time decisions that cope with unexpected events which are due to complex work flows (e.g., interruption, change of work). Therefore, they developed analytical formulae for the transient-state behavior. Lin *et al.* (1992) put that "*due to the lack of viable modeling techniques, abnormal situations at the operational level in a dynamic environment cannot be handled by shop floor production control systems that are only applicable to steady-state performance*". Abate & Whitt (1994) analyze the transient behavior of the $M/G/1$ workload process. Their main results describe the time-dependent probability that the server is busy, the time-dependent moments of the workload process and the covariance function of the stationary workload process. Stoop (1996) states that long term performance targets are not always realistic for short term performance, due to the uncertainty of the system. Therefore, he advocates to use separate targets for short and long term performance. Tan (1999) investigates, using a Markov reward model, the transient behavior of a production line, namely, the mean and variance of the number of parts produced in a given time period conditioned on an arbitrary initial condition. With respect to queuing systems, Whitt (1999) states that information should be exploited about the system state (including the number of customers in the system ahead the current customer) to better manage the systems.

Summarizing, these studies conclude that considering only steady-state behavior is not adequate to design and control production lines that operate in a rapidly changing, dynamic and uncertain environment.

As stated earlier, we fully endorse these approaches: short-term behavior should be included in the control models. However, an underexposed aspect of these studies is the robustness of the models: to what extent are the models capable of ensuring the desired output in case the actual environment differs from the assumed environment? After all, in a lot of real-life situations the actual environment differs from the assumed environment. Besides, the processes within the repair shops are complex and only a part of them can be explained by the logistic regression model from Chapter 3 and the analytical queuing model from Chapter 5. Furthermore, if also the random noise is considered which is inherent to the processes, it needs no further explanation that the *direction* of the outcomes of the control decisions might be known, but the *exact* outcomes are unpredictable. With respect to this particular case, we know what happens if part of the workload is subcontracted: the delivery performance will increase. We also are capable of making estimations about the *amount* of increase. But besides the analyzed uncertainty about the processes, also events may occur which are not modeled (variation in speed of work, learning curve effects, last minute working overtime, et cetera).

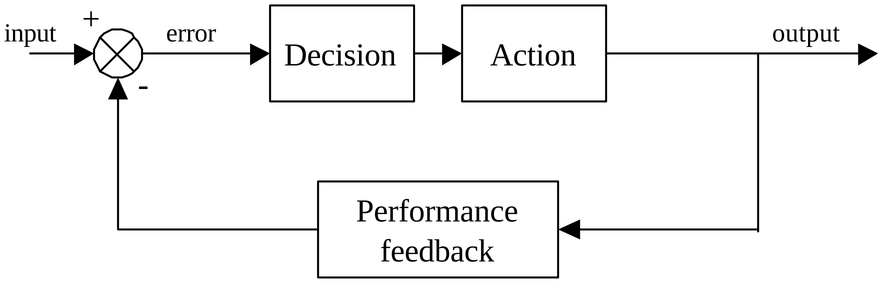


Figure 6.1. Block diagram representation of a closed-loop system with performance feedback.

Clearly, we are faced with the following situation. At every point in time at which we want to make a decision, we have information available about the current state of the system and the performance at that time. Furthermore, the queuing model can predict part of the future processes. As will be shown in the remainder of this chapter, these ingredients make it possible to use concepts from control theory, such that also corrections are made for the events that are not modeled.

A useful concept obtained from control theory is a closed-loop (or feedback) system (see Figure 6.1). According to Marshall (1978), a closed-loop system has to maintain an output equal to a desired value (i.e. the input) despite outside fluctuations or disturbances. By comparing input to output and using a valve or similar device, this difference should be controlled and kept minimal. The output of the valve (or the control policy) usually is a function of the current state of the system. Furthermore, this type of control loop is capable of dealing with the nature of dynamic processes. Holt *et al.* (1960) state that it is a common control principle in many different real-life contexts to switch on a feeding process at the point in time when a certain state variable reaches a predefined level (i.e. the *trigger* level), and to turn it off when the variable reaches some other level. This is for example the case for a heating system. Bishop (1975) states that control systems generally consist of:

- an input or reference level, that is a quality characteristic of the products being controlled (e.g. desired delivery performance);
- an output or controlled variable, that is a quality measure of the characteristics being controlled (e.g. actual delivery performance);
- a controller or decision-making device, that is where adjustments are made using the control valves (e.g. subcontracting);
- perturbations or noise, which are the results of environmental effects over which no direct control, at least within the control system itself, is available (e.g. arrival of emergency projects, deviations between actual and estimated processing times);
- a feedback device, that is the device that measures some feature of the output and transforms this information for comparison with the input (e.g. performance measurements);
- a comparator or detector, that actually compares the recorded output with the input (e.g. human observer, information system).

All these elements can be found in Figure 6.1. The quality of the control decisions heavily depends on the chosen control rule. The control rule describes the relation between on the one hand the difference between the input and output signal(s) and on the other hand the control decision. As discussed by Fortmann & Hitz (1977), a commonly used concept is the proportional-plus-integral-plus-derivative (PID) control action (or three-term controller). According to this concept, the controller output at time t, say $u(t)$, is defined as

$$u(t) = k_p e(t) + k_d \frac{de(t)}{dt} + k_i \int_0^t e(t)dt \tag{6.1}$$

where k_p, k_d and k_i are constants and $e(t)$ the error signal. Clearly, the output $u(t)$ contains information about the current error, the change in error and the accumulation of the past errors and should be used as a base for the control decision. This control rule usually does not contain any detailed process knowledge. In most cases it is sufficient to know only rough process knowledge. Consider the example of a car. Suppose that the car driver wants to drive according to the speed limits. He or she only uses the (rough) information which says that when the gas pedal (resp. brake) is pressed, the car drives faster (resp. slower). Each car driver knows that this is sufficient. Detailed process information about how far exactly to press the gas pedal and brake is not necessary. This is due to the fact that the output variable (i.e. the current speed) can be read real-time on the speedometer in the car. This latter can be regarded as the feedback device. Note that the car driver only uses the proportional part of (6.1) when actually driving. Now suppose that the car driver is driving in the inner city and that he or she approaches the highway. Clearly - assuming that he or she still wants to drive according to the speed limits - the input signal changes, because the driver is allowed to drive at a higher speed. To bridge this gap, the gas pedal is pressed. At this stage, the change in error becomes important. Accelerating too fast makes that the driver is in danger of bumping into the car in front of him or her: the acceleration should be controlled. Finally, the integral term becomes important in case the driver has the objective to drive a certain average speed during a given time interval. This objective can for example be valid in case the car driver wants to control the average speed between two consecutive tollgates to avoid a ticket for exceeding the speed limit. In that case, past errors should be taken into account. Notice that in all the above described cases, no detailed process model has to be available for effective decision making!

Some systems can best be modeled as a *discrete event* model, when the decisions are only taken at some fixed points in time (e.g. at the beginning of every day or week). According to Bishop (1975) the PID control action (6.1) needs to be transformed into another representation for discrete event systems:

$$u(t) = k_p e(t) + k_d \left(e(t-1) - e(t)\right) + k_s \sum_{i=0}^{t} e(t-i) \tag{6.2}$$

where k_p, k_d and k_s are constants and $e(t)$ the error signal. This representation clearly has the same features as the representation (6.1).

Principles of control theory have received little attention in scientific literature on production control. Simon (1952) was to first to use control theory in production and inventory control, and discussed the application of linear deterministic control theory to production control, by considering a continuous-time model and using Laplace transform techniques. He expected quite a lot from the application of classical control theory to the practice of production-inventory control design. Axsäter (1985) concludes that the application of control theory in production control has achieved considerable interest in the sixties (see e.g. Adelson 1966, Elmaghraby 1966, Diezel & Eilon 1967, Bessler & Zehna 1968) but this interest waned in the following decades. Since then, every now and then, studies come up which emphasize the promising results of the application of classical control theory. Bertrand (1980) concludes after analyzing a diffusion department, that classical control theory seems to provide a solid initial solution for the design of production-inventory control systems. Using a control theory approach, Grubbström & Wikner (1996) show that a typical inventory control system can be described by difference or differential equations, corresponding to the so-called ‘sawtooth’ patterns of a typical inventory process. Recently, White (1999) discusses the problem of inventory management and the application of system dynamics.

From the literature review in this section we adopt the feedback concept for the design of the control decisions for the maintenance organizations. Feedback turns out to be useful in applications where the system state has a dynamic and stochastic character.

Furthermore, the feedback loop compensates for differences between the input and the output (i.e. the error), irrespective by which event this error has been caused. Even in case the error has been the result of events that are not included in the model. This latter observation is extremely important. Already in an earlier stage of this thesis, we concluded that is has been impossible to model the processes within the repair shop at a very detailed level! Applied to our control problem, the past delivery performance (that can be fed back to the decision makers) should be a base for the subcontracting decision rule. Like the role of the gas pedal in the car driver case, the decision maker may decide to subcontract in case the fed back delivery performance is poor and not to subcontract in case the delivery performance is satisfactory. Of course, subcontracting should be done gradually: the worse the delivery performance, the more jobs should be subcontracted.

The main problem that occurs when the subcontracting decision is based on the past performance, is the existence of the subcontracting lead time. After all, the effectuation time of subcontracting makes that the influence of subcontracting at time t on the delivery performance cannot be measured until time $t + L_0$. Furthermore, at time t also earlier decisions taken with respect to projects with due dates at $t+1, t+2, \ldots, t+L_0-1$ are in the 'pipeline'. Therefore, in the remainder of this chapter we create an artificial output signal that includes both past performance and estimates about the performance of the projects in the pipeline These estimates are based on the analytic process model that has been developed in the previous chapter. From a conceptual point of view, this artificial output signal can be compared to e.g. the economic inventory level that is used in inventory management. This artificial signal cannot be observed in real-life, but is a measure for the signal that we actually want to control: the delivery performance. The combination of feedback and anticipation should produce a delivery performance that both has a satisfactory steady state behavior and is robust at the short term as it includes current state information.

6.3 Performance indicators

The performance for the maintenance organizations is two-dimensional. The main performance indicator for the maintenance organizations is the due date performance, i.e. the fraction of projects that is delivered before or at their due dates. Subcontracting is recognized to be the main control decision to influence this performance. Besides, we have observed that the utilization of the internal capacity resources is another important performance measure. Given the internal and subcontracted workload, the pulling decision is concerned with ensuring a capacity utilization that is as high as possible. In order to be able to design a subcontracting control rule, in Subsection 6.3.1 we derive for each project the probability of being completed in time, whereas Subsection 6.3.2. is addressed to the capacity utilization in the different repair shops.

First, we make some assumptions regarding the subcontracting procedure. In Chapter 4 we already noticed that the subcontracting decision has a longer effectuation time than the pulling decision. In order to avoid to unnecessarily complicate the model, we assume that the subcontracting lead time equals L_0 and the pulling time equals L_1 time units, logically with $L_1 < L_0$. Furthermore, we assume that the subcontractors deliver with 100% certainty at the end of the subcontracting lead time, regardless the amount of work that is subcontracted at the same time.

6.3.1 Delivery performance

As we have learnt from control theory, the control decision should be based on both past performance (for compensating undershoot and overshoot) and the current state of the system (for anticipating). This ‘current state’ comprises both the projects currently on hand and the knowledge about the projects which will arrive in the (near) future. Therefore, at time t, we consider the time interval $\left[t - t_R ; t + t_H\right]$, where t_R is the review period and t_H the planning horizon (with $L_0 \leq t_H$). To be able to construct feedback about the projects in this time interval, the (expected) performances of these projects need to be evaluated. As the performance depends on the (dynamic) workload in the repair shops, we first derive a dynamic expression for the workload. Let $\Lambda(t)$ be the set containing all projects A_i in the organization satisfying *at least one* of the following conditions at time t:

- unfinished at time *t* and having a due date d_i before $t - t_R$ (type I in Figure 6.2);
- either finished or unfinished and having a due date d_i within the interval $[t - t_R; t + t_H]$ (type II in Figure 6.2);
- either finished or unfinished and having an earliest possible start date r_i within the interval $[t - t_R; t + t_H]$ (type III in Figure 6.2);
- either finished or unfinished and having both its earliest possible start date r_i before $t - t_R$ and its due date d_i after $t + t_H$ (type V in Figure 6.2).

Note that some projects satisfy both the second and third condition (type IV in Figure 6.2). The projects A_i in $\Lambda(t)$ concern both the preventive projects and the repairable projects which have become ‘planned’ as the result of a class transition due to the criticality of the levels of ready-for-use repairables. For the latter, these due dates have been set at $t + t_H$. Such a project is represented by project type IV in Figure 6.2.

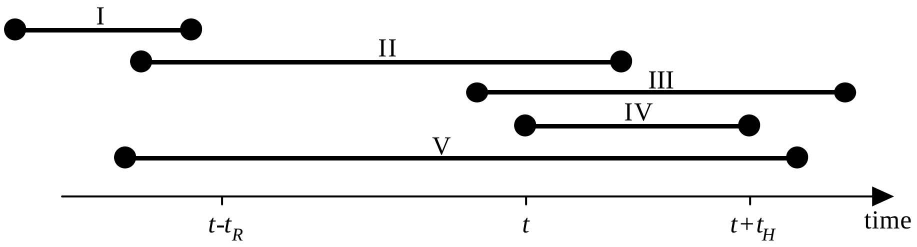


Figure 6.2. Different types of projects in $\Lambda(t)$.

In a next step, we derive the workload of the projects (expressed as amount of working hours) in a dynamic way. At time *t*, for each project $A_i \in \Lambda(t)$, the expected remaining project processing time $E(X_i(t))$ and standard deviation $s(X_i(t))$ is available, where $X_i(t)$ is the random variable describing the remaining processing time of project *i* at time *t*. We assume that the $X_i(t)$’s are independently distributed.

Finally, it should be noted that $\Lambda(t)$ is a dynamic set. Compared to $\Lambda(t)$, the set $\Lambda(t+\Delta t)$ (with $\Delta t > 0$) may differ in four ways:

- projects with a due date d_i before $t - t_R$ and which are finished in the time interval $[t, t+\Delta t]$ are in $\Lambda(t)$ but not in $\Lambda(t+\Delta t)$;
- projects with a due date d_i in $[t - t_R; t - t_R + \Delta t]$ and which are completed before $t + \Delta t$ are in $\Lambda(t)$ but not in $\Lambda(t+\Delta t)$;
- projects with an earliest start date r_i in the interval $[t + t_H; t + t_H + \Delta t]$ are in $\Lambda(t+\Delta t)$ but not in $\Lambda(t)$;
- parameters $E(X_i(t))$ and $s(X_i(t))$ of some projects that are both in $\Lambda(t)$ and $\Lambda(t+\Delta t)$ might have been altered in case operators have been working on these between time *t* and $t + \Delta t$.

Without loss of generality, in case the number of projects in $\Lambda(t)$ is denoted by $N(t)$, we assume that all projects in $\Lambda(t)$ are numbered and sequenced in such a way that the following holds: $d_1 \leq d_2 \leq \ldots \leq d_{N(t)}$.

The assembly character of the projects results from the fact that projects are divided among one or more repair shops. In case there are *J* multi-project repair shops which operate in parallel, we consider for every project A_i its sub-projects A_{ij} in repair shop *j* (*j* = 1,...,*J*). Note that a sub-project A_{ij} can also be a so-called empty-set, in case repair shop *j* is not involved in project *i*. Furthermore, let $d_{ij} = 1$ if repair shop *j* is involved in project A_i. Otherwise $d_{ij} = 0$. Let $\Lambda_j(t)$ be the set of the sub-projects in this repair shop *j*. The start dates and due dates of these sub-projects are inherited from the parent-project A_i. Finally, at time *t*, let $E(X_{ij}(t))$ and $s(X_{ij}(t))$ respectively be the expectation and the standard deviation of the remaining processing time of the sub-project of parent-project *i* in repair shop *j*. We assume that the processing times of the different sub-projects are independently distributed.

At the next lower level, we have observed that sub-projects consist of one or more jobs. If we consider the sub-project A_{ij} in repair shop *j*, this sub-project consists of the jobs $a_{ij1},\ldots,a_{ij,n_{ij}}$, with n_{ij} the number of jobs in sub-project A_{ij}. Again, for each job it holds that the earliest start and due dates are inherited from the parent-project and that the expectations and standard deviations of their remaining processing times are given by $E\left(X_{ijk}(t)\right)$ and $s\left(X_{ijk}(t)\right)$, where $k = 1,\ldots,n_{ij}$. Also these processing times are assumed to be independently distributed.

Besides, repairable projects might be available. At organizational level, assume that there are $M(t)$ repairable projects available at time t. Because a repairable needs capacity of only one repair shop, it is more important to consider repairable projects at the repair shop level than at the organizational level. Therefore, we denote the set of available repairable projects in repair shop j at time t, by $\Omega_j(t)$. Note that the sets $\Omega_1(t),\ldots,\Omega_J(t)$ are piecewise disjoint. Their union contains $M(t)$ projects. Each repairable project contains a single failed repairable which has to be transformed into a ready-for-use condition. At time t, for each project in $\Omega_j(t)$ we have information about the first and second moment of the processing time.

To get a clear view of the workload and how it is phased in time, assumptions need to be made about the sequence in which projects are executed. Although the method that will be described is applicable under the use of any chosen sequencing rule, we assume that the projects are processed according to the EDD rule. That is, projects are processed in order of increasing due dates (if possible). This rule can be considered as fair and gives a small variance of the project lateness. An overview of different (elementary) sequencing rules is given by e.g. Baker (1975). It should be noted that for the remainder of the analysis it does not matter what sequencing rule has been chosen: without highly complicating the analysis, also other rules could have been chosen. For example, one could think of *synchronization*, that is synchronizing the different sub-projects belonging to the same parent-project as much as possible. This decision rule is widely used in production systems with parallel shops. Although we are aware that this rule is also applicable in this situation, we do not take a closer look at it. The main reason for this is the fact that later on, the control structure will be validated at repair shop level, where decisions cannot be made based on the progress in other shops. So, synchronization is not relevant in this case.

The project set description together with the chosen EDD sequencing rules enables us to compare demand for capacity with the availability of capacity at different points in time. In deterministic environments capacity requirements often are graphically depicted in a non-cumulative way: for each period, (a so-called time bucket), the available and needed capacity are depicted. Also the planners at SEWACO seemed to work in this way. It is our strong conjecture that this is not the most suitable way in stochastic environments. The stochastic processing times make that it is not possible to assign projects to time buckets beforehand. After all, delays of previous projects may cause that a project is processed in a later bucket than the one it is assigned to. Therefore, a cumulative approach is more appropriate, where we depict for each point in time the amount of work that has to be completed before that time. This is exactly the approach we have followed in Figure 2.7 in Chapter 2. To make this approach operational, the following notation is introduced. At time *t*, let $C_{j,\Delta t}(t)$ (with $\Delta t \geq 0$) be the random variable describing the completion date of a meta-project containing the remaining workload of all projects in $\Lambda_j(t)$ with a due date smaller than or equal to $t+\Delta t$. Furthermore, let $C'_{j,\Delta t}(t)$ (with $\Delta t \geq 0$) be the random variable describing the completion date of the meta-project containing the workload of all projects in $\Lambda_j(t)$ with a due date smaller than or equal to $t+\Delta t$, augmented with the repairable projects in $\Omega_j(t)$.

At this stage we are able to construct at time t for each project A_i in $\Lambda(t)$ with $d_i \in [t-t_R; t+t_H]$ the probability of completing it in time, say $p_i(t)$ as follows:

$$p_i(t) = \prod_{\substack{j=1 \\ d_{ij}=1}}^{J} P\left(C_{j,d_i-t}(t) \leq d_i\right). \tag{6.3}$$

The individual probabilities $P\left(C_{j,d_i-t}(t) \leq d_i\right)$ can be approximated with the model presented in Chapter 5, in the following way. At time *t*, consider the unfinished sub-project *i,j* with a due date at time d_i, with $d_i > t$. Firstly, of all unfinished projects in shop *j* with a due date before d_i, the workload is projected at time *t*. The expectation and the variance of the processing time of this meta-project is calculated by accumulating the expectations and variances of the (remaining) processing times of all projects with a due date before d_i. The probability density distribution of the workload of this meta-projects is approximated by fitting a mixture of two Erlang distributions, using the expectation and variance of the meta-project. This gives - among other things - the number of initial *phases* (i.e. the *n* in *T*(*n*,*m*)), like we have introduced in Chapter 5. Secondly, information about the emergency jobs (initial number in the repair shop, arrival intensity and mean processing times) has to be derived (e.g. by analyzing historical data). Thirdly, the queuing model in Chapter 5 is used to calculate $P\left(C_{j,d_i-t}(t) \leq d_i\right)$. Repeating this procedure for all repair shops involved in project *i*, enables us to calculate the probability $p_i(t)$ according to (6.3). The value of $p_i(t)$ for the projects with a due date in the interval $[t-t_R;t]$ is either equal to zero or one. For each project with a due date in the interval $(t;t+t_H]$, it holds that $p_i(t)$ is either one (in case it is already finished at time *t*) or somewhere between zero and one (in case it is not yet finished at time *t*). Note that representation (6.3) does not use information about the earliest start dates r_i. We assume that as soon as a project has been finished, the next project is available for processing. In some cases this may not be true! At that time, the pulling decisions have to make that other work becomes available.

Finally, some comments have to be made about the interdependencies between the different repair shops. Assume that the management of the maintenance organization has set the norm for the delivery performance at α (with $0 \leq \alpha < 1$). Furthermore, in case of a ‘fair’ control policy, we have seen in Chapter 3 that the probability of completing a certain project in time should not depend on its project characteristics, nor on its process characteristics. The assembly nature of the projects makes that in case more than one repair shop is involved in a project A_i, the probability of completing a sub-project A_{ij} should be higher than α in order to ensure a project performance of α. In this way, in an artificial way, *internal due dates* are created: the more repair shops involved in a project, the more time slack is needed in the different repair shops. We come back to this later, when we aggregate the different $p_i(t)$’s into one performance measure that can be influenced by the subcontracting decision.

6.3.2 Utilization rates

The due date performance is the key performance indicator. The utilization rate of the repair shops is subordinate to this key indicator. However, given that the performance indicator is satisfactory, the utilization rates should be as close to one as possible. This subsection is aimed at the construction of an indicator that can be used to control the utilization rates of the capacity resources in the repair shops. Using the knowledge that is obtained so far, we are able to construct an indicator which is an expression of the resource demand and thus a good indicator for the utilization of the resources. Assume that the pulling lead time equals L_1 time units. This means that in case it is decided to pull repairables at time *t*, these become available in the repair shops at time $t + L_1$. Consequently, it should be decided to pull repairables in case idle time is expected soon after time $t + L_1$. Therefore, we look at the resource demand in the interval $[t; t + L_1 + \Delta t]$. This demand for capacity resources in repair shop *j* induced by the projects in $\Lambda_j(t)$, say $E_j(t)$, for the planning horizon $[t; t + L_1 + \Delta t]$ is approximated by the following expression:

$$E_j(t) \approx \frac{E\left(C_{j,L_1+\Delta t}(t)\right) - t}{L_1 + \Delta t}, \qquad j = 1, \ldots, J \tag{6.4}$$

It should be noted that this formula is only an estimate. The expression $E\left(C_{j,L_1+\Delta t}(t)\right)$ is the expected completion date of some meta-project: in a *c*-server shop this is the time at which the *c*-1 operators are idle and the *c*-th operator finishes his or her job. The quality of this approximation depends on the number of parallel repair men in the shop and the processing times characteristics. In a similar way as (6.4), estimates can be derived for the demand for capacity resources in repair shop j, for the planning horizon $[t; t + L_1 + \Delta t]$, in case also the repairable projects in $\Omega_j(t)$ are considered, say $E'_j(t)$, in the following way:

$$E'_j(t) \approx \frac{E\left(C'_{j,L_1+\Delta t}(t)\right) - t}{L_1 + \Delta t}, \qquad j = 1, \ldots, J \tag{6.5}$$

It should be noted that both (6.4) and (6.5) can be larger than one. Therefore, it is a good measure for the pressure of work in the different repair shops.

6.4 Control decisions

Now that we have derived performance indicators for individual projects and individual repair shops, these indicators need to be aggregated into one signal output signal, that should be used in the feedback loop. This output signal should be the base for the subcontracting and pulling decisions. These control decisions should bring the output signal in accordance with the reference level, if necessary.

6.4.1 Subcontracting decision

Seen from a control point of view, the projects with due dates in the time interval $[t - t_R ; t + t_H]$ can be assigned to different interval classes. Some projects are already finished: their performance cannot be influenced anymore. Other projects still have to be completed and thus interventions are possible for these. Assume at time *t*, that for a certain repair shop *j*, the set $\Lambda_j(t)$ contains the projects $A_{1,j}, \ldots, A_{11,j}$ as shown in Figure 6.3. Based on the due dates, these eleven projects can be assigned to exactly one of the following five time intervals.

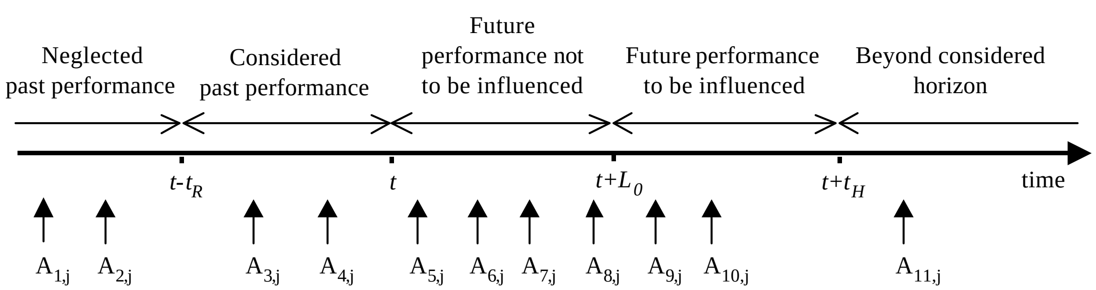


Figure 6.3. Time window with different states and projects pointing their due dates in shop *j*.

- **Neglected past performance** $\left(-\infty; t-t_R\right)$. The projects with a due date before $t-t_R$ (i.e. $A_{1,j}$ and $A_{2,j}$) are not considered for the construction of the output signal in the control system. Because $A_{1,j}$ and $A_{2,j}$ still have to be completed, they still influence the performance of the other projects currently on hand.
- **Considered past performance** $\left[t-t_R; t\right)$. Projects with a due date in this interval should have been completed before time *t*, but currently they are not necessarily completed. Because their due dates have expired, we can determine whether these were completed in time or not. Consequently, at organizational level, for these projects holds $p_i(t)=0$ or $p_i(t)=1$. In Figure 6.3, these are depicted by project 3 and 4.
- **Future performance not to be influenced** $\left[t; t+L_0\right)$. Projects with a due date in this interval do not have to be completed yet. However, some already may be completed. If the latter is the case, for these projects holds that $p_i(t)=1$. If this is not the case, the value $p_i(t)$ can be approximated by the use of (6.1). Important characteristic of each project with due dates in this time frame is that its value $p_i(t)$ cannot be increased by subcontracting (part of) the other projects in the same interval, without decreasing the performance of other projects with due dates in this interval. For example, the EDD rule makes that the performance of project A_7 (i.e. $p_7(t)$) can be improved by subcontracting (part of) the workload of projects $A_{5,j}$ or $A_{6,j}$. However, the fixed subcontracting lead time makes that the subcontracted work returns at $t+L_0$, which is after the due dates d_5 and d_6. And thus, these two projects definitely are completed beyond their due dates, when subcontracted Therefore, we do not consider the projects with due dates in $\left[t; t+L_0\right)$ as candidates for being subcontracted.
- **Future performance to be influenced** $\left[t+L_0; t+t_H\right]$. Due to the fixed subcontracting lead time L_0, the latest possible time at which project A_i can be subcontracted without being delivered late equals d_i-L_0. Therefore, at time *t*, only projects with a due date of $t+L_0$ are candidates for subcontracting. For projects with a due date larger than $t+L_0$ holds that the subcontracting decision can be taken at a later point in time. In Figure 6.3, only project $A_{8,j}$ is candidate to

be subcontracted at time *t*. Subcontracting (part of) $A_{8,j}$ increases not only the performance of project $A_{8,j}$, but also the performance of projects $A_{9,j}$, $A_{10,j}$ and $A_{11,j}$. The performance of projects with due dates in this interval can be increased, without decreasing the performance of other projects.

- **Beyond considered horizon** $(t+t_H;\infty)$. Projects with a due date in this interval fall beyond the considered horizon. Furthermore, taking into account the EDD sequencing rule, this projects does not influence the performance of the projects with due dates in the horizon $[t-t_R;t+t_H]$.

We base the delivery performance indicator on all projects with a due date in $[t-t_R;t+t_H]$ and use a decision function that is aimed at keeping this performance indicator in accordance with a. This performance indicator (i.e. output signal) $\bar{p}(t)$ is defined as:

$$\bar{p}(t)=\frac{\displaystyle\sum_{\substack{i=1\\ d_i\in[t-t_R;t+t_H]}}^{N(t)}\left(\prod_{\substack{j=1\\ d_{ij}=1}}^{J}P\left(C_{j,d_{j-t}}(t)\le d_i\right)\right)}{\left|\{i|d_i\in[t-t_R;t+t_H]\}\right|},\qquad(6.6)$$

with $|S|$ the cardinality of set *S* (i.e. the number of elements in *S*). This indicator can be compared to the speedometer in the earlier described car driver example. At time *t*, if this value is under the threshold value a, the value $\bar{p}(t)$ should be increased by subcontracting one or more jobs a_{ijk} with a due date of $d_i=t+L_0$, if any available. The way in which the subcontracting decision depends on the output signal is described in the control or decision rule. The decision rule that we suggest for the remainder of this study is based on an iterative approach. Suppose that the output signal is below the threshold a. In each iteration one job is subcontracted, until the (updated) output signal is at least a. We come back to this rule in the next chapter.

It is clear from (6.6) that the output signal depends on the values for t_R and t_H. The choice of these parameters depends upon preferences of the management about the allowed fluctuations in the performance. The values t_R and t_H cannot be derived analytically, but need to be established by experiments. However, at this stage we should also make some statements about what would happen in case these periods are shortened or extended.

Concerning the review period t_R, we state that the longer this period is, the more historic projects are evaluated by the control mechanism in the decision making process. As a result, the influence of the subcontracting decision becomes smaller. This has serious consequences for the short term performance. In case the past performance exceeds (resp. is lower than) the target value a, (almost) nothing (resp. everything) is subcontracted, which may result in a very erratic short term performance. On the other hand, if t_R is short, the control rule seems solely concerned with the future delivery performance. Irrespective the past performance, the control mechanism intends to always keep the expected future performance above a. As a result, the long term performance is always larger than the target a (in case the suggested control actions indeed can be taken, i.e. there are jobs with can be subcontracted). This may cost an unnecessarily large amount of money.

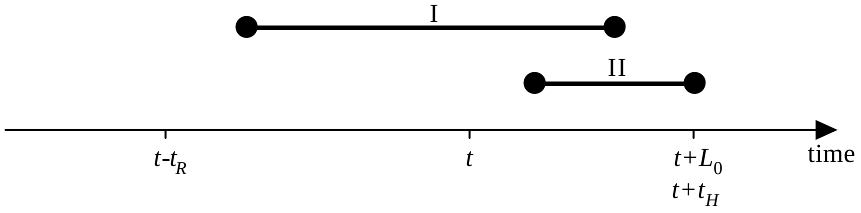


Figure 6.4. Example of a situation where planning horizon equals subcontracting lead time.

With respect to the planning horizon t_H, a similar argumentation is applicable. In case the shortest possible planning horizon is chosen (note that the shortest possible planning horizon equals the subcontracting lead time L_0), the subcontracting decision maker may be surprised by projects entering the planning horizon. This can be illustrated by an example. Consider the situation in Figure 6.4. Assume that at the time project II enters the planning horizon (i.e. its due date equals $t+L_0$ and thus its workload is included in calculating the system output for the first time) the output signal drops below α. However, despite the fact that project II has already been arrived, its earliest possible start date is beyond time t: no control actions can be taken to improve the system output. In case the planning horizon would have been larger that the subcontracting lead time, the workload of project II would have been considered earlier. Then, part of project I could have been subcontracted to improve the system output signal. Summarized, increasing the planning horizon may increase the flexibility. On the other hand, the queuing model developed in Chapter 5 only considers already arrived planned projects and expectations about possible future arrivals of emergency projects. In case t_H becomes too large, also new planned projects may arrive between time t and t_H due dates between t and t_H. Clearly, the analytic model does not compensate for these projects, such that predictions made by the model may differ from the simulation outcomes.

As a final remark concerning the control rule, we have to make a statement about the speed of response. One can speak about a *rapid response* control rule if it only takes one period to compensate the difference between the input and output signals (Bishop, 1975). That is, if at day t the output signal is below a, a sufficient number of jobs can be subcontracted such that the (updated) output signal becomes above a. However, a rapid response may not always be applicable, because of the following reasons. Firstly, it may not be possible to compensate within one period, because even subcontracting all the jobs with a due date at $t + L_0$, may not be enough: the expected future output is still below a. Secondly, there may be no jobs at all with a due date at $t + L_0$, such that no jobs can be subcontracted at all. And thirdly, there may be jobs with a due date at $t + L_0$ for which it holds that their earliest start dates is beyond t such that these also cannot be subcontracted (like we have seen in Figure 6.4). However, it should be kept in mind that this response time holds for the artificial output signal (6.6). This signal cannot be observed in real-life. The signal that can be ‘observed’ in real-life is the delivery performance over a certain past period. Clearly, the latter is not influenced if jobs are subcontracted at time t. It takes L_0 more periods before the effects of subcontracting on the past performance can be measured. This again emphasizes the need to include the look ahead period in the control rule!

6.4.2 Pulling decision

Once the subcontracting decision has been taken at time t and indeed jobs are subcontracted, the sets $\Lambda_j(t)$ (j = 1,...,J) are altered. The performance measure $\overline{p}(t)$ may still be below a in case not a sufficiently large number of subcontracting candidates had been available. However, it can also be the case (even without subcontracting) that the performance measure $\overline{p}(t)$ exceeds a. In that case, apparently projects already have enough slack. Moreover, it is likely that idle time occurs if no countermeasures are taken. In case the resource demand (6.4) (and thus the expected future utilization rate) drops below some satisfactory level, idle time can be prevented by pulling failed repairables. Because the pulling lead time equals L_1, pulling repairables increases the resource demand in $[t;t+L_1+\Delta t]$ to (at most) $E'_j(t)$, according to (6.5). It should be noted that as soon as repairables are pulled and carried out, the performance of other (future) preventive projects decreases. After all, these preventive projects are postponed. Therefore, while pulling repairables, the decrease of the performance measure $\overline{p}(t)$ needs to be monitored!

6.5 Summary

In this chapter we have given an overview of control theory and its application to production control. Our conjecture is that due to the dynamics, uncertainty and interaction between the different projects, control theory is an appropriate way to control the processes. Moreover, we have applied control theory in such a way that - besides feedback - also the future state of the system is considered. Before the stated conjecture can be tested in real-life, the subcontracting and pulling decisions need to be geared to the SEWACO situation. This will be subject of the next chapter.

7

The Royal Netherlands Navy Case III - Making the Control Structure Applicable

7.1 Introduction

At this stage of the thesis, we have answered the first and second research question. Firstly, we have identified the project and process characteristics which should be available to planning and control departments in order to be able to control the processes and the project delivery dates. Secondly, we have incorporated this information in a hierarchical planning and control structure, for which we have designed the subcontracting and pulling decision functions. Moreover, these decisions also describe *how* to use the project and process characteristics. We now have come to the stage at which we want to investigate the applicability and the consequences of the control model in the real-life naval maintenance shops. This chapter solely addresses the problem of making the control model applicable for the SEWACO situation.

Testing the control structure can be done in various ways. We briefly discuss two extreme methods. Firstly, a computer simulation can be carried out while varying (highly idealized) arrival and order patterns. Secondly, the control structure could directly be implemented in the real-life situation. The first method has the advantage that all possible mixtures of arrival and orders patterns can be analyzed. As a result, this method produces generic knowledge, but still may lack applicability in the real world, because not all characteristics from real-life are modeled in the control model. The second method indeed gives accurate answers concerning the difference between the actual and expected performance derived from the analytic model. However implementing this method may be very costly and may generate undesired side effects. Furthermore, managers do like to first have some knowledge about the expected costs and revenues before they will start using the decision procedures. Besides, this can be a very time consuming method. To overcome both problems we choose for an approach which uses the advantages of both methods and is ‘as close to real-life’ as possible without affecting it. In our approach we apply the presented models on historical real-life naval data. In other words, we repeat the historical projects of the maintenance organization in a simulation model and investigate what would have happened in case the designed control structure would have been applied to these data. Besides, we are aware that we did not model every detail in the maintenance processes. Therefore, the use of historical data offers us the possibility to check to what extent the control model is capable of dealing with events that are not modeled in detail. The feedback character should take care of this.

In the remainder of this chapter, we discuss what the framework for testing should look like and discuss the choices and assumptions we have to make. Moreover, we make the control model applicable for testing it with SEWACO data. Subsequently, we have a framework that can be used to test the role of the various parameters in the model and investigate the resulting subcontracting behavior of the organization.

7.2 A framework for testing

Ideally, given the fact that we have chosen for a framework in which we use computer experiments with empirical data to validate and evaluate the control structure, we want to test the model by repeating the historical processes and to make decisions based on the designed control structure for all repair shops at SEWACO. Afterwards, the diagnosis model from Chapter 3 should be applied at experiment outcomes with respect to process and project characteristics and the (simulated) performance. If attention has adequately been paid to the various process and project characteristics in the control model (or technically stated: if the $\hat{b}_j$'s associated with the process and project characteristics do not significantly differ from zero), we indeed may conclude that the control structure does have the necessary properties to control the delivery performance. Unfortunately, as will be seen below, this approach is not applicable in our situation, due to a lack of data. For that reason, we have chosen for a different approach.

The subcontracting decision is based on process characteristics of the repair shops under consideration and project characteristics of the projects on hand. This (historical) information can be extracted from the Business Planning System (BPS). On the other hand, the pulling decision is based on processing times, inventory levels of ready-for-use repairables and failure rates (see Schneeweiss & Schröder, 1992). The latter two cannot be extracted from the BPS, but should be extracted from another information system. Creating this information can only manually be done and therefore is too time consuming. As a consequence, no use can be made of the pulling decision in the framework. The only information we have about failed repairables is the information concerning the failed repairables that are already sent from the main depot to the maintenance organization and therefore are stored in the BPS. This has serious consequences for the framework. We already have shown that subcontracting goes hand in hand with an increased risk of capacity idle time. We also have seen in the previous chapter, that idle time has a high probability of going together with a delivery performance that is above the satisfactory threshold value. Pulling repairables makes that both low-priority projects are carried out and the expected future delivery performance is reduced to the (satisfactory) threshold performance. Summarizing, we have got the subcontracting decision which may *increase* the performance, but lack the pulling decision which may *decrease* the performance, such that it is expected that the observed simulated delivery performance fluctuates *above* the threshold value, rather than *around* the threshold. Despite the fact that the delivery performance also is controlled using just the subcontracting decision (after all, subcontracting brings the performance to at least the threshold), it is unlikely that the logistic regression model introduced in Chapter 3 gives $\hat{b}_j$'s that do not significantly differ from zero. Recognizing these problems, we take the following approach for the validation.

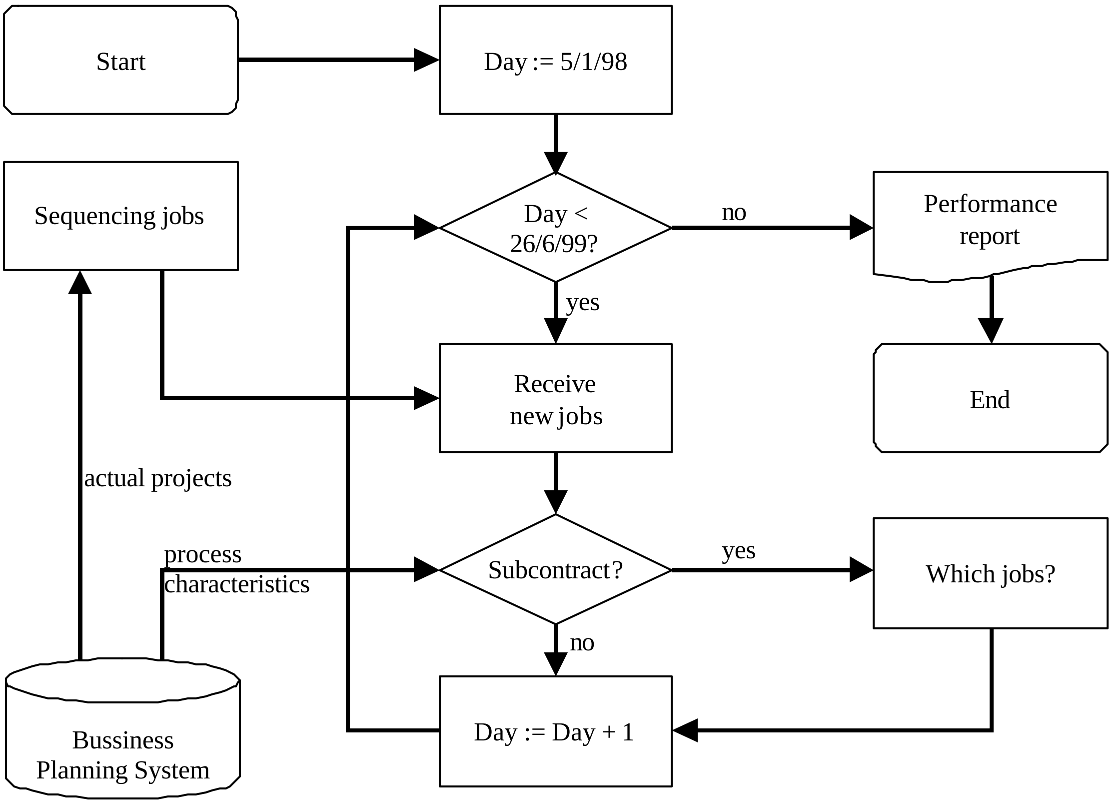


Figure 7.1. A flow diagram of the framework for testing the control structure.

For a selected set of repair shops with varying project and process characteristics, we simulate the period from January 5, 1998 until June 26, 1999. During this simulation, the projects arrive according to their historical arrival date and are released to the shop (see Figure 7.1). On a daily base we use the formulae from Chapter 6 to analyze whether subcontracting is necessary. If so, we use an algorithm to select which job(s) to subcontract. At the end of the simulated period, a performance report is derived, which is subject of further analysis. This framework can be used to investigate for various types of repair shops (and thus various mixtures of process characteristics) the consequences of subcontracting on several performance criteria. In the next chapter, this framework is being used to carry out the simulation for three carefully selected repair shops. These three shops are selected, such that the differences between the process characteristics of the three repair shops are large. This makes the knowledge obtained from the experiments as generic as possible.

The experiments are kept as close as possible to the real processes. We mainly focus on the subcontracting decision level, whilst realizing that other factors which did not have received any attention so far, play a role now. Besides, the designed models and decision procedures ask for a certain availability and format of information. This is not always available: either it is not yet in the required format, or it cannot be created at all from the available BPS data. Therefore, in order to make the control structure applicable and operational for testing, we have to make assumptions and choices. These will be the subject of the next section.

7.3 Assumptions and choices

In this section we subsequently discuss in more detail the assumptions and choices we make regarding i) the relation between the various decision levels, ii) the reliability of the subcontractor, iii) the uncertainty of the processing times, iv) the arrival pattern of emergency jobs, v) the availability of capacity resources and vi) the interaction with the other repair shops.

7.3.1 Relation between different decision levels

In Chapter 4 we have discussed the various decision levels in the control structure. Furthermore, we have stated that the main lack of scientific knowledge concerns the integration of the subcontracting decision and the internal capacity planning. This has been pointed out as the main contribution of this thesis and therefore also of this framework. However, we also have seen that the different control levels interact: decisions taken at a certain decision level create conditions for the lower levels. In the remainder of this thesis we consider the decisions at higher level (norms for delivery performance, amount of repairables, availability of internal resources) as given. This data of the naval case cannot be altered. For the lower level decision function (capacity allocation), the conditions are created by the subcontracting level. The resulting performance is fed back to the subcontracting level and partly base for the future subcontracting decisions. As discussed earlier, for this capacity allocation level we assume that all jobs are processed in an EDD sequence, as much as possible.

7.3.2 Subcontracting assumptions

Concerning the subcontracting decision, we assume that the subcontracting lead time (i.e. the elapsed time between making the subcontracting decision and the time at which the technical systems are returned) is fixed at L_O working days. Furthermore, we assume that the subcontractors deliver with a 100% reliability. That is, in case a job is subcontracted at some time t, it is always returned before or at time $t + L_O$. Finally, we assume that the capacity at the subcontractor is infinite or at least sufficiently large to process all the jobs that are offered by the naval maintenance organization. This ‘perfect reliability assumption’ can be relaxed without increasing the complexity of the models. For the moment, such a relaxation does not increase the obtained insights.

7.3.3 Uncertainty of processing times

Analyses in previous chapters have shown that the processing times of jobs are stochastic. Furthermore, it has been shown by (ex-post) calculating the coefficients of correlation between predicted and real processing times, that the various repair shops face with different levels of uncertainty. This is incorporated in our model by using the expectation and the standard deviation of the (remaining) repair times of the jobs on hand. In case the BPS data is used for testing the model, we face the problem that the BPS only provides the expected processing times, rather than also some measure of uncertainty. We have already seen in Chapter 2 (see Table 2.2), that the level of repetition for individual jobs is rather low. Besides, it is a rather time-consuming task to search for identical historical jobs and use these to estimate the standard deviations: this can only be done manually. To overcome this problem of the unavailability of standard deviations, we assume that all jobs within one repair shop have Erlang distributed repair times with (not necessarily) different expectations but identical shape parameters k (i.e. the ratio of the squared expectation and the variance of the repair time at job level). This assumption may be justified by the fact that all jobs within one repair shop are extracted from a set with a very limited number of different technical systems, each requiring more or less the same technical skills. This reduces the problem of determining the standard deviation for all individual jobs based on historical data to the less complicated problem of determining for each repair shop the parameter k based on the historical data which is available in the BPS. The latter problem can be solved in the following way. Consider a repair shop for which we know for N finished (and not necessarily identical) jobs, the actual processing times $X_1,\ldots,X_N$ and estimations for these processing times $\overline{X}_1,\ldots,\overline{X}_N$, which are defined as the average processing times of historical processing times. Subsequently, define for $i = 1,...,N$ the normalized processing times $Z_1,\ldots,Z_N$ in the following way:

$$Z_i = \frac{X_i}{\overline{X}_i}. \tag{7.1}$$

As a result, the value $1/k$ can be estimated by

$$\hat{E}(1/k) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{Z_i}{\overline{Z}} - 1\right)^2 \tag{7.2}$$

Note that if k is estimated by simply taking the value $\hat{E}(1/k)^{-1}$, it is very likely that this results into a non-integer value. Rounding off is necessary: k should be set equal to either the largest integer smaller than $\hat{E}(1/k)^{-1}$ or the smallest integer larger than $\hat{E}(1/k)^{-1}$. As a result, an error cannot be avoided. However, it should be noted that the larger the value of $\hat{E}(1/k)^{-1}$, the smaller the effects on the model outcomes.

At the time the subcontracting decision is taken for a particular project (that is L_0 days before its due date) it is most likely that part of the released jobs are already in progress: some working hours are spent on these. The fact that the processing time of a job is now represented by an Erlang-k distributed random variable enables us not only to estimate the standard deviation of the repair times, but we are also capable of calculating the expectation and the standard deviation of the *remaining* processing time for jobs in process. Pure technically speaking, the coefficient of variation of the *remaining* processing time asymptotically increases from $1/k$ to 1 (that is: the uncertainty increases) as the hours already spent on a job increase from zero to infinity. This neglects the real-life phenomenon that during the execution of a job, increased insights in the actual condition of the technical system (and thus processing time) may come into light which *reduce* this uncertainty. Although this asks for a careful and more in-depth study regarding this behavior, we have to make assumptions about this behavior in order to make the control structure applicable for testing. Therefore, in the remainder of this study, we assume that the increased insights fully compensate for the increase in uncertainty. Consequently, the ratio between the expected remaining processing time and its variance is (artificially) kept stable at k, regardless the hours already spent on the job.

These considerations enable us to make at any point during the simulation, estimations for first and second moments of the remaining workload. To minimize the effects of the assumptions that are put on the processing time distributions, we do not draw the actual processing times from an Erlang-k distribution function, but use the historical actual processing times instead. Only in case the project has not been finished in real-life at the end of the period for which data is available, we draw the actual processing time from the Erlang-k distribution, given the already processed hours.

7.3.4 Arrival process of emergency projects.

In Chapter 5, we have seen that the needed capacity slack for a preventive project is partly based on the arrival pattern and processing times of emergency projects. The higher the fraction of capacity consumed by emergency projects, the more slack is needed. Emergency projects arrive according to some irregular pattern, driven by breakdowns of technical systems. This behavior is purely random. Of course, at the time the subcontracting decision should be taken, the processing times of the emergency project which *might* arrive during the execution of the preventive projects are also uncertain because no ideas exist about which technical system might fail. Therefore, we assume that both the interarrival times and processing times of emergency projects are exponentially distributed, respectively with intensity parameters l and $1/m_E$. These assumptions seem to be justified when historical data is used to estimate the coefficients of variation of the interarrival times and processing times for the emergency projects. To give the reader an idea about their values, for 9 repair shops these coefficients of variation are shown in Table 7.1. It appears that the value of these coefficients of variation vary in the neighborhood of 1, which is the value of the coefficient of variation of an exponentially distributed random variable.

7.3.5 Availability of capacity resources

The availability of the capacity resources is determined by the availability of the operators in the repair shop. The number of operators within a repair shop is stored into the BPS. Their availability fluctuates during time because of holidays, illness, education, reduction of working hours, et cetera. Analyses have shown that it is justified to assume that 65% of the yearly available working hours per operator (that is 52 times 5 times 8 hours) is actually spent on processing jobs. Therefore, in the remainder, we assume that each operator works 52 times 5 days a year, for 0.65*8 = 5.2 hours a day.

Repair shop	Variation coefficient processing times	Variation coefficient interarrival times
640E	1.254	1.165
650A	1.099	0.709
680F	1.115	1.055
680G	0.753	0.766
740M	1.227	1.066
760S	1.103	1.325
770F	1.215	1.504
840A	1.396	1.125
840D	0.963	0.771

Table 7.1. Arrival and processing characteristics of emergency projects for 9 of the 75 repair shops.

7.3.6 Interaction with other repair shops

A typical maintenance project consists of all kinds of jobs that have to be processed by a certain number of repair shops. As we have seen in the previous chapter, the probability of completing the parent-project before its due date depends on the progress of its sub-projects in all the repair shops involved. Moreover, this probability is calculated as the product of the probabilities of completing the individual sub-projects in the involved repair shops before the due date. We also have noticed that we run the simulation model at repair shop level: we investigate the subcontracting decisions for various repair shops, not taking into account the *actual* progress of the other sub-projects. To justify this choice, assumptions have to be made about the progress in the other repair shop. Consider a project which is carried out in n repair shops. We want to investigate the control policy and its consequences within only one of these shops. For each of the shops, it is possible to compute the probability of completing the sub-project in time. In case we observe a single-project in a multi-repair shop situation, the product of these n probabilities should at least equal a. Assuming that each repair shop controls its processes in such a way that the probability of completing a sub-project in time equals at least $a^{1/n}$. Translating this to the multi-project, multi-repair shop situation, at the time the subcontracting decision should be taken, we assume that on average the probability of completing an entire project in time equals a. Not using actual progress information of the other n-1 involved repair shops, we assume that the product of the probabilities of these n-1 sub-projects equals $a^{(n-1)/n}$. Of course, in case experiments would have been carried out with multiple repair shops at the same time, the actual progress in the different repair shops would have been used in controlling the projects.

7.4 Subcontracting decisions

Now that we have discussed the choices and assumptions we made in the framework, the subcontracting decision still has to be made operational. The subcontracting decision, which we assume to be made on a daily base, concerns two main questions:

(i) is subcontracting necessary and if so, (ii) which jobs have to be subcontracted. These questions are answered in the remainder of this section.

7.4.1 Subcontracting necessary?

The question whether subcontracting is necessary or not can be answered based on an indicator containing both historical as well as future projects. At time t, at least one job with a due date at time $t + L_0$ (if available), should be subcontracted in case the delivery performance indicator

$$\overline{p}(t) = \frac{\sum_{\substack{i=1 \\ d_i \in [t-t_R; t+t_H]}}^{N(t)} \left(\prod_{\substack{j=1 \\ d_{ij}=1}}^{J} P\left(C_{j,d_j - t}(t) \leq d_i\right) \right)}{\left|\left\{ i \middle| d_i \in [t - t_R; t + t_H] \right\}\right|} \tag{7.3}$$

is smaller than a. We have seen this formula before in Chapter 6. Note, that (besides the control parameters t_R and t_H which will be derived in the next chapter) this value can be calculated using the assumptions and choices we have made in Section 7.3.

7.4.2 Which jobs to subcontract?

Once it has been established that subcontracting is necessary, the question remains which jobs to subcontract. Based on our assumptions, at day t only jobs with a due date at day $t + L_0$ (and that are currently not yet finished and not yet in process) are candidates for being subcontracted. In case the set of these jobs is not empty, jobs are selected in an iterative way and subcontracted as long as the updated value $\overline{p}(t)$ is smaller than a or the set of candidates is not empty. In the latter case, it is possible that all candidates are subcontracted while still having a $\overline{p}(t)$ that is smaller than a. The main question is to point out in each iteration the job which has to be subcontracted. For that purpose we use a marginal analysis. Referring to the notation used in the previous chapter, we assume that the total subcontracting costs for a single job a_{ijk} that belongs to the set of subcontracting candidates, are proportional to its expected remaining processing time $E(X_{ijk}(t))$. Consequently, if these costs are denoted by $TC(a_{ijk})$, the following holds:

$$TC(a_{ijk}) = c_{\mathrm{I}} E(X_{ijk}(t)). \tag{7.4}$$

where c_{I} are the subcontracting costs per expected hour of processing time. For each job in the candidate set, it holds that the delivery performance $p_i(t)$ of its parent-project increases in case it is subcontracted. This increase can be calculated based on the formula (6.1), once with the job internally processed and once without the job internally processed. Furthermore, for each job this increase can be divided by the total costs to subcontract that particular job, resulting in an increase in probability per unit costs. Now we subcontract that job with the highest increase in probability per unit costs.

7.5 Summary

In this chapter we attempted to gear the designed control model and its decisions to the SEWACO situations. Due to lack of data and our system boundaries, we are forced to make assumptions and choices regarding the real-life processes at SEWACO, while keeping in mind to stay as close as possible to reality. Now that these assumptions and choices are made and grounded, the model is ready to be used for the actual validation and evaluation questions. These questions are to be answered in the next chapter.

Finally, it should be emphasized that this chapter has not presented a guide for implementation. The various assumptions and choices that have been made indicate that more in-depth investigation is necessary before the control structure can be implemented.

8

The Royal Netherlands Navy Case IV - Validation and Evaluation of the Control Structure

8.1 Introduction

Now that the control structure has been made applicable for carrying out experiments with real-life SEWACO data, it can be used for the validation and evaluation. The validation concerns the question to what extent the control mechanism does what it has to do (is the actual delivery performance of the simulation results in accordance with the required performance?), whereas the evaluation concerns the question to what extent the control mechanism provides outcomes that are acceptable and feasible for application in real-life (e.g. what percentage of the total workload should - on average - be subcontracted in order to achieve satisfactory performance?).

The validation will be executed at two different stages. Firstly, we validate the control mechanism, consisting of the queuing model embedded in the feedback loop. The question to be answered is whether or not this mechanism provides the delivery performance that is used as a target value. Or in other words: is the mechanism capable of doing what it should do? As this thesis is oriented towards the role of analytical models in real-life situations, this is the main question of this chapter. Note that the control mechanism is considered as a black box: in case the model does provide acceptable outcomes, we still do not know whether this improvement (in comparison to the current delivery performance) should be contributed to the feedback character, the queuing model, or a combination. Furthermore, in case a fit is found, we do know that the mechanism gives - at least for the tested repair shops - a controlled performance. This question will be discussed in Section 8.4.

Once this question has be answered, we descend one level in the control mechanism and take a closer look at the role of the queuing model in the control mechanism. It is well known that feedback is capable of providing performance improvement without the use of an explicit process model. Perhaps, the knowledge that a poor past performance can be compensated with subcontracting and a more than satisfactory performance can be compensated by less or no subcontracting might by itself also be capable of providing an improved performance. Therefore, the second question is to what extent the improved performance (if this is indeed the answer to the first question) indeed is achieved by the use of the queuing model in the control loop. For methodological reasons, we should also carry out experiments with a control mechanism which also uses feedback but has no explicit process model. Only in case the queuing model outperforms this simple model and gives satisfactory results, we are allowed to claim that the queuing model is an acceptable representation of the processes within the repair shops. Furthermore, we then can claim that the queuing model indeed is necessary to provide a robust performance. This will be the subject of section 8.5.

Besides the validation question, we pay attention to the evaluation question. For use in real-life situations it is not only necessary to know whether or not the control mechanism delivers the required performance, but even more interesting is the question whether or not the outcomes are feasible in terms of applicability. In case (on average) an unacceptable large part of work should be subcontracted, the management may consider the effects of the control structure as unacceptable costly. Thus in the evaluation process, we take a closer look at the resulting subcontracting behavior.

Before we come to the validation and evaluation question, we first describe the three repair shops that we use for the validation. Furthermore, we discuss the output measures that we monitor during the simulations. Subsequently, we choose appropriate lengths for the review period t_R and planning horizon t_H to make the control mechanism operational. Not until then, the validation and evaluation can be carried out.

8.2 Selection of repair shops and output measures

For generic results, the control structure should be tested with various repair shops having different process characteristics. Because we want to stay close to real-life, we use real-life data. As a result, we can only lean on the actual process and project patterns in the repair shops, rather than constructing various types of (idealized) order patterns. To partly overcome the disadvantage of not being able to create all different types of arrival and project patterns, and to optimally use the reflection of reality in the data, we select our repair shops from the set of candidates by ensuring that the differences in process and project characteristics are large. Repair shops need to be chosen such that different demand mixtures, project characteristics and shop sizes are distinguished. Based on this consideration we have chosen for the shops 640E, 770F and 860D, with characteristics as depicted in Table 8.1.

Shop	Frac. Planned	Frac. Em.	Frac. Rep.	Mean project lead time (work days)	Mean job processing time (hrs)	Erlang *k* factor	Operators
640E	0.59	0.22	0.19	104	43.4	1	10
770F	0.71	0.24	0.05	148	24.1	10	3
860D	0.46	0.53	0.01	95	36.2	4	3

Table 8.1. Characteristics of the three selected repair shops.

Obviously, the number of explanatory variables (the project and process characteristics) exceeds the number of observations (that is, the number of repair shops). Furthermore, we do not even consider the (second and higher order) interactions that can be defined based on these process and project characteristics, which would make that there are even more possible explanatory variables. For that reason, it is clear that it is impossible to aim for the construction of relationships between on the one hand the explanatory variables and on the other hand the resulting delivery performance and subcontracted hours of work. However, we still are able to check for the three shops in Table 8.1 (and thus for a varying set of process and project characteristics) to what extent it is possible to indeed control the processes.

For each experiment that we report on in this chapter, a specific set of input parameters (review period, planning horizon, subcontracting lead time, etcetera) needs to be selected. It is to be expected that different sets of input parameters produce different outcomes. To get a clear and objective insight into the performance, at the end of each experiment a report is created which displays for each day t during the simulation:

- percentage of projects delivered in time during the last year before day t, that is f_t;
- short term robustness, (i.e. absolute difference between the required service level a and delivery performance during the last 50 working days), that is STR_t;
- internal resource utilization (i.e. fraction of capacity spent on jobs);
- subcontracted hours of work, that is SHW_t;
- Expected remaining workload in hours, that is $ERWL_t$.

8.3 Review period and planning horizon

At the last step before the actual validation, appropriate values for the lengths of the review period t_R and planning horizon t_H need to be chosen. As extensively described in Chapter 6, these values respectively determine to what extent the subcontracting decision should be based on past and expected future performance. The choice of these parameters depends upon preferences of the management about the fluctuations they allow in the performance and cannot be derived analytically: experiments are needed. Before carrying out these experiments, two statements should be made. At first, we remind the reader that the long term performance of the SEWACO is measured over a year. Consequently, the decisions that are taken, given the actual conditions in the shops, should contribute to this long term performance.

For that reason, we select values for t_R and t_H, such that $t_R + t_H = 260$ (remind that a year has 52 times 5 working days). Secondly, it is our conjecture that different repair shops require different values for t_R and t_H, because of their variety in process and project characteristics. Because it is impossible to analytically derive appropriate combinations of (t_R, t_H), experiments are carried out with the combinations (50,210), (70,190), (90,170) and (110,150) or more as long as no satisfactory combinations are found. In each of these experiments a fixed subcontracting lead time of L_0 = 50 working days (i.e. 10 weeks) is assumed. Furthermore, for each repair shop and (t_R, t_H) combination, we simulated two different long term service level requirements: $a = 0.5$ and $a = 0.9$. For each parameter setting, 10 replications are carried out. The simulations run from January 5, 1998 until June 25, 1999. Results (i.e. delivery performance over the last simulated year - i.e. f_{last} - and the average short term robustness) are summarized in Table 8.2, Table 8.3 and Table 8.4. From these results, we have to choose acceptable values for (t_R, t_H). Before we come to that point, we first explain the entries in these tables in more detail by an illustration.

(t_R,t_H)	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
(210,50)	0.572	0.329	0.960	0.076
(190,70)	0.600	0.335	0.972	0.077
(170,90)	0.638	0.333	0.972	0.077
(150,110)	0.642	0.336	0.972	0.077

Table 8.2. Experimental results for repair shop 640E with α = 0.5 and α = 0.9.

(t_R,t_H)	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
(210,50)	0.330	0.239	0.727	0.157
(190,70)	0.338	0.233	0.953	0.051
(170,90)	0.357	0.224	0.953	0.051
(150,110)	0.382	0.214	0.953	0.052
(130,130)	0.412	0.198	0.953	0.051
(110,150)	0.462	0.167	0.953	0.052

Table 8.3. Experimental results for repair shop 770F with α = 0.5 and α = 0.9.

(t_R,t_H)	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
(210,50)	0.327	0.252	0.966	0.058
(190,70)	0.485	0.243	0.966	0.064
(170,90)	0.510	0.244	0.964	0.059
(150,110)	0.440	0.203	0.966	0.064

Table 8.4. Experimental results for repair shop 860D with α = 0.5 and α = 0.9.

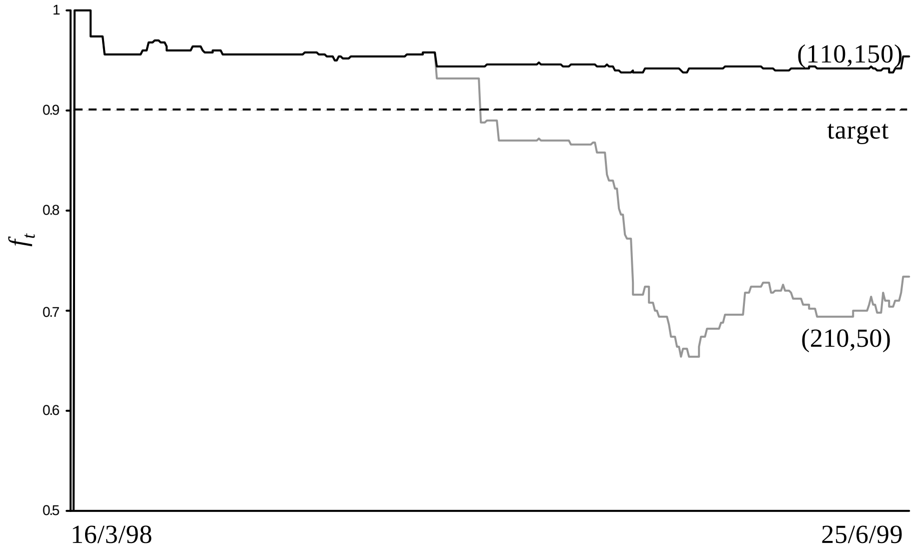


Figure 8.1. Long term performance f_t in repair shop 770F with anticipating (110,150) and without anticipating (210,50).

As an illustration, we show for two instances in Table 8.3, namely (210,50) and (110,150), with a required service level of $a = 0.9$, the behavior of f_t and STR_t. Table 8.3 shows that the outcomes for these two instances differ a lot. Note that in the first case, the planning horizon does not go beyond the subcontracting lead time. Analyzing the separately created list with subcontracted jobs shows that the (210,50) instance is surprised by a project at the end of July 1998. Because its lead time (that is $d_i - r_i$) is smaller than the subcontracting lead time, it is not possible to subcontract it. For the (110, 150) instance holds that this project is seen 100 days earlier. Moreover, during these 100 days other projects are subcontracted in order to create slack for that particular project. Figure 8.1 shows that it has successfully been done. The lists also show that for the (210,50) instance from July 1998 on, all possible jobs are subcontracted to get the long term performance back above the target level. However, even these efforts are not sufficient. In Figure 8.2 it is shown that the shapes of the short term robustness (logically) also differ a lot. Because the short term robustness is only measured over the last 50 working days, the response of 'subcontracting all you can' is much more rapid. These figures show that anticipation is necessary to create a more flexible control environment.

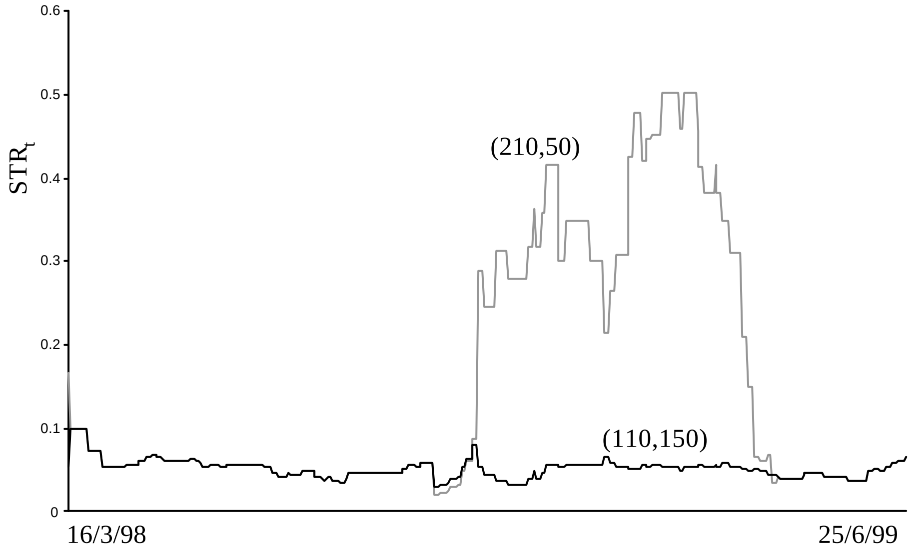


Figure 8.2. Short term performance STR_t in repair shop 770F with anticipating (110,150) and without anticipating (210,50).

Now that we have described the consequences of choosing a typical set of review and horizon parameters, the final task in this section is to select values t_R and t_H. These will be used in the remainder of this chapter. Relevant indicators for this choice are a low average STR_t and controlled subcontracting costs, both for the cases $a = 0.5$ and $a = 0.9$. Carefully observing Table 8.3, Table 8.2 and Table 8.4, makes that we choose for the settings that are displayed in Table 8.5. Note that the simulation outcomes are rather insensitive to a change in (t_R, t_H). This insensitivity makes that - although the combinations are only tested for two values of a and one value of L_0 - we assume that these combinations are also acceptable for instances with other values for a and L_0.

Repair shop	t_R	t_H
640E	190	70
770F	110	150
860D	170	90

Table 8.5. Review period and planning horizon for the different repair shops.

8.4 Validation of the control mechanism

In Chapter 3, we have seen that the project and process characteristics are not adequately being used in the current planning and control procedures at SEWACO. Moreover, we have established what the role of each characteristic should be in the planning and control procedures. Subsequently, we have developed a control mechanism that incorporates these characteristics. Deduction has led to the hypothesis that this control structure adequately deals with the characteristics and (thus) should be capable of achieving a controlled performance. As a first step in the validation process, it has to be shown that the control mechanism (that is the queuing model embedded in the feedback-based control rule) indeed is capable of doing what it should do: achieving the required performance, complying the different project and process characteristics. This question will be answered by carrying out different experiments. In case that for various repair shops and various service levels it holds that the requirements equal the actual simulation outcomes, it can be concluded that the control structure does what it has to do: the control mechanism is validated. If this is not the case, arguments need to be found to explain the differences. For each repair shop we carry out various experiments. We vary the requested service from 0.0 to 0.95 with step size 0.05 both for a subcontracting lead time of $L_0 = 10$ and $L_0 = 50$. We are aware that a subcontracting lead time of 10 working days is very unrealistic. However, from the methodological point of view it is very interesting to see how outcomes vary in case a short and long pipeline of future decision outcomes are considered.

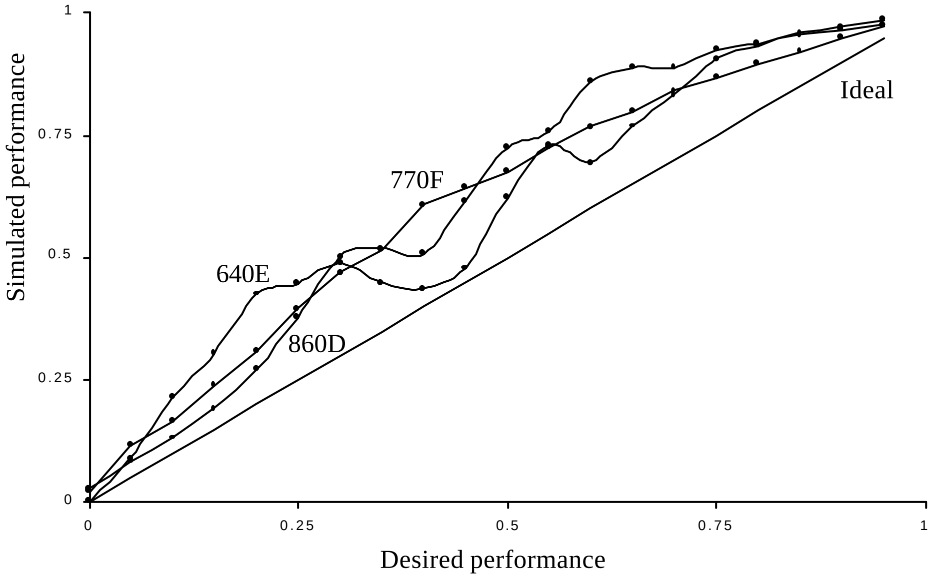


Figure 8.3. Desired versus simulated performance for the three different repair shops with a subcontracting lead time of 10 working days.

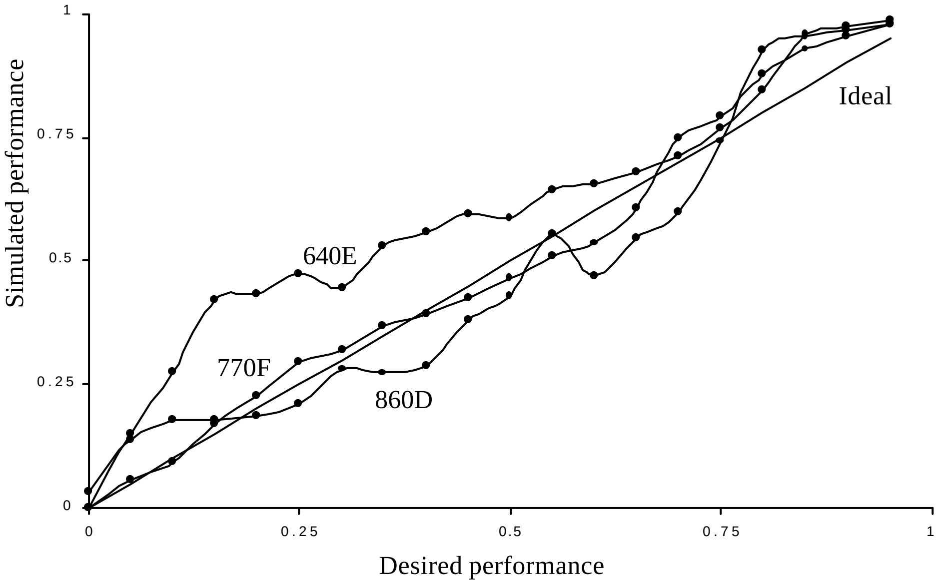


Figure 8.4. Desired versus simulated performance for the three different repair shops with a subcontracting lead time of 50 working days.

For each instance, 10 replications are carried out. We first look at the differences between the desired and the actual outcomes. Subsequently, we describe the robustness of the control system. The accuracy of the control mechanism is shown in Figure 8.3 and Figure 8.4. These depict the desired versus the (simulated) actual performance and therefore show to what extent the model does what it has to do. The experiment outcomes are represented by the dots, which are - for reasons of clarity - connected by a (smoothed) line. At first glance, it appears that the actual performance behaves more or less like the desired performance. If a closer look is taken at these figures, several noteworthy phenomena come into light.

Firstly, we see that in case $L_0 = 10$, all actual outcomes are above the desired outcomes, whereas this is not the case for $L_0 = 50$. A plausible explanation may be the different *speeds of response* for both cases. It has already been explained that due to the lack of an operational pulling decision, the control mechanism is capable of *increasing* the delivery performance by subcontracting, rather than also *decreasing* it by the pulling decision. In the latter case, nothing can be done, except ‘allowing’ that the performance is temporarily above the required performance. In case the system output signal falls under the target performance, subcontracting is necessary. Because only projects can be subcontracted which have a lead time that is longer than the subcontracting lead time, the number of potential subcontracting candidates is a decreasing function of the subcontracting lead time (see Table 8.6). Therefore, in case $L_0 = 50$, it turns out that it may occur that a temporarily lower output signal should be accepted, because the set of subcontracting candidates is empty. This can be summarized by stating that the speed of response increases as the subcontracting lead time decreases.

Repair shop	Percentage jobs with lead time smaller than 10 days	Percentage jobs with lead time smaller than 50 days
640E	1%	17%
770F	4%	9%
860D	1%	8%

Table 8.6. Job lead time characteristics for the three repair shops in relation to the subcontracting lead time of 10 and 50 working days.

Secondly, looking closer at Figure 8.3, we see for $L_0 = 50$, that one shop (640E) performs always better, one shop (770F) performs sometimes worse and sometimes better and one shop (860D) performs for the most cases worse than desired. In the 640E case we see - particularly for low values of a, that is $0.05 \leq a \leq 0.45$ - that the simulated line significantly differs from the ideal line. This can be explained by analyzing the list with subcontracted jobs. At the start of the simulation there is a large backlog. As a result, a lot of jobs are subcontracted to increase future performance. These control actions create that much slack, that the last year of the simulation, no jobs are subcontracted at all! In that case, pulling repairables would have been a perfect control action to both reduce the simulated outcomes to the desired outcomes and increase the inventory levels of ready-for-use repairables. Secondly, concerning the 860D repair shop, we see that (for most instances) the performance is worse than for the ideal situation. Again looking carefully at the list of subcontracted jobs (for example considering the instance where we require $a = 0.8$), it follows that during the last 5 months of the simulation, all possible candidates for subcontracting indeed were subcontracted. The subcontracting model even advises to subcontract more, but these jobs simply are not available. Finally, when considering the repair shop 770F, it follows that the simulated and desired performance more or less coincide. Referring to Table 8.1, this can be explained from the fact that projects have (relatively) long lead times and thus are more likely to be subcontracting candidates.

One final remark has to be made regarding these simulation results, when comparing these to the real-life performance measured in Chapter 2. Assume that in the model a service level of 0% is claimed: i.e. the subcontracting valve is not being used, just as this currently is the case in the SEWACO situation. The real-life performance in 1998 appeared to be around 40%, whereas the simulation model provides a performance that is between 0% and 5% (see Figure 8.3 and Figure 8.4). This can be explained by the EDD sequence that is used in the experiments. In real-life, this sequencing rule has not been used at all. As is shown in Chapter 2 a lot of jobs are delivered way too early, while others are completed way beyond their due dates. Clearly, the short term pressure at repair shop level is ‘solved’ by advancing jobs that still can be delivered in time (some kind of a Last In First Out rule). Our main reason for choosing the EDD rule is that it provides a lateness with a low variance: the problems are more or less equally spread among the various projects and therefore can be considered as a ‘fair’ rule. The usage of another sequencing rule can easily be implemented in the control structure. Although such a rule may provide a higher performance, it is very likely that this goes together with an increased standard deviation of the lateness, which makes that the control mechanism becomes less robust.

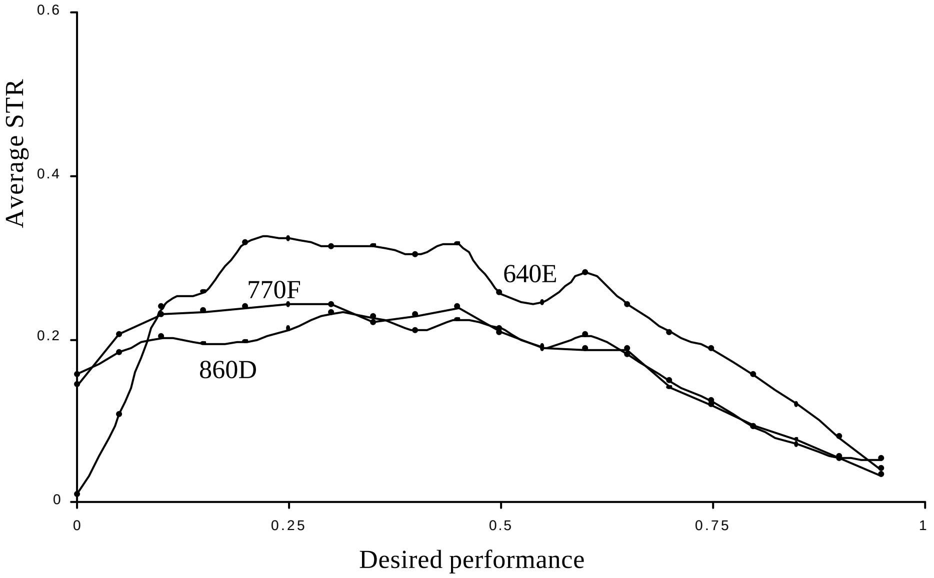


Figure 8.5. Average short term robustness for the three different repair shops with a subcontracting lead time of 10 working days.

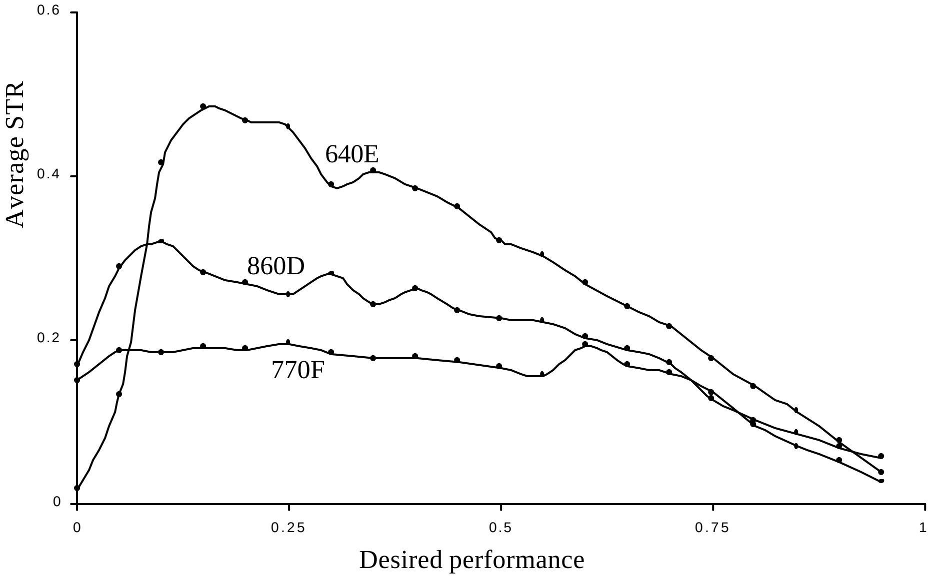


Figure 8.6. Average short term robustness for the three different repair shops with a subcontracting lead time of 50 working days.

As extensively argued before, managers are not only interested in steady-state performance. Also short term robustness is of their interest. At the beginning of this chapter, we have made the short term robustness operational by calculating STR_t: the absolute value of the difference between the target delivery performance and the service level measured over the last 50 working days. These values are aggregated by averaging them over the simulation period. Their values are shown in Figure 8.5 and Figure 8.6. Note that the lower the value of the average *STR*, the more robust the model. After all, in case the average *STR* is zero, at every point in time the service level over the last 50 days equals the target service level. The figures tell us that in both cases the control mechanism is least robust for a desired delivery performance that is between 0.2 and 0.3. Furthermore, we see that for values that are relevant in real-life - that is higher than 0.8 - the average *STR* decreases, which means that the robustness increases. As a final remark, we state that the model is very likely to become more robust in case the pulling decision has also become operational. In that case the overshoot in delivery performance can be compensated with a shorter response time, which makes that the average *STR* decreases.

The analyses in this section lead to the following conclusion. From the case description in Chapter 2 and the case analysis in Chapter 3 it has become clear that the actual real-life delivery performance is poor. Moreover, we were able to designate the characteristics which are considered as relevant in literature for these production situations, but which are not adequately used in the current control structure. Part of the poor performance could be explained by this misuse. In Chapter 5, these characteristics are used to build an explicit process model, which is combined with principles from queuing theory in Chapter 6. In this section it turns out that if the control mechanism is used, indeed the performance could be controlled (see Figure 8.3 and Figure 8.4). Especially, for relevant delivery performance requirements, that is a performance of 80% or higher, it turns out that the performance is not only controlled, but also robust with respect to the short term behavior. This conclusion holds for all the three different repair shops, and thus for varying settings of process and project characteristics. This means that at least for these settings - although it is very likely that this holds for more repair shops - the performance can be controlled. As a result, we conclude that the control mechanism is indeed capable of doing what it should do.

8.5 Validation of the analytical queuing model

We have established that the control mechanism is capable of doing what it should do (i.e. controlling the delivery performance). This control mechanism has been considered as a black box containing two main elements: feedback and an explicit process model (which is the analytic queuing model). However, it is well known that also control mechanisms with feedback and just simple knowledge (e.g., "in case we subcontract more, the delivery performance increases more") are capable of achieving performance improvement. Therefore, it is still not clear whether or not the explicit process model is a valid reflection of the real-life situation at the naval repair shops. This section contributes to the validation of the explicit process model. For that reason, we repeat the experiments described in this chapter after replacing the explicit process model by a simple model. This model is only based on the assumption that the delivery performance can be increased by subcontracting: the poorer the delivery performance (compared to the threshold value), the more work should be subcontracted. Furthermore, whereas the explicit process model also evaluates decision outcomes that are still in the pipeline, the simple model does not consider this pipeline over the lead time. After all, without an explicit process model, we are not able to evaluate and quantify the future outcomes of these decisions. In case the control mechanism with the explicit process model outperforms the mechanism with the simple model (in terms of long term delivery performance and short term robustness), we conclude that the explicit process model is an appropriate and satisfactory representation of the real-life situation.

8.5.1 Simple model

For methodological reasons, the simple model has to be chosen as simple as possible. Only in that way, the real strength of just applying feedback is best shown. Furthermore, the simple model should solely be based on past performance, because it lacks the sophisticated process knowledge to judge the expected performance of the subcontracting decisions which are in the pipeline (i.e. the projects with a due date between time t and $t + L_0$). At time t, the fraction of projects that is being delivered in time, measured during the last t_R time units, is denoted by f_{t,t_R}. At that time, there is a probability P_t of subcontracting a project with a due date at $t+L_0$. Or stating otherwise: for every project with a due date at $t+L_0$ (if any), a random value is drawn from the uniform (0,1) distribution. If this draw is smaller than P_t, the entire project is subcontracted. Otherwise, it is internally processed. The feedback character is represented by the following update procedure:

$$P_t = \begin{cases} 0 & \text{if} \quad P_{t-1} + d(a - f_{t,t_R}) < 0 \\ P_{t-1} + d(a - f_{t,t_R}) & \text{if} \quad 0 \le P_{t-1} + d(a - f_{t,t_R}) \le 1 \\ 1 & \text{if} \quad P_{t-1} + d(a - f_{t,t_R}) > 1 \end{cases} \tag{8.1}$$

with a the target delivery performance and d the control parameter that should be sufficiently small. Initially, P_0 has to be chosen. In the experiments in this chapter, we set $P_0 = 0.5$. Note that indeed the past performance influences the probability of subcontracting, but the model does not contain explicit detailed process information.

To make the model operational, appropriate values need to be found for t_R and d. To make sure that the feedback character differs as little as possible from the model with an explicit process model, the same values for t_R are used as we did in the previous section. The parameter d (which is also assumed to be different for each of the three shops) will be chosen based on experiments, because it cannot be derived analytically. For each repair shop we start with an experiment with d = 0.1 and we repeat increasing it with step sizes of 0.1 until an acceptable value has been found. A value of d turns out to be acceptable in case the simulated delivery performance more or less equals the required performance a (if possible at all) and the short term robustness also is satisfactory. Again, a subcontracting lead time is used of 50 working days. Results are tabled in Table 8.7, Table 8.8 and Table 8.9. These tables show that for some repair shops values for d exist for which the actual performance is at least in the direction of the desired performance (640E, 860D with $a = 0.9$). The tables also show that for some shops there may not be an acceptable value for d at all (with regard to shop 770F, experiments with larger values for d are omitted, but give identical results). This could be considered as a first clue that indeed explicit process knowledge is necessary to control the processes. For the experiments in the remainder of this chapter, we choose $d = 0.1$, $d = 0.7$ and $d = 0.4$ for the shops 640E, 770F and 860D, respectively.

δ	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
0.1	0.508	0.291	0.917	0.130
0.2	0.640	0.304	0.894	0.136
0.3	0.622	0.285	0.834	0.143
0.4	0.573	0.284	0.812	0.150

Table 8.7. Experimental results for repair shop 640E with $\alpha = 0.5$ and $\alpha = 0.9$.

δ	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
0.1	0.235	0.280	0.560	0.275
0.2	0.291	0.266	0.493	0.309
0.3	0.285	0.270	0.532	0.293
0.4	0.303	0.266	0.518	0.299
0.5	0.313	0.260	0.529	0.292
0.6	0.318	0.256	0.530	0.289
0.7	0.350	0.240	0.546	0.286
0.8	0.333	0.250	0.529	0.290

Table 8.8. Experimental results for repair shop 770F with $\alpha = 0.5$ and $\alpha = 0.9$.

δ	f_{last} (α=.5)	Average STR (α=.5)	f_{last} (α=.9)	Average STR (α=.9)
0.1	0.290	0.287	0.835	0.134
0.2	0.240	0.287	0.845	0.139
0.3	0.298	0.290	0.959	0.118
0.4	0.311	0.284	0.941	0.125
0.5	0.247	0.276	0.962	0.109
0.6	0.246	0.277	0.941	0.126

Table 8.9. Experimental results for repair shop 860D with $\alpha = 0.5$ and $\alpha = 0.9$.

8.5.2 Long term performance and robustness

To establish to what extent the control mechanism with a simple process model is capable of doing what it should do, experiments are carried out for the three shops with two different subcontracting lead times, ($L_0 = 10$ and $L_0 = 50$), and a required performance varying from 0 to 0.95. Results are shown in Figure 8.7 and Figure 8.8.

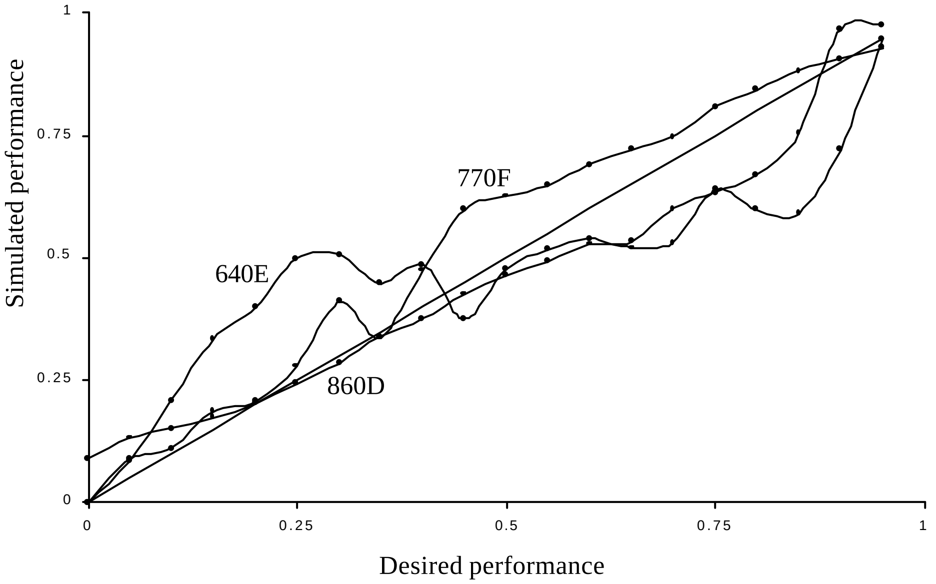


Figure 8.7. Desired versus simulated performance for the three different repair shops with a subcontracting lead time of 10 working days.

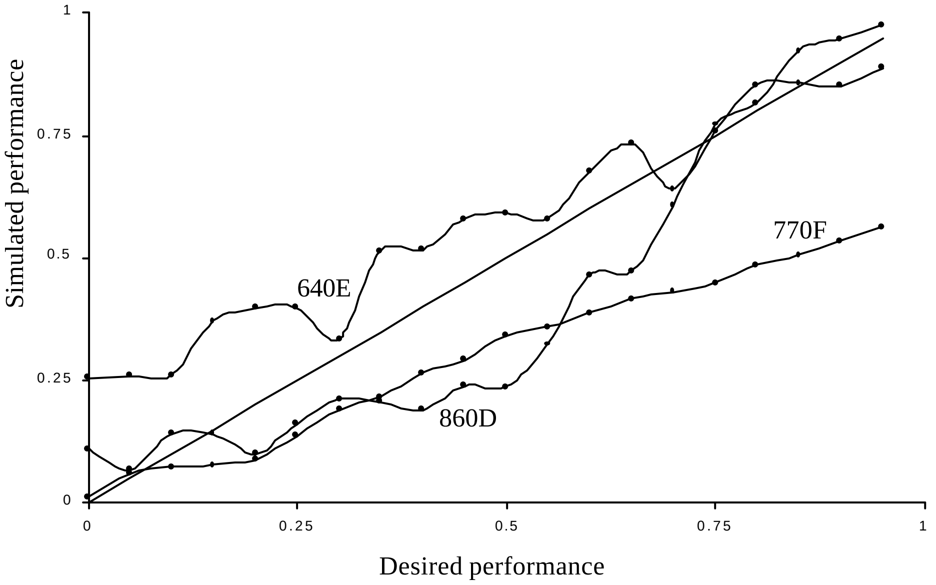


Figure 8.8. Desired versus simulated performance for the three different repair shops with a subcontracting lead time of 50 working days.

Observing and comparing these figures, we can draw the following conclusions. At first, we see that for the case $L_0 = 10$, some experiments provide an outcome that is below the target value. From Table 8.6 we have learnt that almost all jobs can be subcontracted in this case. As a result, the unavailability of subcontract candidates could be no reason for subcontracting less than minimally required. Clearly, the control mechanism does not provide the right output signals. Particularly the curve that represents repair shop 640E fails for relevant choices of a. Secondly, looking at the results for $L_0 = 50$, we immediately see that the curve representing the 770F shop, structurally differs from the desired performance. Moreover, these differences are unacceptable. Clearly, the workload in this shop has characteristics which makes it impossible to control the performance without using an explicit process model. Thirdly, if we compare the results obtained with the simple model to the results obtained with the explicit process model, we see that the explicit model outperforms the simple model with respect to the number of outcomes that are *above* the threshold value. Because of the lack of a pulling decision, we cannot prevent that the outcomes become (far) above the threshold. However, with the subcontracting decision, we should be able to prevent the outcomes falling below the target probability. Clearly, the analytic model is more capable of efficiently pointing out the right projects that need to be subcontracted.

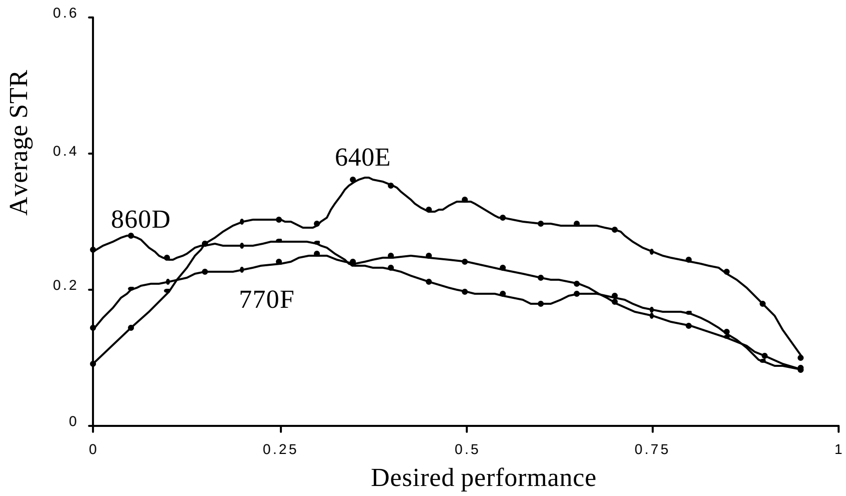


Figure 8.9. Average short term robustness for the three different repair shops with a subcontracting lead time of 10 working days.

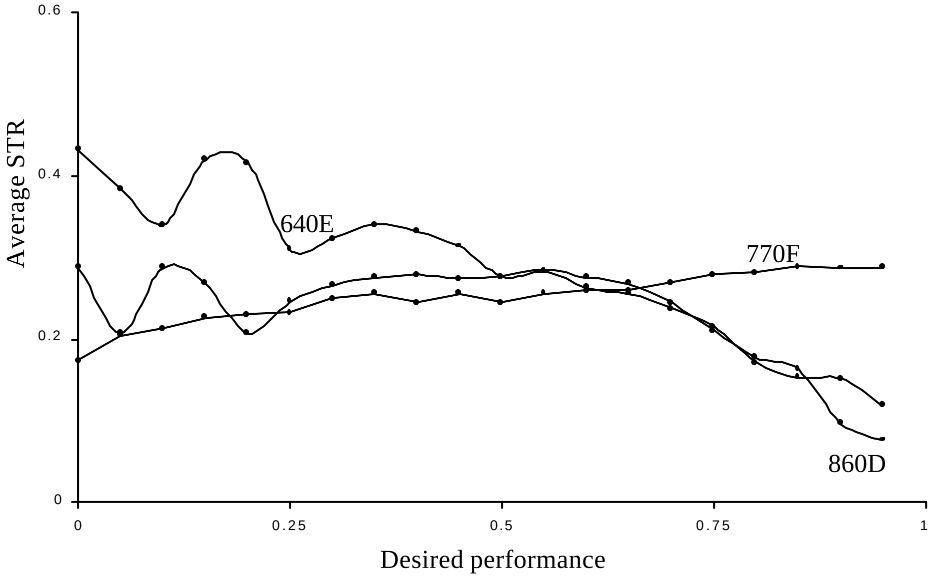


Figure 8.10. Average short term robustness for the three different repair shops with a subcontracting lead time of 50 working days.

When comparing the average short term robustness for the case $L_0 = 10$ (Figure 8.9) to the robustness for the case $L_0 = 50$ (Figure 8.10), at first glance we see that the simple model gives a more robust performance for small lead times. This can be explained by the fact that the shorter the lead time, the shorter the ignored pipeline. Furthermore, we see that the robustness for repair shop 770F with a lead time of 50 working days is very poor. This was also concluded from Figure 8.8. When comparing these figures to the robustness of the analytical model (i.e. Figure 8.5 and Figure 8.6) we see that for small values of a there is not a great difference between the models in terms of the short term robustness. However, when we increase the desired performance and come to relevant requirements (i.e. $a \geq 0.8$), we see that the explicit process model outperforms the simple model.

8.5.3 Conclusion

In this section we have taken a closer look at the validation of the explicit process model. When comparing it to the simple model with no explicit process knowledge, we see that the control mechanism with explicit process model is more capable of doing what it should do. Both for the cases with a short and a large lead time, we see that the explicit process model is more accurate: the number of cases where the outcome is at least as high as the desired performance is substantially larger for the control mechanism with an explicit control model. Moreover, we see that this difference in accuracy increases as the subcontracting lead time increases, which can be explained by the fact that the explicit process model explicitly evaluates the state of the system for the next L_0 time periods, whereas the simple model does not. Concerning the robustness of the control mechanism we conclude that the control mechanism with the explicit process model is more robust with respect to the delivery performance than the simple model.

The main question that is addressed in this section is the validation of the explicit process model. Obviously, the experiments have shown that the explicit process model contains information that is more valuable than just using a rough conjecture of the process. Moreover, using the explicit model results into outcomes that are both more accurate and robust. Therefore, we conclude that the analytic model is an acceptable representation of the real-life processes at the maintenance organization.

8.6 Evaluation of the control mechanism

In the introduction we already have stated that we are not only interested in the *validation* of the control mechanism, but also in the *evaluation* of this mechanism. In the evaluation phase we take a closer look at the implications of the control mechanism. This mainly concerns questions like the costs of subcontracting (what part of the workload needs to be subcontracted to achieve a satisfactory performance?) and the resulting subcontracting behavior (what is the subcontracting intensity and how much has to be subcontracted each time?). These questions will be answered by experiments. Note that this evaluation merely addresses experimental results, rather than already implications for implementation.

Again, we look at both the analytic and the simple model to establish the advantage of using an explicit process model. As we have seen that the analytical model is more accurate in pointing out the right projects to subcontract, we also are interested to what extent less jobs should be subcontracted to achieve a certain performance, as compared to the simple model.

In the remainder of this section we look for every repair shop at the curve depicting the relationship between the fraction of the workload (expressed in working hours) that is being subcontracted and the achieved performance. For the construction of these figures we make use of the experiments that are carried out in the previous part of this chapter. For each combination of repair shop and subcontracting lead time (10 and 50 days), we plot for all requested probabilities the actual fraction of subcontracted workload against the achieved performance. Results are presented in the next sub-sections.

8.6.1 Repair shop 640E

At first glance, we see from the experimental results for the explicit process model (see Figure 8.11) and the simple model (see Figure 8.12), that in both cases the delivery performance is an increasing function of the fraction of subcontracted workload. Looking first in more detail at the results for the explicit model, it follows that for fractions smaller than 0.2, the model with a long lead time is more effective than the model with a small lead time. After all, in case a certain fraction of subcontracted workload is considered, it appears that a higher performance is achieved when the lead time is longer. At first sight this may be strange: increasing the lead time (i.e. increasing the uncertainty) gives better results. However, if a closer look is taken at the type of projects that are subcontracted, we see that in the case of a lead time of 50 working days, projects are subcontracted with substantially smaller processing times. Thus, the model with a longer lead time delivers more projects in time for the same amount of money. The same argumentation is applicable to the control mechanism with a simple model. We are aware that this is only an observation for this case and not generally valid! But clearly, characteristics of the projects on hand play an important role.

Comparing the explicit and simple model, it follows that the shapes of both curves are more or less equal. However, from the previous section we have learnt that the performance becomes more robust in case an explicit process model is being used (for the same amount of money, as follows from Figure 8.11 and Figure 8.12).

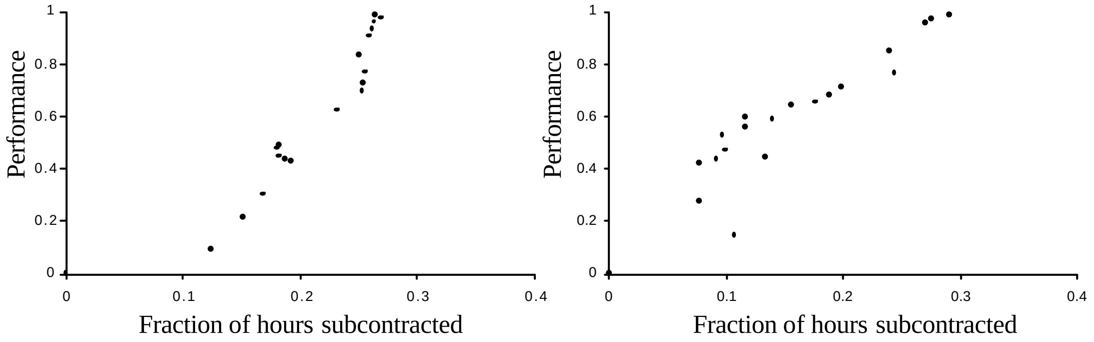


Figure 8.11. Cost-performance relationship for repair shop 640E obtained with explicit process model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

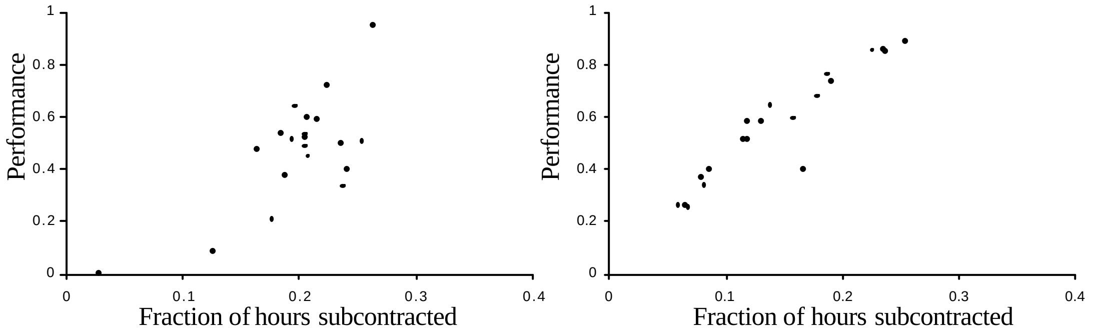


Figure 8.12. Cost-performance relationship for repair shop 640E obtained with simple model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

8.6.2 Repair shop 770F

For repair shop 770F, we already have seen that an explicit process model is necessary to control the processes in case the subcontracting lead time increases. This also follows from Figure 8.14. Clearly, the process and project characteristics of this shop make that explicit process knowledge is necessary in order to achieve a controlled performance. Furthermore, in accordance with our expectations, the subcontracted hours of work increase in case the subcontracting lead time increases, both for the explicit and simple model (see Figure 8.13 and Figure 8.14). Moreover, considering a lead time of 50 working days, we see that the explicit process model provides the same performance for less money.

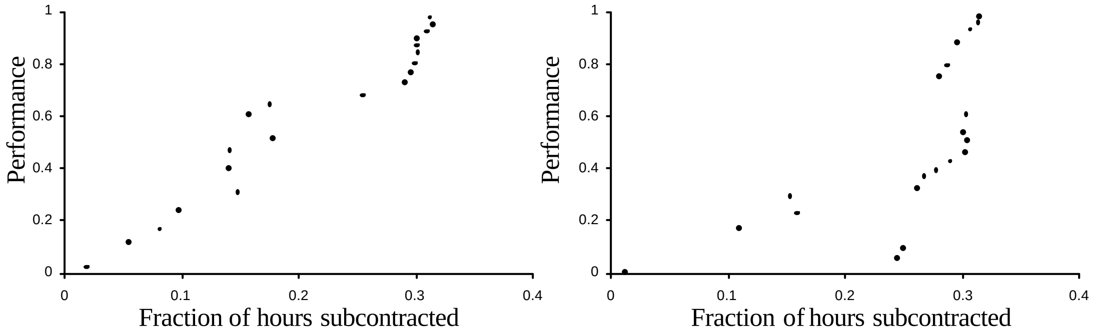


Figure 8.13. Cost-performance relationship for repair shop 770F obtained with explicit process model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

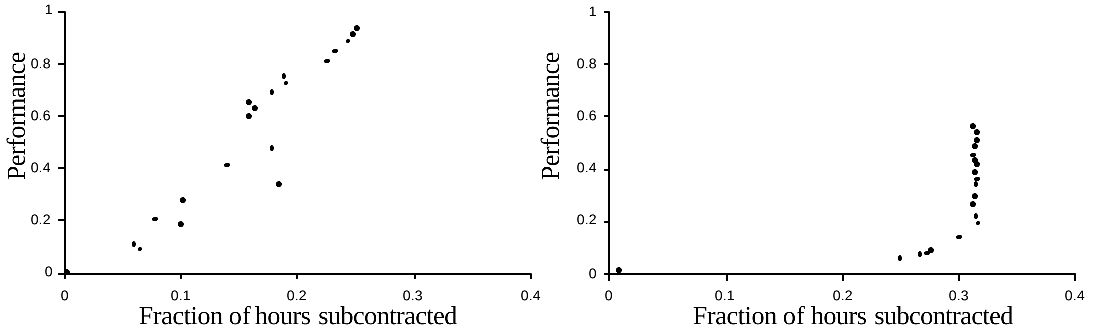


Figure 8.14. Cost-performance relationship for repair shop 770F obtained with simple model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

8.6.3 Repair shop 860D

Comparing the results obtained with the explicit process model (Figure 8.15) to the results obtained with the simple model (Figure 8.16), a surprising resemblance is found. After all, the models give both for a small lead time of 10 working days and a long lead time of 50 working days, nearly equal cost-performance curves. Furthermore, it appears that subcontracting becomes more expensive as the subcontracting lead time increases. This is according to our expectations. Referring to the figures in the previous section, we can conclude that in the case of repair shop

860D, the use of an explicit process model makes that for the same amount of money a more robust performance is obtained: the amplitudes of overshoot and undershoot are smaller.

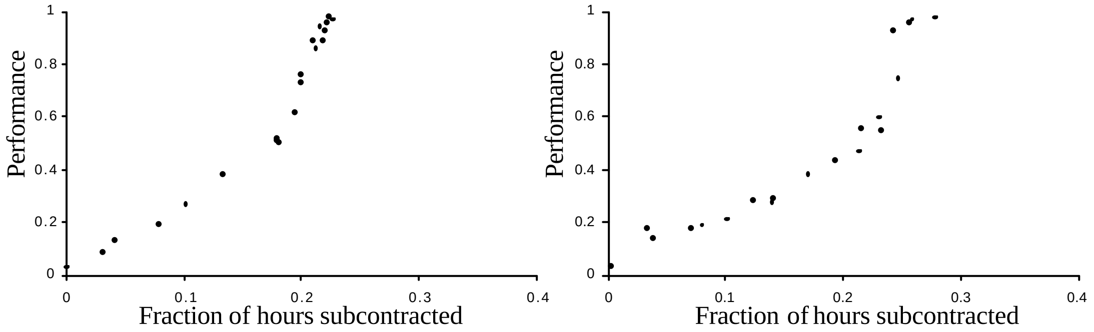


Figure 8.15. Cost-performance relationship for repair shop 860D obtained with explicit process model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

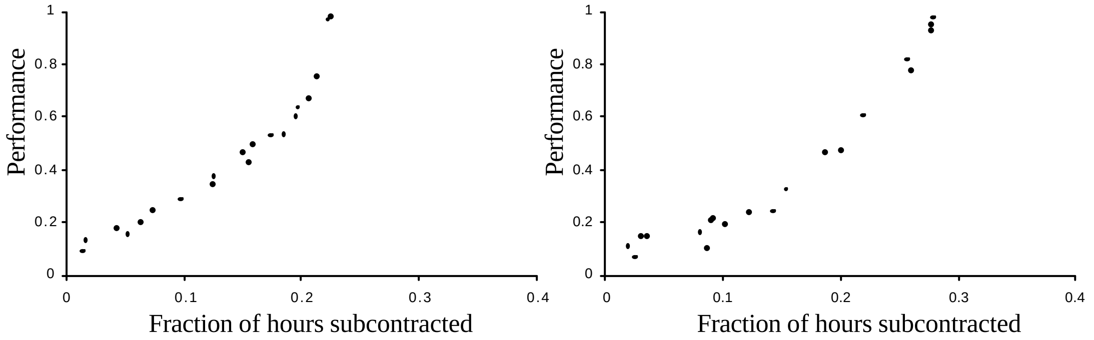


Figure 8.16. Cost-performance relationship for repair shop 860D obtained with simple model for subcontracting lead time of 10 (left hand) and 50 (right hand) working days.

8.6.4 Conclusion

At first glance, we do not see that much resemblance between the relationships for the different repair shops, besides the fact that the performance is an increasing function of the fraction subcontracted. Of course, this is to be expected: the less jobs kept internally, the easier it becomes to achieve a higher performance. Clearly, based on these experiments, we can conclude that each repair shop - and thus each combination of process and project characteristics - gives a different shape of the cost-performance curve. Remarkable resemblance between the three cases is that although their shapes differ a lot, in all cases it is sufficient to subcontract up to 30% of the workload in order to get a satisfactory performance that is above 95%, when a realistic subcontracting lead time of 50 working days is considered. Of course, these cost-performance relationships depend on the characteristics of the workload. It cannot be said that these figures are also valid for the other repair shops that have not been under consideration in this chapter.

The main question in this section is to check whether the results are satisfactory for application in real-life. From the figures in this section we have seen that - using the fair EDD sequencing rule - in case the subcontracting valve is not being used, a very poor performance is obtained. For all the tested repair shops, this performance is between 0% and 5% and can be considered as dramatic. Therefore, it is concluded that subcontracting is necessary. In case the subcontracting valve indeed is being used, we see that - due to the shape of the cost-performance curves - on average an increase of 1 percent point in the fraction subcontracted causes an increase in performance that is higher than 1 percent point. Moreover, each repair shop is capable of achieving an excellent performance in case about 30% of the workload is being subcontracted. Clearly, the amount of work that should be subcontracted turned out to be controlled. It turned out that the explicit process model achieves a more robust performance when compared to the simple model. In some cases, this more robust performance even costs less money when compared to the simple model.

Finally, some comments have to be made about the assumptions and choices that have been made for the testing framework in Chapter 7. We have assumed that the internal capacity levels have been fixed at the levels that are currently used at SEWACO. It may be very well possible that re-dimensioning the internal and external capacity results in lower overall costs than the cost-performance curves in this section show. However, because we are primarily interested in validation of the control rule and the curves of the costs-performance relationships, we do not consider this trade-off. But it needs to be emphasized that this is a very interesting topic for further research once the above stated questions have been answered.

8.7 Side effects of control mechanism

The results of the evaluation show that some percentage of the workload has to be subcontracting in order to achieve a satisfactory performance. In this section, we discuss two side effects of subcontracting. We take a closer look at the capacity utilization and the resulting subcontracting behavior. This latter question concerns the intensity of subcontracting. The discussions are only based on the results obtained with the control rule containing the explicit process model.

8.7.1 Capacity utilization

A high delivery performance usually goes together with a low capacity utilization. After all, we have seen that the delivery performance is an increasing function of the subcontracted workload and thus a decreasing function of the internal workload. The environment in which we are operating offers us an excellent way to control not only the delivery performance, but also the capacity utilization: failed repairables can be pulled. The designed control structure even explicitly contains a decision function that aims at achieving a sufficiently high utilization. In that way, both the internal resources are used more efficiently and the inventory levels with ready-for-use repairables increase: two birds are killed with only one stone. Although the needed information for the pulling decisions is not available in this study, we still can investigate the resulting internal capacity utilization due to subcontracting. Moreover, this means that the actual internal capacity utilization would only have been higher in case the pulling decision indeed (successfully) is used.

The capacity utilization is defined as the percentage of the total available capacity (that is the number of operators times 5.2 hours times the number of simulated working days) that has been spent on processing jobs. Given the values that are listed in Table 8.10, there does not seem to be any problem at all. Especially not if one reminds that the pulling decision is not even used in these experiments. However, if a closer look is taken at the dynamic behavior of the expected remaining workload per operator in the shops, another view comes up.

Requested Service	Capacity utilization 640E	Capacity utilization 770F	Capacity utilization 860D
0.95	97.7%	100%	97.2%
0.90	99.9%	100%	99.0%
0.85	99.9%	100%	99.2%
0.80	100%	100%	99.8%
0.75	100%	100%	99.8%
0.70	100%	100%	100%
0.65	100%	100%	100%

Table 8.10. Requested service versus capacity utilization in the three repair shops considering a subcontracting lead time of 50 working days.

In Figure 8.17, we see that the expected remaining workload per operator for shop 640E is - except for the last days of the simulation period - above zero. However, it appears that the workload fluctuates around a trend with a negative slope: the operators in the shop are in severe danger of becoming idle shortly after the simulation period. Note however, that the expected remaining workload does not include future projects that are already known, but not yet available to be processed. Furthermore, we see that the expected remaining workload fluctuates in time, as could be expected from the dynamic and stochastic environment in which the shops are operating.

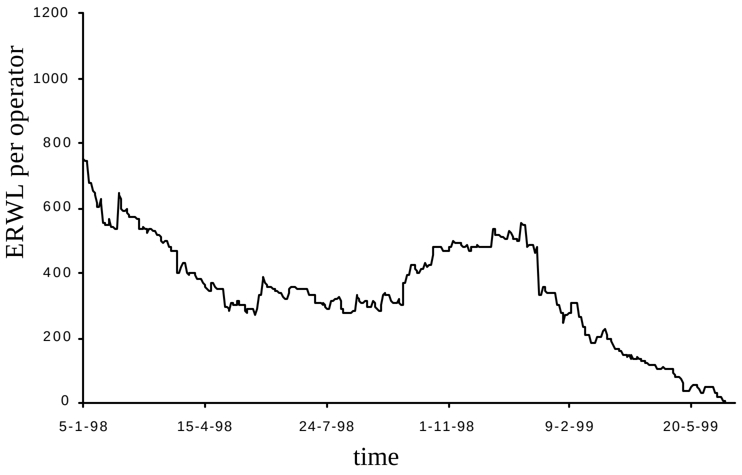


Figure 8.17. Expected remaining workload for the 640E repair shop when requesting a 95% service level and a subcontracting lead time of 50 working days.

For repair shop 770F, a totally different dynamic workload behavior comes up (see Figure 8.18). Obviously, here is absolutely no danger of getting idle in the near future. On the contrary: it looks like that the remaining workload per operator fluctuates around a slightly increasing trend.

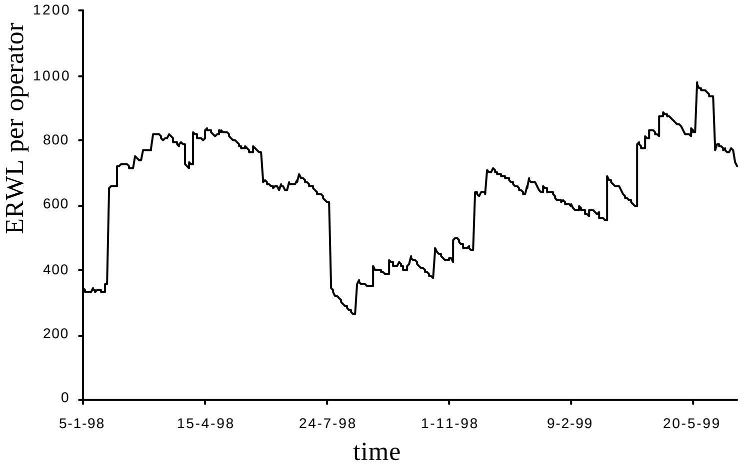


Figure 8.18. Expected remaining workload for the 770F repair shop when requesting a 95% service level and a subcontracting lead time of 50 working days.

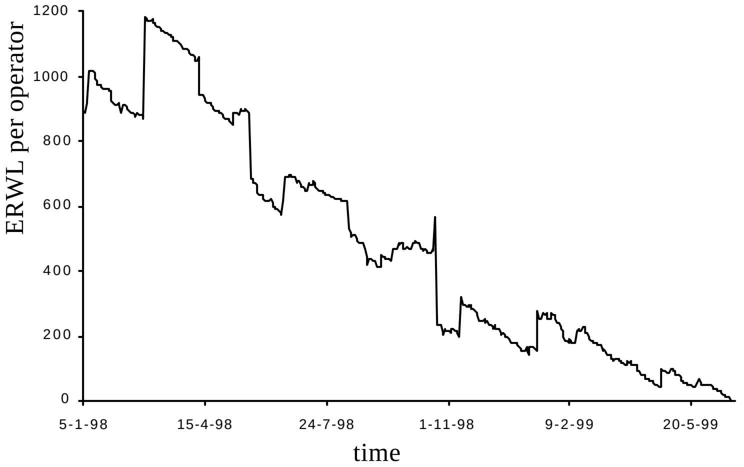


Figure 8.19. Expected remaining workload for the 860D repair shop when requesting a 95% service level and a subcontracting lead time of 50 working days.

Repair shop 860D seems to be in real trouble (see Figure 8.19). It is very likely that the operators become idle in the near future. Further investigation is necessary to examine the influence on the utilization rate for this shop. In case the utilization rate becomes unacceptable low, it should be considered to exchange internal capacity for an increased subcontracting budget. This might result in lower total costs.

Summarizing, we see that the dynamic behavior of the workload significantly differs for the three different repair shops. For shops 640E, 860D the above sketched scenarios are in conformity with the expectations. After all, where in the current real-life situation all jobs are internally processed, in the new situation part of this workload is being subcontracted: less work is being processed by the same number of operators. Whereas in some cases the new level of capacity utilization is acceptable, this may not be the case for all repair shops. These structural problems need to be escalated to the next higher echelon in the control structure. Re-dimensioning the internal capacity (i.e. decreasing the internal capacity) may be an adequate decision to achieve again both a satisfactory performance and a satisfactory utilization of internal capacity. If the decrease in costs of the internal capacity exceeds the increase in subcontract budget, this is an effective decision.

8.7.2 Subcontracting behavior

As a final topic of this chapter, we investigate the subcontracting behavior. The model puts no restrictions on the subcontracting pattern. We only assume that both the subcontractor’s internal capacity is at least sufficient to process all the jobs offered by the maintenance organization and that all technical (sub-)systems are returned in time with a hundred percent reliability. As a result, of course, the subcontracting behavior (in terms of when and how much to subcontract) is determined by the model. To investigate this behavior, we carried out experiments for the three different repair shops with the parameter setting $a = 0.85$ and $L_0 = 50$.

In Figure 8.20, we see the subcontracting behavior for repair shop 640E. This behavior turns out to be quit lumpy: periods of hardly any or no subcontracted jobs are followed by periods with a lot of subcontracted jobs. Furthermore, in a substantial number of cases, the subcontracted workload comprehends a large amount of work. The resulted pattern can be explained by looking more carefully at the initial workload. This shop was faced with a large backlog at January 5, 1998. As a result, a lot of jobs needed to be subcontracted, which also created a lot of slack for other future projects. Clearly, it takes about a year to consume this slack: at the beginning of 1999 again many jobs are subcontracted.

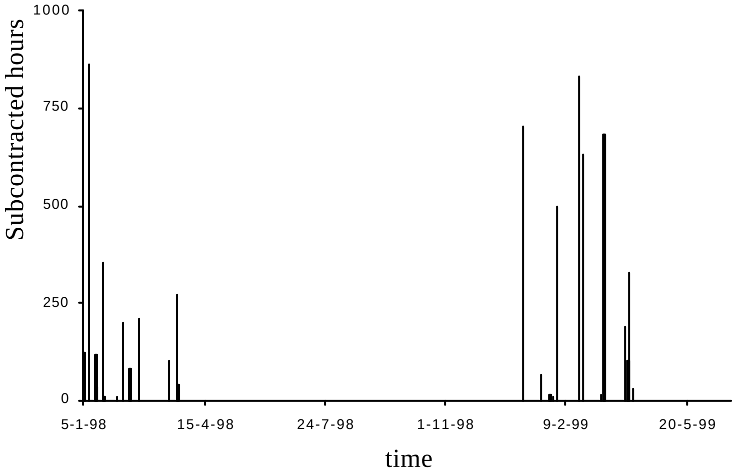


Figure 8.20. Number of hours subcontracted per day for shop 640E in case of a desired performance of 85% and a subcontracting lead time of 50 working days.

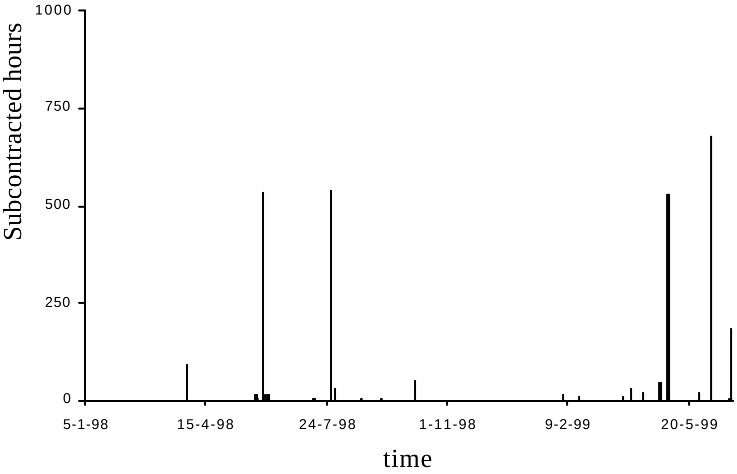


Figure 8.21. Number of hours subcontracted per day for shop 770F in case of a desired performance of 85% and a subcontracting lead time of 50 working days.

In Figure 8.21, the subcontracting pattern is shown for repair shop 770F. This pattern looks like the pattern for shop 640E, although the intensity is lower. This latter fact can be explained from the fact that the overall workload for 770F is much smaller than the workload for 640E. The pattern in Figure 8.21 can be explained by the characteristics that we have seen in Chapter 2. Apparently, there is no backlog at January 6, 1998: no jobs are subcontracted at the beginning of the experiment. We already have seen that peaks in demand occur before the summer and at the end of the year. The two first large peaks of about 500 hours are subcontracted as soon as the output signal includes the expected delivery performance at the end of the year. That is, t_R working days before the end of 1998. The same argument is applicable to the pattern for shop 860D in Figure 8.22. Exactly L_0 working days before December 31, 1998, about 1000 hours of work are being subcontracted.

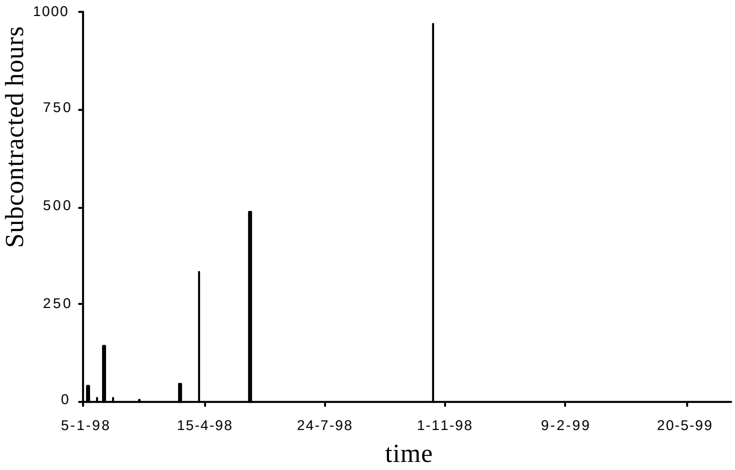


Figure 8.22. Number of hours subcontracted per day for shop 860D in case of a desired performance of 85% and a subcontracting lead time of 50 working days.

Summarizing, all figures show that every now and then, large projects should be subcontracted in order to achieve a satisfactory performance. This may cause serious problems. Subcontractors may not be willing or not be capable to accept large projects, because they may have the same capacity inflexibility as SEWACO. If the subcontractor still accepts the jobs (perhaps, because of an attractive price), the problem has only been moved and thus the delivery performance will not improve. Therefore, SEWACO benefits by a careful selection procedure. Several types of solutions can be thought of to solve this problem. Firstly, there may be more organizations with the same core business (e.g. original equipment manufacturers, maintenance organizations of allied naval forces) which are faced with the same problems of a fluctuating demand for capacity. This means that periods of capacity shortage at SEWACO may coincide with periods of capacity surplus at the other organizations. As a result, capacity requirements can be smoothed for both organizations by subcontracting. Secondly, it may be possible to split the work package that needs to be subcontracted into smaller packages. After all, usually a large project contains several technical systems that all need to receive maintenance. Smaller packages may be easier to subcontract by dividing it among various subcontractors. We come back to this problem in the next chapter.

8.8 Summary

In Chapter 2 we have described the processes at a real-life maintenance organization. The environment of this organization has been classified as dynamic and stochastic. Although the performance was found to be uncontrolled and unsatisfactory at the time this research started, we found several indications that the performance could be controlled when applying appropriate decision functions. After all, about 60% of the workload has been classified as “planned”. The developed control structure is aimed at controlling the performance. We are aware that part of the processes are not and could not be modeled. Furthermore, the dynamic aspect of the processes is important. To overcome these difficulties, principles from control theory are used. It turns out that a control mechanism containing an explicit control model and a feedback loop is capable of producing a robust delivery performance. This indicates that the control rule is indeed sufficiently rich: the right characteristics are used in the right way.

With respect to the control mechanism we have seen that the process model is an important factor. Experiments have been conducted both with an explicit process model (i.e. the developed queuing model in Chapter 5) and with a simple process model. Comparing these results, we can draw some very important conclusions. At first, the capability of achieving the required performance increases in case an explicit process model is used. Using the simple process model, the required and achieved performance varied a lot for some instances. Secondly, the inclusion of an explicit process model has only minor advantages in terms of costs (if any at all). However, the advantages in terms of robustness are considerably. Obviously, the stated conjecture that shop information should be used in order to achieve a satisfactory short term robustness, is correct. And finally, it has been shown that in case the subcontracting lead time increases, it is more advantageous to use the explicit process model. This is due to the fact that the queuing model explicitly considers the already made subcontracting decisions for which the outcomes are in the ‘pipeline’.

9

Behavior of the Control Structure under Relaxed Assumptions Regarding Demand and Resource Structure

9.1 Introduction

The control structure that we have introduced in Chapter 4 and for which we have developed the subcontracting and pulling decisions in Chapter 6 has been validated and evaluated for three different repair shops at SEWACO. We have shown that the delivery performance can be controlled for the three tested repair shops, regardless of the project and process characteristics. This process can be considered as part of the internal validation of the control model. With this internal validation, we have shown that the structure is at least valid for repair shops with project and process parameter values for the demand and capacity characteristics as they appear in the SEWACO case. In this chapter, we take a closer look at production situations with relaxed demand and capacity assumptions. Using the knowledge obtained in this thesis so far, we point out to what extent it is likely that the developed control structure still is applicable under these relaxed assumptions. And if not, we suggest possible adaptations for the structure. It should be emphasized that the suggested directions for control structure adaptations are not theoretically validated with experiments, nor do we have exact estimations for the improvement on the performance indicators. Instead, a deductive approach - based on reasoning - is used in this chapter.

In the remainder of this chapter, we first consider production situations which do not structurally differ from the considered situations in this thesis, apart from having one dominant demand class (expressed in terms of percentage of hours to be spent on a class). Subsequently, we take a closer look at instances for which at least one of the characteristics described in Chapter 1 no longer applies. First, situations are considered with less demand classes. Secondly, situations are considered which have four instead of three demand classes. And finally, situations are under scrutiny in which the subcontracting characteristics are altered.

9.2 Dominant demand classes

The control structure has been validated with real-life data obtained in three different repair shops. The actual mixtures of the demand classes in these shops can be considered as relevant. After all, these mixtures occur in real-life. However, also more extreme mixtures of demand classes - for example in which one demand class dominantly is available - satisfy the characteristics introduced in Chapter 1. Because these differ from the analyzed real-life case, the conclusions made in the previous chapter may not be extrapolated to the case with an extreme mixture of demand. In what can be considered as a start of a sensitivity analysis, it is described for three different cases to what extent the control structure still may be valid and effective and what adaptations (if any) need to be accomplished to ensure a satisfactory performance. In each of these cases, one demand class is dominantly present.

9.2.1 Dominant fraction of emergency projects

Consider a repair shop in which the emergency projects are dominantly available (e.g. 80% of the available working hours are spent on emergency projects). This situation drastically differs from the analyzed repair shops 640E, 770F and 860D, which respectively face an emergency project share of 22%, 24% and 53%. This situation may - for example - occur in a repair shop which is mainly entrusted with the maintenance of electronic equipment. This type of equipment usually has a constant failure rate, which makes that shortening the interval between two consecutive preventive overhauls does not reduce the number of equipment breakdowns. As a result, a high share of emergency projects cannot be reduced by choosing for a different maintenance concept. This demand mixture has the following three consequences for the controllability of the processes and the projects in the shop.

Firstly, because emergency jobs have short allowed lead times, the percentage of all jobs that are potential candidates for subcontracting (i.e. for which the job's time until delivery exceeds the subcontracting lead time) decreases. For that reason, the speed of response of the control mechanism decreases, like we have seen in the previous chapter. Secondly, the arrival intensity of emergency projects increases. As a result, the queue length increases both in terms of the expectation and the variance of the working hours in the queue. Furthermore, the increased arrival intensity also increases the influence of chosen sequencing rules on the operational availability of the technical systems. Thirdly, the increased uncertainty makes that relatively more slack should be assigned to the preventive projects. This results in more idle time, because of the absence of large inventory levels of non-ready-for-use repairables. Consequently, subcontracting goes not only together with subcontracting costs, but also with unavoidable idle time costs. A delivery performance that is satisfactory may only be achievable in case unacceptably high costs are made.

These consequences may have serious and negative consequences on the delivery performance. To overcome these problems, we advocate the following adaptations of the control structure. Firstly, an additional performance indicator should be created, which should be used in addition to the delivery performance. Because of the increased influence of the handling of emergency projects on the operational availability, a response time (i.e. emergency project waiting time plus service time) related indicator should be added. This indicator, for example the fraction of technical systems currently not available because of failures, can be controlled by sequencing rules. Secondly, to decrease the value of this new indicator and to increase the operational availability, corrective projects arising from the same technical system need to be clustered. Subsequently, the newly clustered projects need to be processed according to the SPT rule. That is, the clustered project with the shortest (expected) processing time needs to be processed first. Thirdly, because of the increased probability of idle time, a new reflection is needed on the costs of subcontracting versus the achieved performance. To what extent do the costs counterbalance the provided service? And how bad is it if - measured at preventive project level - the tardiness is drastically reduced but still above zero? And finally, the management should consider to split the class of emergency projects into two or more sub-classes, in order to still be able to control the processes as much as possible. A possible classification in this area distinguishes two types of emergency projects: one type is labeled as ‘critical’ and the other as ‘non-critical’. Projects are assigned to one of these types, such that postponement of the critical projects endangers the safety of the personnel at the ships whereas the postponement of non-critical projects only postpones certain assignments of a mission. Of course, the critical projects have absolute priority over the non-critical projects. Related research on this topic (see e.g. Cobham 1954, Kesten & Runnenburg 1957, Takács 1964, Conway *et al.* 1967, Stanford 1997) shows that the expectation and variance of the highest priority emergency project throughput times decrease by this adapted classification. In the next section, we come back to the technical consequences of this newly added class.

Summarizing, the dominant presence of the emergency class heavily increases the uncertainty and thus complicates the control problem. In terms of countermeasures, the control structure can be adjusted and extended in such a way, that the maintenance organization may be able to cope with this increased complexity. Further research is necessary to judge the effectiveness of the control structure.

9.2.2 Dominant fraction of preventive projects

In case the preventive projects are dominantly present (that is over 80% of the workload comprehends preventive projects), there are three main effects for the planning and control. Firstly, we gain control options (compared to the mixtures considered in Chapter 8), because relatively more jobs are potential candidates for subcontracting. Secondly, the fraction of emergency projects decreases, which decreases the uncertainty and increases the controllability. Finally, we lose control options due to the fact that the fraction of repairables decreases, which makes that the production system becomes less controllable.

Overall, it seems that the control structure is still applicable, without any substantial adaptations. However, a trade-off between costs and performance should be considered that is similar to the one discussed in the previous section, before stating that the results indeed are satisfactory (or not).

9.2.3 Dominant fraction of repairable projects

Finally, we discuss the topic of a dominant presence of repairable projects. This situation for example occurs in a repair shop where (almost) all technical sub-systems are repairables, with procurement costs that are relatively low. Consequently, a sufficient number of copies of the repairables can be kept (i.e. depot, ships, repair shops) to allow that all preventive and emergency projects are carried out by repair by replacement. As a consequence, most of the time is spent on the recovery of the failed repairables. The main difference - in terms of control - between the analyzed situation in Chapter 8 and this situation is that for the latter situation most of the projects are inventory level driven instead of customer order driven.

This has several implications for the control structure that should be used for effective decision making. Of course, the operational availability is still the key priority. To what extent preventive projects promptly are delivered is no longer a satisfactory performance indicator. The operational availability is better expressed by the fill rate: the fraction of requests for repairables that can be met from stock. Clearly, the focus shifts from controlling the length of the queue with preventive projects to controlling (or balancing) the inventory levels of the ready-for-use repairables. Two decision functions are crucial then. Firstly, the choice of the sequencing decision concerning the repair of non-ready-for-use repairables has a large influence on the fill rate (see e.g. Hausman & Scudder 1982, Albright 1988, Lee 1989, Schneeweiss & Schröder 1992). These studies show that the sequencing decision should be based on - among other things - the number of ready-for-use repairables, failure rates and repair times, and should aim at ensuring sufficiently high inventory levels of ready-for-use repairables after L_0 time units. Once a sequence - which may change dynamically over time due to new failures - has been established, the second decision function becomes effective: subcontracting. The queuing model can be used to compute to what extent these repairables indeed can be restored in the following L_0 time units. If not, part of these repairables should be subcontracted.

Although the focus has changed for such problems, it seems that the control structure is still applicable with some small adaptations. Other performance indicators are used, and the repairables sequencing decision has become more important. The queuing model can still be used to determine whether subcontracting is necessary or not. However, further research should tell whether the decisions are still effective.

9.2.4 Summary

As a very brief start of the internal validation, we have discussed more extreme demand mixtures than that we have considered in the validation of the control structure at the real-life repair shops. Without claiming to be complete, the discussions suggest that the control structure - with some adaptations - is applicable in the adjusted cases. However the influence of the different decision functions on the overall performance, heavily depends on the considered demand mixture. The suggested adaptations can be used as a guideline for further research to establish the effectiveness of the adaptations.

9.3 Demand classification

Another interesting topic comes up if the question arises to what extent the obtained knowledge still is useful in case the demand classification is altered. Throughout this thesis we have considered classes with preventive, corrective (or emergency) and repairable projects. Moreover, the construction of the control structure is entirely based on these three classes and their characteristics. In this section we consider three production situations with only two classes and two situation with an additional fourth class.

9.3.1 Omitting emergency projects

The emergency projects have played a role of disturbing factors in this study. They randomly arrive and immediately consume capacity. As a result, the progress of planned preventive projects is disturbed. Furthermore, we have seen that are no typical control actions that can be taken with respect to the emergency projects to influence the overall delivery performance.

For a situation where only planned and repairable projects are encountered, this means that no control possibilities are lost. After all, only planned and repairable projects are candidates for subcontracting. On the other hand, we do lose uncertainty, because we no longer face the random arrival of emergency projects. Consequences for the control structure can be found at two different levels. Firstly, at the operational level, less slack needs to be assigned to planned projects. Secondly, we see that the two-dimensional queuing model that has been developed in Chapter 5, can now be reduced to a one-dimensional queuing model: horizontal transitions are no longer possible. Or, stated in the terms that are introduced in Chapter 5, the state of a repair shop can now be described by the random value $T(n,m)$, where m is always zero.

Summarizing, we see that omitting the class of emergency projects reduces both the complexity of the algorithms and the uncertainty in the (future) work load. As a result, there seems to be no problem in using the original control structure still. Moreover, the reduced uncertainty may even result in lower subcontracting costs.

9.3.2 Omitting preventive projects

Consider the case where no preventive maintenance is carried out. For example, this is the case in an environment where the technical systems have a constant or a decreasing failure rate. In the first case, preventive maintenance does not increase the reliability of the technical systems, whereas in the second case, preventive maintenance even reduces the reliability of the systems. As soon as a technical system fails, the non-ready-for-use repairable is taken out and replaced by a ready-for-use repairable. This is regarded as an emergency project. In case there are no emergency projects in the shop, time is available for restoring the non-ready-for-use repairables into a ready-for-use condition. Each failure goes together with exactly one exchange of repairables and with one repairable project that is carried out at a later point in time. Assuming that the repair takes (much) more time than the exchange, the larger part of the resource capacity is spent on repairable projects.

Summarizing, we see that the problem is quite similar to the problem that we have discussed in Section 9.2.3, where the class of repairables is dominantly present. For results, we refer the reader to that section.

9.3.3 Omitting repairable projects

Finally, we consider the situation in which we are faced with only the classes of preventive and corrective maintenance. The lack of repairables makes that the pulling decision can no longer be used to prevent idle time. Furthermore, the class transition function that is defined in Chapter 4 to transfer repairables to the preventive class is no longer necessary. Of course, the subcontracting decision is still applicable to this situation. The only substantial problem seems to be that no longer a high utilization rate can be achieved because of the lack of repairables. As we have discussed in Section 9.2.1, also in this situation the management should reflect on the costs of subcontracting and idle time versus the achieved performance, before deciding whether or not the use the control structure.

9.3.4 Including an additional class of projects

Another type of problem comes up in case we add a new type of demand class to the production situation. The production situation observed in this research concerns three types of demand classes. To control the performance with respect to each of the three classes, in fact only decisions can be taken to control the length of the queue of the preventive projects. The workload can be reduced by subcontracting and the workload can be increased by pulling repairables (see Figure 9.1). These decisions partly depend on the characteristics of the other two buffers. In terms of queuing theory, each repair shop can be seen as a shop with three buffers. Moreover, the buffer of the preventive projects has a ‘finite’ capacity. The exact size of this buffer depends on the projects (and their due dates, workload and uncertainty) that are currently on hand. In this section we first look at the consequences of including a demand class which has no additional control possibilities. Clearly, there is nothing else to do than just taking these projects into account when controlling the buffer with preventive projects. Secondly, we look at a situation with an additional demand class *with* additional control possibilities.

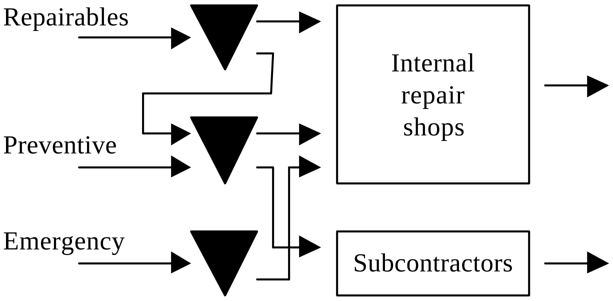


Figure 9.1. Schematic overview of the job flows.

Suppose that we face a fourth class as described in Section 9.2.1. Thus, emergency projects are divided into the classes ‘critical’ and ‘non-critical’. Indeed this new class does not introduce new additional control options. Therefore, the use of the developed control structure still seems to be justified. Only a technical adaptation is required in order to still be able to control the processes. Regarding the subcontracting decision, sufficient slack should be created for both types of future emergency projects. This decision can again be based on the analytical queuing model. Because it is very unlikely that both types of corrective projects have equal arrival and repair intensities, the underlying process model should be adapted towards a three-dimensional Markov representation. In this new model, the state (a,b,c) respectively describes the number of preventive, critical corrective and non-critical corrective phases left in the system. Note that this will not drastically increase the complexity of the computations.

Another type of problem occurs in case we add a new class with not only different characteristics, but with also different control possibilities. To illustrate this situation, we add such a new class and discuss the implications for the control structure. Suppose that the added fourth class consists of corrective projects that can be subcontracted with an almost zero lead time. This is the case if there are some systems which are that common in industry, that repair personnel can be found and hired on a daily base. Furthermore, assume that these systems have a constant or decreasing failure rate. As a result, preventive maintenance is not carried out for these systems and thus, these systems are not found in the queue with preventive maintenance. Finally, we assume that subcontracting is more expensive than processing the projects internally. So, strategic outsourcing is not very attractive.

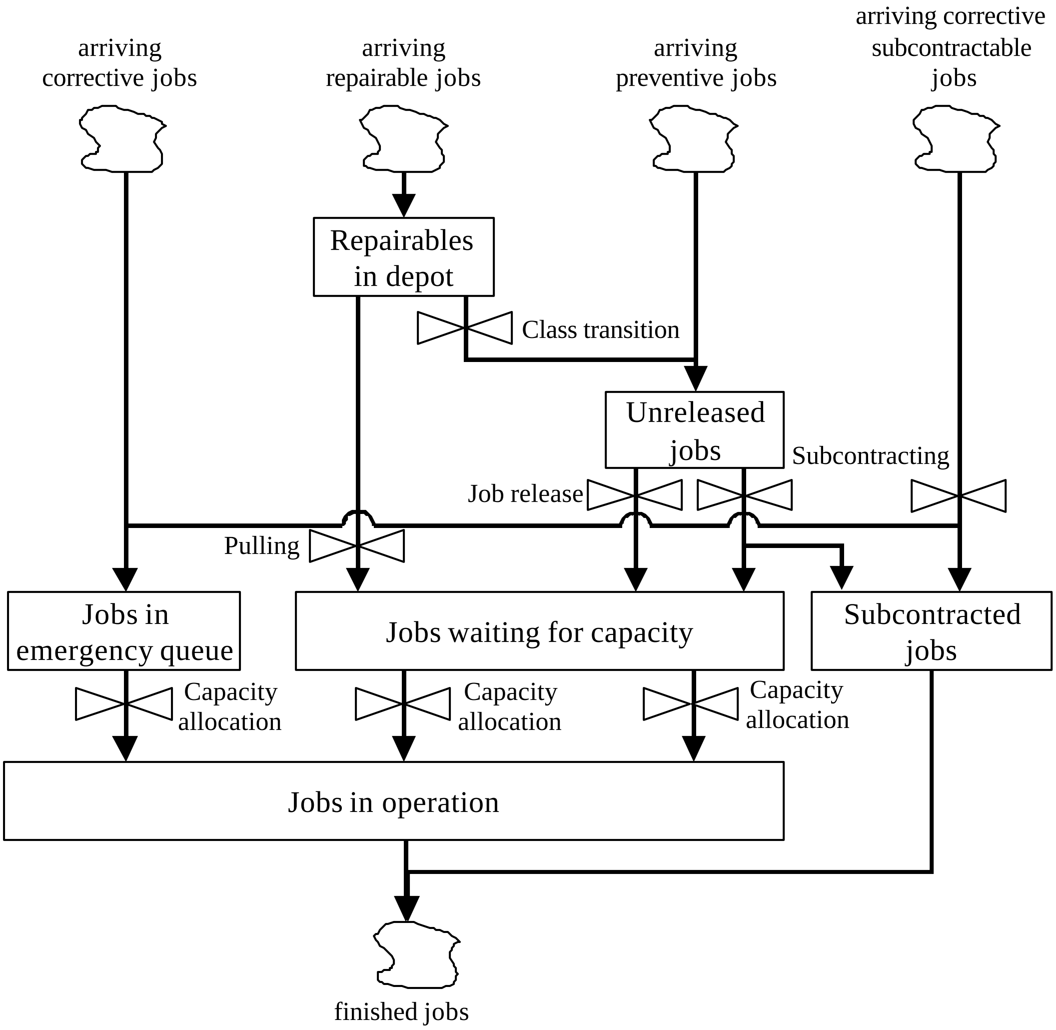


Figure 9.2. Basic decision functions in relation to the job flows for the 4 class situation.

This class adds a new decision function to the hierarchy of already defined functions: *subcontracting corrective projects* (see Figure 9.2). The consequences for the control structure depend on the costs of subcontracting preventive and subcontracting corrective projects. Assume in the remainder of this section that the costs of subcontracting one time unit of corrective (resp. preventive) work are denoted by C_C (resp. C_P) and that these both are higher than the costs of the internal capacity resources per time unit.

First, assume that $C_P < C_C$. At job level, the organization prefers subcontracting preventive jobs over corrective jobs, if necessary at all. Concerning the control structure, two main implications need to be regarded. Firstly, at the time the medium term subcontracting decision should be taken for preventive projects, sufficient slack should be created for both types of future corrective projects. This decision should be based on the analytical three-dimensional Markov representation, with states (a,b,c) with a,b, c respectively the number of preventive, emergency and subcontractable emergency phases in the shop. Secondly, every time a subcontractable corrective projects enters the system, it should be decided whether subcontracting is necessary or not. In case such a project enters the system and the number of subcontractable projects in the shop is larger than or equal to the number of operators in the shop, it needs to be subcontracted: after all, the operational availability of the technical systems is the overall performance indicator. In case such a project enters the system and at least one operator is idle, it needs no further comment that this project is processed internally. In case the previous two scenarios do not apply and operators are busy with preventive projects, the short term subcontracting decision becomes a little more complex. An appropriate approach would be to carry out a what-if scenario: what happens to the system output signal $\overline{p}(t)$ in case the project will be processed internally (and therefore a preventive job is interrupted)? And what happens in case it is subcontracted? Based on these two signals compared the to system reference signal (i.e. the target service level), it can be decided whether or not subcontracting is necessary.

Now, assume that $C_P > C_C$. At job level, the organization prefers subcontracting corrective jobs over preventive jobs, if necessary at all. Concerning the control structure, again the subcontracting decisions should be taken for two different terms. The medium term subcontracting decision function and the underlying process model are not affected by the introduction of this new class. Slack should only be created for the already existing class of corrective projects. After all, if indeed subcontractable corrective projects arrive, it is cheaper to subcontract these than having subcontracted the same amount of preventive jobs beforehand. Moreover, because the effectuation time of the short term subcontracting decision is substantially shorter than the medium term subcontracting decision, the short term decision goes together with a reduced uncertainty. As a result, the two-dimensional Markov model is still appropriate for the medium term subcontracting decision making. Concerning the short term subcontracting, the same procedures are applicable as in the case in which subcontracting corrective projects is more expensive.

Finally, it should be noted that the complexity for the planners increases: they need to be capable of switching their focus between the short and medium term, while maintaining a satisfactory overall performance.

9.3.5 Summary

We have discussed situations in which demand classes are both omitted and added. In the case of omitted classes we have seen that with some slight adaptations the control structure seems to be applicable. This also holds for the case that a class of subcontractable corrective projects is added. The main conclusion is that the impact of the different decisions heavily depends on the mixture of the different demand classes. Because nothing is known about the effectiveness of these adaptations, further research is needed to validate and evaluate the proposed adaptations.

9.4 Subcontracting characteristics

In this section, several restrictions will be imposed on the possibilities for subcontracting. In this research, we have tried not to exclude factors that are relevant in real-life. However, we have standardized some parameters in order not to relapse into an overwhelming amount of choices and assumptions regarding the parameter values in the testing framework. The main restrictions that we have set are the following. Firstly, we have assumed that the subcontracting lead time is equal for all jobs, regardless the type of jobs. This also means that all (preventive and repairable) jobs can be subcontracted. Secondly, we have assumed that the subcontractor delivers with a hundred percent reliability: all jobs are delivered at or before the agreed due dates. Thirdly, we have assumed that the subcontracting lead time is independent of the work content of the jobs that are subcontracted in one batch. Therefore, a very interesting question is to what extent the control structure still is valid in case these restrictions (alternately) are relaxed.

9.4.1 Subcontracting lead time

The developed model assumes that all jobs can be subcontracted and that all jobs have equal subcontracting lead times. In real-life, as discussed earlier, such an ideal situation never occurs. For some jobs there might be technological, financial or political reasons why they cannot be subcontracted (or in terms of the control model: their subcontracting lead time is infinite), whereas other jobs might have longer or shorter subcontracting lead times than the assumed lead time of L_0 time units. This situation has the following impact on the control structure. Recall that the probability of completing project i in time, as estimated at time t, is denoted by $p_i(t)$. In the original assumed setting, this probability could only have been influenced once, namely at L_0 time units before the due date of project i. In the new situation, this probability could be influenced more than once, depending on the number of different subcontracting lead times of the jobs in project i.

Assume that for a certain project, jobs are clustered based on their subcontracting lead times in n clusters. In each cluster, jobs have (more or less) equal subcontracting lead times (see Figure 9.3). After the latest subcontracting moment (i.e. the subcontracting moment of job type n) has passed, the value of $p_i(t)$ should be such that the output signal provided by the control mechanism is above a. Of course, the decision maker has a preference to postpone its decisions, in order to keep as many as possible control options. This preference exists because uncertainty and lead time are positively correlated: reducing the lead time also reduces uncertainty. On the one hand, we see that the availability of information increases as time passes by (and thus the uncertainty decreases). On the other hand, the number of control options also decreases as the due date comes closer. Obviously, the planner has to make a trade-off between these phenomena. At each subcontracting time, he or she has to establish to what extent the remaining subcontracting moments offer sufficient flexibility to ensure a satisfactory contribution to the system output signal. This can be done by a what-if analysis at each subcontracting moment: what happens to the output signal if these jobs are not subcontracted? Is it still possible to achieve a satisfactory system output signal?

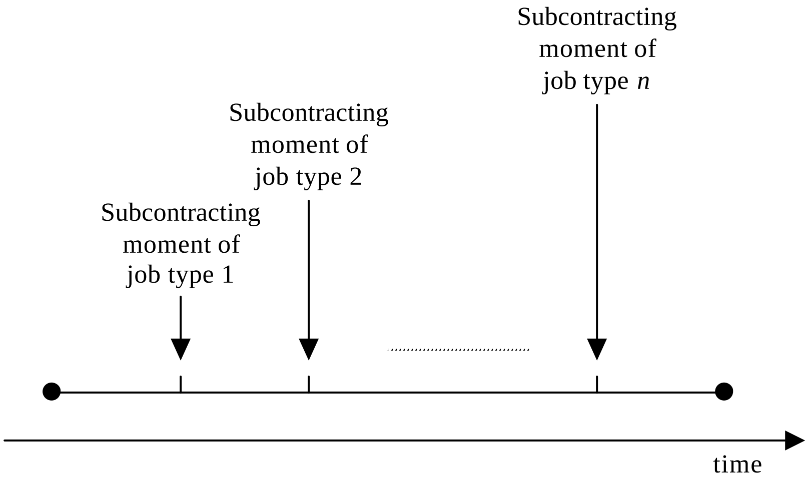


Figure 9.3. A project with *n* subcontracting moments.

It is clear that this situation is closer to reality than the situation studied in this research. The main consequence is the increased complexity of the subcontracting decision. The planner should use the analytical model for what-if scenarios to also monitor the progress *within* a project.

9.4.2 Subcontractor's reliability

Until now, we still have assumed that the subcontractor returns all the subcontracted jobs in time, with a 100% reliability. In reality, this will not be the case. Icmeli-Tukel

& Rom (1998) found that subcontractors in general are reliable deliverers. Therefore, it is assumed that this reliability is likely to be (a little) below hundred percent, say 90% or 95%. This has consequences for the control structure in terms of the cost-performance relation, rather than in terms of adaptations of the models. Whereas a perfectly reliable subcontractor makes that the maintenance organization could redeem uncertainty by subcontracting an entire project (after all, then the probability of delivering it in time equals 1), this is no longer applicable in this situation. It is true that the same models can be used as developed in this thesis without hardly any adaptations. However, subcontracting becomes a less powerful tool.

Assume that the subcontractor's reliability is denoted by a_S (with a_S < 1) and that the value of a_S is independent of the amount of work that is being subcontracted. In case an entire project is subcontracted, it is returned in time with probability a_S. Ergo, in case a project is partly subcontracted and partly internally processed, it is completed in time with probability $a_S p_i(t)$, which is even smaller than a_S. At first glance, it appears that this has serious consequences for the cost-performance relationship. However, looking at an aggregate level, subcontracting part of the workload also creates slack for the projects which have due dates beyond the subcontracted projects. So, subcontracting still is an effective tool to increase the overall delivery performance. However, if subcontracting is necessary, one project should be sacrificed to gain the improved delivery performance.

Summarizing, we see that the relaxation of the perfect reliability assumption does not give any problems for the technical part of the control structure: the structure is still *valid*. However, concerning the *evaluation*, it turns out that the subcontracting costs are likely to increase. After all, the overall subcontracting cost function is a decreasing function of the subcontractor reliability. If this reliability is no longer satisfactory, the maintenance organization should strive for either improved agreements or new subcontractors.

9.4.3 Relation between workload and lead time

As a last assumption, we have stated in this study that the subcontracting lead time is independent of the number of jobs that is subcontracted at the same time. This corresponds to the assumption that the capacity levels are sufficiently high to deliver at least the offered jobs in time. However, in the evaluation phase it appeared that the subcontracting behavior is very lumpy: periods of no subcontracting at all are followed by periods with a lot of subcontracting. These projects usually require so much capacity, that it may be questioned whether the subcontractor can deliver within the agreed lead time.

Firstly, from the point of view of the maintenance organization, the number of hours subcontracted may be considered as huge, compared to the internal capacity in a repair shop (on average 4 operator per shop in the SEWACO case). Now, consider the original equipment manufacturer as a subcontractor. They have much more capacity of the needed skills than the maintenance organization has. Therefore, they might not consider the subcontracting quantities as insurmountable. Furthermore, only the large projects turn out to be bottlenecks: for small projects it is assumed that it is no problem to process them within the agreed subcontracting lead time. An additional advantage of the large projects is that their allowed internal lead time usually is also large. As a result, this lead time could be chopped in 2 or more periods of L_0 time units. In that situation, the project can be subcontracted in pieces. Note however that the subcontracting decision for part of this decision should now be taken earlier, such that the uncertainty increases, which may result in higher average subcontracting costs. However, in Chapter 8 it turned out that the subcontracting costs are rather insensitive to changes in the subcontracting lead time: increasing the lead time from 10 to 50 workings days resulted only in a small increase of the subcontracting costs. Therefore, it is our conjecture that the control structure still provides satisfactory cost-performance curves.

9.5 Summary

Although the control structure has been developed for the maintenance organizations satisfying the conditions as described in Chapter 1, some interesting conclusions can be drawn with respect to the control of production situations with relaxed demand and capacity assumptions. Without claiming to be complete - after all, no validation and evaluation studies are carried out in this chapter - we have seen that the control structure and its underlying models also can be used for situations in which the subcontracting characteristics are different and for situations with a different demand classification. The solutions suggested in this chapter should be the guidelines for further research.

10

Conclusions and Further Research

10.1 Introduction

This thesis is about the production control of maintenance organizations consisting of multi-project repair shops. These shops face a lumpy and stochastic demand, comprehending three different demand classes, whereas their capacity levels (especially on the short term) are rather rigid. Furthermore, the fact that projects compete for resources of more than one repair shop makes that the capacity complexity of the control problem is large. As a result, the processes in the individual shops are hard to control and - obviously - the delivery performance is also hard to control. This has been illustrated by various scientific articles about studies in project management (even in more deterministic environments) which conclude that indeed in a lot of project driven environments the delivery performance is (far) below some satisfactory norm. Furthermore, the case-study at a real-life situation at a naval maintenance organization has shown that this poor performance also holds for this organization. In the remainder of this chapter, we answer the three research questions that are posed in the first chapter.

10.2 Crucial workload characteristics

The first research question concerns the problem of identifying to what extent it is possible to improve the planning and control of the maintenance resources and the resource requirements of SEWACO. Furthermore, what characteristics are crucial for achieving this performance improvement? Like in many other real-life situations, we have seen that it is difficult to establish the exact way in which the resources and the resource requirements are matched. This is mainly due to the fact that various situational factors play a role, such as the level of experience of the planners, individual (assumed) knowledge about the characteristics of the maintenance projects and the repair shops, expectations regarding available capacity, etcetera. Scientific knowledge about planning and control has mainly been achieved in small-scale theoretical models. A lot of the characteristics in these theoretical models can be observed in real-life. However, in view of the huge complexity of the real-life production control situations, it will be clear that it makes no sense to try to develop optimal order planning and capacity planning policies and compare these with the actual policies in order to search for ways to improve the performance. Therefore, in this thesis, using a large database with data on real-life projects and resources, and scientific literature about order planning performance, we show that logistic regression can be used to answer the first research question. It turns out that paying attention to the complexity of the individual projects and the uncertainty in the repair shops is crucial for achieving performance improvement. The complexity of a project can be described by its estimated processing time, the assigned lead time and the number of repair shops involved. The uncertainty of a repair shops can be described by the mixture of the three different demand classes and the predictability of the workload of the projects in the shop.

10.3 Framework for control

The second research question concerns the design of a suitable framework for control for the maintenance organizations under scrutiny. Observing the complexity of the planning and control problems in the maintenance organization, we conclude that a hierarchical control structure may be a suitable type of control structure. Such a structure consists of various decision levels which create conditions for the lower decision levels. These functions have to make sure that the required and available capacity are matched in such a way that a satisfactory delivery performance can be achieved. In general, the following three decision levels can be observed in production organizations: i) capacity determination, ii) order acceptance and due date assignment and iii) capacity allocation. Because of the tight relation between production and maintenance, for maintenance organizations it is not possible to adjust (at the short term) the maintenance schedules. After all, postponing projects directly decreases the operational availability of the technical systems. Consequently, all projects need to be accepted, together with the due dates that are given by the customers. Due to this inflexibility and the dynamic and stochastic environment, the workload may be uncontrolled. To compensate for this inflexibility, two other decision functions should be used: subcontracting (preventive projects) and pulling (low priority work). Because of the short term inflexibility of the capacity levels, these are the capacity adjustment functions with the shortest effectuation time. The subcontracting decision is concerned with the task to identify possible future delivery problems. Subsequently, by subcontracting (part of) this workload, capacity slack is created which increases the delivery performance. However, the uncertainty of emergency arrivals and processing times make that it is very well possible that this slack is not fully consumed. Clearly, subcontracting in stochastic situations goes together with an increased risk of resource idleness. This can be prevented by the pulling decision. This decision is responsible for identifying capacity surpluses in the near future and for pulling failed repairables from the depot to the repair shops. The interaction of these decision functions should contribute to a delivery performance and capacity resource utilization rate that both are satisfactory. In essence, it fills the gap that results from absence of the order acceptance and due date assignment function. Finally, experiments with real-life data have shown that the delivery performance of maintenance organization can indeed be improved if subcontracting and pulling decisions are used, which pay attention to the uncertainty and complexity characteristics that have been derived.

10.4 Knowledge transfer from theoretical models to real-life

The final research question concerns the way in which the knowledge of theoretical models indeed can be applied to improve the performance of real-life processes, assuming that it is not possible to mathematically model the processes at a detailed level. Even if the process models in the control structure explicitly pay attention to the identified capacity and complexity characteristics, it is still very likely that a gap exists between the theoretical process models and the real-life situation. However, these models at least show the directions and the power of the different relationships between the different variables. To transfer this knowledge to the real-life control problem, classical control theory has shown to be of tremendous help. This especially holds for the concept of feedback, which has proven to be a powerful technique, because it also compensates for performance deviations which are due to variables that are not contained in the model.

In terms of the subcontracting decision function, the planner should subcontract more (resp. less) in case the performance is less (resp. higher) than the desired performance. However, in this case the control action (i.e. subcontracting) has a relatively low speed of response. After all, the effects of subcontracting become only effective after the projects are completed. Therefore just giving feedback on current performance is not sufficient. The low speed of response makes that also decision outcomes that are in the pipeline need to be considered. To be able to cope with this pipeline, the subcontracting control rule should not only contain a feedback loop, but also an anticipation loop. The latter uses an explicit process model to evaluate the future outcomes of recent decisions. Experiments show that even in case of a low speed of response of the control action, the combination of feedback and anticipation provides a satisfactory delivery performance: a satisfactory steady-state delivery performance as well as a satisfactory transient-state robustness. Furthermore, it is shown that the necessity to include an anticipation loop (and thus the necessity to have an explicit process model) increases as the subcontracting lead time increases.

10.5 Further research

This study has contributed to the answers to the questions that are posed in Chapter 1. During the execution of this research, several topics came up which are worthwhile to be considered for further research.

In this thesis we have considered various interacting decision levels. After all, decisions taken at a certain level create conditions for the decisions at a lower level. In this thesis, the subcontracting decision level has been addressed as the main topic for research. At the next higher level, that is the capacity determination, we have assumed a certain fixed internal capacity. Once the behavior of subcontracting has been studied, a next relevant question is the interaction between the capacity determination and the subcontracting level. In what way should the internal capacity be re-dimensioned in order to achieve a satisfactory performance at lower total costs?

Secondly, despite the waned interest for research on classical control theory in production control during the last centuries, we have shown that the combination of pure feedback and the use of an explicit process knowledge is promising. Further research on this topic is strongly recommended in order to achieve knowledge about the possibilities of applying classical control theory in a much wider variety of production situations.

References

Abate, J. and W. Whitt (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, **10**: 5-88.

Abate, J. and W. Whitt (1994). Transient behavior of the M/G/1 workload process. *Operations Research*, **42**: 750-764.

Adam, N.R., J.W.M. Bertrand and J. Surkis (1987). Priority assignment procedures in multi-level assembly shops. *IIE Transactions*, **19**: 317-328.

Adam, N.R., J.W.M. Bertrand, D.C. Morehead and J. Surkis (1993). Due date assignment procedures with dynamically updated coefficients for multi-level assembly job shops. *European Journal of Operational Research*, **68**: 212-227.

Adelson, R.M. (1966). The dynamic behaviour of linear forecasting and scheduling rules. *Operations Research*, **17**: 447-462.

Agresti, A. (1990). *Categorial Data Analysis*. Wiley, New York.

Albright, S.C. and A. Soni (1988). Approximate steady-state distribution for a large repairable item inventory system. *European Journal of Operational Research*, **34**: 351-361.

Atkinson, K.E. (1989). *Introduction to Numerical Analysis* (2nd ed.). Wiley, New York.

Axsäter, S. (1985). Control theory concepts in production and inventory control. *International Journal of Systems Science*, **16**: 161-169.

Baker, B.N., D.C. Murphy and D. Fisher (1983). *Factors Affecting Project Success. Project Management Handbook*. Van Nostrand-Reinhold, New York

Baker, K.R. (1975). *Introduction to sequencing and scheduling*. Wiley, New York.

Baker, K.R. (1984). Sequencing rules and due-date assignments in a job shop. *Management Science*, **30**: 1093-1104.

Bertrand, J.W.M. (1980). Analysis of a production-inventory control system for a diffusion department. *International Journal of Systems Science*, **11**: 589-606.

Bertrand, J.W.M. (1983). The effect of workload dependent due dates on job shop performance. *Management Science*, **29**: 799-816.

Bertrand, J.W.M. and D.R. Muntslag (1993). Production control in engineer-to-order firms. *International Journal of Production Economics*, **30**: 3-22.

Bertrand, J.W.M. and V. Sridharan (1993). *Subcontracting policies for low volume component manufacturing*. Research Report TUE/BDK/LBS93-07. Technische Universiteit Eindhoven.

Bertrand, J.W.M. and J.C. Wortmann (1981). *Production control and information management for component manufacturing shops*. Elsevier, Amsterdam.

Bertrand, J.W.M., J.C. Wortmann and J. Wijngaard (1990). *Production control: a structural and design oriented approach*. Elsevier, Amsterdam.

Bessler, S.A. and P.W. Zehna (1968). An application of servomechanisms to inventory. *Naval Research Logistics Quarterly*, **15**: 157-168.

Bishop, A.B. (1975). *Introduction to Discrete Linear Controls: Theory and Application*. Academic Press, New York.

Bitran, G.R. and A.C. Hax (1977). On the design of hierarchical production planning systems. *Decision Sciences*, **8**: 28-55.

Bitran, G.R., E.A. Haas and A.C. Hax (1981). Hierarchical production planning: a single stage system. *Operations Research*, **29**: 717-743.

Boer, R. de (1998). *Resource-Constrained Multi-Project Management*. Ph.D. Thesis, University of Twente.

Chen, V.Y.X. (1994). A 0-1 goal programming model for scheduling multiple maintenance projects at a copper mine. *European Journal of Operational Research*, **76**: 176-191.

Cho, D.I. and M. Parlar (1991). A survey of maintenance models for multi-unit systems. *European Journal of Operations Research*, **51**: 1-23.

Cobham, A (1954). Priority assignments in waiting line problems. *Operations Research*, **2**: 70-76.

Conway, R.W., W.L. Maxwell and W.M. Miller (1969). *Theory of Scheduling*. Addison-Wesley, London.

Corder, A.S. (1976). *Maintenance Management Techniques*. McGraw-Hill, London.

Crawford, S., B.L. MacCarthy, J.R. Wilson and C. Vernon (1999). Investigating the Work of Industrial Schedulers through Field Study. *Cognition, Technology & Work*, **1**: 63-77.

Dempster, M.A.H., M.L. Fischer, L. Jansen, J. Lageweg, J.K. Lenstra and A.H.G. Rinnooy Kan (1981). Analytical evaluation of hierarchical production systems. *Operations Research*, **29**: 707-716.

Dzielinski, B.P., C.T. Baker and A.S. Manne (1963). Simulation tests of lot size programming. *Management Science*, **9**: 229-258

Dzielinski, B.P. and R.E. Gomory (1965). Optimal programming of lot sizes, inventory and labor allocations. *Management Science*, **11**: 874-890.

Dekker, R. (1995). Integrating optimisation, priority setting, planning and combining of maintenance activities. *European Journal of Operational Research*, **82**: 225-240.

Dekker, R. (1996). Applications of maintenance optimization models: a review and analysis. *Reliability, Engineering and System Safety*, **51**: 229-240.

Diezel, D.P. and S. Eilon (1967). A linear production - inventory control rule. *Production Engineering*, **46**: 93 - 104.

Dijkhuizen, G.C. van (1998). *Maintenance meets production: on the ups and downs of a repairable system*. Ph.D. Thesis, University of Twente.

Eekhout, L.L.M., A.T.I. Hazenoot and W.H. Lutgert (1998). 500 *Jaar Marine* (in Dutch). De Bataafsche Leeuw, Amsterdam.

Eilon, S. and I.G. Chowdhurry (1976). Due dates in job shop scheduling. *International Journal of Production Research*, **14**: 223-237.

Elmaghraby, S.E. (1966). *The Design of Production Systems.* Reinhold Publishing Corporation, New York.

Embleton, P.R. and P.C. Wright (1998). A practical guide to successful outsourcing. *Empowerment in Organizations*, **6**: 94-106.

Enns, S.T. (1994). Job shop lead time requirements under conditions of controlled delivery performance. *European Journal of Operational Research*, **77**: 429-439.

Feiler, A.M. (1972). Project risk management. In: M. Ogander (ed), *The practical application of project planning by network techniques*. Halsted Press, New York, 439-461.

Fortmann, T.E. and K.L. Hitz (1977). *An Introduction to Linear Control Systems.* Marcel Dekker, Inc, New York

Fry, T.D., M.D. Oliff, E.D. Minor and G.K. Leong (1989a). The effects of product structure and sequencing rule on assembly shop performance. *International Journal of Production Research*, **27**: 671-686.

Fry, T.D., P.R. Philipoom and R.E. Markland (1989b). Due date assignment in a multi-stage job shop. *IIE Transactions*, **21**: 153-161.

Gaver, D.P. (1962). A waiting line with interrupted service, including priorities. *Journal of the Royal Statistical Society,* **B24**: 73-90.

Geraerds, W.M.J. (1992). The EUT maintenance model. *International Journal of Production Economics*, **24**: 209-216.

Giebels, M.M.T, E.W. Hans, M.P.H. Gerritsen and H.J.J. Kals (2000). *Capacity planning for make- or engineer-to-order manufacturing; the importance of order classification*. Presented at the 33rd CIRP international seminar on manufacturing systems, Stockholm.

Giebels, M.M.T. (2000). *EtoPlan: A Concept for Concurrent Manufacturing Planning and Control*. Ph.D. Thesis, University of Twente.

Gits, C.W. (1992). Design of maintenance concepts. *International Journal of Production Economics*, **24**: 217-226.

Gits, C.W. (1994). Structuring maintenance control systems. *International Journal of Operations & Production Management*, **14**: 5-17.

Grubbström, R.W. and J. Wikner (1996). Inventory trigger control policies developed in terms of control theory. *International Journal of Production Economics*, **45**: 397-406.

Gupta, M. and D. Zhender (1994). Outsourcing and its impact on operations strategy. *Production and Inventory Management Journal*, **35**: 70-76.

Hagle, T.M. and G.E. Mitchell (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, **36**: 762-784.

Hartley, J.F. (1994). Case studies in organizational research. In: C. Cassell and G. Symon (ed), *Qualitative Methods in Organizational Research*; *a Practical Guide*. Sage Publications, London.

Hausman, W.H. and G.D. Scudder (1982). Priority scheduling rules for repairable inventory systems. *Management Science*, **28**: 1215-1232.

Hax, A.C. (1978). Aggregate production planning. In: J.J. Moder and S.E. Elmaghraby (ed), *Handbook of operations research models and applications.* Van Nostrand-Reinhold Co, New York.

Hax, A.C. and H.C. Meal (1975). Hierarchical integration of production planning and scheduling. In: M.A Geisler (ed), *Studies in Management Sciences*, Vol. I, *Logistics*. North-Holland - American Elsevier: 53-69.

Holt, C.C., F. Modigliani, J.F. Muth and H.A. Simon (1960). *Planning Production, Inventories and Work Force*. Englewood Cliffs, N.J. Prentice-Hall.

Hosmer, D.W. and S. Lemeshow (1989). *Applied Logistic Regression*. Wiley, New York.

Hunt, S. (1991). *Modern Marketing Theory: Critical Issues in the Philosophy of Marketing Science*. South-Western Publishing, Cincinnati.

Icmeli-Tukel, O. and W.O. Rom (1998). Analysis of the characteristics of projects in diverse industries. *Journal of Operations Management*, **16**: 43-61.

Kamien, M.I. and L. Lode (1990). Subcontracting, coordination, flexibility and production smoothing in aggregate production planning. *Management Science*, **36**: 1352-1363.

Kanet, J.C. and J.C. Hayya (1982). Priority dispatching with operation due dates in a job shop. *Journal of Operations Management*, **2**: 67-175.

Keizers, J.M. and J.W.M. Bertrand (1998), Due date and shop load control in multi project repair shops with subcontracting capabilities. Proceedings of the 5th EurOMA conference, Dublin: 262-267.

Keizers, J.M, I.J.B.F. Adan and J. van der Wal (1998). *A Queuing Model for Due Date Control in a Multi Server Repair Shop*. Working Paper WP32, Research Institute BETA, Technische Universiteit Eindhoven. Submitted for publication.

Keizers, J.M., J.W.M. Bertrand and J. Wessels (2000). *Diagnosing Order Planning Performance at a Navy Maintenance and Repair Organization using Logistic Regression*. Working Paper WP39, Research Institute BETA, Technische Universiteit Eindhoven. Submitted for publication.

Kesten, H. and J. Th. Runnenburg (1957). Priority waiting line problems. Proceedings of the Koninklijke Nederlandse Academie van Wetenschappen. Series A, Mathematical sciences, **60**: 312-326.

Kim, D.S. and J.M. Alden (1997). Estimating the distribution and variance of time to produce a fixed lot size given deterministic processing times and random downtimes. *International Journal of Production Research*, **35**: 3405-3414.

Kleinrock, L. (1975). *Queueing Systems, Vol. I: Theory*. Wiley, New York.

Kralj, B. and R. Petrovic (1995). A multi-objective optimization approach to thermal generating units maintenance scheduling. *European Journal of Operational Research*, **84**: 481-493.

Lasdon, L.S. and R.C. Terjung (1971). An efficient algorithm for multi-item scheduling. *Operations Research*, **19**: 946-969.

Lee, H.L. and K. Moinzadeh (1989). A repairable item inventory system with diagnostic and repair services. *European Journal of Operational Research*, **40**: 210-221.

Levy, J.B. and G.K. Tayi (1989). Analysis of project scheduling strategies in a client - contractor environment. *Naval Research Logistics*, **36**: 69-87.

Lin, L. and J.K. Cochran (1990). Meta-models of production line transient behaviour for sudden machine breakdowns. *International Journal of Production Research*, **28**: 1791-1806.

Lin, L., J.K. Cochran and J. Sarkis (1992). A Meta-model based decision support system for production control. *Computers in Industry*, **18**: 155-168.

Manne, A.S. (1958). Programming of economic lot sizes. *Management Science*, **4**: 115-135.

Marshall, S.A. (1978). *Introduction to Control Theory*. The Macmillan Press, London.

Maxwell, W.L. and M. Mehra (1968). Multiple-factor rules for sequencing with assembly constraints. *Naval Research Logistics Quarterly*, **15**: 241-254.

McKay, K.N., F.R. Safayeni and J.A. Buzacott (1995a). Schedulers and planners: what and how can we learn from them. In: D.E. Brown and W.T. Scherer (ed), *Intelligent scheduling systems*. Kluwer, Dordrecht, 41-62.

McKay, K.N., F.R. Safayeni and J.A. Buzacott (1995b). 'Common sense' realities of planning and scheduling in printed circuit board production. *International Journal of Production Research*, **33**: 1587-1603.

Meal, H.C. (1984). Putting production decisions where they belong. *Harvard Business Review*, **62**: 102-111.

Melnyk, S.A. and R.B. Handfield (1998). May you live in interesting times... The emergence of theory-driven empirical research. *Journal of Operations Management*, **16**: 311-319.

Menard, S. (1995). *Applied Logistic Regression Analysis*. Sage Publications, London.

Miles, M.B. (1979). Qualitative data as an attractive nuisance: The problem of analysis. *Administrative Science Quarterly*, **24**: 590-601.

Ministry of Defense (2000). http://www.mindef.nl/.

Morris, P.W. and G.H. Hough (1987). *The Anatomy of Major Projects*. Wiley, New York.

Muckstadt, J.A. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory systems. *Management Science*, **20**: 472-481.

Patton, J.D. (1980). *Maintainability and Maintenance Management*. Instrument Society of America.

Paz, N.M. and W. Leigh (1993). Maintenance scheduling: issues, results and research needs. *International Journal of Operations & Production Management*, **14**: 47-69.

Philipoom, P.R., R.S. Russell and T.D. Fry (1991). A preliminary investigation of multi-attribute based sequencing rules for assembly shops. *International Journal of Production Research*, **29**: 739-753.

Pinedo, M. (1995). *Scheduling: Theory, Algorithms and Systems*. Englewood Cliffs, Prentice-Hall.

Raaymakers, W.H.M. (1999). *Order Acceptance and Capacity Loading in Batch Process Industries*. Ph.D. Thesis, Technische Universiteit Eindhoven.

Reynolds, P.D. (1971). *A Primer in Theory Construction*. MacMillan, New York.

Rustenburg, W.D., G.J. van Houtum and W.H.M. Zijm (2000a). Spare parts management at complex technology based organizations: an agenda for further research. To appear in *The International Journal of Production Economics*.

Rustenburg, W.D., G.J. van Houtum and W.H.M. Zijm (2000b). Spare parts management for complex technical systems: resupply of spares under limited budgets. To appear in *IIE Transactions*.

Ryan, T.P. (1989). *Statistical Methods for Quality Improvement*. Wiley, New York.

Scarf, P.A. (1997). On the application of mathematical models in maintenance. *European Journal of Operational Research*, **99**: 493-506.

Schneeweiss, C.A. and H. Schröder (1992). Planning and scheduling the repair shops of the Deutsche Lufthansa AG: A hierarchical approach. *Production and Operations Management*, **1**: 22-33.

Schneeweiss, C.A. (1999). *Hierarchies in Distributed Decision Making*. Springer, Berlin.

Schonberger, R.J. (1981). Why projects are “always” late: a rationale based on manual simulation of a PERT/CPM network. *Interfaces*, **11**: 66-70.

Schrady, D.A. (1967). A deterministic inventory model for repairable items. *Naval Research Logistics Quarterly*, **14**: 391-398.

Shapiro, R.D. (1986). Managing supplier relations effectively: International comparisons. *L'enterprise Logistique*, II. Printemps.

Sherbrooke, C.C. (1968). Metric: a multi-echelon technique for recoverable item control. *Operations Research*, 16: 122-141.

Siegel, G.B. (1971). *An investigation of job shop scheduling for jobs with assembly constraints*. Ph.D. Thesis, Cornell University.

Silver, E.A. and C.N. Fiechter (1995). Preventive maintenance with limited historical data. *European Journal of Operational Research*, **82**: 125-144.

Simon, H.A. (1952). On the application of servomechanism theory in the study of production control. *Econometrica*, **20**: 247-268.

Simon, H.A. (1967). The business school: a problem in organizational design. *Journal of Management Science*, **4**: 1-16.

Stanford, D.A. (1997), Waiting and interdeparture times in priority queues with Poisson- and general-arrival streams. *Operations Research*, **45**: 725-735.

Stoop, P.M. (1996). *Performance Management in Manufacturing: A Method for Short Term Performance Evaluation and Diagnosis*. Ph.D. thesis. Technische Universiteit Eindhoven.

Takács, L. (1964). Priority queues. *Operations Research*, **12**: 63-74.

Tan, B. (1999). Variance of the output as a function of time: Production line dynamics. *European Journal of Operational Research*, **117**: 470-484.

Tansey, R.M., M. White, R.G. Long and M. Smith (1996). A comparison of loglinear modeling and logistic regression in management research. *Journal of Management*, **22**: 339-358.

Taylor, R.W. (1996). A linear programming model to manage the maintenance backlog. *Omega*, **24**: 217-227.

Tijms, H.C. (1994), *Stochastic Models: An Algorithmic Approach*, Wiley, New York.

Venkatesan, R. (1992). Sourcing: to make or not to make. *Harvard Business Review*, **70**: 99-107.

Waard, A.J. de (1999). *Re-engineering Non-Flow Large Scale Maintenance Organisations*. Ph.D. Thesis, University of Twente.

Wiers, V.C.S. (1996). A quantitative field study of the decision behaviour of four shop floor schedulers. *Production Planning and Control*, **7**: 383-392.

White, A.S. (1999). Management of inventory using control theory. *International Journal of Technology Management*, **17**: 847-860.

Whitt, W. (1999). Predicting queueing delays. *Management Science*, **45**: 870-887.

Yin, R.K. (1981). The case study crisis: some answers. *Administrative Science Quarterly*, **26**: 58-65.

Zijm, W.H.M. (1996). Reengineering production systems: the Royal Netherlands Naval Dockyard. *Annals of the CIRP*, **45**: 427-432.

Appendix A: Capacity Requirements and Availabilities at SEWACO

In Chapter 2, we have looked at the match between capacity availability and capacity requirements, seen at an aggregate level. To show that the mismatch between availability and requirements is common within SEWACO, we show these figures also for a sub-maintenance organization (Underwater Systems), a repair centre (number 770) and a repair shop (number 770F). In each of these figures, the available cumulative capacity, the cumulative capacity requirements measured against the original due dates and the cumulative capacity requirement measured against the adjusted due dates, are shown.

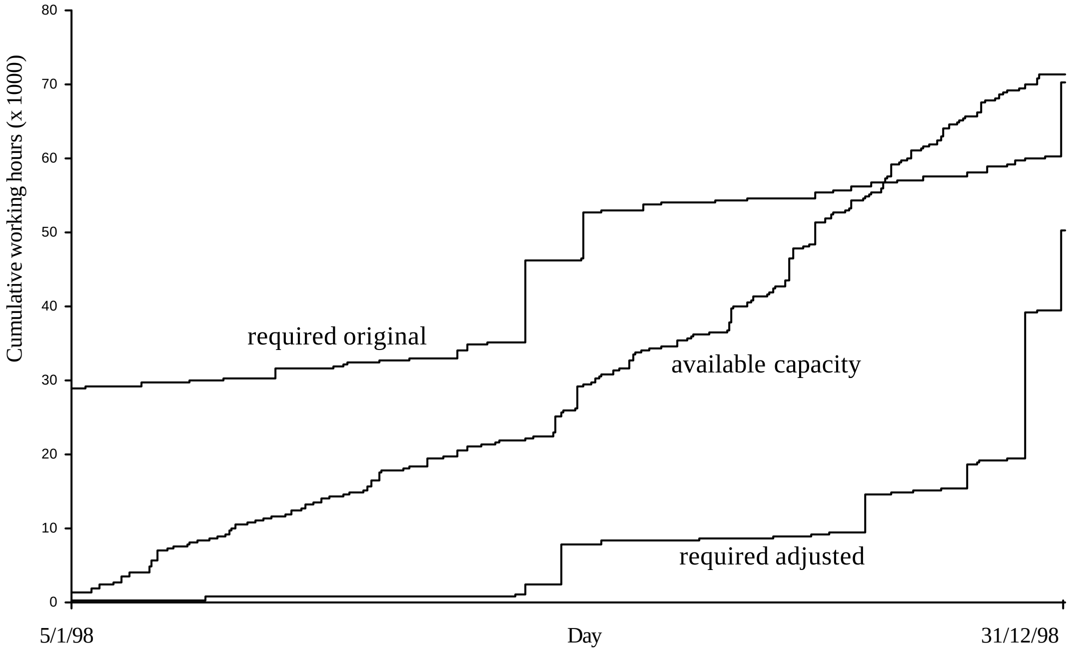


Figure A.1. Capacity requirements and availability for Underwater Systems in 1998.

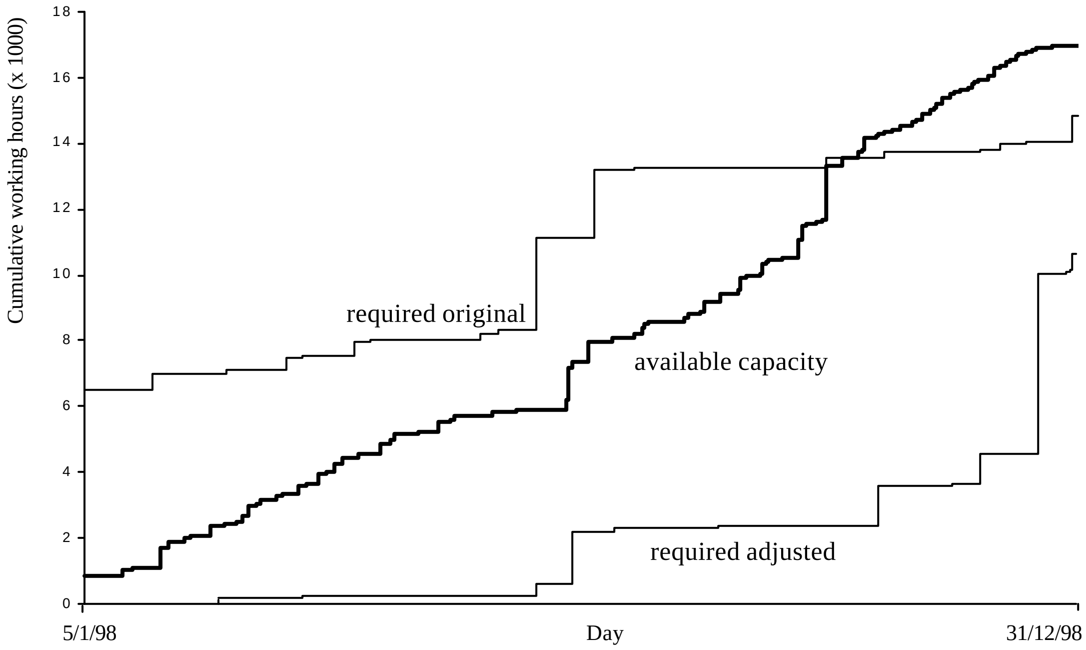


Figure A.2. Capacity requirements and availability for repair centre 770 in 1998.

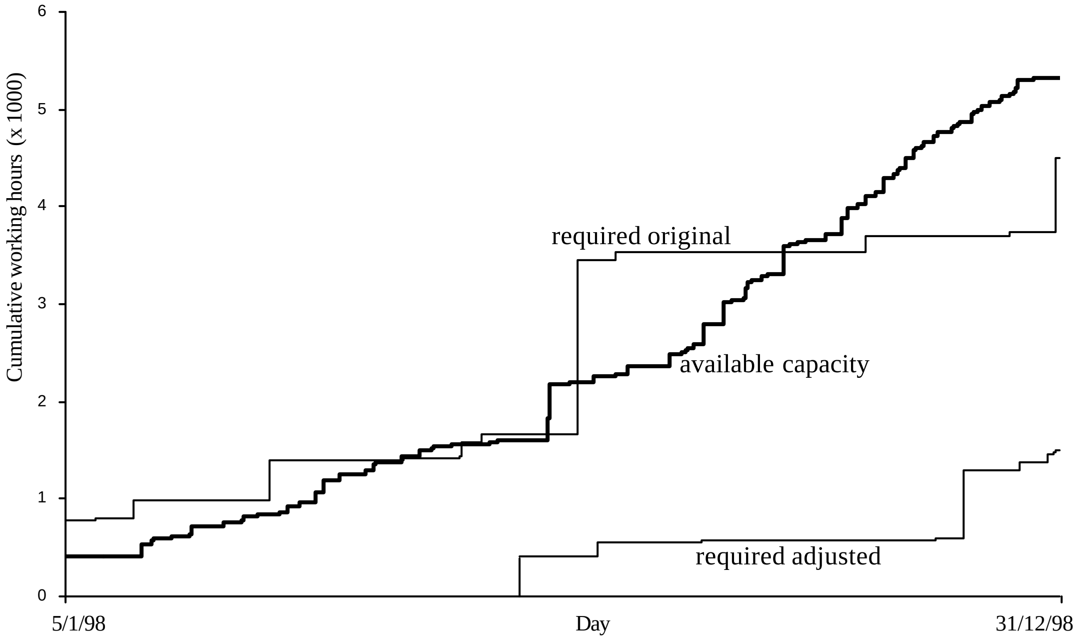


Figure A.3. Capacity requirements and availability for repair shop 770F in 1998

Appendix B: Actual Delivery Performance of the SEWACO Repair Shops

Repair shop	Performance (original due date)	Performance (adjusted due date)	Repair shop	Performance (original due date)	Performance (adjusted due date)
640E	28.6%	67.3%	660M	55.1%	92.5%
650A	35.5%	84.1%	660W	44.6%	86.3%
650B	32.7%	72.2%	670L	61.7%	79.4%
650C	36.6%	68.8%	670M	51.5%	68.0%
650E	48.3%	75.7%	680E	22.6%	53.6%
650M	41.0%	76.1%	680F	34.3%	71.1%
660E	53.3%	76.3%	680G	35.2%	68.8%
660F	36.4%	90.0%	680H	25.8%	73.5%
660H	33.3%	77.1%	680S	50.0%	73.3%
660K	65.5%	81.7%	680V	50.0%	100.0%
660L	38.5%	67.4%			

Table B.1. Delivery performance for the 21 AWS repair shops in 1998.

Repair shop	Performance (original due date)	Performance (adjusted due date)	Repair shop	Performance (original due date)	Performance (adjusted due date)
942A	74.5%	97.9%	963B	80.2%	94.7%
942B	80.0%	82.1%	963C	70.0%	93.9%
952A	60.5%	96.9%	972A	67.9%	89.3%
953A	86.0%	95.3%	973A	80.4%	94.2%
962A	54.5%	85.4%	974S	80.1%	98.3%
962B	50.0%	91.1%	974V	76.3%	97.7%
962C	50.5%	95.2%	982A	71.4%	96.8%
963A	79.7%	96.5%			

Table B.2. Delivery performance for the 15 SW repair shops in 1998.

Repair shop	Performance (original due date)	Performance (adjusted due date)	Repair shop	Performance (original due date)	Performance (adjusted due date)
740A	40.0%	91.7%	760B	20.2%	34.1%
740F	41.6%	66.3%	760K	34.4%	48.5%
740M	30.1%	54.2%	760P	28.6%	39.8%
740O	33.7%	60.2%	760S	23.9%	52.8%
740W	39.5%	71.0%	760V	16.7%	33.3%
750B	16.1%	34.1%	770F	42.0%	65.8%
750E	30.8%	41.8%	770K	42.3%	84.0%
750S	15.9%	34.3%	770M	45.8%	63.0%
750T	34.1%	42.0%	770W	31.8%	65.2%
750W	15.0%	34.1%			

Table B.3. Delivery performance for the 19 UWS repair shops in 1998.

Repair shop	Performance (original due date)	Performance (adjusted due date)	Repair shop	Performance (original due date)	Performance (adjusted due date)
840A	34.6%	49.0%	840K	23.9%	64.0%
840B	27.8%	56.7%	840M	20.8%	66.7%
840C	37.5%	72.7%	840N	100%	100.0%
840D	42.9%	62.8%	840T	57.1%	86.3%
840E	27.3%	66.7%	840W	5.0%	50.8%
840F	14.3%	60.0%	850R	55.1%	76.1%
840G	35.9%	58.8%	860B	25.0%	80.3%
840H	36.1%	58.7%	860D	33.0%	73.6%
840I	48.3%	77.1%	860M	30.6%	46.5%
840J	31.1%	59.7%	860N	22.5%	79.8%

Table B.4. Delivery performance for the 20 C3 repair shops in 1998.

Appendix C: Algorithm for the Makespan Moments

```
ALGORITHM: COMPUTE_MAKESPAN_MOMENTS
Input:                  $n, m, c, l, m_E, m$
Output:                 $E(T(n,m))$ and $E(T(n,m)^2)$
Stability condition:    $l / m_E < c$
(Step 0) $i := 0$;
            $\boldsymbol{j}_0(0) = (1,\ldots,1)^T$;
            $\boldsymbol{j}_0'(0) = (0,\ldots,0)^T$;
            $\boldsymbol{j}_0''(0) = (0,\ldots,0)^T$;
(Step 1) WHILE $i \le n$ DO
            BEGIN
                $i := i+1$;
                $\boldsymbol{j}_i(0) = [A_i(0)]^{-1} D_{i-1}\boldsymbol{j}_{i-1}(0)$;
                $\boldsymbol{j}_i'(0) = [A_i(0)]^{-1}[D_{i-1}\boldsymbol{j}_{i-1}'(0) - A_i'(0)\boldsymbol{j}_i(0)]$;
                $\boldsymbol{j}_i''(0) = [A_i(0)]^{-1}[D_{i-1}\boldsymbol{j}_{i-1}''(0) - A_i''(0)\boldsymbol{j}_i(0) - 2A_i'(0)\boldsymbol{j}_i'(0)]$;
            END
(Step 2) IF $m < c$ THEN
            BEGIN
                $E(T(n,m)) = -(\boldsymbol{j}_i'(0))_{m+1}$
                $E(T(n,m)^2) = (\boldsymbol{j}_i''(0))_{m+1}$
            END ELSE
            BEGIN
                $E(T(n,m)) = -(m-c+1)\boldsymbol{a}'(0)(\boldsymbol{j}_i(0))_c - (\boldsymbol{j}_i'(0))_c$
                $E(T(n,m)^2) = (m-c+1)(m-c)\boldsymbol{a}''(0)^2(\boldsymbol{j}_i(0))_c$
                        $+ (m-c+1)\boldsymbol{a}''(0)(\boldsymbol{j}_i(0))_c$
                        $+ 2(m-c+1)\boldsymbol{a}'(0)(\boldsymbol{j}_i'(0))_c + (\boldsymbol{j}_i''(0))_c$
            END
```

Appendix D: Fitting a Mixture of Two Erlang Distributions

Consider the random variable T with mean $E(T)$ and second moment $E(T^2)$, and hence with variance $s^2(T) = E(T^2) - (E(T))^2$. Let c_T denote its coefficient of variation, defined as $c_T = s(T)/E(T)$. If c_T is less than 1, then following Tijms (1994), a mixture of Erlang(k) and Erlang(k-1) distributions can be fitted to these moments, with density function

$$f(t) = pm^{k-1}\frac{t^{k-2}}{(k-2)!}\mathrm{e}^{-mt} + (1-p)m^k\frac{t^{k-1}}{(k-1)!}\mathrm{e}^{-mt}, \qquad t \geq 0,$$

with k such that

$$\frac{1}{k} \leq c_T^2 \leq \frac{1}{k-1}$$

and where the parameters p and m are defined by

$$p = \frac{kc_T^2 - \sqrt{k(1+c_T^2) - k^2c_T^2}}{1+c_T^2}, \; m = \frac{k-p}{E(T)}.$$

Summary

This thesis is about the design of a framework for the planning and control of the projects and processes in a maintenance organization. This framework should enable the decision makers to assign maintenance jobs to internal resources (operators) and external resources (subcontractors) in such a way that the fraction of projects that are completed before their due dates exceeds a satisfactory threshold value. The organizations that are considered have specific demand and capacity characteristics. Three different demand classes can be identified: (1) large preventive projects requiring a large variety of capacity resources, (2) corrective projects having short repair times and random arrival patterns and immediately requiring repair capacity and (3) the recovery of failed repairables that are used for repair-by-replacement activities. The common factors in these three types are that the repair times are random and that the due dates are given by the production department. The maintenance department has to treat these dates as given and fixed. Capacity in the maintenance organization is clustered in different repair shops. Each shop has its own skills. The shops work in parallel. A project is split into smaller sub-projects which are each carried out in the various repair shops. The volume-flexibility of these shops is rather restricted. Only at the medium term it is possible to (temporarily) increase the capacity levels by subcontracting.

As an illustration, Chapter 2 describes the processes and projects of SEWACO. This company is a maintenance organization of the Royal Netherlands Navy and satisfies the characteristics of the type of maintenance organization that are under scrutiny in this thesis. It is responsible for the maintenance of the technical sensor, weapon and command systems of the Navy. An initial performance scan for the year 1998 shows that less than 40% of the projects is being completed before or at its due date. Furthermore, it is shown that a subcontracting decision function has to be

initiated to be capable of controlling the workload. However, for a well working subcontracting decision function, knowledge is necessary about the process. In the current way of planning and controlling, it appears that the perception of the planners about the processes does not coincide with the reality of a dynamic and stochastic workload. As a consequence, making predictions about future capacity requirements (and thus the need for subcontracting) becomes a hard task, which explains the poor performance.

Chapter 3 discusses more precisely this perception and the gap with real-life. This perception is partly represented by the procedures which are currently used. However, the exact procedures which are used by the planners - and also the correctness of these procedures - are hard to observe and to analyze. These procedures are merely a combination of explicitly formulated control rules, expert knowledge and ad hoc control rules. Scientific knowledge provides useful control knowledge for production situations that slightly differ from the situation under scrutiny in this thesis. In combination with a diagnose tool that has been developed for this purpose and which has been applied to one year of information about the SEWACO projects and resources, it can be shown that the following project characteristics are currently not adequately being used by the planners when determining the need for capacity:

- estimated processing time of a project;
- agreed project lead time;
- number of repair shops involved in a project.

Furthermore, it turns out that the knowledge about the mixture of the three demand classes in the repair shops involved (i.e. process characteristics) is not adequately being used.

The knowledge obtained so far in the first chapters is used in Chapter 4 t to present a hierarchical production control structure. This structure contains the following key decision functions: (1) adaptation of capacity, (2) class transition, (3) work order release, (4) subcontracting, (5) pulling repairables and (6) allocation of jobs to operators. From known research it follows that a particularly lack of knowledge exists about the use of the subcontracting function as an operational instrument to match the (expected) internal workload with the available capacity.

In order to design this subcontracting decision, Chapter 5 presents a small-scale mathematical model which relates the project and process characteristics to the probability of completing a single project in time. The internal workload in a repair shop is described using a two-dimensional Markov model. The first dimension describes the remaining preventive workload, whilst the other describes the remaining (higher prioritized) corrective workload in the shop. Subsequently, this model is used to approximate the remaining makespan of a certain project, taking into account the characteristics that are identified in Chapter 3.

In Chapter 6, the subcontracting decision function is actually presented. Two key principles are adhered to. A first assumption is that managers are no longer interested in the steady state behavior of the performance indicators (such as delivery performance). Moreover, they are being helped with a controlled performance over a *finite* time horizon. The current state of the shops should be included in the decision making process in order to achieve this. Secondly, it should be stated that the model that is developed in Chapter 5 is not a complete representation of reality. Knowledge about classical control theory makes it possible to incorporate both principles in the control rule, using the descriptive Markov model as a descriptive process model which relates the control action (i.e. subcontracting) to the resulting change in performance. Through feedback, the past performance influences the need to subcontract in the next period. The worse the performance, the higher the necessity for subcontracting. In this way, errors that are linked to the fact that the model could not represent all possible real-life events, can be compensated for.

The control structure has been designed to support the planning and control processes for a particular type of maintenance organizations. SEWACO has been used as an example of these organizations. To validate the control structure, real-life data from SEWACO is used to apply some experiments. Because this thesis is mainly devoted to the design of a subcontracting rule, choices and assumptions have to be made concerning the used procedures at the other decision levels. Furthermore, some necessary information may not (yet) be available in the real-life situation at SEWACO. Chapter 7 describes the assumptions and choices that are made concerning the other decision levels and the lack of some data. These are used to provide a framework, that can be used to carry out experiments with the subcontracting decision function at repair shop level. For this purpose, a database with one and a half year of data about projects and resources is available.

Subsequently, Chapter 8 is devoted to the actual validation and evaluation of the control structure for three different repair shops with different project and process characteristics. It appears that - independent of the project and process characteristics - it is possible to achieve a certain a-priori desired delivery performance. Therefore, it can be concluded that the process is controlled and validated. Furthermore, it is shown that the results are due to the combination of the feedback loop and the analytic Markov model. Moreover, the influence of the Markov model becomes more important as the subcontracting lead time increases. In the evaluation phase, two aspects come to light. Firstly, it appears that the elasticity of the subcontracting rule is rather high. In all three repair shops, only 30% of the workload needs to be subcontracted in order to increase the delivery performance from 0% to about 95%. Secondly, the time phased pattern of subcontracted jobs heavily fluctuates in time. Periods in which hardly any projects are subcontracted are followed by periods in which a lot of projects are subcontracted.

To complete the validation process, Chapter 9 pays attention to production systems which slightly differ - in terms of demand and capacity characteristics - from the maintenance organizations that are considered in the previous chapters. It appears that only minor adaptations of the control structure are needed to be able to control these new situations. However, the influence of the various decision functions in the control structure heavily depends upon the mixture of the different demand classes.

Finally, Chapter 10 presents the conclusions of the research. The main conclusion is that the designed subcontracting decision rule appears to be capable of controlling situations in which no short-term capacity volume flexibility exists. Besides, it can be concluded that the use of information about the current state of the repair shop, in combination with a feedback loop, is necessary to achieve a performance that both has a satisfactory transient-state and steady-state performance. Finally, the longer the subcontracting lead time, the more wisely it is (from a performance point of view) to base the subcontracting decision on the current state of the system.

Samenvatting

Het onderzoek dat beschreven wordt in dit proefschrift richt zich op het ontwerp van een raamwerk voor de planning en beheersing van de projecten en processen die worden uitgevoerd binnen onderhoudsbedrijven. Dit raamwerk moet de beslissers in staat stellen om de interne capaciteit (d.w.z. de onderhoudsmonteurs) en de externe capaciteit (door middel van uitbesteding) zodanig in te zetten dat het percentage projecten dat tijdig wordt opgeleverd, overeenkomt met een vooraf gestelde eis. De bedrijven die beschouwd worden in dit onderzoek laten zich beschrijven door een aantal karakteristieken aangaande de capaciteit en vraag naar hun diensten. Aan de vraagkant zijn drie soorten projecten te onderscheiden: (1) grote preventieve projecten, die veel en meerdere soorten reparatie capaciteit vragen, (2) correctieve projecten, die een korte bewerkingstijd hebben maar onverwacht aankomen en onmiddellijk capaciteit vragen en (3) het herstellen van repareerbare delen die gebruikt kunnen worden voor *repair-by-replacement* activiteiten. Kenmerk van al deze projecttypen is dat de bewerkingstijd een stochastisch karakter heeft en dat de opleverdata worden bepaald door de vooraf gemaakte productieschema's. Het onderhoudsbedrijf dient deze als gegeven beschouwen. De capaciteit binnen het onderhoudsbedrijf is geclusterd in verschillende bewerkingsgroepen. Iedere groep heeft andere kennis en kunde. Een groep bestaat uit een aantal monteurs met onderling dezelfde vakkennis. Deze groepen werken parallel. Een project valt uiteen in een aantal deelprojecten die elk door één groep uitgevoerd moeten worden. De volume-flexibiliteit van de capaciteit in de groepen is beperkt. Slechts op middellange termijn kan de capaciteit (tijdelijk) vergroot worden door uitbesteding.

Hoofdstuk 2 geeft ter illustratie van de beschouwde bedrijven een beschrijving van de interne processen en procedures van het SEWACO bedrijf van de Koninklijke Marine in Den Helder. Dit bedrijf is verantwoordelijk voor het onderhoud aan de sensor-, wapen- en commandosystemen van de Marine. Uit een prestatiemeting over het jaar 1998 blijkt dat minder dan 40% van de onderhoudsprojecten tijdig wordt opgeleverd. In dit hoofdstuk wordt aangetoond dat uitbesteding noodzakelijk is om aan het vaak grillig vraagpatroon te kunnen voldoen. Echter, voor een goed werkende uitbestedingsfunctie, is kennis omtrent de processen van cruciaal belang. De huidige wijze van plannen en beheersing leert dat de planners vaak een perceptie van het proces hebben, dat niet overeenkomt met de dynamische en stochastische werkelijkheid. Derhalve is het moeilijk - zo niet ondoenlijk - om goede voorspellingen te maken omtrent de toekomstige capaciteitsbehoefte (en dus de behoefte aan uitbesteding). Dit verklaart ook mede de slechte leverprestaties.

Hoofdstuk 3 gaat dieper in op deze perceptie en het verschil met de werkelijkheid. Deze perceptie wordt mede gerepresenteerd door de gebruikte beslisregels. Echter, de precieze beslisregels - en dus de deugdelijkheid ervan - die gehanteerd worden zijn moeilijk te achterhalen. Deze regels zijn vaak mengelingen van expliciet geformuleerde procedures, expert kennis en ad hoc maatregelen. De wetenschappelijke literatuur biedt voldoende kennis omtrent de proces- en projectkarakteristieken die nodig zijn voor de planning en beheersing van werkplaatsen die één of meerdere kenmerken gemeenschappelijk hebben met de bewerkingsgroepen die hier onderzocht worden. Een speciaal hiervoor ontwikkeld diagnose instrument leert ons aan de hand van 1 jaar aan SEWACO gegevens, dat de volgende projectkarakteristieken niet of niet voldoende gebruikt worden bij het maken van ramingen voor toekomstig capaciteitsbeslag:

- verwachte bewerkingstijd van een project;
- afgesproken doorlooptijd van een project;
- aantal bewerkingsgroepen betrokken bij een project.

Daarnaast blijkt ook dat de kennis omtrent de mengeling van de drie soorten projecten in de betrokken bewerkingsgroepen (proceskarakteristieken) niet of niet voldoende gebruikt wordt.

Aan de hand van de kennis die in de eerste hoofdstukken opgedaan is, wordt in hoofdstuk 4 een ontwerp gepresenteerd voor een hiërarchische productie-beheersingsstructuur met de beslissingsfuncties: (1) aanpassing capaciteit, (2) klasse overgang, (3) werkorder vrijgave, (4) uitbesteden, (5) repairables inbesteden en (6) toewijzing van taken aan operators. Literatuuronderzoek wijst uit dat er vooral een gebrek bestaat aan wetenschappelijke kennis omtrent het gebruik van de functie uitbesteding als operationeel instrument om de interne werklast te balanceren, overeenkomstig de beschikbare capaciteit.

Om de uitbestedingsfunctie te kunnen ontwerpen, wordt in hoofdstuk 5 een kleinschalig wiskundig model gepresenteerd dat de relatie legt tussen de karakteristieken van een project enerzijds en de kans om het betreffende project voor de gevraagde leverdatum op te leveren anderzijds. Het model beschrijft de werklast in een bewerkingsgroep middels een tweedimensionaal Markov proces. Eén dimensie is een indicatie voor de hoeveelheid resterend preventief werk in de bewerkingsgroep, de andere dimensie een indicatie voor de hoeveelheid correctief werk in een bewerkingsgroep. Vervolgens geeft dit model de kansverdeling van het gereedheidsmoment als functie van de twee dimensies en de in hoofdstuk 3 geïdentificeerde factoren.

In hoofdstuk 6 wordt vervolgens daadwerkelijk de uitbestedingsregel ontworpen. Hierbij worden twee uitgangspunten aangehouden. Ten eerste wordt aangenomen dat managers niet geïnteresseerd zijn in acceptabele *steady-state* prestaties, maar meer in een acceptabele prestatie over een *eindige* horizon. Om dit te bereiken dient de huidige status van de bewerkingsgroep gebruikt te worden in de beslisregel. Een tweede uitgangspunt is dat het model geen volledige representatie is van de werkelijkheid. De kennis omtrent meet- en regelsystemen biedt ons tenslotte de mogelijkheid om deze twee uitgangspunten te incorporeren in de beslissingsregel, gebruikmakend van het beschrijvende model uit hoofdstuk 5. Feedback zorgt er nu voor dat de prestatie in de voorbije periode mede bepaalt in welke mate uitbesteding nodig is. Op deze manier kunnen fouten die optreden omdat bepaalde gebeurtenissen niet gemodelleerd zijn, gecompenseerd worden. Naarmate de (voorspelde) prestatie slechter is, zal er dus meer uitbesteed worden.

De beslissingstructuur is ontworpen om de planning en beheersing te ondersteunen voor de klasse van onderhoudsorganisaties waar ook de SEWACO onder valt. Om het model te valideren worden gegevens gebruikt van het SEWACO bedrijf. Echter, voordat tot de uiteindelijke validatie over gegaan kan worden, worden in hoofdstuk 7 de keuzes en aannames gemaakt en verdedigd die nodig zijn vanwege het feit dat sommige gegevens niet voorhanden zijn en omdat op bepaalde beslissingsniveaus vereenvoudigde regels gehanteerd zijn. Dit levert een raamwerk op waarmee voor een bewerkingsgroep experimenten uitgevoerd kunnen worden met de ontworpen beheersingstructuur. Hiervoor is een database aanwezig met anderhalf jaar aan project en proces gegevens aanwezig.

Vervolgens wordt in hoofdstuk 8 het raamwerk daadwerkelijk gevalideerd en geëvalueerd voor drie verschillende SEWACO bewerkingsgroepen met sterk verschillende karakteristieken. Het blijkt dat - onafhankelijk van de karakteristieken - het mogelijk is om de vooraf gewenste leverprestatie te behalen. Derhalve mag geconcludeerd worden dat het proces beheerst is en het raamwerk dus gevalideerd. Voorts wordt aangetoond dat de resultaten een combinatie zijn van zowel het gebruik van het analytische Markov model alsook van de feedback lus. Tevens blijkt dat de rol van het Markov model belangrijker wordt naarmate de doorlooptijd van het uitbesteden langer is. In de evaluatiefase komen twee belangrijke punten naar voren. De elasticiteit van de uitbestedingsregel is groot. In alle drie de bewerkingsgroepen blijkt dat door 30% uit te besteden, de leverprestatie van iets meer 0% tot circa 95% verhoogd kan worden. Daarnaast blijkt dat het uitbestedingspatroon buitengewoon patroon grillig is. Perioden waarin nauwelijks wordt uitbesteed, worden afgewisseld met perioden waarin veel werk wordt uitbesteed.

Om te validatie te voltooien, wordt in hoofdstuk 9 aandacht besteed aan productiesituaties die net iets verschillen van de type organisaties die in dit proefschrift behandeld zijn. Het blijkt dat met kleine aanpassingen het ontworpen raamwerk ook voor deze gevallen gebruikt kan worden. Echter, de invloed van de verschillende beslissingen op de uiteindelijke prestatie blijkt met name af te hangen van de mengeling van de verschillende soorten onderhoudsprojecten.

Hoofdstuk 10 presenteert tenslotte de conclusies van het onderzoek. De belangrijkste conclusie is dat de ontworpen uitbestedingsregel in staat is om in een situatie waarin de capaciteit op de korte termijn zeer inflexibel is, te komen tot een beheerste prestatie. Verder blijkt dat het gebruik van informatie omtrent de huidige status van de bewerkingsgroep in combinatie met een feedback lus een prestatie levert die zowel voldoet aan de lange termijn eisen, alsmede ook robuust is op de korte termijn. Tenslotte blijkt dat hoe langer de uitbestedingsdoorlooptijd is, hoe verstandiger het is - in termen van de behaalde leverprestatie - om de huidige status van de bewerkingsgroep te incorporeren in de uitbestedingsfunctie, vergeleken met beslisregels die van deze informatie geen gebruik maken.

Curriculum Vitae

The author of this thesis was born on November 18, 1972 in Nijmegen. In 1991 he received his Gymnasium diploma from the Dominicus College in Nijmegen, after which he started his study Econometrics at the Rijksuniversiteit Groningen. He received his Masters Degree in 1996 after a research project about the travelling salesman problem. In 1996 he started as a Ph.D. student at the Technische Universiteit Eindhoven. His research project has been carried out in close cooperation with a maintenance organization of the Royal Netherlands Navy. This thesis concludes the research. From September 2000 he is employed as a consultant at Arthur D. Little in Rotterdam.

STELLINGEN

behorende bij het proefschrift

Subcontracting as a Capacity Management Tool in Multi-Project Repair Shops

van

Joris Keizers

I

Gegeven de dwingende eisen ten aanzien van de beschikbaarheid van productiemiddelen, heeft een onderhoudsafdeling in het algemeen geen vrijheid in het nemen van de orderacceptatie en levertijdafgifte beslissingen. De inflexibiliteit die dit met zich meebrengt ten aanzien van de werklastbeheersing kan gecompenseerd worden de beslisfuncties uitbesteden (van preventief werk) en inbesteden (van repareerbare delen).

(Dit proefschrift)

II

Beschouw een project bestaande uit een gegeven aantal parallelle bewerkingen met identiek en exponentieel verdeelde bedieningstijden, dat aankomt bij een verwerkingssysteem met meerdere parallelle loketten. Voorts kan de verwerking van dit project verstoord worden door spoedbewerkingen met een Poisson aankomstproces en identiek en exponentieel verdeelde bedieningstijden. In dat geval is het mogelijk om de momenten van de doorlooptijdverdelingsfunctie van het project exact uit te rekenen. Indien de bedieningstijden niet identiek en niet exponentieel verdeeld zijn, is het mogelijk om zeer acceptabele benaderingen voor de momenten van de doorlooptijdverdelingsfunctie te vinden.

(Dit proefschrift)

III

Om te komen tot relevante kennis omtrent de beheersing van productiesystemen, dienen onderzoekers hun aandachtsgebied te verschuiven van *steady-state* gedrag naar *transient-state* gedrag van productiesystemen.

(Dit proefschrift)

IV

De kennis omtrent de klassieke meet- en regeltechniek is zeer bruikbaar bij de ontwikkeling van logistieke beheersingsconcepten, mits de te beheersen outputgrootheden gedefinieerd zijn over een eindige horizon.

(Dit proefschrift)

V

Beslissers hebben percepties omtrent de bedrijfsprocessen die niet noodzakelijk overeenkomen met de werkelijkheid. Derhalve kan onderzoek onjuiste conclusies opleveren wanneer dit onderzoek enkel gebruik maakt van gegevens uit interviews met de beslissers.

VI

Door de komst van ERP systemen en de drastische kostenreductie van computergeheugen, slaan organisaties steeds meer gegevens op. De bedrijfskunde dient van deze ontwikkeling gebruik te maken om het empirisch onderzoek binnen bedrijfskunde verder gestalte te geven.

VII
Ondanks het multidisciplinaire karakter van bedrijfskunde, worden binnen dit vakgebied nog te vaak onderzoeksobjecten eenzijdig bekeken.

VIII
Voordat bedrijfskundigen kans gaan maken op een Nobelprijs, zal de bedrijfskundige eerst poëtischer of vredelievender moeten gaan schrijven.

IX
Televisiequizzen van de commerciële omroepen werken inkomen-nivellerend. Daardoor heeft de commerciële televisie ook een sociaal economische functie.

X
Het aanbieden van informatie over RSI via internet is een armzalig idee.

XI
In het huidige internationale topvoetbal zijn de onderlinge kwaliteitsverschillen tussen elftallen zeer klein, waardoor arbitrale dwalingen een grote invloed hebben op de einduitslag. Daarom dienen voetballers slechts kansspelbelasting te betalen over de premies behaald tijdens de EK, WK en Champions League.

XII
Het is moeilijker een ingezonden brief in de Donald Duck, dan een artikel in een wetenschappelijk tijdschrift gepubliceerd te krijgen.