



Dissertation

**Selecting cost-minimal Delivery Profiles
and Assessing the Impact on
Cost and Delivery Schedule Stability
in Area Forwarding Inbound Logistics Networks
in the Automotive Industry**

Dipl. Wirt.-Inform. Tim Schöneberg

Schriftliche Arbeit zur Erlangung des akademischen Grades
doctor rerum politicarum (dr. rer. pol.)
im Fach Wirtschaftsinformatik

eingereicht an der
Fakultät der Wirtschaftswissenschaften der
Universität Paderborn
Paderborn, im März 2013

Acknowledgments

This thesis is a result of a longsome process. It began when I had the opportunity to write my diploma thesis under supervision of Prof. Dr. Leena Suhl and Dr. Jens-Peter Kempkes at the Decision Support & Operations Research Lab (DSOR) at the University of Paderborn in cooperation with the Daimler research facility in Ulm. During that time I got a glimpse at the interesting work of PhD candidates involved in projects that combined both practice and theory. Given this impression, the desire was formed to go through this process myself. I would like to start by thanking Prof. Dr. Leena Suhl for giving me the opportunity to work in her research group and to write this thesis. Her ever-friendly nature and the productive and enjoyable atmosphere she created at the DSOR Lab were key success factors for my work. During my time in the Daimler AG founded graduate class in the International Graduate School of Dynamic Intelligent Systems I had many colleagues that supported me in terms of both fruitful discussions and motivation. Here I would like to thank Prof. Dr. Wilhelm Dangelmaier and Prof. Dr. Achim Koberstein for ensuring a decent quality. From the DSOR team, I would like to thank in particular Kostja Siefen, my ever joking office mate, and the other team members who as a team always ensured that I was enjoying my work. I would also like to thank my colleagues at the Daimler Research Lab in Ulm, especially Joanna Schyroki and Ralf Bihlmaier. From the Daimler AG several persons are to mention as they contributed to the work in this thesis, these are Rainer Volk, Tobias Roesser and Manfred Richter from Daimler Trucks in Kassel, Helmut Röther from the Global Service & Parts team in Germersheim and Carsten Haink from the environmental protection research team. When my time at the DSOR Lab was finished my thesis, however, was not. So it is up to the motivation given by friends and relatives and my new team manager at Volkswagen, Jörg Knop, that it has been finished by now. Finally, I wish to sincerely thank my parents, siblings and grand-parents for their love and support in all my private and professional belongings. Most importantly, I could not have done this without Kristina, the girl on my side that shares everything with me both in good times and in bad.

Abstract

Automotive manufacturers in Europe use area forwarding based inbound logistic networks to obtain cost advantages in the inbound logistics section. Thereby, effective control of the ongoing logistics operations is necessary to gain the edge over competition. Delivery schedules that are frequently generated by the automotive manufacturers are used to control the material flow in the area forwarding networks. In doing so, different delivery schedule generation approaches can be used to balance between the different objectives of cost reduction and delivery schedule stability. A promising approach discussed in literature and successfully applied in retailer business are delivery profiles. When chosen wisely, this control rule is said to reduce both logistic cost and schedule instability. In this thesis, a method to select cost minimal delivery profiles under the consideration of area forwarding networks in the automotive industry is presented and its impact on both cost and delivery schedule stability in a rolling horizon environment is assessed in a case study. To identify the aspects of the problem setting that have to be considered, a description of the planning processes in the automotive industry and the operational order lot sizing in particular is given. In doing so, two types of delivery schedule generation mechanisms, algorithmic approaches and rule-based approaches are pointed out. An appropriate solution algorithm which uses a decomposition technique to overcome runtime issues is developed. A mixed integer formulation and heuristic algorithms, a sequential algorithm and a genetic algorithm that can be used in the solution algorithm are presented. The model and the solution algorithms are then extended to a two-stage stochastic program in order to consider demand uncertainties in the solution process. A large scale industry case study is then used to assess the impact on both cost and delivery schedules. A comparison with state-of-the-art algorithmic delivery schedule generation approaches is conducted to enlighten the pros and cons of both approaches. The method to select cost minimal delivery profiles is novel, and the case study provides useful insights for possible applications in practice.

Keywords: Automotive, Supply Chain, Material Flow, Logistics, Area Forwarding Inbound Logistic Networks

Contents

1	Introduction	1
1.1	Goals of the thesis	2
1.2	Structure of the thesis	2
2	Problem statement	3
2.1	Logistics	3
2.1.1	Area forwarding based inbound logistic networks	4
2.1.2	Inventory	9
2.2	Automotive supply chains	12
2.2.1	Material flow control	13
2.2.2	Planning process in automotive supply chains	16
2.2.3	Iterative planning in a rolling horizon	20
2.3	The operational order lot-sizing planning problem	21
2.3.1	Objectives and decisions	21
2.3.2	Constraints	23
2.3.3	Algorithmic and rule-based delivery schedule generation	24
2.4	Assessing the impact of cost minimal delivery profiles	29
3	Literature Review	33
3.1	Models and algorithms covering important aspects related to the selection of cost-minimal delivery profiles	33
3.1.1	Models for lot-sizing problems	34
3.1.2	Models for joint replenishment problems	37
3.1.3	Models for purchasing quantity discount problems	40
3.1.4	Models for minimum cost network design and flow problems	47
3.1.5	Considering uncertainty of demand	54
3.2	Assessing the impact on realized logistics cost	57

3.2.1	Simulating a rolling horizon environment to benchmark planning methods	58
3.2.2	Architectural approaches to benchmark simulation environments	61
3.3	Assessing stability of the generated delivery schedules	63
4	Outline of the Required Work	71
4.1	Selecting cost-minimal delivery profiles for area forwarding inbound logistic networks	71
4.1.1	Performance issues	72
4.1.2	Considering uncertainty	72
4.2	Assessing the impact of cost-minimal delivery profiles in a rolling horizon environment	72
4.2.1	A simulation framework for operational order lot-sizing planning methods	73
4.2.2	Measuring delivery schedule stability	74
4.3	Targeted contributions	74
5	Selecting cost-minimal and robust delivery profiles	77
5.1	Summary of the given decision problem	77
5.2	Exploiting the problem structure	78
5.3	Preprocessing	81
5.3.1	Determination of resulting orders	82
5.3.2	Computation of inventory related cost factors	83
5.3.3	Freight computation	84
5.3.4	A primal packing heuristic	91
5.4	Main leg model formulation	95
5.5	Primal heuristics	101
5.5.1	A local search heuristic	102
5.5.2	A genetic algorithm	105
5.6	Consideration of demand uncertainty	108
5.6.1	Preprocessing	110
5.6.2	Adapted model formulation	110
5.6.3	Modified solution algorithm for the stochastic case	114
5.6.4	A simplified model formulation	115

5.7	Scenario generation	119
5.7.1	A forecast deviation oriented scenario generation approach	119
5.7.2	A demand distribution oriented scenario generation approach . .	124
5.7.3	Scenario reduction	125
6	An Evaluation Framework for Delivery Schedule Generation Approaches	129
6.1	Simulation Approach	129
6.1.1	Representation of time	131
6.1.2	Representation of the underlying base system	132
6.1.3	MRP system logic	135
6.1.4	Interface between MRP system logic and delivery schedule generation approaches	136
6.1.5	Simulation procedure	137
6.2	Performance indicators	141
6.2.1	Assessing the realized cost	142
6.2.2	Assessing the stability of the generated delivery schedules	143
7	A case study from an automotive company	147
7.1	Experimental Design	148
7.1.1	Description of the examined areas	149
7.1.2	Considered delivery profiles	150
7.1.3	Testing environment	151
7.1.4	Considered alternatives	152
7.2	Analysis of the selected delivery profiles	153
7.3	Algorithmic performance	155
7.3.1	Evaluation of the decomposition approach	155
7.3.2	Evaluation of the generic model formulation	158
7.3.3	Evaluation of the simplified model formulation for the stochastic case	158
7.3.4	Evaluation of the heuristic algorithms	159
7.4	Evaluation of monetary effects	163
7.4.1	Expected costs	164
7.4.2	Realized costs	167
7.4.3	Value of perfect information	172

7.5	Stability of the generated delivery schedules	176
7.6	Inventory behavior	181
7.7	Summary of the case study results	186
7.7.1	Algorithmic results	186
7.7.2	Predictability and quality of applications in a rolling horizon . .	187
7.7.3	Implications for applications in practice	187
8	Summary and Conclusion	191
8.1	Summary of the achieved contributions	191
8.2	Outlook on further research	193
	Bibliography	195

List of Figures

2.1	Overview of logistic processes and subprocesses.	5
2.2	A typical area forwarding based inbound logistic network.	6
2.3	Exemplary discounting schemes for a single route.	8
2.4	An exemplary automotive supply chain.	13
2.5	Components of a delivery schedule and the according standards according to Klug [2012].	15
2.6	Overview of planning processes in automotive supply chain.	17
2.7	Transformation from gross dependent demand to net dependent demand.	20
2.8	Example of the lot-sizing investment issue.	26
2.9	Translation from net dependent demands into delivery schedule based on a delivery profile rule. In this example 'W003' from Table 2.2 is used.	28
3.1	A typical time-space network representation.	49
3.2	Example network model instance for delivery profile selection.	54
3.3	Basic architecture of simulation-based benchmarking environments (based on Mönch [2007], p. 1383).	63
5.1	Segregation of goods in an area forwarding inbound logistics network allowing decomposition of pre leg runs and full load runs from different suppliers.	80
5.2	Overview of the proposed solution algorithm.	81
5.3	Impact of load density on vehicle capacity use.	93
5.4	Overview of heuristic solution procedure for main leg problem.	103
5.5	Hamming neighborhood for a hamming distance of one and two suppliers.	104
5.6	Overview of the genetic algorithm.	106
5.7	Problem specific example for crossover and mutation operators.	107
5.8	Possible changes to a demand forecast.	121

5.9	Overview of the forecast deviation oriented scenario generation procedure.	122
5.10	Different scenario trees resulting from multiple time-slice scenario pool sizes on each level.	123
5.11	Demand entry quantity and distance between two demand entries as derived properties of the demand situation.	124
6.1	Entities of the base system, the plant master reference data.	134
6.2	Entities of the forecast data, the plant transaction data.	135
6.3	Interface between the simulation approach and the delivery schedule generation approaches.	137
6.4	Overview of the simulation procedure.	139
6.5	The area between two cumulative quantities used as a figure to describe the changes between two delivery schedules.	145
7.1	Data separation into training set and test set.	149
7.2	Distribution of selected delivery profiles per configuration.	154
7.3	Solution progress over time in comparison between local search and genetic algorithm for two area instances.	162
7.4	Expected procurement volume for the different sources of demand information and real consumption.	168
7.5	Cost components for the realized cost.	172
7.6	Distribution of parts based on the percent degree of underestimation. . .	181
7.7	Number of escalation processes, percentage of orders without escalation processes and inventory holding costs for each configuration.	183
7.8	Inventory level over time for three selected configurations.	184

List of Algorithms

1	The preprocessing algorithm.	81
2	The packing heuristic algorithm.	95
3	Modified solution procedure for the stochastic case.	115
4	Algorithm to generate scenarios based on demand patterns.	125
5	Fast forward reduction algorithm. Notation is adopted from Heitsch and Römisch [2003].	127
6	Simultaneous backward reduction algorithm. Notation is adopted from Heitsch and Römisch [2003].	128
7	Calculation of net dependent demands including lot ceiling, safety stock quantity and safety lead time.	140

LIST OF ALGORITHMS

List of Tables

2.1	Conditions under which safety lead time helps to protect against demand uncertainty.	12
2.2	Exemplary delivery profile transformation rules.	27
5.1	Preprocessing output parameters requiring an additional subscript for the stochastic case.	110
7.1	Overview of the general properties of the examined areas.	151
7.2	Overview of the delivery profiles considered in the case study.	152
7.3	Time required to preprocess the given area instance with the algorithms described in Section 5.3 for the deterministic case or Section 5.6.1 for the stochastic case with 10 scenarios respectively.	156
7.4	Overview of the runtime for the generic model formulation for the deterministic case.	159
7.5	Overview of the runtime for the stochastic case with mixed scenarios for the simplified model formulation.	160
7.6	Overview on the heuristic runtimes and solution quality for the given area instances.	161
7.7	Improving solutions found during the genetic algorithm procedure grouped by their source.	163
7.8	Expected costs and expected savings for all configurations.	165
7.9	Overview of the realized total cost of each area as derived from the simulation study.	169
7.10	Savings prediction error of the different configurations.	173
7.11	Realized savings in comparison with the optimal savings that could have been achieved if all information had been available at planning time. . .	175
7.12	Overview of key stability indicators for the different configurations. . . .	178

7.13	Overview of key stability indicators for parts that do not have delivery profile W11111 assigned in the initial forecast configuration.	179
7.14	Figures on end-of-simulation inventory for the different configurations. .	185

1 Introduction

The automotive industry plays an important role in the German economy. "About 20 % of annual German gross domestic product in the last decade was earned with the product "automobile" " (Becker [2006], p. 218), and it may be stated that "there are about 5.3 million people in Germany today who make a living, directly or indirectly, from cars"(Becker [2006], p. 218). Driven by the challenges of globalization and the necessity to offer more and more models and options to meet the customers demand, automotive manufacturers have created global supply networks with thousands of suppliers. Maintaining a smooth and cost-efficient flow of goods from supplier facilities to the automotive manufacturers plant is therefore one of the capabilities vital to survival in active competition. Area forwarding networks are a concept widely used among automotive manufacturers to run the necessary logistic operations. Due to a focus on the core competences operations of those area forwarding networks are carried out by logistics service providers, whereas control of the material flow remains in the hands of automotive manufacturers. The automotive manufacturers' ability to control the material flows has not been used to its full extent in practice. It has been addressed in a vast number of publications and programming solutions that may in general be divided into two categories. On the one hand, there are algorithmic approaches that propose to dynamically and frequently adapt a plan to imminent changes. On the other hand, rule based approaches try to determine a fixed rule that can be applied to derive a plan if necessary. Whereas the former have been more intensively studied in literature, the latter are more often demanded in practice. Three reasons can be cited for this. First, most algorithmic planning approaches demand rather complex software implementations to be integrated into company-wide applications, a step which in most cases is expensive and time-consuming. Second, algorithmic planning approaches are hard to understand for practitioners that barely have time to read up on the details behind these algorithms and models. Last but not the least, algorithms presented in the literature focus on cost reductions, whereas other goals, especially the desire for a stable and reliable plan remain undiscussed. In addition, an unproved bias among

practitioners tends to view algorithmic approaches as further enhancing the instability given in the supply chain. A control rule that has recently been studied by both researchers and practitioners is the so-called delivery profile or replenishment epoch. Delivery profiles provide a set of days on which deliveries are allowed and neglect deliveries on other days, thereby controlling the material flow. Despite their ease of use and their applicability to the underlying problem setting, no planning approach exists to determine optimal delivery profiles for suppliers in area forwarding inbound logistic networks. Furthermore, no insights are available on whether or not delivery profiles help to increase the stability of delivery schedules in an automotive manufacturer's day-to-day operations. This gap will be closed in this thesis.

1.1 Goals of the thesis

The thesis has three major goals. First, to develop a planning method capable of determining cost-minimal delivery profiles for area forwarding based inbound logistic networks under special consideration of an automotive environment. Market demand uncertainty should thereby be incorporated in the decision making process. Second, a method of analysis of both control rules and algorithmic planning methods for delivery schedule generation in a rolling horizon planning environment in respect to costs, robustness of the solutions and the impact on delivery schedule stability will be developed. The third goal is to investigate the outcome of the planning method for delivery profiles in a rolling horizon production planning environment and compare the results with state-of-the-art algorithmic planning approaches.

1.2 Structure of the thesis

In Chapter 2 the planning problem and its relevant aspects will be described. Chapter 3 contains an outline of previous research on similar topics. After summing up the most important findings from literature, the gap between existing literature and requirements of this thesis are analyzed and the missing steps are shortly depicted in Chapter 4. In the following, Chapter 5 presents the solution approach to determining cost-minimal delivery profiles and Chapter 6 gives details on the evaluation method. To prove the validity of the solution approach a case study with industry applications is given in Chapter 7.

2 Problem statement

This Chapter will first give a short overview of logistics in general and area forwarding based inbound logistic networks in particular. Automotive supply chains in particular will then be explained, and a description offered of how supply chain operations are planned, along with an indication of how the operational order lot-sizing planning problem integrates into the big picture. Thereafter, the operational order lot-sizing planning problem for area forwarding networks will be depicted and the different components of the problem, including the decisions to be made, the boundaries to those decisions and the resulting difficulties will be explained. Algorithmic and rule-based delivery schedule generation for the operational order lot-sizing problem will then be distinguished. The Chapter will close with a summary of the planning problem that arises when delivery profiles are used as delivery schedule generation rule.

2.1 Logistics

Logistics may be defined as "the management of all activities which facilitate movement and the co-ordination of supply and demand in the creation of time and place utility" (see Heskett et al. [1973]). These activities include transformations in time (transport) and space (storing) as well as changes in composition of objects (handling) (see Button et al. [2011], p. 250).

According to the *Supply Chain Operation Reference* (SCOR) model (see Stewart [1997]), a company's business process may be divided into the five major sub-processes: plan, source, make, deliver and return. Logistics is a necessary support process for the business processes source, make, deliver and return, because it ensures "the positioning of resource at the right time, in the right place, at the right cost, at the right quality" (Chartered Institute of Logistics and Transport (UK), 2005, cited in Rushton et al. [2006], p. 6) which are necessary to maintaining the core business processes. Rushton et al. [2006] defines logistics as "the efficient transfer of goods from the source of supply through the place of manufacture to the point of consumption in a cost-effective way

while providing an acceptable service to the customer” and thus lays emphasis not only on the content of logistic operations but also on its goals.

According to Gudehus and Kotzab [2009], logistics may be divided into four fields of operation:

- Inbound logistics (or procurement logistics)
- Production logistics (or internal logistics)
- Outbound logistics (or distribution logistics)
- Reverse logistics (or disposal logistics)

While the first three fields (inbound logistics, production logistics and outbound logistics) are existential parts of a production supply chain, reverse logistics deals with recirculation of produced goods. This issue is not relevant to this work and will thus not be considered. Figure 2.1 depicts the different fields of operation and the relations between them. At the beginning of a company’s logistics operations stands the supplier’s goods-issuing department. This is the point at which the raw goods leave the supplier’s system and are transferred to the company’s control. Inbound logistics is responsible for transport of goods to the incoming goods department warehouse (process source). At this point production logistics takes over responsibility for the material flow in the company’s production system (process make), until the finished goods are placed in the outgoing goods warehouse. From here finished goods are delivered to the customer (process deliver). This task is fulfilled by distribution logistics. The red-dotted rectangle in Figure 2.1 delimits the scope of this thesis, in which only inbound logistics and the incoming goods department warehouse will be considered.

2.1.1 Area forwarding based inbound logistic networks

One approach widely adopted in the automotive industry is the use of area forwarding based inbound logistics networks. The main idea behind area forwarding logistics networks is that of bundling inbound transports from multiple suppliers in accordance with their spatial arrangement in order to increase vehicle use in less-than-truckload (LTL) transport. According to the spatial distribution, suppliers are segregated into several *consolidation areas*. For each consolidation area a consolidation center for pure cross-docking operations exists in a central location. Figure 2.2 depicts a typical area

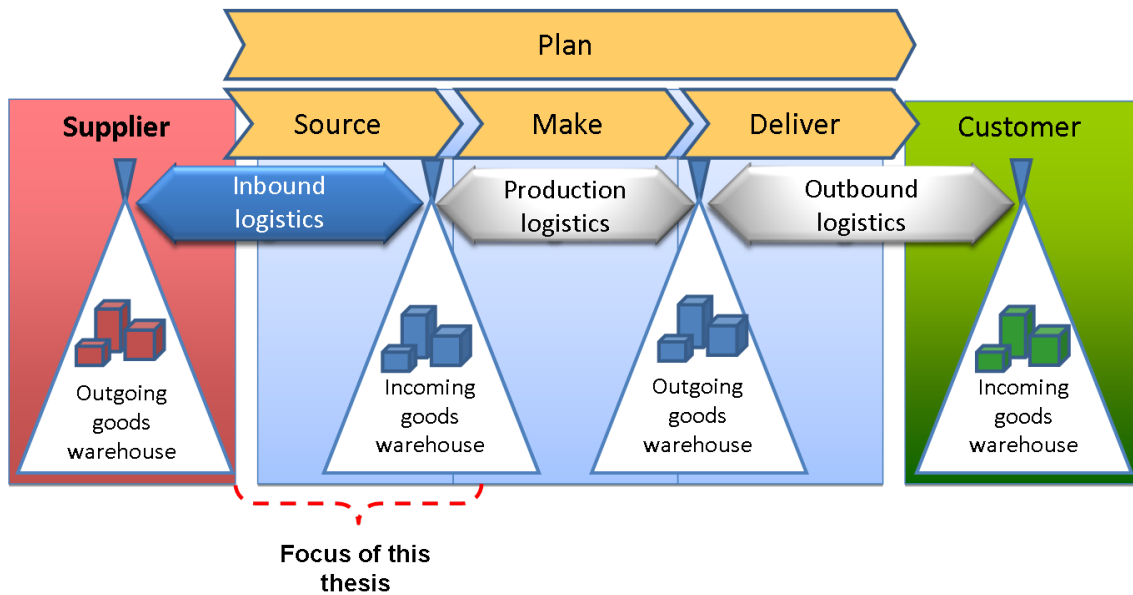


Figure 2.1: Overview of logistic processes and subprocesses.

forwarding based inbound logistics network. When the goods are picked up from a supplier within the area there are two possible follow-up steps. If the goods from a single supplier fill a vehicle completely, a full truckload (FTL, the green line in Figure 2.2) transport from the supplier directly to the unloader (in this case the OEM) takes place. This step is called the *full load run*. Otherwise, the goods are brought to the consolidation center. This step is called the *pre leg run* and is represented by the blue line in Figure 2.2. In the consolidation center goods from different pre leg runs from different suppliers within the same area will be cross-docked and then transported to the target location. This transport step, depicted by the yellow line, is called the *main leg run*. Due to cross-docking operations there can be only a single LTL transport from the consolidation center to the unloader, but several additional FTL transports. This helps to increase the average vehicle use on the main leg run, which in most cases is the longest distance within the supply network.

Logistic service providers A logistic service provider (LSP) is a company that provides logistic services to other companies. These services may include both operative and administrative services from the field of transport, handling, storing and special ser-

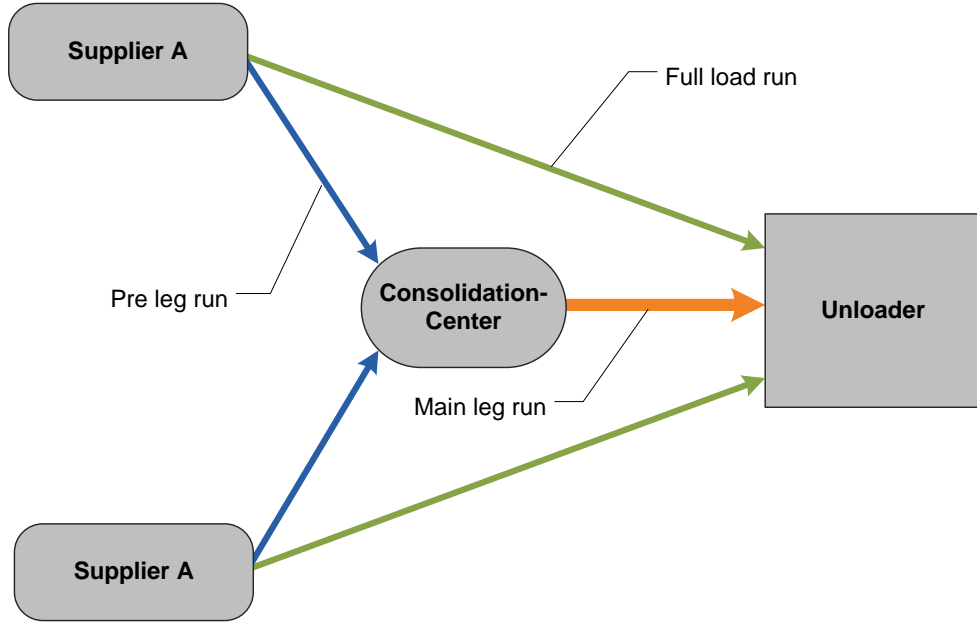


Figure 2.2: A typical area forwarding based inbound logistic network.

vices (e.g. filling and bottling or waste removal) or compounded logistic services which combine multiple services into a value-creating chain (see Gudehus and Kotzab [2009], p. 805 ff). Recently, "more and more companies assign their logistic demand to service providers" (Gudehus and Kotzab [2009], p. 803) to focus on their core competencies. Cost reduction in general and a change in cost structure are the most important reasons for a company's decision to outsource logistic services to a LSP (see Blecker et al. [2007], p. 42). As the logistics services are the LSP's core competency, the LSP can achieve economies of scale and realize synergies by utilizing its resources better and may pass these cost reductions to the customer (see Gudehus and Kotzab [2009], p. 809). Chou et al. [2009] points out that "auto-makers and parts suppliers maintain their own warehouses" while the LSPs "are contracted by the manufacturers to help manage all aspects of intermediary logistic tasks" (Chou et al. [2009], p. 583). For each consolidation area one LSP is selected to operate the consolidation center and execute all transport. In comparison with an open competitive bidding for each single transport relation this approach reduces management overheads on the OEM's side and brings more opportunities for synergy effects on the LSP's side. Further informations about

the LSPs options to run the logistic network can be found in Crainic [2000].

Tariff systems The services offered by the LSP have to be paid for by the company that uses the services on negotiated conditions. While some services can clearly be accounted for (e.g. packing parts into a load carrier may be accounted per part), other operations can include synergy effects which could result in rebates for the customer. Due to its cost structures consisting of a large fixed block for vehicle usage and driver payment and a variable part of fuel cost, this holds true especially for truck transport services. A common way to pass incentives to the customer is to use a tariff system for a transport relation. "The tariff system is negotiated on a mid-term base (usually between one and two years) and defines the price for the services of the logistics service provider" (Schöneberg et al. [2011], p. 217). To reflect the different conditions in consolidation areas (e.g. more urban districts require more city traffic and a region within a mountain range involves more up- and downwards driving), a separate tariff discounting scheme can be put up for each consolidation area. For a single route or distance the structure of the tariff system is usually based on load measures, e.g. weight, load metres or the number of freight pieces. For each value of the load measure, a specific price is given by the discounting-scheme. Usually, the discounting-scheme provides an incentive for a higher vehicle use, which reflects the internal cost structure of the LSP, as use-independent costs like the driver's payment, tolls, etc. occur independently of vehicle use level. Considering the discounting scheme, we can distinguish between an all-units discount and incremental discounts. In the first case (green line in Figure 2.3) the price of all-units is adapted according to the discount if a certain value of the underlying measure is exceeded. In the latter case (red line in Figure 2.3) only the price of additional units of the underlying measure are discounted. Incremental discounting schemes can additionally include a base price for each rebate level. A piecewise constant discounting scheme is a special case of the incremental discounting scheme with a base price and without variable cost per unit of measurement. A flat-rate discounting scheme is in turn a special case of a piecewise constant discounting scheme with only one rebate level.

Synergy effects As stressed in the paragraphs above, there are synergy effects for both the LSP and the unloader in an area forwarding based inbound logistics network because the LSP's synergy effects are partly passed on to the unloader. In detail, these

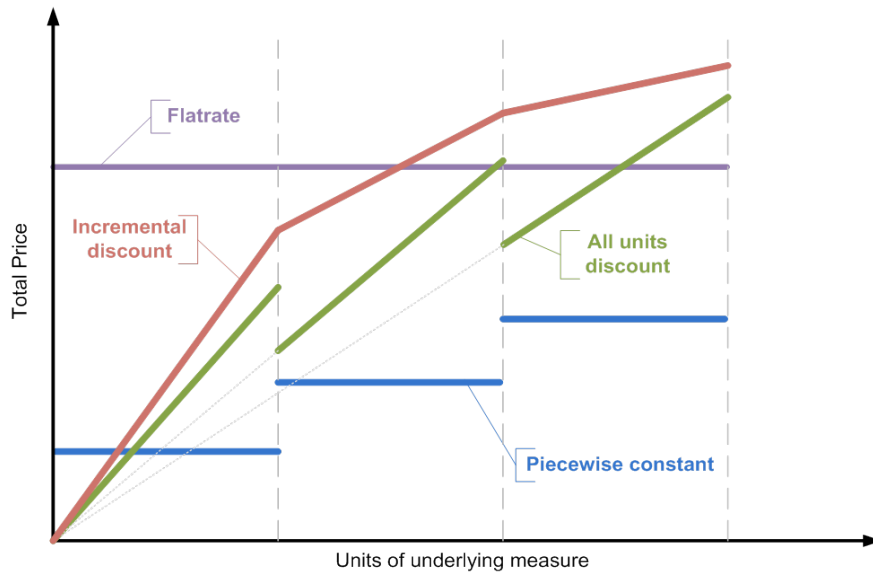


Figure 2.3: Exemplary discounting schemes for a single route.

synergy effects can be achieved by two leverages, the consolidation of goods from one supplier to make use of the more efficient full-load runs and increase vehicle use in pre leg runs on the one hand and consolidation of goods from multiple suppliers in the main leg run. These two leverages can be pulled by time-based consolidation on top of the spatial consolidation lying in the structure of area forwarding based networks.

Load carriers Load carriers are boxes, cases or pallets used to bundle several parts. Load carriers are used to ease the handling of goods and to protect them during transport. As load carriers are the objects that determine the space usage within a vehicle and can also be responsible for a considerable share of load weight, it is necessary to consider load carriers. Load carriers can be diversified into so-called set load carriers and non-mixed load carriers. While set load carriers carry different parts at the same time, non-mixed load carriers carry only parts of the same type at once. These attributes need not to be fixed for the load carrier's lifetime, e.g. a load carrier may have the physical ability to carry different parts, but could still be used in non-mixed mode. Set load carriers are often used in Just-In-Time or Just-In-Sequence environments to carry all parts required for a single manufacturing step at once or to

provide the material for a predefined sequence of jobs. As Just-In-Time and Just-In-Sequence deliveries are not within the focus of this thesis, only non-mixed load carriers will be considered. Another possible way to classify load carriers is to segregate them into reusables and non-reusables. Reusable load carriers are typically made of hardened materials (e.g. steel or wood), while non-reusable ones are paperboard containers. In most industrial applications reusable load carriers are preferred for ecologic reasons. Only in case of high prices for returning reusable load carriers (e.g. shipping relations with a high price on return path) or high correlation between weight and price (e.g. air freight) are non-reusable load carriers used.

2.1.2 Inventory

Between each pair of fields of operation, inventory in warehouses is used as a buffer which provides the possibility of decoupling two fields of operation (see Bose [2006], p. 4). In addition to the function of decoupling, inventory serves as a protection against fluctuation in demand and unreliability in supply (see Müller [2003], p. 3). In some cases inventory may also be used to protect against rising prices (see Müller [2003], p. 4). Inventory can also be used to decrease setup or purchasing cost by using economies of scale (see Axsäter [2006], p. 2). The buffering function of inventory can also have positive influence on freight cost, which will be discussed in detail in section 2.3. In summary it may be said that inventory serves two major goals, namely protection against uncertainty and creation of opportunities for lot-sizing in both input and output of inventory.

Types of inventory According to Shah [2009] inventory may be divided into different parts depending on its intended purpose:

- **Cycle inventory** is used to leverage economies of scale in production or procurement.
- **Safety stock** as protection against uncertainties in demand and supply.
- **Decoupling stock** used to enable decoupling between two fields of operation.
- **Anticipation inventory** used for speculative reasons, e.g. if rising prices are expected.

- **Pipeline inventory** may be divided into work-in-progress and transit inventories. It comprises materials that are currently worked on or that are currently being transported from one location to another.
- **Dead stock** is inventory which is not used at all, e.g. because it is obsolete or there is no demand.

Inventory cost Inventory also has an effect on the cost side. First, materials and parts in the inventory have to be bought. The capital invested in goods cannot be used to support other business operations. Thus it may be said that opportunity cost occur. If goods in the inventory are bought on credit, interest has to be payed. Both cases lead to inventory cost dependent on the value of goods in the inventory (see Kapoor and Kansal [2004], p. 133). This value can be measured in different ways, either by pegging the rate to the prime rate or by using so-called hurdle rates, which reflect expected return values on investment for capital deployed (see Bowersox et al. [2007], p. 136). Second, warehouses have to be built and operated, which leads to investments and cost for energy and personell. "These can be paid either in the form of rates charged by an outside firm offering such services or through internal costs generated from the particular operational activity system adopted in the company controlled warehouse" (Kapoor and Kansal [2004], p. 172f). There are different techniques of allocating this cost to different products, e.g. the space occupied measured in square or cubic meters or the amount of storage slots used in an automated storage system (see Kapoor and Kansal [2004], p. 134). Third, obsolescence cost can occur if goods are stored too long and cannot be used afterwards. Even though goods in the automotive industry do not decay in the short or medium terms, they can in fact decay. Additionally, frequent improvements in construction patterns can cause changes to parts and materials. If these are still in inventory when the improvement occurs, it could be that they cannot be used afterwards due to incompatibility or security reasons. Fourth, inventory has to be insured "as protection against inventory losses such as fire and theft" (Kapoor and Kansal [2004], p. 133).

Service quality The success of protection against uncertainty in demand and supply through inventory can be measured by service levels. There are three measures which are summed up under the term service level.

- The α -service level, which is orientated on the event of a stock-out. It is defined as "the probability that an incoming order can be fulfilled completely from stock" (Stadtler and Kilger [2008], p. 53).
- The β -service level, which is quantity-orientated and "is defined as the proportion of incoming order quantities that can be fulfilled from inventory on-hand" (Stadtler and Kilger [2008], p. 53).
- The γ -service level, which is time- and quantity orientated. In addition to the β -service level, it considers the time it takes to balance the backlog again. It can be defined as:

$$\gamma\text{-service level} = 1 - \frac{\text{mean backlog at end of period}}{\text{mean demand per period}}$$

(see Stadtler and Kilger [2008], p. 53).

Each of these values can be used to measure the level of service which an inventory provides for the following field of operation. Depending on emphasis, an appropriate value can be selected.

Inventory management Inventory management can be seen as a necessary coordination mechanism between two systems which are decoupled by an inventory. This applies as well to coordination between suppliers and recipient as well as coordination between inbound and production logistics. The role of inventory management is "to maintain a desired stock level of specific products or items" (Toomey [2000], p. 1). Desired in this case means that stock is high enough to fulfill the goals of inventory described above while at the same time remaining as low as possible in order to minimize inventory holding cost. To control the level of safety inventory two parameters are widely used, the *safety lead time* (SLT) and the *safety stock quantity* (SSQ). Safety lead time is a time-based safety parameter and represents the number of periods (usually measured in working days) a part will be ordered before the demand due date. A safety lead time value of two means that all parts are ordered two days earlier than they are required (see Swamidass [2000], p. 655). Safety stock quantity is a quantity-based safety parameter and represents the number of parts which should always be in the inventory, independent of whether or not a demand was predicted. In perfect circumstances safety stock quantities will always remain in inventory and will not be

	Quantity is higher than expected	Quantity is met	Quantity is lower than expected
Due date is earlier than expected	$\Delta t < SLT \wedge \exists D_{t2} : t2 \leq t1 + SLT \wedge FD_{t2} \geq D_{t1}$	$\Delta t < SLT$	$\Delta t < SLT$
Due date is met	$\exists D_{t2} : t2 \leq t + SLT \wedge D_{t2} \geq D_{t1}$	uncritical	uncritical
Due date is later than expected	$\exists D_{t2} : t2 \leq t + SLT \wedge D_{t2} \geq D_{t1}$	uncritical	uncritical

Table 2.1: Conditions under which safety lead time helps to protect against demand uncertainty.

planned for consumption (see Toomey [2000], p. 47). Both methods have advantages and disadvantages, as pointed out in van Kampen et al. [2010]. "A safety lead time is the more effective strategy for coping with supply variability" (van Kampen et al. [2010], p. 7478) while "holding a safety stock is to be preferred in coping with uncertainties in demand information" (van Kampen et al. [2010], p. 7478)). In contrast to safety stocks a safety lead time builds up a dynamic buffer which exists only if a demand is forecasted. This in turn means that safety lead time can only help to satisfy an unpredicted demand if one of the conditions listed in Table 2.1 holds true. At the same time this feature offers the opportunity to reduce safety inventory for parts that are seldom used. Conversely, a safety stock will also protect against unforecasted demand, but it requires more total safety inventory if all parts are secured by safety stocks.

2.2 Automotive supply chains

Automotive supply chains consist of one or more car retailers, one *Original Equipment Manufacturer* (OEM) and its plants (including final assembly plants and component plants) and multiple suppliers. Each supplier can have its own suppliers, which leads to a so-called supply network or supply chain. Suppliers in the supply chain can be divided into tiers according to their position within the value creation process as depicted in Figure 2.4. A tier-one supplier delivers his goods directly to the OEM, while a tier-two supplier delivers his goods to a tier-one supplier and so on. Within a supply chain, material flow from the source (tier n) to the sink (customer), described as the

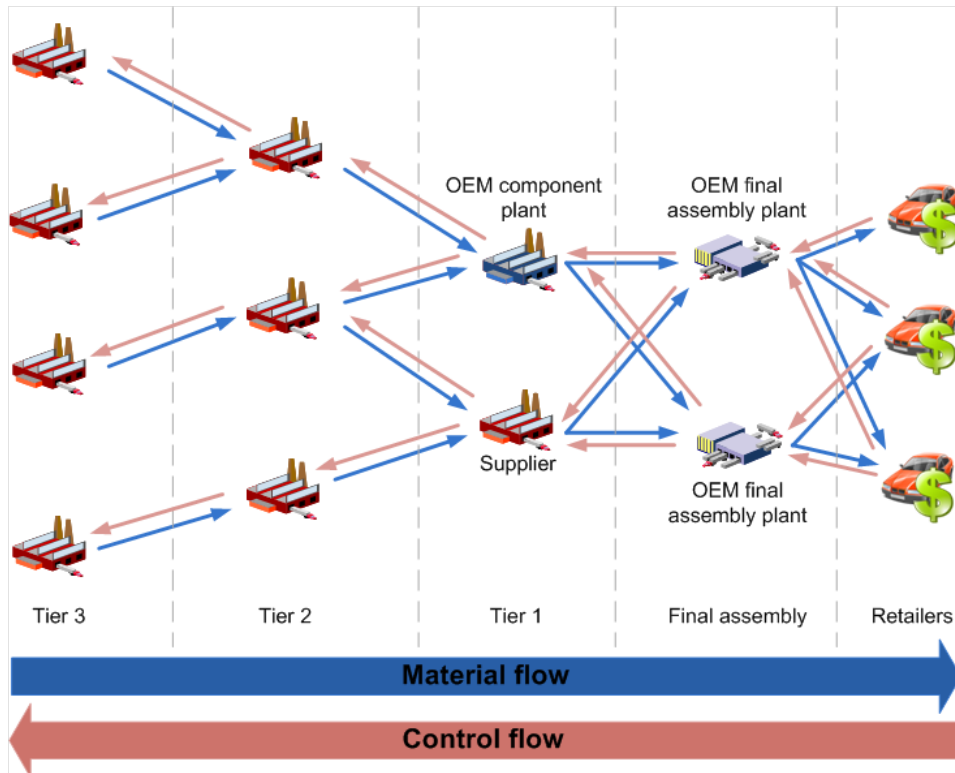


Figure 2.4: An exemplary automotive supply chain.

downstream direction. At each stage of the supply chain materials are transformed into more complex materials by production processes. At the same time control information flows from the customer to tier n supplier, in an upstream direction. Organizational boundaries might not reflect the role of a plant within the supply chain. A plant which organizationally belongs to the OEM but in fact produces components, e.g. engines or gearboxes, can also be seen as a supplier for the final assembly plant.

2.2.1 Material flow control

The material flow can be controlled either by a pull- or push-system. "In a pull (or make-to-order (MTO)) system, finished products are manufactured only when customer require them" (Ghiani et al. [2004] p.4). Production jobs are used to determine the demand at a production stage and the demand is then communicated to the next production stage. By contrast, push (or make-to-stock (MTS)) systems have the in-

ventory located at the end of each production stage. If material is required, the next higher production stage fulfills the demand from the supplier's inventory, and the lower production stage is responsible for filling up the inventory again. While pull-based systems can be harder to implement and require a higher level of control, they provide the advantage of a reduced supply chain inventory, shorter lead times and less total system cost (see Simchi-Levi et al. [2003] p. 122 ff). This advantage is especially important in the automotive industry, where a high variability of parts exists which would necessitate a huge inventory of expensive finished goods at the supplier's outgoing goods warehouse. A common approach with push-systems is the vendor managed inventory (VMI). In VMI environments the automotive company has no opportunity to control the material flow and thus no operational order lot-sizing is necessary. In this work only pull-systems will be discussed as most automobile production networks are pull-based systems in which OEMs control the material flow.

Delivery schedules To share information about required materials with the suppliers the OEM sends out *Delivery Schedules* at a regular interval. These contain an array of orders for each required part. An order consists of a part code to identify the part, a quantity and a date. According to Klug [2012] there are three established standards for the electronic transmission of delivery schedules. The most popular in Europe is the standard given by the *German Automotive Manufacturers Association* (Verband der Automobilindustrie, VDA). Another important standard has been defined by the *Odette* (Organization for Data Exchange by Tele Transmission in Europe) organization. The Odette standard is an extension of the existing VDA standard and is used mainly by European automotive manufacturers and suppliers. Over the years the Odette standard has been continuously developed towards an Edifact subset. Edifact is the third important standard. It was defined by a working group of the United Nations and the European Union during the 80s and accredited by the International Organization for Standardization (ISO) in 1987 (see Klug [2012], p. 248) in the ISO Norm 9735. Edifact was originally designed for electronic data interchange among trade partners independently of an specific industrial sector and covers more than 220 types of message, whereof only a small subset is relevant to the automotive industry (see Klug [2012], p. 249). As depicted in Figure 2.5 the delivery schedule is divided into three separate parts. The first is defined in VDA norm 4905 or Odette DELINS respectively and provides the supplier with a weekly *forecast* for the next six to eight-

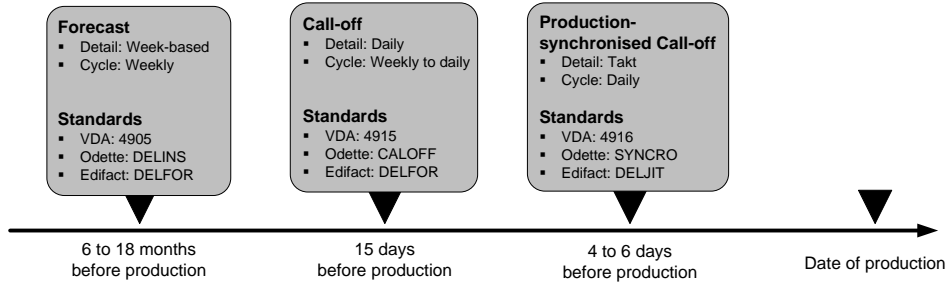


Figure 2.5: Components of a delivery schedule and the according standards according to Klug [2012].

een months. Delivery schedules are updated and sent out at regular short term interval and updated according to the current MPS at least every week. The second part, called *call-off*, is defined in VDA norm 4915 or Odette CALOFF and contains a daily delivery schedule for the next 15 days which is updated in planning cycles between one day and one week. Both forecast and call-off are integrated in Edifact DELFOR if the Edifact standard is applied. The third part is the so-called *production-synchronized call-off* defined in VDA norm 4916, Odette SYNCRO and Edifact DELJIT, which contains detailed scheduling information about the production sequence of the next days. The detail level of this production-synchronized call-off is one takt level. In addition to this information, a delivery schedule also contains the cumulative quantity of materials that has passed the incoming goods department since a fixed date, e.g. the first day in the calendar year. The cumulative quantity can be compared with the cumulative quantity of goods which left the suppliers outgoing goods department. The difference between the two cumulative quantities indicates the quantity of goods which is still in transport. In order to reduce fluctuation in delivery schedules and maintain the supplier's ability to create his own production schedule according to the delivery schedule an allowed level of fluctuation is negotiated between the individual supplier and the OEM. Klug [2012] states that these fluctuations typically cover a bandwidth of $\pm 20\%$ for the forecast part and $\pm 5\%$ for the call-off part. In addition, the first few days are considered as a *frozen zone* (see Graf [2006], p. 432). Orders within the frozen zone may not be changed in the next iteration. The duration of the frozen zone is to be negotiated between the supplier and the OEM.

Replenishment lead times The term lead time describes "the interval between when an order is placed and when it is received" (Basinger [2006], p. 45). Lead times are defined for each part in negotiations between supplier and buyer. Orders in the delivery schedule which are within the lead time may not be changed in a future delivery schedule. The lead time is equivalent to the frozen zone in VDA 4905.

Escalation processes If a demand is forecast to be within the lead time, an escalation process is started. The material controller responsible for the specific parts inventory contacts the supplier and asks whether it is possible to deliver the parts in time even though the lead time is not met. If the request is declined, the production job is blocked and an internal rescheduling is required. This two-stage process is defined as an standardized process and carried out each time a part bottleneck occurs (see Graf [2006], p. 447). Internal rescheduling may lead to non-optimal master production schedules - which were optimal before rescheduling. Thus an escalation process may impose high costs. In addition to costs due to required rescheduling, other parts for the same production job cannot be used and therefore increased inventory results, which leads to additional cost for warehousing and tied capital.

2.2.2 Planning process in automotive supply chains

The complex production process requires a systematic planning approach. In this section the existing planning process will be discussed. Figure 2.6 gives an overview of the planning process. In a first step the sales planning defines the gross primary demand forecast. Production planning then defines a master production schedule. Based on the master production schedule, materials requirement planning determines the net dependent demands. The net dependent demands are used as input for the operational order lot-sizing to create the delivery schedule.

Sales planning To sell cars to customers most automotive manufacturers offer two sales channels based on two shop-like concepts. On the one hand there are company-owned sales subsidiaries and, on the other retailers which are independent of the automotive company itself, but have contracts with the automotive company to sell only cars of brands of a single automotive company. A customer can choose to either buy a preconfigured car which is immediately available from the retailer or to configure

Production planning It may be said that "the production system in a car assembly plant usually comprises the four stages pressing of metal or aluminium sheets, welding the body-in-white from the moulded sheets in the body shop, painting it in the paint shop and final assembly, where painted body, engine, transmission and the further equipment are brought together or built in" (Günther and Meyr [2009], p. 347). Body shop and paint shop consist of large machinery and robotic production lines and are operated in a typical lot production mode. In "the final assembly in vehicle and vehicle component plants where variant flow production with low automation and high labor intensity exists" (März et al. [2010]) there is a high cost sensitivity depending on the right production sequence. In the *Master Production Scheduling* (MPS) a production schedule for the next several weeks is created. The schedule determines on which days which production orders should be processed. The MPS schedule is recreated regularly, e.g. every day for the short-term horizon of the next two weeks and every week for the medium-term horizon of the next several weeks. Due to its labor-intensive design the final assembly line is the most expensive part of the production process. Master production scheduling accordingly lays an emphasis on the efficient use of the final assembly line.

Sequencing and Resequencing Based on the MPS schedule a production sequence for the production orders of one day is created. The production sequence defines the order in which the production orders should pass the final assembly line. If a production order cannot be fulfilled for various reasons ranging from missing parts to problems with the paint of the vehicle body, a re-sequencing takes place. The production order is taken from the current spot in the production sequence and is exchanged with another production order or moved to another day.

Materials requirement planning Procurement plays an important role in the automotive industry, as nearly 80 percent of a car's value corresponds to parts provided by suppliers. Parts from various suppliers are required in final assembly. Due to the highly customizable product there is a considerable variance in the required parts. Meyr [2004] states that BMW offers up to 10^{32} variants of a single car. Due to the high product variance the demand for specific parts is highly dynamic even if the number of produced vehicles may be almost constant. These circumstances demand vital cooperation between suppliers and OEMs. At the beginning of the materials require-

ment planning, only primary demands in form of production jobs are known and the question is which materials will be required to satisfy the primary demands. This planning step can be supported by so-called *Materials Requirement Planning* (MRP) systems. Depending on the position within the supply chain, the production jobs originate either from the next higher production stages delivery schedule or from sales forecasts. To evaluate which parts are required for which production orders, a production order-specific *Bill-of-Materials* (BoM) is used. This step is called bill explosion. The resulting information on material quantities has to be combined with a due date which results from the due date of the final product given by the MPS schedule and the production lead time. If multiple production steps are necessary, e.g. if a material will be processed first into a semi-manufactured product, this step will be repeated at the next lower production stage until quantities and due dates for purchased materials are available. These quantities and due dates for purchased materials are called *gross dependent demands*. In a second step gross demands are charged up against current inventory and fixed orders and then modified according to preset safety parameters. As this part of the process in the MRP system is of importance for the understanding of the problem itself, the basic mechanisms will be described shortly. In a first step the expected demand at the end of the last period within the replenishment lead time is derived. In so doing, the following steps are repeated for each period in the replenishment lead time. As depicted in Figure 2.7 the units in stock at the beginning of the period are charged up against the expected gross demand for that period and the fixed orders expected to arrive in that period are then added. The result is the expected units in stock at the end of the period, which is then used as a starting point for the following period. For periods after the replenishment lead time the desired inventory level is determined by adding up the safety stock quantity and all gross dependent demands within the safety lead time. The idea is that, in every planning period, the inventory should always be capable of fulfilling the gross dependent demands of the number of days given by the safety lead time parameter. In addition, the safety stock quantity should protect against unexpected fluctuation in demand. In the third step the desired stock level is compared with expected units in stock at the end of the period. If the desired stock level is lower than the sum of the initial stock level less gross dependent demands, a net dependent demand is created. These are then used as input for the operational order lot-sizing. For the remainder of the computation process the units

Period	T = 1	T = 2	T = 3	T = 4	T = 5	T = 6	T = 7	T = ...
T ≤ Replenishment lead time			T > Replenishment lead time					
(Expected) units in stock at the beginning of the period	20 ⊖	25 ⊖	20	57	42	52
Gross dependent demand	10 ⊕	20 ⊕	10	30	15	15	25	...
Consider Fixed orders with a replenishment lead time of 2 periods	15 ⊖	15 ⊖	-	-	-	-	-	-
Net dependent demands			47 ⊖	15 ⊖	25 ⊖
(Expected) units in stock at the end of period	25	20	10	27	27	37		
(Desired) units in stock at the end of period according to safety parameters given	42	52	57	42	52
- Safety stock of 12 units	12	12	12	12	12	12	12	12
- Safety lead time of 2 periods	30	40	45	30	40

Figure 2.7: Transformation from gross dependent demand to net dependent demand.

in stock at the beginning of the following period are considered to equal the desired stock level in the previous period, as it is assumed that the net dependent demands will be ordered as desired.

Operational order lot-sizing planning In a last planning step, operational order lot-sizing, a delivery schedule for the net dependent demands is created. The delivery schedule prescribes on which day which amount of goods should be ordered. More details on operative order lot-sizing will be given in section 2.3.

2.2.3 Iterative planning in a rolling horizon

The planning steps described above are repeated regularly. Each time new information is available the complete planning process is repeated. One repetition of the planning

process will hereafter be referred to as *planning cycle*, whereas the time between two planning cycles is called *planning cycle time*. Only a short fragment of the plan created in one planning cycle is executed, and the remainder of the plan is discarded as a new plan is generated in the next planning cycle. Thus only the part of the plan that covers the planning cycle time is relevant to execution. If, for example, a plan is created every day, only the part of the plan that covers the next day will be executed: thereafter the remainder of the plan will be replaced by the newly generated plan. If the planning cycle time were longer, e.g. one week, the part of the plan to be executed would also cover one week. This method is one of the chief reasons for fluctuations in the plan. Small changes in the upper stages can result in huge differences at the end of the supply chain. This issue is known as the *bullwhip effect* initially discovered by Simon [1952] and explored by Forrester [1958]. Recent research by Lee et al. [1997] has identified demand forecast updating as one of the primary reasons for the bullwhip effect along with order batching, price fluctuation and rationing.

2.3 The operational order lot-sizing planning problem

The major question to answer in operational order lot-sizing is "Which parts should be ordered on which days and in what quantity such that all capacities are sufficient, all demands can be fulfilled and total costs are minimal?". The question already discloses the decisions to be made, the constraints and a part of the objective.

2.3.1 Objectives and decisions

The objectives of operational order lot-sizing are twofold. On the one hand there is the target to *minimize total cost*; on the other hand there are two soft targets which focus on the robustness of the overall process, namely *increasing delivery schedules stability* and *increasing the service level towards production*.

Minimize total cost The cost parts which can be directly influenced by the planning problem are inventory and freight costs. These are mutually contradictory of each other. If orders are accumulated over time, freight cost will be reduced due to higher vehicle use, but inventory cost will increase because stocks will be built up. If an order is placed every day, inventory cost will decrease as there is no inventory except that for safety inventory, but freight cost will increase because vehicle use may be low. The

goal is to find a balance between the different cost parts such that the total cost is minimized.

Stability of delivery schedules As the delivery schedule (and thus the output of the operational order lot-sizing) is used as input for the supplier's production planning, it is important to provide stability from one repetition to another. Otherwise an unstable series of delivery schedules induces higher production cost and lower service levels due to unfulfilled obligations. Lower service levels require higher safety inventory and a greater number of reschedules on the unloader's side. This leads to higher cost for production and inventory. Aside from the resulting bullwhip effect (see section 2.2.3), the fluctuation in delivery schedules makes it hard for the supplier to maintain his production planning adapted to constantly changing requests, and keep it cost-effective at the same time.

Maintaining a high service level towards production As described in section 2.1 inventory between two fields of operation is used as decoupling point. Inbound logistics stands at the beginning of the value chain and ends at the incoming goods department warehouse where production logistics takes over control. If the task of providing raw materials and parts in the incoming goods department warehouse is not fulfilled, production logistics cannot provide production with necessary materials and thus a production stop or rescheduling is necessary. A rescheduling or even a production stop causes high cost. From the inbound logistics perspective the incoming goods department warehouse can be seen as the customer and production logistics stock withdrawal orders can be seen as customer demands. The degree of satisfaction of customer demands is typically measured in service levels (see 2.1.2 for details).

Conflicting goals The goals described above are in partially mutual conflict with each other. A delivery schedule that is highly optimized for low cost may differ strongly from the last delivery schedule and thus reduce delivery schedule stability. Thus it is necessary to find a balance between cost and stability similar to the case of the lot-sizing problem itself, where a balance between inventory and setup cost has to be found. In contrast to the original lot-sizing problem, not all targets can be quantified in monetary values. In addition, stability of delivery schedules and service level towards production

cannot be measured within a single iteration of operational lot-sizing planning as they depend on a series of delivery schedules.

Decisions To achieve these goals multiple decisions have to be made. One part of these decisions are made explicitly, whereas others are made implicitly and can be inferred from the former.

A decision to be made explicitly is to determine which part should be ordered on which days and in what quantity. The result of the operational order lot-sizing is a delivery schedule which contains this information.

Depending on the delivery schedule several implicit decisions can be inferred. First, due to the tariff conditions described in 2.1.1, transport modes are also decided on when the decision on amount and date of delivery is made. Second, the number of load carriers is implicitly decided when setting up the delivery schedule as it can be derived from the parts order quantity. Third, it can be inferred from the delivery schedule and the net demands which parts will be stored in inventory for what duration.

2.3.2 Constraints

These decisions cannot be made without considering certain constraints imposed by the nature of the problem structure and its environment. These constraints are of two kinds. On the one hand capacity restrictions on the vehicular capacity, available storage area and incoming goods personnel have to be considered. On the other hand, all demands have to be satisfied.

Vehicular capacity has two limits: On the one hand, a vehicle has limited space inside its cargo hold, so there is a volume constraint which can be measured in either volume (e.g. cubic meters) or flat area (e.g. load meters). On the other hand, there is a boundary on the maximum load a vehicle may carry. The weight constraint can be expressed in a weight measure (e.g. kilogram). A vehicle can be seen as full if an additional load carrier would increase weight or volume such that one of the two constraints is exceeded. The inventory is usually limited for constructional reasons and can be measured in either the available storage area (in square meters), available storage space (in cubic meters) or the number of storage slots available for load carriers. As the existing warehouse is a long-term investment and increasing its capacity is expensive, the capacity can be seen as fixed and thus has to be considered a constraint.

Each time a truck arrives incoming goods personnel have to complete administrative tasks (e.g. checking invoice) and operational tasks (e.g. unloading the goods and transferring them into the warehouse). For each of these tasks a specific time is required. In order to balance this workload, the maximum amount of available working time in the incoming goods department can be limited, which introduces a constraint on human resources.

Due to the necessity to satisfy all production demands, orders must not be placed in an earlier period than that at which the net dependent demand was initially placed. This means in turn that a time-based consolidation of goods can only be achieved by ordering parts earlier than they are required and storing them in inventory until the day of consumption.

2.3.3 Algorithmic and rule-based delivery schedule generation

Two leading planning concepts can be applied to this problem. On the one hand, each time a delivery schedule has to be created an algorithm which uses the forecasted demands to obtain an optimal delivery schedule can be run. This method is hereafter referred to as algorithmic delivery schedule generation. Alternatively, a rule can be set up in advance and can then be used to create a delivery schedule each time a new schedule is required. In the remainder the following definitions will be used to distinguish between algorithmic and rule-based delivery schedule generation approaches.

Definition 1. *Algorithmic delivery schedule generation refers to approaches which use a mathematical model or an algorithm to compute a delivery schedule whenever a new forecast is given. The planning problem is to create a delivery schedule. The schedule or parts of it are then transferred to the MRP system.*

Definition 2. *Rule-based delivery schedule generation refers to approaches which set up a control rule in advance. The control rule specifies how to create a delivery schedule when a new forecast is given. The planning problem in this case is to find the control rule. The control rule will be transferred to the MRP system and the application of the rule to the forecast is executed by the MRP system.*

The proposed advantage of an algorithmic delivery schedule generation approach is that it can react more effectively to changing conditions. In theory this approach promises better results in the sense of cost reduction. This advantage is believed to come at

the price of a higher fluctuation in the plans because with each new set of information a completely different delivery schedule may be cost-optimal. Figure 2.8 shows an issue concerning algorithmic delivery schedule generation for operational order lot-sizing problems in a rolling horizon. To achieve a cost reduction an investment in form of inventory has to be made to achieve a higher rebate level. As goods are ordered earlier, depicted by the slashed lines in Figure 2.8, additional capital commitment has to be accounted. If the investment is made according to a calculation depending on a forecast which changes afterwards, it may be useless and thus not reduce cost, but conversely increase total cost and waste storage space which is no longer available for further improvements. In the example given in Figure 2.8 two demands forecast on Monday and Tuesday of the third week are crossed out in red, meaning they are removed from the forecast. As the parts would already have been bought following the delivery schedule, the investment would be useless and increase total cost. Apart from this issue, the more important drawback is the nervousness of the created delivery schedules reported by practitioners. Common models and algorithms for delivery schedule generation do not consider previous plans. In a rolling horizon planning environment only the first part of the generated schedule is actually carried out and the rest of the delivery schedule is discarded when new information becomes available. This leads to less reliable delivery schedules and is a major source of the issues described in Section 2.3.1. Another issue addressed by practitioners is ease of understanding. While a rule-based delivery schedule generation approach is based on a rule that can easily be understood, an algorithmic delivery schedule generation approach's behavior cannot be foreseen and is hard to reconstruct except for simple examples. This may lead to logistic planners' refusing the use of the algorithmic delivery schedule generation approach due to a lack of trust (see Arnott and Dodson [2008], p. 764 f). In the following different control rules for delivery schedule creation will be discussed.

Fixed lot size One possible control rule is to determine a fixed lot size which will be used for all orders of a specific product. This method is also known as the (s,Q) (see Ghiani et al. [2004], p. 132) ordering policy. The determined lot sizes are often orientated on the maximum quantity of parts which fits into a single load carrier. The basic principle of this method is to find a lot size for which the order overhead cost, consisting of freight cost and handling cost, can be balanced with inventory cost. It has been widely explored in numerous publications. It requires frequent review to

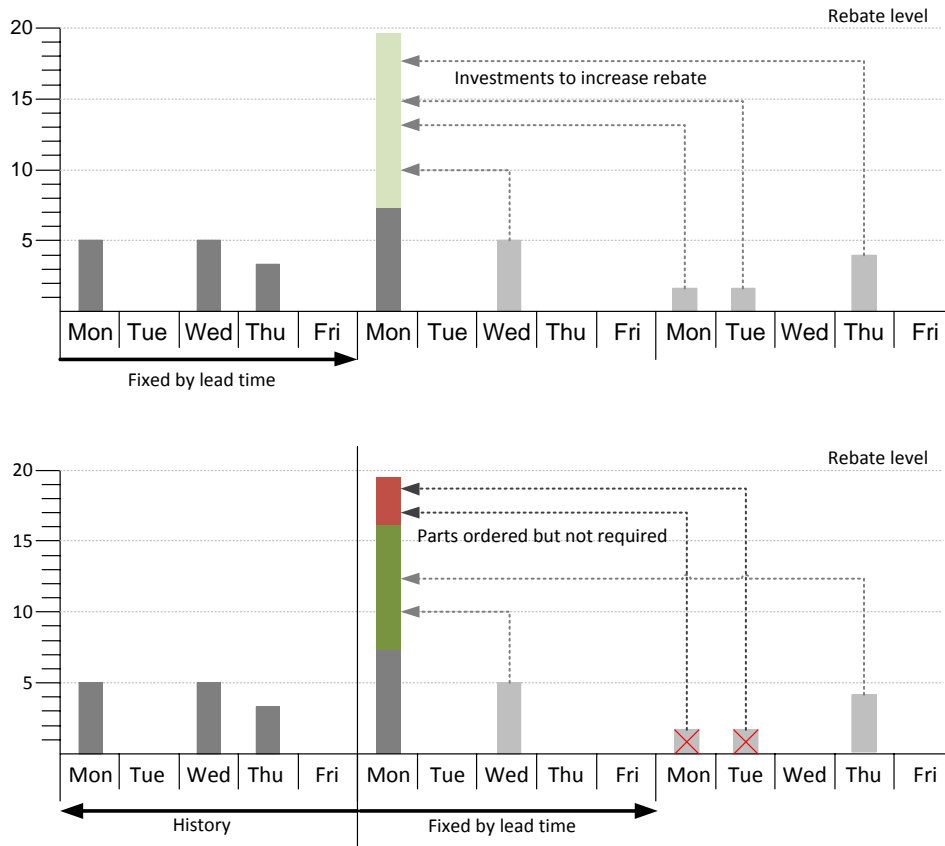


Figure 2.8: Example of the lot-sizing investment issue.

maintain efficiency in dynamic environment and, due to its focus on a single part, this approach cannot provide benefits from synergy effects among multiple parts from one supplier or even among multiple suppliers. However, the control rule may be used in combination with other control rules to deny orders of less quantity than one filled load carrier which can increase handling overheads in warehouses and incoming goods processing.

Fixed replenishment cycle time This control rule, also known as (t, S) (see Ghiani et al. [2004], p. 132) ordering policy, assigns a replenishment cycle to each part and sets up an order each time the cycle has passed. The quantity of the order will be determined

Delivery profile	Mon	Tue	Wed	Thu	Fri	Frequency
11111	✓	✓	✓	✓	✓	Weekly
10000	✓	-	-	-	-	Weekly
10100	✓	-	✓	-	-	Weekly
10101	✓	-	✓	-	✓	Weekly
010002	-	✓	-	-	-	Bi-weekly
10000M	✓	-	-	-	-	Monthly

Table 2.2: Exemplary delivery profile transformation rules.

dynamically depending on the demands forecast until the next replenishment day. Gudehus and Kotzab [2009] differentiates between single-article cycle time strategy, where an order is placed for a single part if it is necessary to do so, and consolidated cycle time strategy where a cost-optimal share of all parts from the same source will be ordered if a single part from a given source requires an order (see Gudehus and Kotzab [2009], p. 316). It may be stated that in comparison with quantity-based approaches, "with cyclic scheduling optimally filled load units and transport means are easier to achieve" (Gudehus and Kotzab [2009], p. 316).

Delivery Profiles The delivery profile control rule, also called common replenishment epochs, see Viswanathan [2001], is an improved variation of the replenishment cycle control rule. The idea behind delivery profiles is to manage the transport frequency or replenishment cycle in such a way that it adapts to the week-oriented production rhythm within a plant. The major advantages over the replenishment cycle are the adaption to the week-oriented production rhythm in a plant on the one hand and the possibility of providing a more detailed schedule accounting for fractional cycle lengths on the other hand. A delivery profile is a control rule that defines a set of days on which a delivery is allowed, and a frequency of repetition (see Table 2.2 for an example). On each day that is not defined as a delivery day a suppliers' delivery is not foreseen in the delivery schedule. This control rule can be applied to the net dependent demands to receive a delivery schedule which fits the delivery profiles' pattern. When the rule is applied all net demands with due dates equal or greater than the delivery day and smaller than the next delivery day will be cumulated on the first delivery day. An example of this behavior is depicted in Figure 2.9. Delivery profile '10100' from Table 2.2 was selected as transformation rule. The delivery profile '10100' allows

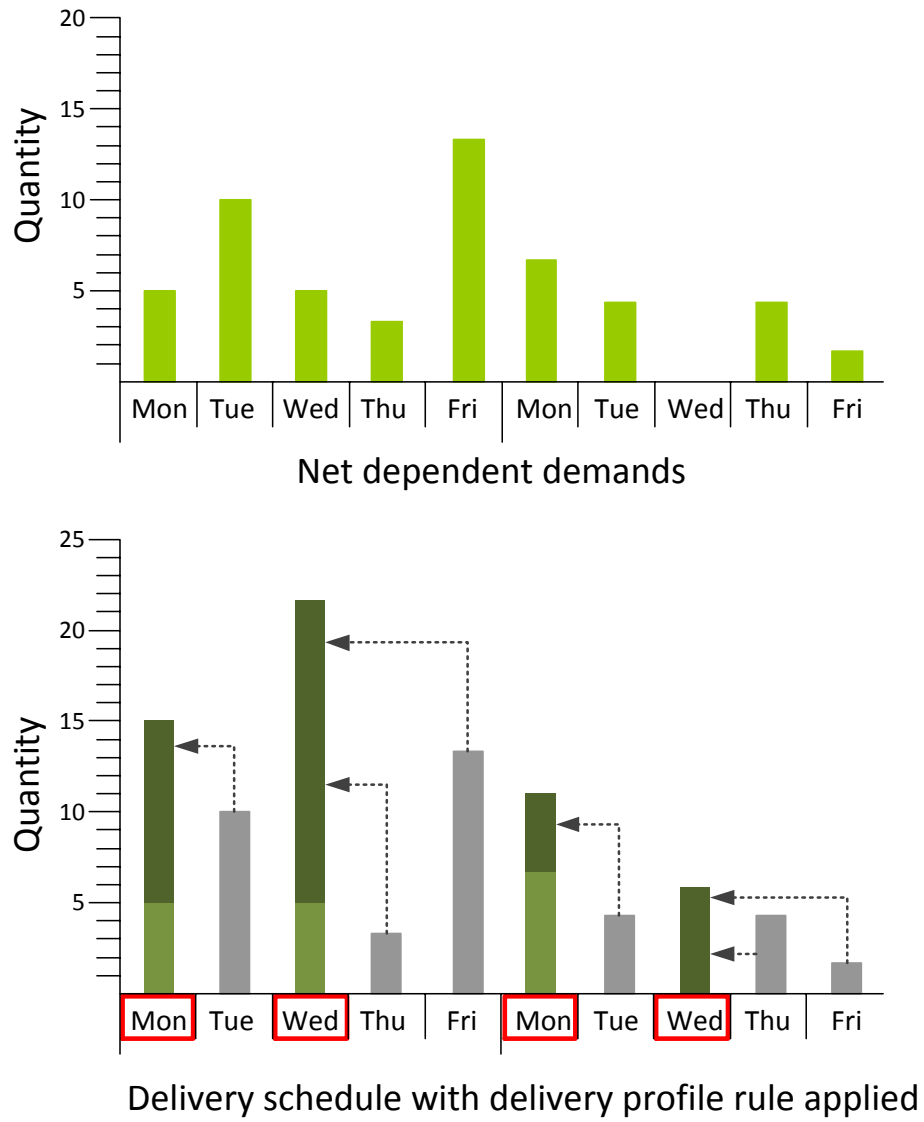


Figure 2.9: Translation from net dependent demands into delivery schedule based on a delivery profile rule. In this example 'W003' from Table 2.2 is used.

deliveries on Monday and Wednesday only. All net demands from Monday and Tuesday of the first week are cumulated to one order on Monday of the first week. Demands from Wednesday, Thursday and Friday of the first week are then cumulated to one order on Wednesday and so on. The goods will then be stored in inventory from their delivery day until they are finally requested by the production department. Common MRP systems provide the possibility to set up delivery profile control rules for each supplier via parameters and automatically apply the delivery profiles when creating delivery schedules.

2.4 Assessing the impact of cost minimal delivery profiles

Recently both researchers and practitioners have stressed the positive aspects of delivery profiles. Viswanathan [2001] considers an environment with a single vendor and multiple buyers, in which the vendor can assign a delivery profile (or common replenishment epochs in his terms) to each buyer. His results show that delivery profiles are a "coordination mechanism wherein it is able to consolidate several replenishment orders and economize on order processing and delivery costs" (Viswanathan [2001], p. 278). Another work on the same topic was that of Hwang [2008], who analyzed a VMI environment with a single supplier and multiple buyers. Hwang [2008] states that using delivery profiles strategy in combination with VMI results in "further savings for most cases with higher joint order cost" (Hwang [2008], p. 205).

In this thesis delivery profiles were chosen for a deeper analysis due to the positive discussions in both practice and the literature. Whereas Hwang [2008] and Viswanathan [2001] consider distribution logistics, this work considers the application to an area forwarding inbound logistic network. Both authors present a situation in which a single supplier delivers his goods to multiple buyers. In the application presented in this thesis the situation is the other way round in the sense that there is only one buyer but multiple suppliers. Applications to both inbound logistics and distribution logistics have in common that a set of supply chain partners has to be coordinated in order to reduce freight cost and that the underlying logistic network offers a structure with the possibility of consolidated runs that can be used to achieve synergy effects. Given these similarities and differences between the considered problem settings, three research questions arise.

- *How can cost minimal delivery profiles be selected for an area forwarding based*

inbound logistic network with complex tariff structures? Selecting cost minimal delivery profiles means making an optimal trade-off between the most important cost factors in the inbound logistic, freight cost, on the one hand and inventory holding cost on the other. Due to the complexity of the tariff systems, the number of different parts and suppliers and the availability of various valid delivery profiles, this task can not be fulfilled manually. An optimization system could be employed that uses a mathematical model to answer the question. In so doing the relevant aspects from the problem setting have to be integrated in such a model.

- *Does the proposed cost advantage hold true in a rolling horizon planning environment?* As a delivery profile is a control rule whose real outcome can be observed only in a rolling horizon planning environment, the developed model can only provide an estimate of the realized cost. Whether or not the proposed cost advantage can be realized in a real-world application can only be determined if a fair comparison with traditional or MRP planning techniques can be drawn. An application in the real world cannot answer this question completely, as it does not allow comparison of two different approaches under the same conditions, because the conditions change during the application and two methods cannot be applied at the same time. Thus an artificial benchmarking environment is required that allows application of different delivery schedule generation approaches in a rolling horizon under exactly the same conditions.
- *Can the stability of the generated delivery schedules be improved?* Likewise, the stability of the generated delivery schedules can only be observed in a rolling horizon planning environment. Thus the measurement should take place in the same artificial environment. Whereas the realized cost is a well-defined performance indicator, it is not clearly defined how the stability of generated delivery profiles can be measured. Thus it is necessary to evaluate different approaches in respect of their explanatory power and to develop a set of performance indicators that allow us comparison of the stability of delivery schedules generated by different approaches.

The remainder of this thesis will be organized as follows. First, the state-of-the-art literature will be reviewed and discussed in respect of its applicability to the given

research questions in Chapter 3. In Chapter 4 it will be pointed out which aspects are not covered by existing approaches and which extensions and development have to be made to bridge this gap. Chapter 5 then presents the chosen solution approach for the selection of cost-minimal delivery profiles. In Chapter 6 the work on the artificial rolling horizon planning environment will be described. Finally, a case study with industrial applications is conducted in Chapter 7. Chapter 8 presents a summary of the thesis. In the conclusion proposals are made for further research.

3 Literature Review

In this Chapter, a summary of the most promising approaches from the literature will be given. Due to the nature of the problem this Chapter is divided into three parts. In the first, models and algorithms that cover important aspects of the determination of cost-minimal delivery profiles will be depicted and their applicability to the underlying problem setting and their capacity to deal with the special conditions of area forwarding logistic networks will be discussed. The second part of the Chapter is dedicated to approaches which can be used to assess the outcome of planning methods in a rolling horizon environment. The Chapter closes with a section focusing on performance indicators to assess the stability of the generated delivery schedules.

3.1 Models and algorithms covering important aspects related to the selection of cost-minimal delivery profiles

To derive a model that selects cost-minimal delivery profiles, multiple cost-relevant aspects of the problem setting have to be considered. First, when choosing delivery profiles order lot sizes will be determined and a trade-off between fixed cost related to an order of a part on the one hand and inventory holding cost on the other has to be found. This part of the problem is dealt with in classic lot-sizing literature (see Section 3.1.1). In addition to fixed cost related to an order of a single part, fixed cost of joint replenishments occur in area forwarding networks. Such a problem structure can be found in the Joint Replenishment literature (see Section 3.1.2) and will therefore be discussed in this Chapter. Moreover, joint order costs are bound to the given discounting scheme: therefore the model has to incorporate these discounting schemes to derive a valid objective function. Modeling of complex discounting schemes has been discussed in the literature on purchasing models, which will be discussed in Section 3.1.3. In area forwarding inbound logistic networks the material flow of the various parts cannot be handled without a consideration of the underlying network. Section 3.1.4 examines how these network structures are modeled in network design

and network flow models, and how these modeling techniques may be applied to the given problem setting. Section 3.1.5 closes with a discussion of approaches dealing with the uncertainty that derives from delivery profiles having to be selected according to a demand forecast, which is unreliable and will be revised in later iterations of the planning cycle.

3.1.1 Models for lot-sizing problems

The traditional lot-sizing problem originates from production planning where "a trade-off between low setup costs (favoring large production lots) and low holding costs (favoring a lot-for-lot-like production where sequence decisions have to be made due to sharing common resources)" (Drexel and Kimms [1997], p.222) has to be found. Due to their partly similar problem structure, production lot-sizing models also find applications in the field of order lot-sizing. In regard to order lot-sizing applications it may be stated that both problem statements have a similar structure. In both cases it has to be decided which quantity should be processed (produced or ordered respectively) in which period. In both cases a larger lot size can reduce the processing cost (setup or freight cost respectively) whereas a smaller lot size may reduce the inventory holding cost.

The first methodological research on the question how lots should be sized was conducted by Harris [1913] and Andler [1929]. They considered an environment with stationary demands and a single item being produced on a single production stage. In their studies they developed the *Economic order quantity* (EOQ) formula, which determines the optimal lot size for the single item case with a single production stage. The EOQ formula to compute the optimal lot size Q^* as defined by Andler [1929] reads as follows:

$$Q^* = \sqrt{\frac{2 \cdot D \cdot C^{Fix}}{C^{Inventory}}}$$

where D is the demand per period, C^{Fix} the fixed cost per order and $C^{Inventory}$ the the inventory holding cost per period.

Wagner and Within [1958] were the first to consider a more realistic environment with non-stationary demands. Dynamic lot-sizing models are an extension of the previously described static lot-sizing model. In this model class demand is considered to be non-stationary. Therefore a constant lot size would not yield optimal results in

such an environment. Thus instead of determining a constant lot size for the whole planning horizon, dynamic lot-sizing models allow determination of a specific lot size for each period. Dynamic lot-sizing models have been extensively studied over the recent decades and have often been applied to order lot-sizing problems. Several heuristic algorithmic approaches to the problem have been developed. The most notable are found in DeMatteis [1968] (Part-Period-Balancing, PPB) and Silver [1979] (Silver-Meal-Heuristic, SM). Over time more and more extensions to the basic problem have been made which lead to various different models and approaches. The basic model set up by Wagner and Within [1958], the *single item uncapacitated dynamic lot-sizing model*, can be noted as follows:

$$\text{Min} \sum_{t=1}^T x_t c_t^p + y_t c_t^s + s_t c_t^h \quad (3.1.1)$$

$$s.t. s_{t-1} + x_t = d_t + s_t \quad \forall \quad t = 1..T \quad (3.1.2)$$

$$x_t \leq \sum_{t'=t}^T d_{t'} y_t \quad \forall \quad t = 1..T \quad (3.1.3)$$

$$x_t, s_t \geq 0 \quad \forall \quad t = 1..T \quad (3.1.4)$$

$$y_t \in \{0, 1\} \quad \forall \quad t = 1..T \quad (3.1.5)$$

The model consists of three variables for each period in the planning horizon ($t = 1..T$) representing production quantity (x_t), setup decision (y_t) and quantity in inventory (s_t). Production quantity and inventory quantity are required to be positive and setup decision is a binary variable. Cost parameters are introduced for each of the variables. c_t^p stands for production cost per produced unit, c_t^s is the parameter for setup cost and c_t^h gives back inventory holding for one product unit in one period. The inventory balance equation 3.1.2 ensures that produced product units are either consumed by demand or put to inventory. It also allows satisfaction of a demand in one period with quantity from inventory accumulated in previous periods rather than from production in the current period. Production may not exceed the sum of demands from the current period until the last period, which is assured by equation 3.1.3.

To categorize the enormous amount of extensions to the basic lot-sizing model their

properties and underlying assumptions can be used. The major assumptions and properties available for classification (see Karimi et al. [2003]) are listed in the following:

- **Demand handling.** Models may either consider a stationary demand which has the same quantity in each planning period or consider a non-stationary demand with variations in quantity over time.
- **Fixed lot size or varying lot size.** In response to variations in demand over time, a varying (dynamic) lot size may provide better results. There are models which determine a varying lot size for each period while other models determine one lot size to be used for all periods.
- **Single item or multiple items.** Some models consider only a single item at once while other models consider multiple items at the same time.
- **Single stage or multiple stages.** Production environments often consist of multiple production stages. While some models are restricted to a single stage, some model are capable of dealing with multiple stages.
- **Capacity or resource constraints.** There are different ways of dealing with existing capacity or resources. They can either be ignored, in which case the model is said to be *uncapacitated* or they can be considered and explicitly constrained, in which case the model is called *capacitated*. Capacity may be considered either to be discrete or continuous.
- **Scheduling.** Scheduling becomes especially important in combination with dynamic lot sizes. If models also deal with scheduling aspects, they are considered to be lot-sizing and scheduling models. This is usually the case if varying lot sizes and multiple products are considered.
- **Planning horizon.** The model can either consider an finite or infinite planning horizon. In most cases, "a finite planning horizon is usually accompanied by dynamic demand and an infinite planning horizon by stationary demand" (Karimi et al. [2003], p. 366). In addition to the length of the planning horizon, its granularity can be used as a measure for classification. It may be differentiated between *small bucket* and *big bucket* problems (see Karimi et al. [2003], p. 366). Small bucket problems consider very short periods within which only one item

may produced at once. In big bucket problems several items can be produced in a single period.

- **Inventory modeling.** Inventory may be constrained by both upper and lower bounds or regarded as unbounded (see Jans and Degraeve [2008], p. 1627). In regard to inventory pricing different approaches can be used. Price-dependent capital commitment, storage slot use cost and fixed charges have to be listed at this point. Additionally, models can be classified according to their handling of inventory shortages. If shortages are allowed, they can either be satisfied by future demand, thus a *backlogging* takes place, or shortages can be punished by *lost sales*. In all other cases shortages are not allowed to occur in the model.
- **Setup.** There is also diversity in the modeling of setup procedures. First of all setup time and setup cost handling can be distinguished. Using setup time as intermediate step allows for more complex cost structures, e.g. modeling the number of necessary setup teams. Setup cost on the other hand involves direct accounting of each setup which takes place. Another differentiation can be made between *simple* and *complex* setup structures (see Karimi et al. [2003], p. 367). Simple setup in this case means that setup time and cost do not depend upon previously produced items or states of other machines, whereas complex setup means that either setup cost or time depends upon the previous state of the machine or on the state of other machines (e.g. joint setups).

For extensive surveys of lot-sizing models and extensions we may refer to Jans and Degraeve [2008], Karimi et al. [2003], Buschkühl et al. [2008], Robinson et al. [2009] and Drexel and Kimms [1997].

3.1.2 Models for joint replenishment problems

Joint replenishment problems are special lot-sizing problems with a complex setup structure which put their focus on cost savings that can be achieved "by coordinating the replenishment of some items" (see Boctor et al. [2004], p. 2667) in multi-item inventory environments. Three types of cost are considered, namely individual ordering cost for one product type, common ordering cost for all orders within one period and inventory holding cost per product unit. We distinguish *static demand joint replenishment problems* (SJRP) and *dynamic demand joint replenishment problems* (DJRP).

Static in this case means that demands are considered stationary: thus the same demand occurs in each period, whereas *dynamic* denotes non-stationary demands, where demand may differ from one period to another. Another differentiation between different models is based on the consideration of uncertainty. Whereas most models focus on a single deterministic demand scenario, some models explicitly consider uncertainty, either in form of demand distributions or demand scenarios. These problems are referred to as stochastic joint replenishment problems. In stochastic JRP replenishment lead times are introduced. These can be either zero, constant for all items, variable for each item or stochastic.

Models for static joint replenishment problems

Models for static joint replenishment problems aim to find a cost-minimal *order frequency* or replenishment cycle for each part in an infinite planning horizon. They identify the optimal cycle time for the (t,S) inventory management policy. According to definition 2, static JRP models can be considered a rule-based planning approach. Static JRP models can be further divided into deterministic and stochastic models. Deterministic models consider a constant demand and do not allow shortages to occur. Stochastic models by contrast consider a stochastically distributed demand. Due to unknown future demand behavior, other decisions are involved. Instead of fixing a parts replenishment cycle, parameters for ordering policies are determined. Goyal and Satir [1989] models the deterministic case with a frequency of replenishment N . All parts can be replenished in accordance with a multiple $k_p > 0$ of the replenishment frequency N . For $k_p = 1$, part p is replenished in every cycle, for $k_p = 3$ in every third cycle and so on (see Goyal and Satir [1989], p. 3). All parts P replenishment factors $k_p \quad \forall p \in P$ form the vector K . Given these assumptions the model can be formulated (following Goyal and Satir [1989] with slight modifications in notation) as:

$$Min \left(N \cdot \left(c^c + \sum_{p \in P} \frac{c_p^i}{k_p} \right) + \sum_{p \in P} \frac{D_p \cdot c_p^h \cdot k_p}{2N} \right)$$

where c^c is the common replenishment cost factor, c_p^i the parts individual replenishment cost factor and c_p^h represents the individual holding cost respectively. The demand of part p is denoted as D_p . Multiple approaches have been suggested to solve the equation

either optimally or heuristically. For a detailed discussion of these approaches the reader may refer to Goyal and Satir [1989].

As mentioned before stochastic static JRP models determine parameters for ordering policies under stochastic demand. Most work focuses on the (S, c, s) policy or coordinated control policy. In this case, if a part p 's inventory reaches the must-order point s_p , a review of all parts takes place and the part p is replenished in such a quantity that the order-up-to level S_p is reached. During the review all parts below a threshold (the can-order-point) c_p will also be replenished so as to raise their inventory to their S_p . Some researchers (e.g. Chern [1974], Simmons [1972]) consider ordering a fixed quantity Q_p for all items in a review process instead of filling up to S_p . The basic model for the stochastic static JRP model ordering policy as given by Goyal and Satir [1989] can be noted as

$$\text{Min} \sum_{p \in P} \left(T_p \cdot c^c + N_p \cdot c_p^i + I_p \cdot c_p^h + B_p \cdot c_p^b \right)$$

where N_p is the expected replenishment frequency and T_p the expected number of replenishments triggered by the part per period. I_p and B_p denote to the expected inventory respectively backorder level of part p . c_p^b is introduced as a cost factor for backorder levels. Goyal and Satir [1989] provide a detailed discussion of extensions and solution approaches to this problem.

For a detailed review of inventory models for joint replenishment the interested reader may be directed to Aksoy and Erenguc [1988] for the single supplier case and to Minner [2003] for the case of multiple suppliers.

Models for dynamic joint replenishment problems

Models for dynamic joint replenishment problems determine the ordering quantities and periods for a set of product types. In addition to ordering cost per product type, a common ordering cost is introduced which has to be paid if an order for any product type is placed in a period. Following Boctor et al. [2004] the deterministic DJRP can be noted as follows:

$$\text{Min} \sum_{t=1}^T z_t c_t^c + \sum_{p \in P} y_{t,p} c_{t,p}^i + s_{t,p} c_{t,p}^h \quad (3.1.6)$$

$$s.t. s_{t-1,p} + x_{t,p} = d_{t,p} + s_{t,p} \quad \forall \quad t = 1..T, p \in P \quad (3.1.7)$$

$$x_{t,p} \leq \sum_{t'=t}^T d_{t',p} y_{t',p} \quad \forall \quad t = 1..T, p \in P \quad (3.1.8)$$

$$z_t \geq \sum_{p \in P} y_{t,p} \quad \forall \quad t = 1..T \quad (3.1.9)$$

$$x_{t,p}, s_{t,p} \geq 0 \quad \forall \quad t = 1..T, p \in P \quad (3.1.10)$$

$$y_{t,p} \in \{0, 1\} \quad \forall \quad t = 1..T, p \in P \quad (3.1.11)$$

$$z_t \in \{0, 1\} \quad \forall \quad t = 1..T \quad (3.1.12)$$

The close relation between lot-sizing and dynamic joint replenishment models can be seen in the model. Equations 3.1.7 and 3.1.8 are extended versions of the constraints 3.1.2 and 3.1.3 from the classical lot-sizing model. The main differences between a multi-item lot-sizing model and the DJRP lie in constraint 3.1.9 and the objective function. Constraint 3.1.9 activates the binary variable z_t associated with common ordering cost (c_t^c) where any order has been placed in period t . In addition to common ordering cost, individual ordering cost ($c_{t,p}^i$) for each product type and inventory holding cost ($c_{t,p}^h$) are considered. In some cases, variable costs per ordered product unit are also included. These are only of interest in the dynamic case and if cost can vary over time, as they do not otherwise affect the optimal solution. Boctor et al. [2004] give an extensive summary and benchmark of solution approaches to the deterministic dynamic JRP. There are also two alternative model formulations presented which have been shown to be faster than the classic formulation.

3.1.3 Models for purchasing quantity discount problems

Models for purchasing quantity discount problems consider an environment in which products can be bought from different suppliers which provide discounting schemes for product prices. The discounting schemes accounted for are similar to tariff systems described in 2.1.1, except that discounts are given based on product unit quantity

rather than a measurement unit. It is therefore necessary to study these purchasing quantity discount models to obtain useful information on how to model discounting schemes.

Benton and Park [1996] categorize the models and approaches to this issue according to three properties. With an analogy to lot-sizing models from which the purchasing quantity discount models originate, demand consideration can be used as the distinguishing property. Demand can be seen as either stationary or non-stationary. The second differentiation can be made based on the underlying point of view. Several researchers consider not only the buyer's point of view but also the supplier's perspective. These models will not be considered in the following because they are not applicable to the given problem setting. Third, the different kinds of discounting scheme can be used. There are models designed explicitly for all-unit quantity discounts or for incremental quantity discounts and also some models that cover both types of quantity discounts¹. Another aspect that was not considered in Benton and Park [1996] is the application in environments with multiple parts or multiple suppliers, as Benton and Park [1996] only consider the single part, single supplier case.

The simplest form of quantity discount model is an extension of the EOQ formula. In case of an all-unit quantity discount the purchasing price for all parts depends on the ordered quantity. It can be said that for an order quantity q the price is C_r^{Unit} when $Q_{r-1}^{Max} \leq q \leq Q_r^{Max} - 1$. "The total cost curve is discontinuous at the price break quantities, so that the EOQs minimizing the total cost for each unit price could not be valid if they are out of the boundaries of the quantity discount interval for the discount price" (Benton and Park [1996], p. 221). Hence, "the optimal lot size is determined at either a valid EOQ or one of the price break quantities in terms of the minimum total costs" (Benton and Park [1996], p. 221). For detailed discussion of solution approaches on this model, see Benton and Park [1996].

When an incremental quantity discount is considered the cost function is both convex and continuous. Hence a derivative of the total cost function can be used to determine the optimal quantity. The total cost function can be put up as

$$TC = \left(\frac{U_r}{\Delta Q_r} + C_r^{Unit} \right) \cdot D + C^{Fix} \cdot \frac{D}{Q_r^{Max}} + \frac{i \cdot (U_r + \Delta Q_r \cdot C_r^{Unit})}{2}$$

¹Benton and Park [1996] do not include models that cover both discounting schemes at once because those models were developed after their review was finished.

where U_r is the marginal purchasing cost for the fraction $\Delta Q_r = Q - Q_{r-1}^{Max}$ of the order quantity which is

$$U_r = \sum_{r=2}^R (Q_r^{Max} - 1) \cdot (C_{r-1}^{Unit} - C_r^{Unit})$$

The optimal lot size can then be determined as

$$Q_r^* = \sqrt{\frac{2 \cdot i \cdot (C^{Fix} + U_r)}{i \cdot C_r^{Unit}}}$$

Chakravarty [1984] has extended the models to consider multiple parts, where a common rebate on all parts is given based on complete invoice value. Both all-units discounting and incremental discounting are considered. A shortest-path model is introduced for the all-units discounting case. In a further extension Benton [1991] has included multiple parts, multiple suppliers and a resource constraint in the case of stationary demands with all-units discount. A heuristic procedure based on that presented by Rubin et al. [1983] and a Lagrangian relaxation were used to solve the problem.

Chaudhry et al. [1993] consider a situation where a buyer has multiple opportunities to source a part from several vendors each offering discounting schemes and has to decide strategically on how many parts to order from which vendor under consideration of qualitative aspects. Two models for the single-period, single-item case are put up, one for all-units quantity discounts and one for incremental quantity discounts.

For non-stationary demands more complex models have to be developed. This is a more recent development as in previous decades computational effort was too high as complex MIP models are required. Even though several single-item lot-sizing heuristics have been developed or adapted for the quantity discount case (see Benton and Park [1996], p. 231 ff for further details), no mixed integer programming model formulation was included in the review. Federgruen and Lee [1990] were the first to extend the dynamic lot-sizing model from Wagner and Within [1958]. A dynamic programming procedure to determine the optimal lot sizes for a single part environment with non-stationary demand rates was developed for both all-units discounting schemes and incremental discounting schemes. Bregman and Silver [1993] have extended the Silver-Meal-Heuristic to consider all-units quantity discounts.

A special case has been analyzed by Xu et al. [2000], who considered an environment with multiple parts in which discounts are based on total invoice value rather than part quantity. This kind of discount is referred to as *Joint Business Volume Discount* (see Xu et al. [2000], 317). A mixed integer linear programming model for this case is introduced. Additionally, a solution method consisting of an heuristic procedure based on lower bounds for distinct item sets is presented.

Tempelmeier [2002] generalizes the problem setting and considers all-units and incremental discounts that may vary over time. This is an interesting approach as it also allows inclusion of special price offers that hold only for a short period. A supplier-specific fixed ordering cost and a restriction to certain delivery periods for each supplier as well as upper and lower bounds on order lot sizes can be integrated. Holding cost is computed upon on the product units real purchase price. Two models accounting for different types of quantity discounts are presented, the *Uncapacitated Multi-Supplier Order Quantity Problem with Time-Varying All-units Discounts* and the *Uncapacitated Multi-Supplier Order Quantity Problem with Time-Varying Incremental Discounts*. As the holding cost in the objective function is based on purchasing cost, the objective function becomes nonlinear in the all-units discount case. To cope with this issue and provide the ability to solve larger real-world problem instances, a two-phased heuristic solution procedure consisting of a construction phase and an improvement phase is presented.

In Reith-Ahlemeier [2002] an extended model for order quantity decisions and supplier selection was presented which includes lower and upper bounds for order quantities at part and supplier level, incremental quantity discounts as well as all-unit quantity discounts, and a resource-based capacity concept. The capacity concept includes suppliers' capacities, buyers' handling capacities and storage capacities. There is also a possibility provided to disable deliveries from suppliers in certain periods. Different model formulations including or excluding certain aspects of the problem are given. This includes a reformulation as a facility location problem (see Reith-Ahlemeier [2002], p. 52). The model becomes quite complex and can hardly be solved with linear programming solvers for real-world problem sizes. Hence three heuristic solution approaches are presented. These include a primal heuristic to create a starting value, a Lagrange-based heuristic and a modification of the Branch and Bound algorithm that uses certain problem-specific properties to reduce runtime.

Indices and sets

$t, \tau = 1..T$	Period t or τ in the planning horizon which ends in period T
$s \in S$	Supplier s from the set of all suppliers S
$p \in P$	Part p from the set of all parts P
$r = 1..R$	Discount level r . It may depend on part p , supplier s and period t

Parameters

d_{tp}	Demand for part p in period t
C_{tpsr}^{Unit}	Unit price of part p in period t from supplier s using discount level r
C_{tpsr}^{Level}	Fixed cost of an order for part p in period t from supplier s using discount level r
$C_{tps}^{Product}$	Fixed cost of an order for part p in period t from supplier s
$C_{ts}^{Supplier}$	Fixed cost of an order in period t from supplier s
Q_{tpsr}^{Max}	Maximum quantity of the interval for discount level r from supplier s in period t for part p

Variables

x_{tpsr}	Quantity of part p ordered from supplier s in period t at discount level r
\tilde{x}_{tps}	Quantity of part p ordered from supplier s in period t
u_{ts}	Indicator variable for an order from supplier s in period t , 1 if an order is placed, 0 otherwise
v_{tps}	Indicator variable for an order of part p from supplier s in period t , 1 if an order is placed, 0 otherwise
y_{tpsr}	Indicator variable for an order of part p from supplier s in period t with discount level r , 1 if an order is placed, 0 otherwise
q_{tp}^{τ}	Percentage of demand of part p in period t ordered for period τ

Stadtler [2006] considers an environment with multiple items and multiple suppliers in which multiple suppliers may offer the same product and non-stationary, deterministic demands. In addition to both incremental and all-units discounting schemes fixed ordering cost which are shared among items from the same supplier are considered in the objective function. Orders may be constrained by handling and storage capacities on the buyer's side and lower and upper bounds for product lot sizes on the supplier's side. The most important contribution was to establish a model formulation that considers both all-units and incremental quantity discounts within a single model. The basic model presented in Stadtler [2006] with slight modifications in notation is presented as follows:

$$\begin{aligned}
 Min \quad & \sum_{t=1}^T \sum_{s \in S} \sum_{p \in P} \sum_{r=2}^{R+1} \theta_t \cdot C_{tpsr}^{Unit} \cdot Q_{tpsr-1}^{Max} \cdot y_{tpsr} \\
 & + \sum_{t=1}^T \sum_{s \in S} \sum_{p \in P} \sum_{r=1}^R \theta_t \cdot C_{tpsr}^{Unit} \cdot x_{tpsr} \\
 & + \sum_{t=1}^T \sum_{s \in S} \sum_{p \in P} \sum_{r=2}^{R+1} \theta_t \cdot C_{tpsr}^{Level} \cdot y_{tpsr} \\
 & + \sum_{t=1}^T \sum_{s \in S} \sum_{p \in P} \theta_t \cdot C_{tps}^{Product} \cdot v_{tps} \\
 & + \sum_{t=1}^T \sum_{s \in S} \theta_t \cdot C_{ts}^{Supplier} \cdot u_{ts}
 \end{aligned} \tag{3.1.13}$$

subject to

$$\sum_{\tau=1}^t q_{tp}^{\tau} = \begin{cases} 1 & \text{if } d_{tp} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall t = 1..T, p \in P \tag{3.1.14}$$

$$\sum_{s \in S} x_{tps}^{\tau} = \sum_{\tau=t}^T d_{\tau p} \cdot q_{\tau p}^t \quad \forall t = 1..T, p \in P \tag{3.1.15}$$

$$x_{tps}^{\tau} = \sum_{r=2}^{R+1} Q_{tpsr-1}^{Max} \cdot y_{tpsr} + \sum_{r=1}^R x_{tpsr} \quad \forall t = 1..T, s \in S, p \in P \tag{3.1.16}$$

$$x_{tps,1} \leq Q_{tps,1}^{Max} \cdot y_{tps,1} \quad \forall t = 1..T, s \in S, p \in P \tag{3.1.17}$$

$$x_{tpsr} \leq (Q_{tpsr}^{Max} - Q_{tpsr-1}^{Max}) \cdot y_{tpsr} \quad \forall t = 1..T, s \in S, p \in P, r = 2..R \tag{3.1.18}$$

$$v_{tps} = \sum_{r=1}^{R+1} y_{tpsr} \quad \forall t = 1..T, s \in S, p \in P \tag{3.1.19}$$

$$v_{tps} \leq u_{ts} \quad \forall t = 1..T, s \in S, p \in P \quad (3.1.20)$$

$$x_{tpsr} \geq 0 \quad \forall t = 1..T, s \in S, p \in P, r = 1..R \quad (3.1.21)$$

$$\tilde{x}_{tps} \geq 0 \quad \forall t = 1..T, s \in S, p \in P \quad (3.1.22)$$

$$q_{tp}^\tau \geq 0 \quad \forall \tau = 1..T, t = \tau..T, p \in P \quad (3.1.23)$$

$$v_{tps} \in \{0, 1\} \quad \forall t = 1..T, s \in S, p \in P \quad (3.1.24)$$

$$u_{ts} \geq 0 \quad \forall t = 1..T, s \in S \quad (3.1.25)$$

The objective function 3.1.13 contains five parts, of which the first two "represent purchase costs at the limits of the purchase intervals plus purchase amounts within the intervals" (Stadtler [2006], p. 728) and the following three parts represent fixed cost associated with a discount level, a part and a supplier respectively. Note the use of θ_t as multiplier for all cost parts in the objective function, which is the interest rate accounting factor. Unlike Tempelmeier [2002] the model also considers capital commitment for fixed cost charges. If this is not intended, θ_t can be removed from the corresponding terms in the objective function. θ_t can be used instead of the traditional lot-sizing inventory pricing and can be computed with different equations depending on the desired meaning, representing either the net present value or capital commitment cost. For the capital commitment cost case the value can be computed for each period as

$$\theta_t = (1 + i)^{T-t+1}$$

where i is the interest rate per period. According to Fleischmann [2001] (see Fleischmann [2001], pp 152), "only the distribution of the inflow over time is to be planned" if "the total inflow [...] is also fixed" (Fleischmann [2001], p. 152). This means that optimal purchasing decisions are independent of the sequence of outflows. The exact capital commitment cost can be calculated after the optimal solution has been derived. Therefore a First-In-First-Out rule can be applied to determine the real unit prices C_{tp}^{Unit} for which a demand d_{tp} in period t has been bought. A capital commitment correction

value can then be computed as

$$\Delta C = \sum_{p \in P} \sum_{t=1}^T (T-t) \cdot d_{tp} \cdot C_{tp}^{Unit}$$

and subtracted from the objective value to obtain the correct total cost value. The constraints 3.1.14 and 3.1.15 cover demand fulfillment. Constraint 3.1.16 secures that the ordered amount x_{tps} "is composed of the lower limit of a specific discount level r plus a continuous amount within that discount level" (Stadtler [2006], p. 730). The link between the indicator variable y_{tpsr} and upper bounds on quantities is created in constraints 3.1.17 and 3.1.18, whereas constraint 3.1.19 ensures that a part may be ordered only in one distinct discount level. In constraint 3.1.20 the indicator variable v_{tps} for an order from a supplier in a certain period is set to 1 if any part is ordered from that certain supplier.

3.1.4 Models for minimum cost network design and flow problems

Network flow and design problems deal with the cost-minimal transport of commodities within a network. Modeling techniques from models for these problems may be helpful to model the area forwarding inbound logistic network which is used for transport in the given problem setting. Multi-commodity minimum cost network design problems consider a network with nodes N and directed arcs A . Within this network a set of commodities K has to be moved from sources to sinks while holding arc capacities. Each commodity k has exactly one source $S_k \in N$ and one sink $T_k \in N$ and a demand d_k . The basic multi-commodity minimum cost network flow problem formulation was developed by Tomlin [1966]. It is stated as (with slight formulation changes from Tomlin [1966])

$$\text{Min} \sum_{(i,j) \in A} \sum_{k \in K} x_{i,j,k} \cdot c_{i,j,k} \quad (3.1.26)$$

$$(3.1.27)$$

subject to

$$\sum_{k \in K} x_{i,j,k} \leq b_{i,j} \quad \forall (i,j) \in A \quad (3.1.28)$$

$$\sum_{j:(n,j) \in A} x_{n,j,k} - \sum_{j:(j,n) \in A} x_{j,n,k} = \begin{cases} d_k & \text{if } n = S_k \\ -d_k & \text{if } n = T_k \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in K, \forall n \in N \quad (3.1.29)$$

$$x_{i,j,k} \geq 0 \quad \forall (i,j) \in A \quad (3.1.30)$$

$$(3.1.31)$$

with $x_{i,j,k}$ representing the flow of commodity k on arc $(i,j) \in A$. The objective 3.1.26 is then to minimize the cost for arc use, given by the cost parameter for arc use $c_{i,j,k}$. Constraint 3.1.28 assures a correct flow balance in the network. For a commodities source the term on the left side becomes positive at the height of demand as there is no ingoing arc available. In case of a sink the term on the left side becomes negative for n as there are only incoming flows. On all other nodes the incoming flow is equal to the outgoing flow. These properties are secured by constraint 3.1.29.

Different types of extension are available for the basic model. One direction pursued in research is to include more complex cost functions, e.g. combining fixed cost value if an arc is used with variable cost for flows on the arcs or including alternative transport modes among the arcs. Another direction is to extend the model formulation by including a time-based aspect and aims at a coordination of flows over multiple periods.

Crainic and Rousseau [1986] modified the model according to a logistic service provider's needs. In addition to considering multiple commodities multiple transport modes were included in the model. Even though it was pointed out that time is an important aspect of logistics service networks, it was not explicitly considered in the model formulation. Rather, a frequency was introduced which indicates how often an arc will be used when the network is operated. This specific use of frequency of operation influences the objective function, but somewhat implies a stationary demand scenario.

Haghani [1996] introduced the technique of using a time-space network to model flows over time. This technique is quite important for the follow-up research as it interconnects the time-based aspect of lot-sizing problems with the network-orientated modeling in network design problems and is therefore depicted in Figure 3.1. For each period in the planning horizon, a separate copy of the underlying network is set up. These networks are then interconnected via arcs that represent a transformation in time. This transformation in time can be achieved either by explicit storage arcs

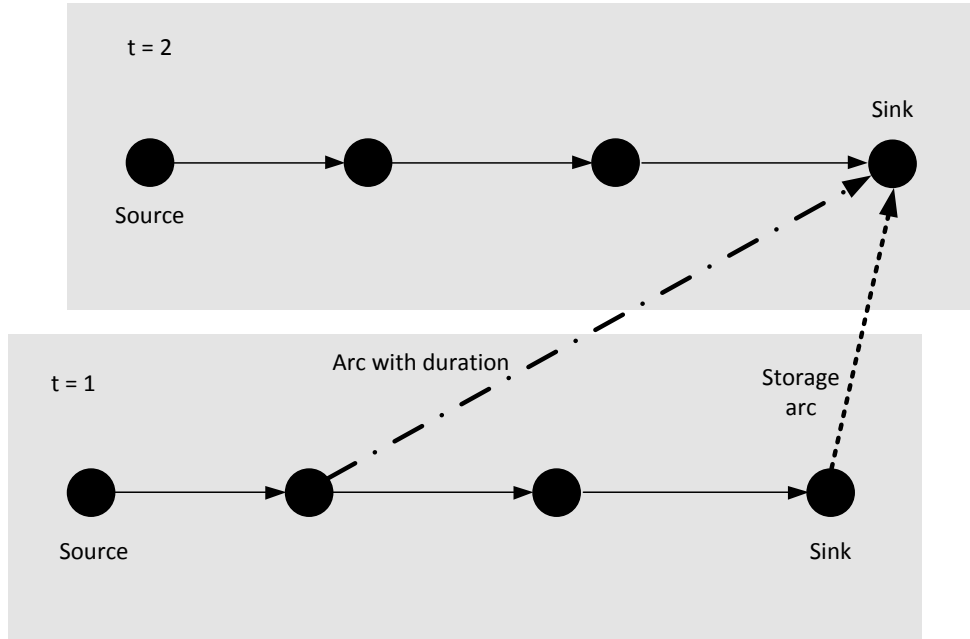


Figure 3.1: A typical time-space network representation.

(green line in Figure 3.1) leading to a transformation in time, but not in space, or by attaching a timespan to an arc from the original network (blue line in Figure 3.1), leading to a transformation in time and space, thus representing transport with a duration. The network presented by Haghani [1996] considers multiple commodities, alternative transport modes on the arcs, each holding separate vehicles and warehousing opportunities. The objective is to minimize "the sum of the vehicular flow costs, the commodity flow costs, the supply or demand carry-over costs, and the transfer costs over all time periods" (Haghani [1996], p. 238), where the term transfer costs refers to a change between different transport modes. This operation is only allowed on certain nodes.

For a detailed review of multi-commodity service network design problems, see Crainic [2000]. There, service network design problems in general and different decisions that have to be made by the carriers at a tactical planning level are discussed. A review of research in the area of service network design is conducted and a proposal on how to categorize the different approaches is offered. The network models are divided into three groups. The first group considers service frequencies explicitly as a decision variable in the model. Models within the second group do not explicitly con-

sider service frequencies but allow derivation of the frequency from the given solution. This approach reduces model complexity because "explicit capacities and a number of complicating constraints" (Crainic [2000], p. 285) are removed from the model. The third group consists of models which explicitly consider time-space networks to cope with a non-stationary demand scenario. At the time of review only a few approaches from the third group were available.

Chen [2005] developed a model that includes multiple commodities, alternative modes and so-called time windows. Time windows describe a relaxed demand constraint. A commodity is seen to be in time if it reaches the sink within a window of opportunity called a time window. Unlike in replenishment environments, the commodity is not accepted if it arrives before the time window opens. If the commodity arrives after the time window, a penalty cost is introduced. For each arc fixed and variable cost are considered. Fixed cost is only applied if a vehicle is used on the arc. If it is necessary to use multiple vehicles on the same arc, additional fixed cost will be added to the objective value. In addition to the model formulation using a time-space network, a heuristic solution approach based on Lagrangian relaxations for the specific problem is presented. Chen [2005] reports that the developed solution algorithm "is more computational efficiency than solving original problems directly" (Chen [2005], p. 51).

Resource based network flow models

Recently, Kempkes and Koberstein [2010], Kempkes [2009] presented a model that combines aspects from time-space multi-commodity network design models with quantity discounting models to reflect a logistics supply chain environment with tariff systems. The model can be used for both internal and external logistics networks at the same time. A resource-based concept is used to model cost factors such as handling cost, freight cost and inventory holding cost. The model also considers multiple vehicles or transport modes to be used. Parts are not seen to be streamed directly through the network but are rather packed into load carriers. Even a repacking between different load carrier types applicable to the same part type is considered. In the following a short form of the model depicting the most important aspects taken from Kempkes and Koberstein [2010] given the notation presented in Table 3.1.4 will be sketched.

Indices and sets

$t = 1..T$	Period t or τ in the planning horizon which ends in period T
$n \in N$	Nodes in the network
$N^S \subset N$	Supplier nodes
$(i, j, t, t') \in A$	Arcs in the network with (i, j, t, t') representing an arc from node i to node j and from period t to period t'
$p \in P$	Parts
$c \in C$	Load carriers
$r \in R$	Resources
$rg \in RG$	Discounting scheme rg
$rs \in RS_{rg}^{RG}$	Discounting scheme levels rs in discounting scheme rg
$r \in R_{rg}^{RG} \subseteq R$	Resources relevant for discounting scheme rg

Parameters

$d_{i,p,t}$	Demand for part p in period t at node i , zero for all nodes without demands
$s_{i,p,t}$	Supply of part p in period t from supplier node $i \in N^S$
$u_{i,j,p,r}$	Use of resource r per part unit when using connection from node i to node j for part p
$u_{i,j,c,r}$	Use of resource r per load carrier unit when using connection from node i to node j for load carrier p
LB_r^R, UB_r^R	Lower and upper bounds on resource use of resource r
$LB_{rg,rs,r}^{RG}$	Lower and upper bounds on use of resource $r \in R_{rg}^{RG}$ for discount level rs in discounting scheme rg
$UB_{rg,rs,r}^{RG}$	
$f_{rg,rs}^{RS}$	Fixed cost charge when using discount level rs in discounting scheme rg
$c_{r,rg,rs}^{RS}$	Variable unit cost factor of resource $r \in R_{rg}^{RG}$ using discount level rs in discounting scheme rg

Variables

$x_{j,i,p,c,t',t}$	Quantity of part p moving in load carrier c along arc $(i, j, t, t') \in A$
$y_{j,i,p,c,t,t'}$	Quantity of carriers c ordered from supplier s in period t

$$\begin{aligned}
 & \text{Min} \sum_{t \in T} \sum_{r \in R} g_r \cdot k_{r,t} \\
 & + \sum_{t \in T} \sum_{rg \in RG} \sum_{rs \in RS^{RG}} \sum_{r \in R_{rg}^{RG}} (c_{r,rg,rs}^{RS} \cdot (k_{r,t,rg,rs}^{RG} - LB_{rg,rs,r}^{RG} \cdot \omega_{rg,rs,t}^{RG})) \\
 & + \sum_{t \in T} \sum_{rg \in RG} \sum_{rs \in RS^{RG}} f_{rg,rs}^{RS} \cdot \omega_{rg,rs,t}^{RG}
 \end{aligned} \tag{3.1.32}$$

subject to

$$d_{i,p,t} = \sum_{j,t':(j,i,t,t') \in A, c \in C} x_{j,i,p,c,t,t'} - \sum_{j,t':(i,j,t,t') \in A, c \in C} x_{j,i,p,c,t,t'} \forall t \in T, i \in N \setminus N^S, p \in P \tag{3.1.33}$$

$$s_{i,p,t} = \sum_{j,t':(j,i,t,t') \in A, c \in C} x_{j,i,p,c,t,t'} - \sum_{j,t':(i,j,t,t') \in A, c \in C} x_{j,i,p,c,t,t'} \forall t \in T, i \in N^S, p \in P \tag{3.1.34}$$

$$k_{r,t} = \sum_{i,j,t':(i,j,t,t') \in A, p \in P, c \in C} (u_{i,j,p,r} \cdot x_{i,j,p,c,t,t'} + u_{i,j,c,r} \cdot y_{i,j,p,c,t,t'}) \forall r \in R \tag{3.1.35}$$

$$\tag{3.1.36}$$

$$k_{r,t} = \sum_{rs \in RS_{rg}^{RG}} k_{r,t,rg,rs}^{RG} \forall rg \in RG, r \in R_{rg}^{RG} \tag{3.1.37}$$

$$y_{i,j,p,c,t,t'} \geq \frac{x_{i,j,p,c,t,t'}}{q_{p,c}} \forall (i,j,t,t') \in A, p \in P, c \in C \tag{3.1.38}$$

$$\sum_{rs \in RS_{rg}^{RG}} \omega_{rg,rs,t}^{RG} \leq 1 \quad \forall rg \in RG \tag{3.1.39}$$

$$k_{r,t,rg,rs}^{RG} \leq UB_{rg,rs,r}^{RG} \cdot \omega_{rg,rs,t}^{RG} \forall t \in T, rg \in RG, rs \in RS_{rg}^{RG}, r \in R_{rg}^{RG} \tag{3.1.40}$$

$$k_{r,t,rg,rs}^{RG} \geq LB_{rg,rs,r}^{RG} \cdot \omega_{rg,rs,t}^{RG} \forall t \in T, rg \in RG, rs \in RS_{rg}^{RG}, r \in R_{rg}^{RG} \tag{3.1.41}$$

$$LB_r^R \leq k_{r,t} \leq UB_r^R \forall t \in T, r \in R \tag{3.1.42}$$

$$x_{j,i,p,c,t,t'}, y_{j,i,p,c,t,t'} \geq 0 \forall (i,j,t,t') \in A, p \in P, c \in C \tag{3.1.43}$$

$$\omega_{rg,rs,t}^{RG} \in \{0, 1\} \forall t \in T, rg \in RG, rs \in RS_{rg}^{RG} \tag{3.1.44}$$

The most important contribution of Kempkes [2009] was to establish a resource-based formulation for discounting schemes. Whereas previous models focused on quantity discounts or business volume discounts, this model formulation allows discounting schemes to be based on multiple resources at the same time. These resources can but do not necessarily have to reflect the parts price. This is especially helpful to modeling tariff systems as described in Section 2.1.1, which are based on measurement units, e.g. weight or volume. The concept works as follows. Rebate groups $rg \in RG$ are introduced, which represent the different discounting schemes. Each rebate group consists of several discount levels $rs \in RS_{rg}^{RG}$. For each discount level, lower ($LB_{rg,rs,r}^{RG}$) and upper bounds ($UB_{rg,rs,r}^{RG}$) are given for each resource $r \in R_{rg}^{RG}$ that is part of the corresponding rebate group. "A resource can only be part of at most one resource group, which means that the sets R_{rg}^{RG} must be pairwise disjunctive" (Kempkes and Koberstein [2010], p. 286). If resource use of all resources within $r \in R_{rg}^{RG}$ lies within the bounds of one discount level (Constraints 3.1.40 and 3.1.41), the discount level is said to be active. If a discount level is active ($\omega_{rg,rs,t}^{RG}$ is set to 1), two cost factors are considered, a fixed cost for the discount level itself ($f_{rg,rs}^{RS}$) and a variable cost ($c_{r,rg,rs}^{RS}$) for resource use within the rebate level. Note that only the utilisation within the bounds of the discounting level is accounted for by the variable cost factor.

The model presented by Kempkes [2009] can to some extent be used to determine delivery profiles, even though it was originally not intended for this task. The major drawback of the model is that it does not consider design decisions. Accordingly, model instances have to be created in a sophisticated manner to incorporate decisions that last over the entire planning horizon. The model instance generation is similar to the approach for modeling transport mode selection (see Kempkes [2009], pp. 66) and will be depicted in the following. Consider an exemplary network that consists of one supplier, a consolidation point and one plant with an incoming goods node and a warehouse. Imagine a decision between two delivery profiles. The network can then be modeled as sketched in Figure 3.2. For every supplier multiple warehouse nodes are included in the network, each representing the choice of a delivery profile. The arcs between the supplier node and the related delivery profile choice nodes have special resource uses associated to restrict multiple choices and to force a choice in the first period. For each of these delivery profile choice nodes a separate copy of the network is created whose endings are then connected to the same demand node. This leads to

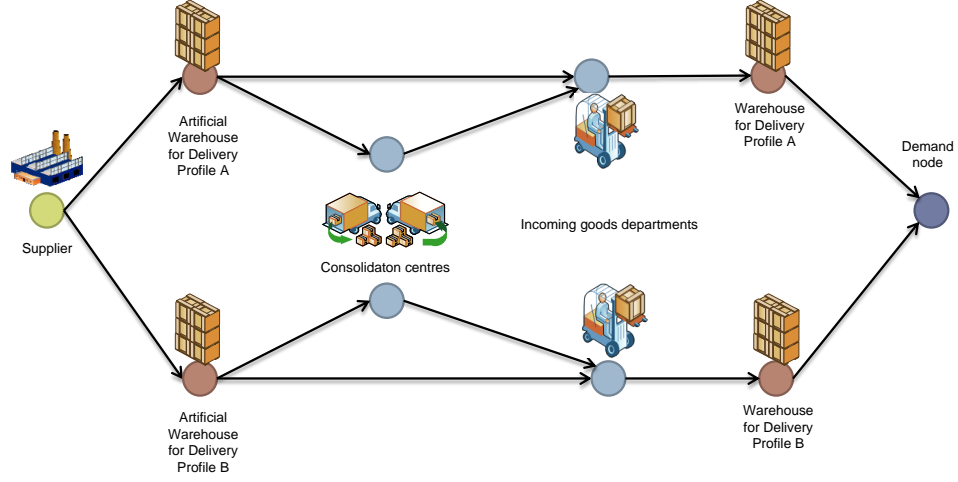


Figure 3.2: Example network model instance for delivery profile selection.

a set of disjunct networks for each delivery profile choice that has only two nodes in common, the supplier's starting node and the demand node. The disjunct treatment of the different networks is necessary to disallow the material flow to change from one delivery profile choice network part to another after the choice has initially been made by routing the material flow to one of these nodes. Each warehouse node within a delivery profile choices subnetwork has resource uses associated with it that disallow a transformation in time for the delivery periods of the corresponding delivery profile. All general resource uses (e.g. warehouse personnel, vehicle weight load) and resource groups (e.g. freight tariffs) have to be common to all subnetworks. This procedure leads to a multiplication of the complete network structure, which in turn results in rapidly increasing solution times. The examples given in Kempkes and Koberstein [2010] cover only five to six periods, and even though it is a very short timing window in comparison with the desired planning horizon of three months, not all instances could be solved to optimality (see Kempkes and Koberstein [2010], p. 293).

3.1.5 Considering uncertainty of demand

In the recent textbook Wallace and Ziemba [2005] it is stated that "Stochastic programming is decision making under risk" (Wallace and Ziemba [2005], p. 3). Even though it is not the only approach to making decisions under risk, reasonable effort

has been made by researchers to explore it and use its techniques to cope with risks in practical decision-making situations. This holds true also for the field of lot-sizing. Sox et al. [1999] gives a review of work on lot-sizing under stochastic demand. In analogy to the deterministic versions of these problems covered in section 3.1.1, two groups of problem are identified, the Stochastic Economic Lot Scheduling Problem (SELSP), and the Stochastic Capacitated Lot Sizing Problem (SCLSP) (see Sox et al. [1999], p. 182). Whereas the SELSP derived from the Economic Lot Scheduling Problem considers time to be continuous and covers an infinite planning horizon with stationary stochastic demand, the SCLSP on the other hand, which is an extension of the Capacitated Lot Sizing Model, uses a discrete time model with finite planning horizon and non-stationary stochastic demand. In both cases it is assumed that demand for different products is uncorrelated. Sox et al. [1999] points out that the SELSP is appropriate for "real time operational control [...] such as the production control of work-in-process inventory" (Sox et al. [1999], p. 182). By contrast, it is argued that the SCLSP problem class is best suited for "MRP-controlled systems in which demand is processed on a periodic basis" (Sox et al. [1999], p. 182). As this work deals with the latter case SELSP models will not be considered in the following. For a summary of work in this area see Sox et al. [1999]. The first approach to solving the uncapacitated, single-item stochastic lot-sizing problem was made by Silver [1978]. An rolling horizon environment is considered in which a product has to be ordered in accordance with frequently released forecasts under consideration of a fixed lead time. Forecast errors are assumed to follow a normal distribution with an average value of zero. A three-stage-heuristic procedure was developed which at first determines when to order, then selects a time period which must be covered by an order, and finally determines the order lot size. Bookbinder and Tan [1988] extended the single-item uncapacitated stochastic lot-sizing model by including a service-level constraint to deal with the probability of stock-outs. The constraint assures that demand can be satisfied in at least α (a parameter given by management) percent of all periods. Three strategies to deal with the uncertainty were defined, namely *Static Uncertainty*, *Dynamic Uncertainty* and *Static-Dynamic-Uncertainty*. The Static Uncertainty strategy refers to fixing all decisions at the beginning of the planning horizon. This results in both fixed replenishment periods and lot sizes. The Dynamic Uncertainty strategy describes the *Wait-And-See* approach known from stochastic programming in which the next period

is planned for after the current period's demand has been realized. The newly developed concept of the Static-Dynamic-Uncertainty determines replenishment periods at the beginning, but determines lot sizes only when the demand uncertainty has been revealed.

The concept of the Static-Dynamic-Uncertainty may be regarded as a rule-based planning approach in accordance with definition 2. It was later adapted by Tarim and Kingsman [2004], and a mixed integer model was formulated that allows determination of an optimal solution for the Static-Dynamic-Uncertainty strategy. The optimal solution was then compared with the heuristic given by Bookbinder and Tan [1988] in different scenarios. It is shown that Bookbinder and Tan [1988]'s heuristic is often close to the optimal solution, but the gap increases with more erratic demands. An extended overview of different model formulations for the single-item case can be found in Tempelmeier [2007]. Several techniques for considering backorders are discussed in the paper. Backorders may be treated either by accounting for backorder costs (see Sox [1997]) or by constraining the α or β service level. The α service level constraint can either be formulated for a single planning cycle (see e.g. Lasserre et al. [1985], Bookbinder and Tan [1988], Tarim and Kingsman [2004]) or for the complete planning horizon (see Tempelmeier [2007]). In Martel et al. [1995] a multi-item capacitated stochastic lot-sizing problem is set up. A solution approach based on a modified branch-and-bound strategy using a piecewise concave approximation is presented. The approach is then tested in a rolling horizon environment. The results show that the piecewise concave approximation provides fast results, making it possible to use it for realistic problem sizes, but still leaves a gap open between the optimal solution and its approximation. The gap depends strongly on the relation between inventory and order costs, as most of the gap originates in a wrong inventory cost approximation. For the case of multiple products, Brandimarte [2006] developed a multi-stage stochastic model. A plant location formulation is used to reduce the integrality gap between the LP formulation and the MIP equivalent. Based on this notation, a special relax-and-fix heuristic is provided. The idea is to fix the binary setup variables sequentially for each period. Thus in each period the setup decisions for previous periods are fixed and the setup decisions for following periods are considered to be relaxed variables. This step is repeated until all setups have been fixed. In addition to the presentation of the solution procedure, different scenario generation methods are discussed. Both the

solution procedure and the different scenario generation techniques are then compared in a rolling horizon simulation with artificial data.

3.2 Assessing the impact on realized logistics cost

To assess the impact of the deployment of cost-minimal delivery profiles on the realized logistics cost a fair comparison of the developed approach and other approaches has to take place. As discussed in Section 2.4 it is therefore necessary to employ an artificial benchmarking environment to conduct the assessment. The term benchmarking has been given different definitions that evolved over time. A widespread use of the term benchmarking "refers to positioning versus best practice where this practice exists in reality" (Valckenaers et al. [2006], p. 668). The idea behind this interpretation of benchmarking is that by comparing a company's key performance indicators with the best values achieved in practice, weaknesses in processes can be detected and the situation improved (see Hanman [1997] for detailed description of this type of benchmarking process). This kind of benchmarking will be described as *best practice benchmarking* in the remainder of this thesis.

A second interpretation often used in operations research literature addresses a comparison of different solution techniques applied to the same problem or model. When it comes to comparison of different solution techniques their outcome is often considered to be the objective value. To make a comparison different objective values of different solution techniques are then compared. When different solution techniques for the same model should be analyzed in respect of the solution quality they provide, this method is definitely applicable and fair. Even different model formulations can be compared if they consider the same aspects in the objective function. These benchmarks often include computational effort and help to find a trade-off between solution quality and runtime. In the remainder these benchmarks will be called *solution quality benchmarking*.

If a planning method's behavior in a rolling horizon planning environment under consideration of uncertainty is to be benchmarked, solution quality benchmarking is no longer sufficient. In a rolling horizon environment, where the MRP system or the planning method can react to changes, more sophisticated approaches have to be used to analyze how the methods adapt to the changes in the environment. When different planning techniques are to be compared, it is important to provide a benchmarking

method which provides a fair environment with clear rules that do not favor a single technique. To provide a fair comparison of different planning methods several requirements have to be fulfilled.

As making a plan and executing a part of it affects future planning situations, it is necessary to assure *repeatability* in the sense that each planning method can start with the same input parameters and form its own version of the future without affecting other planning methods' planning situations. Experiments have to be repeated "to study different systems in identical environments or the same system in different environments" (Fowler and Rose [2004], p. 470). This also leads to the request for the ability precisely to *control* the environment's behavior and all parameters. This is an essential requirement if behavior in special environmental circumstances is to be analyzed in detail. Detailed analysis only makes sense if the required details are represented in the environment. Thus an *integration of relevant aspects* is necessary. It is also important adequately to model reality and find the correct level of detail, without adding too much noise due to unnecessary aspects, while at the same time not leaving important aspects behind.

At the same time it is crucial to *avoid risks*, thus requiring an environment in which the planning methods can be tested in a 'sandbox' without affecting the real situation. Another aspect is *compression of time*, as it is desired to conclude a comparison within a reasonable time, even if long period ranges are considered. According to Fowler and Rose [2004], simulation provides all these strengths, which makes it a good choice as underlying technique for the benchmarking environment required for this thesis (see Fowler and Rose [2004], p. 469 f.). Thus the next section will depict different approaches that have been applied to benchmark planning methods within simulation environments. The literature on performance indicators for the benchmark will then be summarized.

3.2.1 Simulating a rolling horizon environment to benchmark planning methods

Banks [1998] defines simulation as "the imitation of the operation of a real-world process or system over time. Simulation involves the generation of an artificial history of the systems and the observation of that artificial history to draw inferences concerning the operating characters of the real system that is represented" (Banks [1998], p. 3).

We distinguish between static and dynamic simulation. Static simulation does not consider time whereas dynamic simulation does. A classical example of static simulation is Monte-Carlo-Simulation in which an experiment is repeated many times. Dynamic simulation can be further divided into continuous and discrete simulation. While continuous simulation segregates time into very small steps so that a 'natural flow' of time is generated, discrete simulation regards time as a sequence of discrete intervals in which actions may occur within the system. A special form of discrete simulation is discrete event-based simulation. In event-based environments not all discrete time-steps have to be simulated. To increase efficiency, "state variables change only at those discrete points in time at which events occur" (Banks [1998], p. 8). Hence the time-steps without events may be skipped and the overall simulation runtime will be reduced.

While it is clear that a static simulation cannot help to discover the behavior within a rolling horizon because it lacks a consideration of time, both continuous and discrete simulation could be applicable. Scholz-Reiter et al. [2005] analyzed advantages and disadvantages of both continuous and discrete simulations for modeling logistic systems with autonomous control. It was stated that "modelling by a discrete-event simulation tool allows a good description of real-world [...] processes" (Scholz-Reiter et al. [2005], p. 416) whereas continuous modeling "describes logistic processes on a higher aggregation level" (Scholz-Reiter et al. [2005], p. 416).

In analysis of challenges arising when developing simulation environments Fowler and Rose [2004] state that "while incorporation of detail may increase the credibility of the model, excessive levels of detail may render a model hard to build, debug, understand, deploy, and maintain" (Fowler and Rose [2004], p. 470). In the field of manufacturing systems research simulation is widely discussed in the literature as a proper solution for benchmarking efforts. To the author's knowledge such approaches in application to operational order lot-sizing do not exist. Even though there are several parallels between them, it will mainly be referred to manufacturing and replenishment simulation approaches in the following.

The first work focusing explicitly on lot-sizing decisions was presented in Dzielinski et al. [1963]. A single-stage make-to-stock production environment was considered. To account for setup cost the number of work forces required to maintain service operations given a specific lot-production sequence was computed. A demand forecast for

the next periods was made based on historical net demand data iteratively and used as input for both a linear programming model and an Economic Order Quantity formula. In the next iteration the first period was seen as executed, inventory was adjusted and the procedure was repeated. Several settings were then evaluated for both the linear programming model and the EOQ formula. Four test cases covering two time aggregation levels (one or two months considered to be a period) with and without safety stocks were used. It could be shown that the linear programming model was superior in each of the operating cost parts including labor, inventory and backlog costs leading to an improvement in total operating cost. Even though the simulation approach was not very sophisticated, this work can be seen as the first simulation of a rolling horizon planning situation.

Another study benchmarking lot-sizing rules was conducted by Callarman and Hamrin [1983]. Three approaches were evaluated, namely the EOQ formula, the Wagner-Within algorithm (see Wagner and Within [1958], WW) and the Part-Period-Balancing heuristic (see DeMatteis [1968], PPB). Instead of deriving forecasts from net demand history, forecasts were explicitly given such that the forecast error could be controlled as desired. Safety stocks were adjusted based on a self-developed method. Four factors were then varied to study the performance of the different algorithms under these conditions. The varied factors were the coefficient of variation in demand, the average time between orders, the desired service level and the forecast error. Even though differences in total cost results were not significant, a slight advantage of the PPB method could be shown. Only when dealing with a rather high forecast error or very short distances between orders was the EOQ method superior to the PPB method. It was also shown that the coefficient of variation in demand and the forecast error had the highest impact on total cost under all four considered factors.

A recent survey of different lot-sizing and scheduling approaches in a rolling horizon environment was made by Simpson [2001]. Nine approaches are included in the study, Wagner-Within, Part-Period-Balancing heuristic, economic order interval (EOI), Silver-Meal-Heuristic (see Silver [1979], SM), least total cost, least unit cost, Groff's algorithm, McLauren's order moment and the maximum part period gain algorithm (see Roll and Karni [2011], MPG). The simulation setting covered 300 periods worked off in a rolling horizon simulation. Three factors were varied to get a glance at the behavior of the different algorithms under different circumstances. These factors included

the length of the planning horizon, the length of the expected order cycle and the demand pattern. For benchmarking purposes an adapted version of the Wagner-Within algorithm that does not allow holding of items in inventory longer than the targeted planning horizon was applied to all 300 periods at once and the optimal solution value was used to establish a lower bound. 3060 simulation runs were made for each method, resulting in a total number of 27450 simulation runs for the study. The results showed a clear advantage of the Wagner-Within algorithm, with MPG, Groff, least total cost and Silver-Meal following shortly after. The longer the planning horizon chosen, the closer the algorithms came to the lower bound. In general it was found that a longer planning horizon improves solution quality up to a certain level. When a certain horizon length is reached the methods stabilize and provide no further improvement. The number of rescheduling messages is used as an indicator for nervousness of the planning approaches. It is pointed out that the very cost-sensitive algorithms at the same time create measurable higher nervousness than do the other approaches.

3.2.2 Architectural approaches to benchmark simulation environments

Simulation has for a long time been used as benchmark method to compare job-dispatching rules. One of the first publications mentioning simulation as benchmark method was Blackstone et al. [1982]. In their work Blackstone et al. [1982] review research on dispatching rules and gives a short section on how simulation may be used to benchmark dispatching rules and which issues are raised by this technique. One of the issues mentioned is the difficulty of avoiding production of *censored data*. *Censored data* refers to the effect that at the end of the simulation some jobs may remain uncompleted. These jobs are chosen differently by different dispatching rules, and thus the effect may favor one method over another if it is not taken care of.

In Valckenaers et al. [2006] a service for benchmarking manufacturing control approaches was proposed. They point out that it is difficult to compare approaches without having a benchmark simulation environment available. To resolve this issue and provide a sound benchmarking platform for different approaches for all researchers a web-based benchmarking service was developed. The service provides access to a previously developed simulation environment that can be configured by a graphical user interface. The simulation environment then builds a computational model of the

production system and emulates the behavior of a factory based on different scenarios from industrial test cases. Researchers can then implement an interface in their manufacturing control implementation and benchmark it via remote control with other approaches.

Mönch [2007] considers different production control approaches in a simulation based benchmark for manufacturing systems. In a requirement analysis modeling and architectural issues are addressed. A generalized control setting is provided that can be applied accordingly to other simulation environments. The underlying environment with its properties is defined as base system B . The base process within this environment given the input X_B and state set Z_B is defined as PB and is described by the mapping:

$$PB : X_B \times Z_B \longrightarrow Z_B \times Y_B$$

where Y_B is the output of the process. The base system is controlled by a control system C which uses a control process PC to control the base process. The control process PC , its state set Z_C and output Y_C can again be denoted as a transformation based on the input X_C :

$$PC : X_C \times Z_C \longrightarrow Z_C \times Y_C$$

Based on this formalised understanding, the simulation approach is divided into two sections as depicted in Figure 3.3. On the one hand there is the simulation model itself that represents the base system. It completely represents the base system B and the base process PB , but has no knowledge of the control system. In his work Mönch [2007] uses a discrete event simulation model based on a test-bed from his specific industrial domain (in this case semiconductor manufacturing) to cover this aspect. On the other hand there is the control system that interacts with the base system and has to be aware of the base systems properties. To allow for these manipulations, an interface between the base system and the control system has to be developed. This interface can either be a direct manipulation of data on the data level of the simulation tool, or it can be a more advanced data layer in between. The latter option was seen as strongly favorable as it does not produce proprietary source code and allows for different control approaches to be plugged in.

A similar approach was presented by Herrmann [2007]. The basic idea was to provide a possibility to simulate a plant and its control before production was actually started. This should allow a benchmark of different planning approaches as well as plant layouts

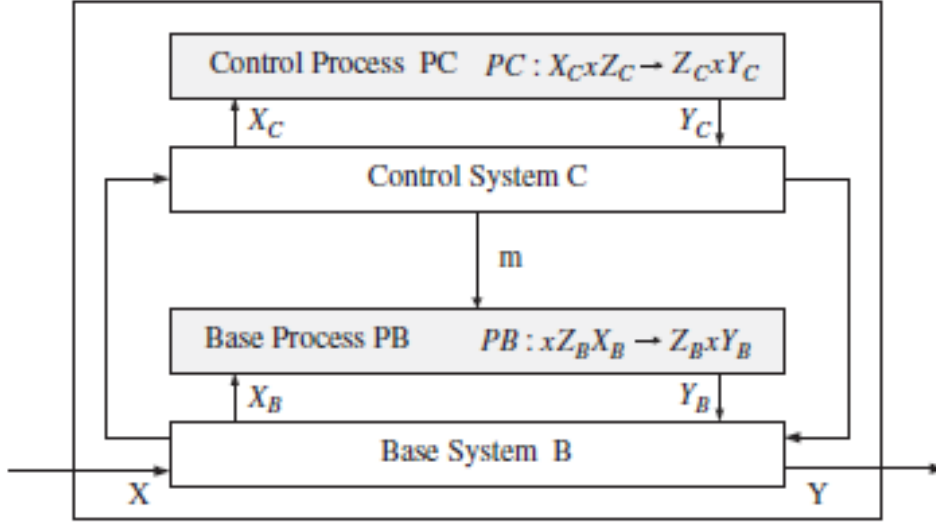


Figure 3.3: Basic architecture of simulation-based benchmarking environments (based on Mönch [2007], p. 1383).

in advance. Instead of coupling a self-developed production control system, the well-known MRP-System SAP was connected to the simulation software eM-Plant via a self-developed middleware. The production process could then be modeled in both SAP and eM-Plant. A stochastic distribution for randomly incoming demands was used to evaluate the outcome of the overall production process and its control settings. Performance indicators are gathered from the simulation model and presented to the user and allow for a comparison with the values targeted within the MRP system. Using the middleware as an interface between control and base process also allows to exchange the MRP system or develop customised planning approaches for benchmarking.

3.3 Assessing stability of the generated delivery schedules

To assess the stability of the generated delivery schedules, it is necessary to define a set of performance indicators that can be compared among the different approaches. A performance indicator should project the stability of the generated delivery schedules to a scalar value on a common scale with a defined order, such that a higher value represents a higher stability and a lower value represents a lower stability. These

t	Period t in the delivery schedule
k	Planning cycle $k = 1..K$
M_k	Start of planning cycle k
Q_t^k	Quantity scheduled for period t in planning cycle k
T	The length of the planning horizon

indicators have to be chosen wisely in order correctly to reflect the performance of the approach. To measure the degree of achievement of the different objectives described in section 2.3.1 different performance indicators have to be used. It is hardly possible to bring all objectives together within one indicator without loss of detail. It would be possible to use a single objective function in which each of the objectives would be included with a weight factor, but this leaves unanswered the question of how to chose the weight factors. In order especially to get a glance on the advantages and disadvantages of each approach it is necessary to measure each objective on its own. Whereas there are pretty clear guidelines on how to measure the realized cost, there is a variety of measures for the stability of the generated delivery schedules. In the following, measures proposed in the literature will be sketched. Most of them come from the area of MRP-system or inventory control research. To improve readability the measures will be given in a consistent notation given in table 3.3.

Blackburn et al. [1986] analyse different strategies in respect of their effect on delivery schedule stability in a rolling horizon multi-stage production environment. Strategies included are freezing the schedule within the planning horizon, safety stocks at the top stage, a lot-for-lot policy after the first stage, a forecast beyond the planning horizon and a planning procedure which includes change cost. Instability is measured as "the number of times an unplanned order was made in period 1" (see Blackburn et al. [1986], p. 418) or an existing order in the first period was "altered either by an increase, decrease or deletion" (see Blackburn et al. [1986], p. 418). This measure is completely focused on the first period in the planning horizon, which is a huge drawback given that that lower production stages use the whole delivery schedule for their production planning. In addition, the first planning period is stable in most environments due to replenishment lead time agreements between suppliers and buyer. This issue could be avoided by extending the measure to the first period after replenishment lead time. Aside from this point, a focus on the first period does not account for changes in the entire relevant short-term planning horizon of the supplier; thus this measure does not

seem applicable to the given problem.

Barrett and Laforge [1991] have made a study to evaluate the effect of the duration of one planning iteration or in other words the re-planning frequency on the stability of delivery schedules. The study evaluates the results in respect of achieved service levels, inventory values and schedule nervousness. The absolute amount of open-order changes is used as a measure of schedule nervousness. Open-order changes include changes in time or quantity as well as adding or removing orders. This measure has two drawbacks. First, it is an absolute value which means that it penalizes systems with many orders and favors systems with only few total orders. Another aspect is that no differentiation takes place between different types of open-order changes, so that a worse-case situation facing high demand underestimation in combination with a shift forward in due dates is valued equally with a small overestimation or shift backward at the end of the planning horizon.

Meixell [2005] presents a study on the effect of setup costs, component commonality, and capacity on delivery schedule stability in supply chains. The study uses the coefficient of variation across schedule quantities for multiple schedule releases and a single production period to measure schedule instability. The coefficient of variation as a measure of delivery schedule instability has the advantage that it is independent of the absolute values because it only covers the relation between standard deviation and mean value. The coefficient of variation can be computed as follows:

$$CV_t = \frac{\sigma_t}{|\mu_t|} \quad (3.3.1)$$

where σ_t is the standard deviation and μ_t the mean value of order quantity for period t among different planning cycles, which can be computed as follows:

$$\mu_t = \frac{1}{K} \sum_{k=1}^K Q_t^k \quad (3.3.2)$$

and

$$\sigma_t = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (Q_t^k - \mu_t)^2} \quad (3.3.3)$$

Apparently this measure does not cover order time shifting explicitly nor does it distinguish between underestimation of demand (quantity increases or order shifting forwards) on the one hand and overestimation of demand (quantity decreases or order shifting backwards) on the other hand. A differentiation between these two cases should be considered, given their different outcomes. An underestimation of demand may cause the supplier to be unable to deliver, whereas an overestimation may lead to an undesired overproduction on the supplier's side.

Pujawan [2004] has developed a model to measure the instability of delivery schedules. For each planning cycle k and each planning period t , an instability value $I(k, t)$ is computed. The model differentiates between different types of change (denoted i in the model) and combines them based on a weight factor w_i . Three types of changes are identified, namely a change in production start time, change in specifications and change in order quantity. An analytical hierarchical process (AHP) is then used to determine the weight factors for the different types of change. If two changes occur at once, only the type with the higher weight factor is considered. The model is denoted as follows (adopted from Pujawan [2004], p. 520):

$$I(k) = \sum_{t=k}^{t+T} I(k, t) \quad (3.3.4)$$

with

$$I(k, t) = \sum_i \sum_k w_i Q^k(i, j, t) \quad (3.3.5)$$

$I(k)$ gives back the total instability in planning cycle k . $Q^k(i, j, t)$ is the quantity of an order j which in planning cycle $k - 1$ was scheduled to be produced in period t and then underwent a change of type i in planning cycle k . Even though this model differentiates between time- and quantity changes, it does not cover the differentiation between underestimations and overestimations. Another issue is that the weighting factors have to be derived from a subjective point of view, so that results are biased by the preferences. This need not necessarily be a disadvantage because it can reflect a company's point of view, but may lead to a weight setting that favors a specific approach over another.

An approach which also focuses on the number of open-order changes was developed by

Ho and Ireland [1998]. To account for the tempestuousness of the impacts of an open-order change, a weighted rescheduling measure is introduced. The approach does not consider changes in order quantity, only rescheduling from one period to another. Nor is there any indication of how added or removed orders will be treated. The measure is given as

$$WR = \sum_{p \in P} \sum_{t=1}^T Q_{tp} \cdot |NDD_{tp} - ODD_{tp}| \quad (3.3.6)$$

with NDD_{tp} as the new due date of product p in period t and ODD_{tp} the old due date respectively. This measure has multiple drawbacks. Aside from not covering changes in quantity or changes in the number of orders, it does not distinguish between orders that were shifted forward or backward. It also treats orders closer to the planning period in the same way as orders further away from the current planning period. In addition, parts with low quantities are favored over parts with high quantities, because their impact is lower as the shift in due dates is weighted with the order quantity. This may lead to a situation where a single high-volume part (e.g. a screw) can increase the measure with few incidents far more than several low volume parts with multiple incidents.

Jensen [1993] uses two stability measures in his study on planning stability of reorder point lot-sizing policies. On the one hand, a setup orientated stability measure is introduced. Setup orientated in this case means that pure changes in quantity are not considered. Only if the quantity changes from zero or to zero, is instability measured. The measure is expressed as

$$\frac{1}{K} \cdot \left(\sum_{\forall k > 1} \sum_{t=M_k}^{M_{k-1}+T-1} |\delta(Q_t^k) - \delta(Q_t^{k-1})| \right) \quad (3.3.7)$$

with

$$\delta(Q_t^k) = \begin{cases} 0 & \text{if } Q_t^k = 0 \\ 1 & \text{if } Q_t^k > 0 \end{cases} \quad (3.3.8)$$

This measure allows evaluation of the number of added or removed orders, but does not differentiate between them. Nor does it consider that additional orders arriving

shortly after the end of the frozen zone will be more difficult to handle than those added in periods close to the end of the planning horizon. On the other hand, a quantity-orientated measure is introduced. It is given as

$$\frac{1}{K} \cdot \left(\sum_{\forall k > 1} \sum_{t=M_k}^{M_{k-1}+T-1} |Q_t^k - Q_t^{k-1}| \right) \quad (3.3.9)$$

Again this measure does not include a differentiation between overestimations and underestimations. Nor does it account for the distance from the planning period to the period with a quantity change and treats changes which are closer to the planning period in the same way it treats orders that are further away.

Zhao et al. [1995] provide a study on lot-sizing rules and master production schedule freezing and their outcome on total cost, service levels and schedule stability. The quantity-related measure from Jensen [1993] is extended for the multi-product case. It then reads as

$$\frac{1}{K} \cdot \left(\sum_{p \in P} \sum_{\forall k > 1} \sum_{t=M_k}^{M_{k-1}+T-1} |Q_t^k - Q_t^{k-1}| \right) \quad (3.3.10)$$

with $p \in P$ being product p out of the set of all products P .

Sridharan and Laforge [1989] try to account for the importance of changes by extending the quantity orientated measure with weight parameters. It was suggested shortly before the study by one of the authors in Sridharan et al. [1988]. The measure is introduced as

$$I = \frac{1}{K} \left(\sum_{\forall k > 1} \sum_{t=M_k}^{M_{k-1}+T-1} |Q_t^k - Q_t^{k-1}| \cdot (1 - \alpha) \cdot \alpha^{t-M_k} \right) \quad (3.3.11)$$

with α being a weight parameter ($0 < \alpha < 1$). As I gives back an absolute value, it can be divided by the average order quantity to gain a relative value. The weight factor α may be adjusted to lay emphasis on more distant changes or to ignore their effect. The smaller α is chosen, the lower is the impact of changes in more distant periods. By including the weight factor the only issue with this measurement approach is the inability to distinguish between underestimations and overestimations. In addition, the study uses eight other measures to give detailed feedback on the kind of schedule

nervousness. These include the order quantity and frequency count of added and canceled orders as well as orders that were increased or decreased in quantity respectively. These eight measures account for all possible changes to a delivery schedule and thus give detailed insight into the structure of change, but do not consider their position in respect of the planning period. In Kadipasaoglu and Sridharan [1997] the measure I was further extended for multi-stage production environments; another weight factor is there introduced for different stages. As this extension is not necessary for the given problem setting, interested readers are redirected to the original source for detailed explanation.

Inman and Gonsalvez [1997] developed a measure based on the percentage deviation between the initial forecast ($k = 1$) and the minimum and maximum scheduled quantity over all planning cycles. It can be formalised as

$$Dev = \max \left\{ \max \left\{ \tilde{Q}^k \quad \forall k = 1..K \right\} - \tilde{Q}^1; \tilde{Q}^1 - \min \left\{ \tilde{Q}^k \quad \forall k = 1..K \right\} \right\} \cdot \frac{1}{\tilde{Q}^1} \quad (3.3.12)$$

with

$$\tilde{Q}^k = \sum_{t=M_k}^{M_{k+1}} Q_t^k \quad (3.3.13)$$

\tilde{Q}^k sums the quantities scheduled in between planning cycle k and cycle $k+1$. Aggregating the periods between two planning cycles compromises precision of the measure, but allows for better comparison of quantity fluctuation in medium terms. The Dev measure is then used to distinguish between stable and unstable parts. Parts are said to be stable if the Dev value falls below a given threshold. Parts with low volumes are excluded from the analysis. Based on this segregation, a stability measure is introduced as the percentage of stable parts in relation to both stable and unstable parts.

$$Stability = \frac{\text{Stable parts}}{\text{Stable parts} + \text{Unstable parts}} \quad (3.3.14)$$

This ratio is useful when dealing with multiple parts. Unfortunately, the stability ratio does not represent the distribution of instability among parts. If the parts within the unstable group are extraordinarily unstable, the stability ratio will give the same

result as if the unstable parts were just marginal above the threshold. To improve its expressiveness the stability ratio may be extended by a division into certain groups based on the parts deviation.

4 Outline of the Required Work

In this chapter the gap between approaches from the literature summarized in Chapter 3 and requirements according to the problem setting given in Chapter 2 is outlined. It will discuss which aspects of the problem setting are covered by existing approaches and which are not. The remainder of this chapter is twofold. First, existing planning approaches and their shortcomings when applied to the problem setting of this thesis are discussed. Second, the steps necessary to assess the impact of the selection of cost-minimal delivery profiles on both realized cost and the stability of generated delivery schedules will be described.

4.1 Selecting cost-minimal delivery profiles for area forwarding inbound logistic networks

The literature on planning approaches for the operational order lot-sizing problem is twofold. On the one hand, a reasonable amount of research has been conducted in the area of rule-based planning approaches, mainly focusing on replenishment frequencies or fixed lot sizes. Even though there are models considering quantity discounting schemes (e.g. Chakravarty [1984], Benton [1991]), these models are not sufficient to cope with the given logistic network and its tariff structures. On the other hand, quite sophisticated approaches exist to model logistics networks and their components. Both the network structure as well as discounting schemes have been discussed in several studies. Kempkes [2009] and Kempkes and Koberstein [2010] present a model that provides enough aspects to model the underlying area forwarding networks and their most relevant properties. But when it comes down to modeling delivery profile selection, which the model was originally not intended for, quite sophisticated model instance formulations have to be used (see Section 3.1.4 for details).

4.1.1 Performance issues

With the given modeling approaches it is hardly possible to solve problem instances from practice within reasonable time and effort. The desired planning horizon is about three months, whereas the examples given in Kempkes and Koberstein [2010] cover only five to six periods, which equals a planning horizon of one week. Even with this limitation to a very short planning horizon, not all instances could be solved to optimality (see Kempkes and Koberstein [2010], p. 293). This leads to the request for a more efficient model formulation and solution algorithm. It has also to be mentioned that if possible, no reduction of the problem itself should take place: thus all cost-relevant aspects should be covered.

4.1.2 Considering uncertainty

Uncertainty in demand forecasts plays an important role in operational order lot-sizing and lot-sizing in general. Several researchers have dealt with this aspect by extending existing models to consider uncertainty. Especially when a rule-based delivery schedule generation approach is to be considered, where a decision on a rule has to be made in advance when information is still unreliable, the necessity for a robust choice seems to be obvious. In the field of algorithmic delivery schedule generation, methods that consider uncertainty have been developed too. Most research in this area focuses on stochastic programming approaches, for which good results have been shown in the past. Stochastic programming models "provide deeper insights because they optimize decisions over multiple scenarios linked together in a single model, each with an associated probability of occurrence" (see Shapiro [2007], p. 443). It should therefore be analyzed if and to what degree stochastic programming can be helpful to deriving delivery profiles with a higher robustness towards the realized cost.

4.2 Assessing the impact of cost-minimal delivery profiles in a rolling horizon environment

Many researchers have pointed out the difference between a priori solutions and objective values, and the actual outcome in a rolling horizon planning environment. Several studies have been carried out in this field, mainly considering production planning problems and production environments. The main aspects that have been considered

in these studies are the impact on realized cost and the stability of the generated delivery schedules and service levels. An environment with tariff discounting schemes and network structures as can be found in the given problem setting has not been considered so far in a comparative study. Even though models for algorithmic delivery schedule generation are very sophisticated, no work has actually been done to test them in realistic circumstances. Nothing is known about their sensitivity to ever-changing forecasts in rolling horizon environments, both in terms of the stability of the generated schedule and realized cost. This gap will be closed by a comparative study of both cost-minimal delivery profiles and promising algorithmic approaches to delivery schedule generation. The impact of cost-minimal delivery profiles can thereby be assessed.

4.2.1 A simulation framework for operational order lot-sizing planning methods

In order to provide a sound analysis of the impact of cost-minimal delivery profiles when employed in a rolling horizon environment in realistic circumstances, a simulation framework has to be set up. Existing simulation frameworks for rolling horizon environments focus on production planning and control rather than order lot-sizing. This leads to the requirement for a benchmarking framework for operational order lot-sizing. Simulation has proven to be the path of choice when coping with a rolling horizon environment. There are architectural approaches (see Mönnch [2007]) that go beyond the specific application of production planning and control. This approach will be transferred to the actual problem setting. A simulation framework for the operational order lot-sizing problem in a rolling horizon has accordingly to be developed. This framework should provide the possibility of exchanging planning algorithms in order to pave the way for a conclusive comparison between rule-based delivery schedule generation approaches like the deployment of cost-minimal delivery profiles on the one hand and algorithmic delivery schedule generation approaches on the other hand. As the practical relevance of the problem setting is obvious, the simulation framework should be capable of operating on data from practice. In addition, relevant aspects of the problem setting, including the network structure, tariff systems, inventory control and forecasts should be covered in appropriate depth.

4.2.2 Measuring delivery schedule stability

Different techniques have been developed in the literature to assess delivery schedule stability, ranging from simple setup orientated measures focusing on single periods to complex algorithms to derive a performance indicator. The diversity of the generated approaches shows that multiple aspects have to be considered to assess the impact of change, including a differentiation between changes in time and in quantity, the distance between the planning period and the source of instability and a differentiation between underestimations and overestimations. None of the existing measures covers all of these aspects simultaneously. A comprehensive set of measures for delivery schedule stability will be developed to overcome these shortcomings.

4.3 Targeted contributions

In summary, it may be said that there will be five major contributions in this thesis, covering aspects from multiple disciplines and fields of research.

1. A deterministic model formulation and efficient solution algorithms for the selection of cost-minimal delivery profiles will be presented.
2. To cope with demand uncertainty a stochastic programming formulation for the selection of cost-minimal delivery profiles will be developed. Different approaches to scenario generation will be evaluated. In addition, a modified solution algorithm for the stochastic model formulation will be presented.
3. A simulation framework for planning methods for the operational order lot-sizing problem in area forwarding based inbound logistic networks with complex tariff structures will be developed.
4. A new measure for delivery schedule stability will be introduced that accounts both for time shifts and quantity changes and additionally accounts for the impact of a change.
5. An assessment based on a case study of the planning techniques under realistic conditions based on data from practice will be conducted, and the impact of the deployment of cost-minimal delivery profiles on both realized cost and the stability of the generated delivery schedules will be analyzed.

The remainder of this thesis will cover the developed solution approaches and is structured as follows. First, a model formulation and solution algorithm for the selection of delivery profiles will be depicted in chapter 5. Model extensions for the consideration of uncertainty and the necessary adoptions of the solution algorithm will then be described. In addition, different approaches to scenario generation will be presented. After the solution procedures have been described, chapter 6 introduces the developed simulation framework, describes its architecture and the performance measures used in the case study. The case study itself will be depicted in chapter 7 and consists of three parts, focusing on runtime, monetary aspects and stability of the generated delivery schedules respectively.

5 Selecting cost-minimal and robust delivery profiles

This chapter starts with a summary of the decision problem that has to be solved in order to determine cost-minimal delivery profiles. It will then be shown how the problem structure can be exploited to reduce computational efforts by using a decomposition approach. Model formulations based on these insights will then be presented. Thereafter a primal heuristic and a meta-heuristic will be described that can provide fast solutions which can then be used as starting values for MIP-Solvers. In the next step the required model extensions for a consideration of demand uncertainty will be discussed. In addition, a model that demands less computational effort but comes at the price of less generous applicability will be presented. Thereafter, a revised solution algorithm for the stochastic case will be depicted. As stochastic programming relies heavily on the scenarios used as input, multiple approaches to generate scenarios will be described.

5.1 Summary of the given decision problem

If a company has decided to use delivery profiles as a control rule for their operational order lot sizing, a tactical decision process has to be set up. In a periodic review process the delivery profiles to be used for the next three months have to be selected. For each supplier one delivery profile out of a set of predefined delivery profiles has to be assigned. The objective is to assign the delivery profiles in such way that the expected total cost of the inbound logistics operations will be minimized. More formally it can be stated that for each supplier from a set of suppliers $s \in S$ a delivery profile from a set of delivery profiles $dp \in DP$ has to be assigned. Let pc be the vector of profile choices $pc_{s_1, dp} \dots pc_{s_{|S|}, dp}$ that holds the delivery profile for each supplier, and let $c(pc)$ be a cost function that estimates the expected cost for a delivery profile choice vector pc . The task can then be formalized to find a delivery profile choice vector pc^* with a minimal value for $c(pc)$, thus $\forall pc \in S \times DP \setminus \{pc^*\} : c(pc) \geq pc^*$, with $S \times DP$ being the set of all possible delivery profile assignment vectors, holds true. The cost function $c(pc)$ can be further divided according to the different cost factors. The main

cost components are freight cost $c^{Freight}(pc)$ and inventory holding cost $c^{Inventory}(pc)$. Thus the cost function $c(pc)$ may be formalized as

$$c(pc) = c^{Freight}(pc) + c^{Inventory}(pc) \quad (5.1.1)$$

These cost components can be further split, as the freight cost $c^{Freight}(pc)$ can be described as the sum of pre leg run cost $c^{Preleg}(pc)$, full load run cost $c^{Fullload}(pc)$ and main leg run cost $c^{Mainleg}(pc)$, whereas the inventory holding cost $c^{Inventory}(pc)$ consist of the cost for warehouse slot usage $c^{Slot}(pc)$ and cost of interest on capital commitment $c^{Interest}(pc)$. Therefore the cost function reads as

$$c(pc) = c^{Preleg}(pc) + c^{Fullload}(pc) + c^{Mainleg}(pc) + c^{Slot}(pc) + c^{Interest}(pc) \quad (5.1.2)$$

Whereas the inventory cost part may be estimated based on the parts values and the expected holding time, the freight cost part of the cost function depends strongly on the tariff system that has been negotiated with the LSPs. As these systems provide synergy effects for consolidated main leg runs of the different suppliers within one area, it is necessary to consider all suppliers within one area at once to find the optimal delivery profile assignment vector pc^* . However, the delivery profile assignments can be determined independently for each consolidation area, as no interlink between the different areas exists.

5.2 Exploiting the problem structure

A delivery profile restricts the supplier's delivery to certain delivery periods. As described in detail in section 2.3.3, the MRP system will gather all net demands with due dates equal or greater than the delivery period and smaller than the next delivery period will be cumulated to an aggregated order on the first delivery period. Considering an environment with deterministic demand this in turn leads to Observation 1.

Observation 1. *If demand is known and a delivery profile has been selected for a supplier, it is predefined which parts will be delivered in which delivery period.*

If a delivery profile will be applied to a given net dependent demand forecast, the demands scheduled between two delivery periods will be ordered jointly on the first of

the two. Therefore parts have to be held on stock from the delivery period until the consumption period. Following this argumentation in combination with Observation 1, we can set up Observation 2

Observation 2. *If demand is known and a delivery profile has been selected for a supplier, it is predefined which parts will be in stock in which period.*

Considering the specific structure of the decision problems for the selection of delivery profiles in area forwarding based logistic networks, as explained in section 5.1, Observation 1 allows us to compute the part of cost function $c(pc)$ of a delivery profiles choice that does not provide synergy effects between different suppliers. As depicted in figure 5.1 pre leg runs and full load runs do not carry a combination of parts from different suppliers. Unlike the main leg run cost, the cost for pre leg runs and full load runs depends only on the choice of a delivery profile of a single supplier. Therefore, the cost of pre leg transport $c^{Preleg}(pc_{s,dp})$ resulting from the assignment $pc_{s,dp}$ of delivery profile dp to supplier s and the respective counterpart $c^{Fullload}(pc_{s,dp})$ may be computed separately for each supplier. Another cost part is the inventory holding cost, which depends on the quantity of parts stored in inventory in each period and specific cost parameters, e.g. parts price or interest rate on bound capital. According to Observation 2 the quantity of parts stored in inventory per period is fixed if a delivery profile is assigned to the supplier that delivers these parts. This allows us to compute inventory holding cost $c^{Inventory}(pc_{s,dp})$ incurred by a delivery profile assignment $pc_{s,dp}$ of delivery profile dp to supplier s . Drawing together these different cost aspects it may be stated that except for the main leg run cost, all cost factors relevant to the problem can be derived for a delivery profile assignment without consideration of other suppliers possible assignments. This property can be exploited to decompose the problem into multiple subproblems and is the foundation of the solution procedure depicted in Figure 5.2. At first, a preprocessing routine, which will be described in detail in Section 5.3, evaluates all possible delivery profiles assignments for each supplier. This includes computation of the cost factors for pre leg runs, full load runs and inventory holding as well as a derivation of parts that will be remaining for main leg runs. The results of this step consist of evaluated delivery profile assignments. In a second step a combination of previously evaluated assignments is selected such that total cost including both main leg run cost and previously evaluated cost factors of selected assignments is minimized. The heuristic procedures proposed in Section 5.5

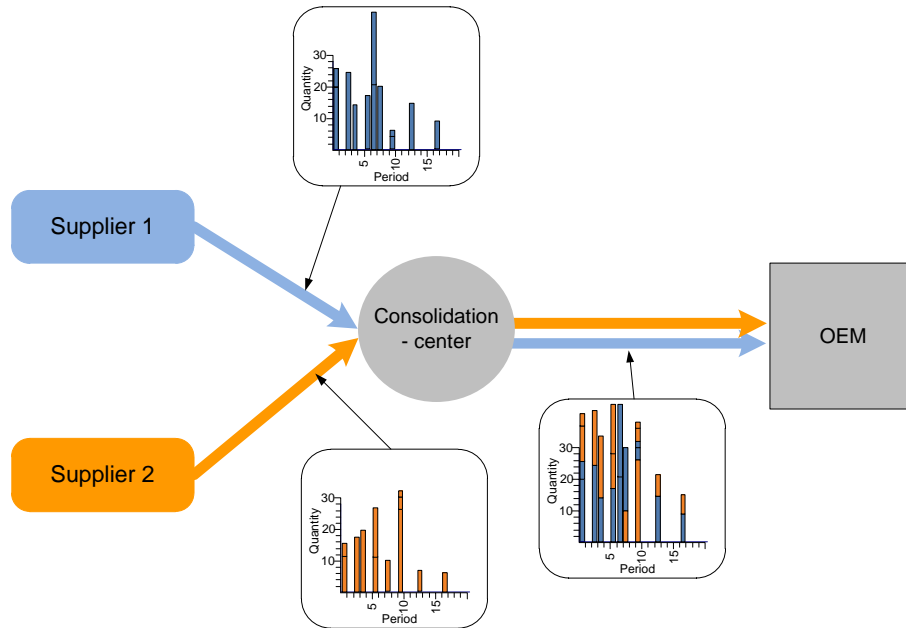


Figure 5.1: Segregation of goods in an area forwarding inbound logistics network allowing decomposition of pre leg runs and full load runs from different suppliers.

and 5.5.2 can be applied to retrieve a primal solution, which can then be used as input for the the model formulation presented in Section 5.4, which can then be solved by standard Mixed-Integer-Solvers.

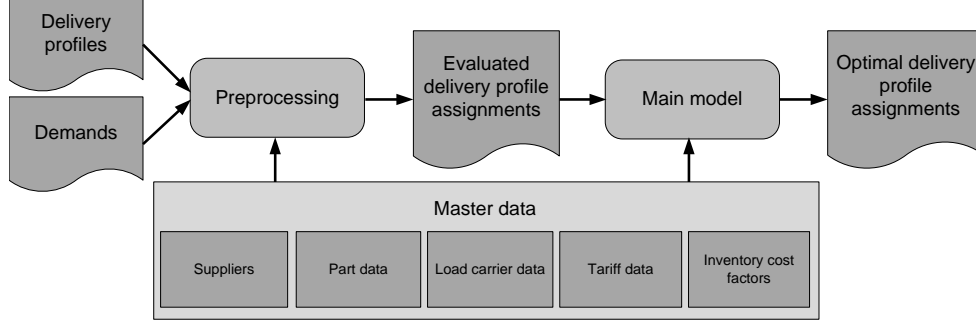


Figure 5.2: Overview of the proposed solution algorithm.

5.3 Preprocessing

In the preprocessing all possible delivery profile assignments $(s, dp) \in S \times DP$ have to be evaluated. Evaluated in this case means that all effects that are directly related to the assignment of a delivery profile to a supplier have to be computed. An overview of the algorithm is given in Algorithm 1. For each combination $(s, dp) \in S \times DP$ of suppliers and delivery profiles, the steps depicted in the following will be executed in order to evaluate the assignment of delivery profile $dp \in DP$ to supplier $s \in S$.

Algorithm 1: The preprocessing algorithm.

```

foreach  $s \in Suppliers$  do
    foreach  $dp \in Delivery\ Profiles$  do
        Determine orders resulting from assignment of  $dp$  to  $s$ ;
        /* See Section 5.3.1 */
        Compute inventory related cost factors;
        /* See Section 5.3.2 */
        foreach  $t$  in  $Periods$  do
            Compute freights;
            /* See Section 5.3.3 */
        end
    end
end

```

5.3.1 Determination of resulting orders

At first the given net dependent demands have to be mapped to the delivery schedule that would result from the delivery profile assignment. Therefore the set of demand entries D is transformed into the set of resulting orders O according to the delivery profile.

$$D \longrightarrow O$$

As described in Section 2.3.3 the application of a delivery profile results in a schedule where orders are placed only on delivery days, and where the order quantity in a delivery period is the sum of net demands with due dates equal or greater than the delivery period and smaller than the succeeding delivery period. Given the set of considered periods T and the set of delivery periods $\hat{T} \subseteq T$, then let $d_{p,t}$ be the demand of part p in period t , and let $\hat{t}_k \in \hat{T}$ be the delivery period of the k -th cycle of the current delivery profile. The ordered amount $O_{p,t}$ of part p in delivery period t can then be denoted as

$$O_{p,t} = \begin{cases} 0 & \text{if } t \notin \hat{T} \\ \sum_{t'=t}^{\hat{t}_{k+1}-1} d_{p,t'} & \text{otherwise} \end{cases} \quad (5.3.1)$$

This computation can be repeated for every part and every period to gain the set of resulting orders.

Handling of fixed material flows sharing the same routes

In certain applications it may be that orders for some parts will not be altered according to the delivery profiles but are included in the area forwarding network flow of goods and must therefore be considered when determining delivery profiles. This may be for several reasons. It may, for example, be that Just-In-Time or Just-In-Sequence delivered parts use the area forwarding network due to their relatively low volumes, or that vendor-managed inventory contracts consider the OEM to pay the freight cost for the delivery to a consignment warehouse and therefore allow use of the area forwarding network as well. Given the case that orders should be considered but not altered by the delivery profiles, the resulting orders for these parts are determined as

$$O_{p,t} = d_{p,t} \quad (5.3.2)$$

In the case where only certain orders of a part are considered to be fixed and should not be altered, these orders will be left untouched, whereas the remaining orders will be treated as described above. The remainder of the procedure does not have to distinguish between parts with fixed orders and parts that may be altered, as after the determination of resulting orders only the resulting orders themselves, but not the delivery profiles are used as underlying information.

5.3.2 Computation of inventory related cost factors

As described in Section 2.1.2 inventory-related costs include interest on bound capital and warehousing cost. Independently of the detailed cost function implemented in practice, computation of inventory cost follows the same procedure in most cases. For each period parts in the inventory are accounted for with different cost factors. These factors may either be multiplied with the parts inventory value or the number of load carriers in the warehouse. The number of parts of type p hold in inventory in period t can be computed as

$$h_{p,t} = \sum_{t'=1}^t O_{p,t'} - d_{p,t'} \quad (5.3.3)$$

The main price-related factor is the interest rate on bound capital. It can be computed for period t and part p as

$$C_{p,t}^{Interest} = h_{p,t} \cdot p_p \cdot \frac{\text{Interest rate}}{\text{Periods per year}} \quad (5.3.4)$$

where p_p is the price of part p . The load carrier-related cost can, like warehousing cost, be computed similarly. A cost factor C_{lc}^{Slot} that covers all cost related to one unit of load carrier lc can be used to set up the following equation:

$$C_{p,t}^{Slot} = \left\lceil \frac{h_{p,t}}{Q_{p,lc}} \right\rceil \cdot C_{lc}^{Slot} \quad (5.3.5)$$

where $Q_{p,lc}$ is the quantity of parts of type p that fits into one load carrier of type lc . By setting up a sum over all parts and periods a total value of inventory cost $C_{s,dp}^{Inventory}$

can be computed for later use.

$$C_{s,dp}^{Inventory} = \sum_{t \in T} \sum_{p \in P} C_{p,t}^{Interest} + C_{p,t}^{Slot} \quad (5.3.6)$$

If limited resources, e.g. boundaries on invested working capital or the number of available storage slots, constrain inventory levels, resource use can be computed in analogy to cost factors. A vector of resource usages for each period can then be obtained by

$$U_{s,dp,t,r}^{Inventory} = h_{p,t} \cdot UI_{p,r}^{Part} + \left\lceil \frac{h_{p,t}}{Q_{p,lc}} \right\rceil \cdot UI_{lc,r}^{Carrier} \quad \forall r \in R, p \in P \quad (5.3.7)$$

where $UI_{p,r}^{Part}$ describes the use of resource r by one unit of part p stored in inventory and $UI_{lc,r}^{Carrier}$ represents the use of resource r by one unit of load carrier lc placed in the warehouse.

5.3.3 Freight computation

Whereas the previous steps were quite easy to compute, it is harder to compute pre leg and full load costs and to derive the remaining orders for the main leg run. These points interact and cannot be computed separately. Two main cases have to be distinguished. In the first case, all parts that have to be ordered in one period fit into a single vehicle. If they do, it can be decided whether the vehicle will be filled, and then a choice made between accounting a single full load run price or a pre leg run price. Otherwise a decision has to be made on which parts to load onto a full load vehicle and which parts to leave in the pre leg vehicle. To obtain an approximation for this decision a MIP-Model can be set up that decides this issue for each period. In the following it will be assumed that logistics service providers try to achieve the best use of the vehicles. Thus the total cost inflicted by pre leg and full load transports is minimized. If another assumption is to be followed, other objective functions could be used. For a worst case scenario wherein the logistics service providers would try to bring the highest possible value to account the objective function may just be inverted from minimize to maximize. Another possibility would be to imply a random distribution of part onto the different vehicles. In this case, instead of solving the optimization model, a randomized solution could be created.

Resource-based modeling approach The optimization model follows a resource-based modeling approach inspired by Kempkes and Koberstein [2010]. The measurement units used to compute the fill level of the vehicle (e.g. weight and volume) are modeled as abstract resources. Given the goods loaded onto the vehicles and the vehicles' rebate levels, resource uses are computed. The use of resources is then linked to cost factors and the tariff system. Two types of cost factor can be assigned to a resource: linear cost functions (e.g. for fuel, which can be bought in any unit) and piecewise-constant cost functions (e.g. for incoming goods personnel, where each additional employee in a shift has to be payed a fixed wage for the shift). This modeling technique allows the model notation to be used for different applications based on the planners preferences. The measurement units can be exchanged without altering the model itself, so volume could be exchanged with load meters etc. Furthermore, additional resources (e.g. carbon-dioxide emissions, fuel consumption, incoming goods personnel, ...) can be integrated easily by adding a new resource. In setting up the model instance, two types of resources have to be distinguished, vehicle-specific resources and shared resources. Vehicle-specific resources that are directly linked to the vehicles' tariff systems (e.g. weight and volume) have to be modeled separately for each vehicle and are thus gathered in subsets R^v for each vehicle v , but must not be used by other vehicles. Shared resources like carbon-dioxide emissions or incoming goods personnel should be modeled as common resources shared by all vehicles.

Using the denotation from Table 5.3.3 the model can be denoted as follows:

Model formulation

$$\begin{aligned} \text{Min} \quad & \sum_{v \in V} C_v^{Vehicle} * v_v^{Active} + \sum_{v \in V, rl \in RL} C_{v,rl}^{Level} * v_{rl,v}^{Level} \\ & + \sum_{r \in R^L} C_r^{Unit} * u_r + \sum_{r \in R^{NL}} C_r^{Unit} * \left\lceil \frac{u_r}{S_r^{Unit}} \right\rceil \end{aligned} \quad (5.3.8)$$

subject to

$$O_p = \sum_{v \in V} o_{p,v} \quad \forall p \in P \quad (5.3.9)$$

$$a_{lc,v} = \sum_{p \in P_{lc}} \left\lceil \frac{o_{p,v}}{Q_{p,lc}} \right\rceil \quad \forall lc \in LC, v \in V \quad (5.3.10)$$

$$u_r = \sum_{p \in P} \sum_{v \in V} U_{p,r}^{part} * o_{p,v} + \sum_{\sum_{v \in V} lc \in LCU_{lc,r,v}^{Carrier} * a_{lc,v}} \quad \forall \bigcup_{v \in V} R^v \quad (5.3.11)$$

$$u_r = \sum_{v \in V} U_{v,r}^{Vehicle} * v_v^{Active} + \sum_{v \in V} \sum_{rl \in RL_v} U_{v,rl,r}^{Level} * v_{rl,v}^{Level} + \sum_{v \in V} \sum_{p \in P} U_{p,r}^{part} * o_{p,v} + \sum_{v \in V} \sum_{lc \in LC} U_{lc,r,v}^{Carrier} * a_{lc,v} \quad \forall r \in R \setminus \bigcup_{v \in V} R^v \quad (5.3.12)$$

$$\sum_{rl \in RL^v} v_{rl,v}^{Level} = v_v^{Active} \quad \forall v \in V \quad (5.3.13)$$

$$UB_{rl,v,r} * v_{rl,v}^{Level} - \epsilon \geq u_r - BigM * (1 - v_{rl,v}^{Level}) \quad \forall v \in V, r \in R_v^v, rl \in RL^v \quad (5.3.14)$$

$$LB_{rl,v,r} * v_{rl,v}^{Level} \leq u_r \quad \forall v \in V, r \in R^v, rl \in RL^v \quad (5.3.15)$$

$$BigM * v_v^{Active} \geq \sum_{p \in P} o_{p,v} \quad \forall v \in V \quad (5.3.16)$$

$$v_v^{Active} \leq \sum_{p \in P} o_{p,v} \quad \forall v \in V \quad (5.3.17)$$

The purpose of the model is threefold. First, freight cost can be determined for both pre leg and full load runs. Second, it can be determined which parts will be transported via the pre leg run and will thus have to be transported in the main leg run too. Third, resource usage of vehicle independent resources can be derived. To achieve these goals different components of the model have to be evaluated. The objective function 5.3.8 is to minimize the sum of four cost terms. The first term corresponds to fixed cost associated with a vehicle's use. The second term reflects the cost of the discounting

Indices and sets

$p \in P$	Set of part types
$lc \in LC$	Set of load carriers
$p \in P_{lc} \subseteq P$	Set of parts which will be delivered in load carrier lc
$v \in V$	Set of vehicles
$rl \in RL^v$	Set of rebate levels when using a vehicle
$r \in R$	Set of resources
$r \in R^L \subseteq R$	Set of resources with a linear cost function
$r \in R^{NL} \subseteq R$	Set of resources with a piecewise-constant cost function
$R^L \cup R^{NL} = R$	The set of resources consists of resources with linear cost function and resources with piecewise-constant cost function
$R^L \cap R^{NL} = \emptyset$	Each resource has either a linear cost function or a piecewise-constant cost function.
$R^v \subseteq R$	Set of resources which are used only by vehicle v and are related to this vehicle's tariff system

Parameters

O_p	Number of part units of part type p which have to be ordered
$Q_{p,lc}$	Maximum quantity of parts of part type p which fit in load carrier lc
$U_{p,r,v}^{Part}$	Usage of resource r by part of type p when using vehicle v
$U_{lc,r,v}^{Carrier}$	Usage of resource r by load carrier of type lc when using vehicle v
$U_{v,r}^{Vehicle}$	Usage of resource r resulting from usage of vehicle v
$U_{v,rl,r}^{Level}$	Usage of resource r resulting from a load equal to rebate level rl when using vehicle v
$C_{v,rl}^{Level}$	Costs of rebate level rl when using vehicle v
$C_v^{Vehicle}$	Base costs of vehicle v
C_r^{Unit}	Costs of one step of resource r
S_r^{Unit}	Step size for costs computation of resource $r \in R^{NL}$
$UB_{rl,v,r}$	Upper bound on resource r for rebate level rl if vehicle v is used
$LB_{rl,v,r}$	Lower bound on resource r for rebate level rl if vehicle v is used
ϵ	A sufficiently small number
$BigM$	A sufficiently large number

Variables

$o_{p,v} \in N_0^+$	Number of ordered part units of part type p delivered in vehicle v
$u_r \in R_0^+$	Usage of resource r
$v_v^{Active} \in \{0, 1\}$	Decision, if vehicle v is used
$v_{rl,v}^{Level} \in \{0, 1\}$	Decision, if rebate level rl is active for vehicle v
$a_{lc,v} \in N_0^+$	Number of load carriers of type lc delivered in vehicle v

level the vehicle falls into according to the underlying tariff scheme. The third and fourth terms sum up resource cost for resources with linear and piecewise constant cost functions respectively. Constraint 5.3.9 ensures that all orders have to be picked up by a vehicle. Constraint 5.3.10 computes the amount of load carriers situated in a vehicle depending on the quantity of parts in it. For resources related directly to the vehicles discounting scheme Constraint 5.3.11 computes the resource use based on the number of parts and load carriers within the vehicle. All other resources are treated in Constraint 5.3.12. In contrast to resources related to the discounting scheme these resources may also depend on whether the vehicle is used and on the vehicle's discounting level. This allows us to model vehicle-independent resources, e.g. incoming goods personnel or common resources that are shared among all vehicles, e.g. carbon-dioxide emissions. Constraint group 5.3.13 ensures that each vehicle has exactly one discounting level if the vehicle is used, and no discounting level is active if the vehicle is not used at all. The Constraints 5.3.14 and 5.3.15 assure that the correct discounting level is selected according to the resource usages for the specific vehicle and that lower and upper bounds are hold. Constraints 5.3.16 and 5.3.17 prevent an inactive vehicle from carrying orders and a vehicle from being activated even though there are no orders placed in it. Even though the ceiling function given in constraints 5.3.10 is non-linear, it can be reformulated as described by Williams [1999]. Therefore a variable $a_{lc,v,p}^{frac} \in R_0^+$ for the fractional number of load carriers of type a required to transport all units of product p within vehicle v has to be introduced for each combination of load carrier, product and vehicle. In addition, a variable $a_{lc,v,p}^{prod} \in N_0^+$ has to be introduced to compute the next integer value for each of the fractional amounts of the individual products. Finally, the product specific integer quantities can be summed to retrieve the value of $a_{lc,v}^{prod} \in N_0^+$. This intermediary step is necessary to prevent different types of parts from being mixed in the same load carrier. The following constraints are required to replace the ceiling function in constraint set 5.3.10:

$$a_{lc,v,p} \geq a_{lc,v,p}^{frac} \quad \forall l \in LC, p \in P_{lc}, v \in V \quad (5.3.18)$$

$$a_{lc,v,p} \leq a_{lc,v,p}^{frac} + 1 - \epsilon \quad \forall l \in LC, p \in P_{lc}, v \in V \quad (5.3.19)$$

$$a_{lc,v} = \sum_{p \in P_{LC}} a_{lc,v,p} \quad \forall l \in LC, v \in V \quad (5.3.20)$$

$sym \in SYM$	Set of all symmetry groups
$r \in R_{sym}^{SYM}$	Resources in the symmetry group sym
$item(R_{sym}^{SYM}, n)$	Resource at position n in symmetry group sym

(5.3.21)

Constraint group 5.3.18 assures that the integer value $a_{lc,v,p}$ is always larger than the fractional value, while constraints 5.3.19 limit $a_{lc,v,p}$ to the next higher integer value. In combination these two constraint sets model the ceiling function. Constraints 5.3.20 then sum up the product-specific ceiled values.

Symmetry breaking

To increase performance when solving the model with a branch and bound algorithm, so-called symmetry breaking constraints can be added to the model. The idea is that several vehicles may have the same underlying discounting scheme and are therefore equally preferable. This will be called a symmetry group in the following. In such case it does not matter which of these equal vehicles will be used exactly, as each vehicle has the same associated capacities and cost factors. This property can be used to reduce the solution space that has to be explored by the branch and bound algorithm by introducing cutting planes. If all vehicles within one symmetry group are equally preferable, it does not make sense to try another vehicle from the group before the first one has been filled. Therefore use of the following vehicles before filling the first can be disallowed by adding the following constraints to the model:

$$u_{item(R_{sym}^{SYM}, n)} \geq u_{Item(R_{sym}^{SYM}, n+1)} \quad (5.3.22)$$

$$\forall sym \in SYM, \forall n = 1..|R_{sym}^{SYM}|$$

The underlying theory was developed by Fahle et al. [2001]. The idea is that symmetric parts of the solution space can be cut off without removing the optimal solution if it is known that one of those parts does not contain the optimal solution. Cutting planes can therefore be added to reduce the solution space when using the branch and bound algorithm.

Leaving out periods with Less-Than-Truckload

In the preprocessing a model instance has to be set up for each combination of delivery profile, supplier and period, thus $S \times DP \times T$ models have to be solved. However, it may be the case that the assignment of a delivery profile dp to a supplier s would result in a load that would fill less than one vehicle in certain periods. In this case it would not be necessary to set up a model instance. To check whether it is necessary to set up a model instance all resources limiting a vehicle's load have to be checked. This can be done in advance by computing the overall resource use for all orders with the following formula:

$$UB_r^{mu} = \sum_{p \in P} U_{p,r}^{Part} \cdot O_p + \sum_{lc \in LC} \sum_{p \in P_{lc}} U_{lc,r}^{Carrier} \cdot \left\lceil \frac{O_p}{Q_{p,lc}} \right\rceil \quad \forall r \in R^{mu} \quad (5.3.23)$$

where R^{mu} represents the set of measurement units used to determine the fill level of a vehicle. Note that $r \in R^{mu}$ will be represented by multiple $r \in R^v$ for different vehicles in the model later on. To estimate the required number of vehicles the upper bounds of the different measurement units have to be divided by the upper bound for a vehicle's load for that measurement unit. This results in multiple upper bounds on the vehicle count depending on the different measurement units. The maximum of these upper bounds can then be used as an absolute upper bound on the number of required vehicles:

$$UB^{Vehicles} = \text{Max} \left\{ \left\lceil \frac{UB_r^{mu}}{\text{Capacity}_r} \right\rceil \quad \forall r \in R^{mu} \right\} \quad (5.3.24)$$

where Capacity_r represents the upper bound of a measurement unit $r \in R^{mu}$ for a vehicle. If $UB^{Vehicles} = 1$, there will be only a pre leg run and no full load run. Thus pre leg cost can be derived directly from a lookup in the discounting-scheme table. All orders O_p will then be considered to be delivered in the main leg run in this period for this delivery profile assignments.

If $UB^{Vehicles} > 1$, an instance of the model has to be set up to identify which part of the goods will be transported in which vehicle. In this case one pre leg vehicle $v_{Preleg} \in V$ will be inserted. In addition, $UB^{Vehicles}$ will be used to determine the number of full load run vehicles $V_{Fullload} \subset V$ to be added to the model instance. For both pre leg run and full load run vehicles, the discounting schemes will be modeled via resources.

In addition, incoming goods department's personnel resources can be modeled to the full load run vehicles. It does not make sense to attach these resources to v_{Preleg} , as the pre leg run never causes resource use in the incoming goods department. After the model has been solved the pre leg run freight cost can be derived from a subset of the objective function that only covers v_{Preleg} and resources from $R_{v_{Preleg}}^v$ and leaves behind all other cost parts. The full load cost can be computed by subsuming the remaining freight cost for $v \in V_{Fullload}$ and the remaining resources. To determine which parts have to be transported in the main leg run, values for $o_{p,v_{Preleg}}$ have to be considered. Resource usages can be derived from u_r values for $r \in \setminus R^v$.

5.3.4 A primal packing heuristic

To reduce solving time primal heuristics can be used to provide a quick solution which offers a lower bound on the objective value and thus reduces the number of nodes that have to be analyzed in the branch and bound algorithm. To cope with the given model a primal packing heuristic has been developed to give a good starting solution for the distribution of goods onto the different vehicles. The heuristic approach is based on the concepts of *load items*, *efficiency* and *density*. A *load item* is considered to be a combination of a load carrier and a set of parts of such quantity that the parts fit into the load carrier, thus $q_{Item} \leq Q_{p,lc}$. Each order is then segregated into multiple load items, trying to fill each load carrier to its capacity, thus $q_{Item} = Q_{p,lc}$. In this way a maximum of one load item can carry less than the quantity of parts, whereas all other load items are completely filled. The load items are then seen as a single object that cannot be changed. This eases the handling of the load as it is no longer necessary to deal with both load carriers and parts.

For most discounting schemes it can be said that one of the measurement units used to determine the discount level is the price-driving measurement unit. In quantity based discounting schemes the price-driver is quantity. In discounting schemes for freight tariffs, weight or load meters are usually used for this purpose. The other measurement units to be considered are used to constrain the load of a cargo to its physical boundaries.

Definition 3. *The **price-driving measurement unit** of a vehicle is the measurement unit that is used to determine the discount level within the tariff discounting scheme related to that vehicle.*

Using the idea of a price-driving measurement unit the efficiency of a vehicle can be defined as follows.

Definition 4. *The **efficiency** of a vehicle describes the ratio between its capacity of the price-driving measurement unit and the price of the highest discount level.*

If defined this way, efficiency describes the magnitude of the price-driving measurement unit that can be loaded onto the vehicle per monetary unit if the vehicle were completely filled. It seems obvious that vehicles with a high efficiency should be loaded first, as they offer the most load opportunities per monetary unit. Using the notation from the model described above, the efficiency of a vehicle can be computed as

$$\text{Efficiency}_v = \frac{C_v^{Vehicle} + C_{v,rl^{Max}}^{Level} + \sum_{r \in RL} C_r^{Unit} \cdot UB_{rl^{Max},v,r} + \sum_{r \in R^{NL}} C_r^{Unit} * \left\lceil \frac{UB_{rl^{Max},v,r}}{S_r^{Unit}} \right\rceil}{UB_{rl^{Max},v,r}^{Price}} \quad (5.3.25)$$

To achieve a high degree of filling it is important to equally balance the fill level of the different resources in a vehicle. If for example one vehicle is completely filled in respect of one resource r_1 (e.g. weight) and another is completely filled in respect of another resource r_2 (e.g. volume), it may be that with an intelligent mixture of their parts, it would have been possible to load one vehicle to capacity and leave the other only partially loaded. This holds true especially for parts with very diverse properties, as they can be found in the automotive industry. A vehicle can take the maximal load when parts with a high density and those with a low density are combined in such way that the overall load density is close to the vehicle's capacity density. As depicted in Figure 5.3 higher or lower values for the density lead to one measurement unit's capacity being met whereas the other measurement unit's capacity is not yet reached. By contrast, the optimal loading strategy orientates to the vehicle's density, and leads to an equally distributed use of both measurement units' capacities. Where there are exactly two capacity boundaries the concept of density can be used. It is adopted from physics, but not limited to its original intention. For the application to a specific problem setting, where logistics service providers use other load measurement units than weight

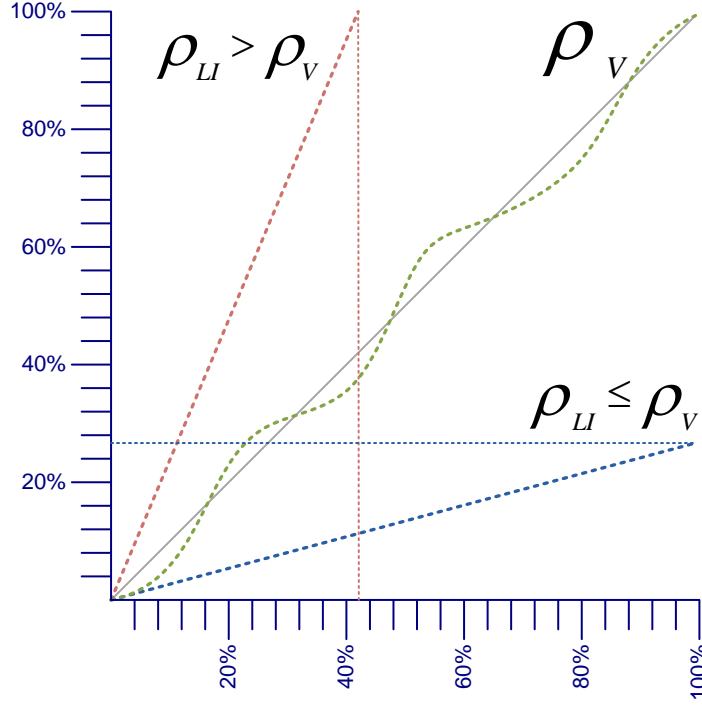


Figure 5.3: Impact of load density on vehicle capacity use.

and volume, it can be adapted to the given measurement units. Unfortunately the density concept is limited to an environment in which there are exactly two bounding measurement units, as density for a multidimensional environment is hard to define. If there is only one important measurement unit, where e.g. weight only is considered, the concept of density is apparently not necessary at all.

Definition 5. The *density* of an object describes the ratio between the price-driving measurement unit and the other measurement unit.

The density of a load item can be computed as

$$\rho^{LI} = \frac{U_{p,r1}^{Part} \cdot q_{Item} + U_{lc,r1}^{Carrier}}{U_{p,r2}^{Part} \cdot q_{Item} + U_{lc,r2}^{Carrier}} \quad (5.3.26)$$

where $r1 \in R^{mu}$ is one measurement unit's resource and $r2 \in R^{mu}$ the other measure-

ment unit's resource. By analogy to the load items' density the vehicle's density can be computed as

$$\rho^V = \frac{\text{Capacity}_{r1}}{\text{Capacity}_{r2}} \quad (5.3.27)$$

An overview of the heuristic packing procedure can be found in Algorithm 2. The load items' densities and the vehicles' efficiencies are first computed. Vehicles are then ordered by their efficiency. Thereafter, the most efficient vehicle is selected to be filled with load items. To achieve an optimal filling it would be best to sort the load elements by the difference between their density and the vehicle's density, and then always select the one with the lowest difference. Unfortunately the remaining density in the vehicle changes with each load item added: thus the list would have to be sorted again each time a load element is added to the vehicle. To improve the loading procedure the load items are separated into two lists, LI^- for load items with $\rho^{LI} \leq \rho^V$ and LI^+ for load items with $\rho^{LI} > \rho^V$. Load items are then taken from LI^- and added to the vehicle. After a load item has been added the vehicle's remaining density $\hat{\rho}^V$ is computed. While load items from LI^- are added, $\hat{\rho}^V$ increases. When $\hat{\rho}^V > \text{Min}(\rho^{L+})$, thus the remaining density has reached the density of the first load item in LI^+ , load items from LI^+ are loaded onto the vehicle. Now the remaining density in the vehicle reduces. When the density of load items from LI^- is reached and thus $\hat{\rho}^V < \text{Max}(\rho^{L-})$ holds true, load items are again taken from LI^- . These steps are repeated until the vehicle's capacity is met or all load items are on the vehicle. If the vehicle is loaded to capacity, thus no further load item can be added, the next vehicle is selected.

When the vehicles have been loaded initially there will be at most one vehicle which could not be loaded to capacity. Further to improve the packing, it can now be checked whether a swap of load items could decrease the partially loaded vehicle's discounting level. To check every possible swap would result in rather high computational effort, thus only promising swaps will be analyzed. A swap is seen to be promising if it can reduce the vehicle's discounting level without worsening another vehicle's discounting level or exceeding the other vehicle's capacity. This can only be the case if the sum of remaining capacities in filled vehicles is larger or equal to the magnitude of excess on the current discounting level of the partially loaded vehicle. If the remaining capacities fulfill the conditions described beforehand, the load items in the partially filled vehicle are compared pairwise with the load items in other vehicles. The swap which leaves

Algorithm 2: The packing heuristic algorithm.

```

Compute load items densities;
Compute vehicles efficiencies;
Order vehicles by their efficiency;
while  $LI \neq \emptyset$  do
     $V^{Fill} \leftarrow Top(V)$ ;
     $LI^- \leftarrow \{le \in LI : \rho^{LI} \leq \hat{\rho}^{V^{Fill}}\}$ ;
     $LI^+ \leftarrow LI \setminus LI^-$ ;
    while  $\exists LI : LI \text{ fits in } V^{Fill}$  do
        if  $\hat{\rho}^{V^{Fill}} \leq Max(\rho^{LI^-})$  then
             $li \leftarrow Top(LI^-)$ ;
             $LI^- \leftarrow LI^- \setminus li$ ;
        end
        else
             $li \leftarrow Top(LI^+)$ ;
             $LI^+ \leftarrow LI^+ \setminus li$ ;
        end
        Add  $li$  to  $V^{Fill}$ ;
        Compute  $\hat{\rho}^{V^{Fill}}$ ;
    end
end

```

the minimal remaining capacity in the filled vehicle will then be carried out. This procedure is repeated until no swap is any longer possible or a lower discounting level has been reached. In the case where only a small number of load items is necessary to reach another discounting level, it may be advantageous to continue swapping after an improvement has been achieved.

5.4 Main leg model formulation

The results from preprocessing are used as input for the delivery profile selection and main leg run model formulation, called main leg model in the following. The main leg model decides the supplier's delivery profile assignments that have been evaluated during preprocessing. An estimation of the main leg run cost is thus used to identify the synergy effects between different delivery profile assignments. The inputs for the main leg model are, on the one hand, the evaluated delivery profile assignments, in-

cluding the associated cost factors, resulting resource uses and the order remaining for the main leg run. On the other hand, main leg run tariff data and resource uses are required as input. The main leg model is an extended version of the model used during preprocessing. These extensions include the assignment of delivery profiles to suppliers, consideration of multiple periods at once and bounds on limited resources. Following the notation given below, the model denotes as follows:

Sets

$t \in T$	Set of time periods.
$s \in S$	Set of suppliers connected to the consolidation center.
$dp \in DP$	Set of available delivery profiles.
$p \in P$	Set of part types.
$lc \in LC$	Set of load carriers.
$p \in P_{lc} \subseteq P$	Set of parts which will be delivered in load carrier lc .
$v \in V$	Set of vehicles
$rl \in RL^v$	Set of rebate levels when using a vehicle.
$r \in R$	Set of resources.
$r \in R^L \subseteq R$	Set of resources with a linear cost function.
$r \in R^{NL} \subseteq R$	Set of resources with a piecewise-constant cost function.
$R^L \cup R^{NL} = R$	Set of resources consists of resources with linear cost function and resources with piecewise-constant cost function.
$R^L \cap R^{NL} = \emptyset$	Each resource has either a linear cost function or a piecewise-constant cost function, but not both.
$R^v \subseteq R$	Set of resources which are used only by vehicle v and are related to this vehicle's tariff system.

Parameters

$C_{s,dp}^{Inventory}$	Inventory related cost factors for assignment of delivery profile dp to supplier s .
$C_{s,dp}^{Preleg}$	Costs for pre leg transport resulting from the choice of delivery profile dp for supplier s .
$C_{s,dp}^{Fullload}$	Costs for full load transport resulting from the choice of delivery profile dp for supplier s .
$U_{s,dp,r,t}^{Profile}$	Resource usage resulting from the choice of the delivery profile dp for supplier s in period t .
$O_{p,dp,s,t}$	Quantity of part units of part type p which have to be ordered in period t if delivery profile dp is chosen for supplier s .
$Q_{p,lc}$	Maximum number of part units of part type p which can be delivered in load carrier lc .
$U_{p,r,v}^{Part}$	Usage of resource r by part of type p when using vehicle v .
$U_{lc,r,v}^{Carrier}$	Usage of resource r by load carrier of type lc when using vehicle v .
$U_{v,r}^{Vehicle}$	Usage of resource r resulting from usage of vehicle v .
$U_{v,rl,r}^{Level}$	Usage of resource r resulting from a load equal to discount level rl when using vehicle v .
$C_{v,rl}^{Level}$	Cost of rebate level rl when using vehicle v .
$C_v^{Vehicle}$	Base cost of vehicle v .
C_r^{Unit}	Cost of one step of resource r .
S_r^{Unit}	Step size for cost computation of resource $r \in R^{NL}$.
UB_r	Upper bound on resource r .
$UB_{rl,v,r}$	Upper bound on resource r for rebate level rl if vehicle v is used.
$LB_{rl,v,r}$	Lower bound on resource r for rebate level rl if vehicle v is used.
ϵ	A sufficiently small number.
$BigM$	A sufficiently large number.

Decision Variables

$pc_{s,dp} \in \{0, 1\}$	Decision whether delivery profile dp is selected for supplier s .
$o_{p,t,v} \in N_0^+$	Quantity of ordered part units of part type p in period t delivered in vehicle v .
$u_{r,t} \in R_0^+$	Usage of resource r in period t .
$v_{t,v}^{Active} \in \{0, 1\}$	Decision whether vehicle v is used in period t .
$v_{t,rl,v}^{Level} \in \{0, 1\}$	Decision whether rebate level rl is active in period t for vehicle v .
$a_{t,lc,v} \in N_0^+$	Number of load carriers of type lc in period t delivered in vehicle v .

Model formulation

$$\begin{aligned}
 \text{Min} \quad & \sum_{dp \in DP, s \in S} \left(C_{s,dp}^{Inventory} + C_{s,dp}^{Preleg} + C_{s,dp}^{Fullload} \right) \cdot pc_{s,dp} \\
 & + \sum_{t \in T, v \in V} C_v^{Vehicle} \cdot v_{t,v}^{Active} + \sum_{t \in T, v \in V, rl \in RL} C_{v,rl}^{Level} \cdot v_{t,rl,v}^{Level} \\
 & + \sum_{r \in R^L} C_r^{Unit} \cdot \left(\sum_{dp \in DP, s \in S} U_{s,dp,r}^{Profile} \cdot pc_{s,dp} + \sum_{t \in T} u_{r,t} \right) \\
 & + \sum_{r \in R^{NL}} C_r^{Unit} \cdot \left[\frac{\sum_{dp \in DP, s \in S} U_{s,dp,r}^{Profile} \cdot pc_{s,dp} + \sum_{t \in T} u_{r,t}}{S_r^{Unit}} \right]
 \end{aligned} \tag{5.4.1}$$

subject to

$$\sum_{dp \in DP} pc_{s,dp} = 1 \quad \forall s \in S \tag{5.4.2}$$

$$\sum_{s \in S} O_{p,dp,s,t} \cdot pc_{s,dp} = \sum_{v \in V} o_{p,t,v} \quad \forall t \in T, dp \in DP, p \in P \quad (5.4.3)$$

$$a_{t,lc,v} = \sum_{p \in P_{lc}} \left\lceil \frac{o_{p,t,v}}{Q_{p,lc}} \right\rceil \quad \forall lc \in LC, v \in V, t \in T \quad (5.4.4)$$

$$u_{r,t} = \sum_{v \in V} \sum_{p \in P} U_{p,r}^{part} \cdot o_{p,t,v} + \sum_{v \in V} \sum_{lc \in LC} U_{lc,r,v}^{Carrier} \cdot a_{t,lc,v} \quad \forall t \in T, r \in \bigcup_{v \in V} R^v \quad (5.4.5)$$

$$u_{r,t} = \sum_{v \in V} U_{v,r}^{Vehicle} \cdot v_{t,v}^{Active} + \sum_{v \in V} \sum_{rl \in RL_v} U_{v,rl,r}^{Level} \cdot v_{t,rl,v}^{Level} + \sum_{v \in V} \sum_{p \in P} U_{p,r}^{part} \cdot o_{p,t,v} + \sum_{v \in V} \sum_{lc \in LC} U_{lc,r,v}^{Carrier} \cdot a_{t,lc,v} + \sum_{dp \in DP, s \in S} U_{s,dp,r,t}^{Profile} \cdot pc_{s,dp} \quad \forall t \in T, r \in R \setminus \bigcup_{v \in V} R^v \quad (5.4.6)$$

$$\sum_{rl \in RL^v} v_{t,rl,v}^{Level} = v_{t,v}^{Active} \quad \forall v \in V, t \in T \quad (5.4.7)$$

$$UB_{rl,v,r} \cdot v_{t,rl,v}^{Level} - \epsilon \geq u_{r,t} - BigM \cdot (1 - v_{t,rl,v}^{Level}) \quad \forall v \in V, t \in T, r \in R^v, \quad rl \in RL^v \quad (5.4.8)$$

$$LB_{rl,v,r} \cdot v_{t,rl,v}^{Level} \leq u_{r,t} \quad \forall v \in V, t \in T, \quad r \in R^v, rl \in RL^v \quad (5.4.9)$$

$$BigM \cdot v_{t,v}^{Active} \geq \sum_{p \in P} o_{p,t,v} \quad \forall v \in V, t \in T \quad (5.4.10)$$

$$v_{t,v}^{Active} \leq \sum_{p \in P} o_{p,t,v} \quad \forall v \in V, t \in T \quad (5.4.11)$$

The objective function 5.4.1 consists of two main parts. The first part reflects cost incurred by the assignment of a delivery profile to a supplier, which have been identified during preprocessing. The second part is similar to the objective function from the preprocessing model. It consists of main leg run freight cost and resource use cost for each period. In contrast to the preprocessing model all periods have to be considered at once, therefore a summation over $t \in T$ takes place. Even though this part of the objective function has the same structure as that from the preprocessing model, other data underly the discounting schemes. Whereas a choice between pre leg run vehicles and full load run vehicles and their corresponding tariff discounting schemes had to be made in preprocessing, different main leg vehicles have to be loaded with the parts ordered on this model. Hence the main leg tariff structure underly the discounting levels in this model. Constraint 5.4.2 forces the selection of exactly one delivery profile assignment per supplier. This is the core decision to be taken on this model. Constraint 5.4.3 connects the delivery profile assignment to the remaining main leg orders computed in the preprocessing. It ensures that the same quantity of parts will be transported in the main leg run in each period as were delivered to the consolidation center in accordance with the selected delivery profile assignment. To compute the number of load carriers per vehicle Constraint 5.4.4 is employed. In analogy to the preprocessing mode Constraint 5.4.5 is used to compute resource use for resources related to the vehicles tariffs. Based on these, Constraints 5.4.7, 5.4.8 and 5.4.9 determine the correct discounting level for each vehicle, whereas Constraints 5.4.10 and 5.4.11 ensure correct vehicles activation. Note that Constraint 5.4.6 is extended not only to consider the multiple periods, but also covers resource usage arising from delivery profile assignments through the term $\sum_{dp \in DP, s \in S} U_{s,dp,r,t}^{Profile} \cdot pc_{s,dp}$. This is necessary to cope with resources shared between both full load and main leg runs, e.g. incoming goods personnel resources. If they were neglected at this stage, the model could no longer guarantee optimality or even provide invalid solutions under consideration of

the shared resources.

By analogy to the preprocessing model, symmetry-breaking constraints can once again be inserted to speed up the branch and bound algorithm. Due to the consideration of multiple periods within the model, they have to be extended by a period index as follows:

$$\begin{aligned} u_{item(R_{sym}^{SYM},n),t} &\geq u_{Item(R_{sym}^{SYM},n+1),t} \\ \forall sym \in SYM, \forall n = 1..|R_{sym}^{SYM}|, t \in T \end{aligned} \tag{5.4.12}$$

5.5 Primal heuristics

In this section two primal heuristics are presented which can be used to solve the main leg model. The first one is a local search algorithm that tries to improve a given solution in a step by step procedure. In so doing a problem-specific selection strategy is used to determine the search direction. The second one is a genetic algorithm that uses biologic analogies to find good solutions but does not use problem-specific considerations to derive the search direction. Both algorithms rely heavily on the packing heuristic described in Section 5.3.4. The main decision to be made in the main leg model is to assign a delivery profile to each supplier. When a specific delivery profile assignment has been selected for each supplier, it has to be decided which parts to load onto which vehicles. The first decision on the delivery profile assignments provides the general conditions for the second decision, the load distribution of parts to vehicles. Hence an optimal part distribution can be seen as a function depending on a given profile assignment solution. Given that the part distribution problem is independent in each period, it can be stated that for each delivery profile assignment solution, a vector of independent optimal solutions for the part-distribution problem can be found. The two subproblems, delivery profile assignment and part distribution, cannot be treated independently, but the latter can be solved quite efficiently with the solution procedure presented for preprocessing (see Section 5.3.4). Therefore the part-distribution problem based on the general conditions given by the delivery profile assignment can be seen as a more complex objective function computation step. This allows us to reduce the whole problem to the selection of an optimal delivery profile assignment vector.

5.5.1 A local search heuristic

Figure 5.4 gives an overview of the heuristic solution procedure. At the beginning a start strategy is used to create a first delivery profile assignment. This delivery profile assignment is then evaluated by solving the part distribution subproblems for each period which result from the selected delivery profile assignment. From both the part distribution and the delivery profile assignments a value for the objective function is derived. A selection strategy is used to analyze the outcome and to decide upon the next delivery profile assignment. After the assignment has been created the subproblems are solved again. This procedure is repeated until a termination condition provided by the delivery profile selection strategy, e.g. a given number of iterations without improvements or a time limit becomes true. A good starting solution may speed up the follow-up process and may also lead to better solution quality. Different starting strategies can be deployed to provide a starting solution. In the current version of the heuristic the following strategies have been implemented and tested:

- **Random selection strategy** is a very simple strategy that picks a random delivery profile for each supplier.
- **Lowest pre leg price strategy** selects the delivery profile with the lowest pre leg run price for each supplier. The idea is that a low pre leg run price is an indicator for fewer parts to be transported in follow-up main leg run, which may thus become cheap as well.
- **Lowest total cost strategy** selects the delivery profile with the lowest sum of pre leg run, full load run and inventory related cost factors for each supplier.
- **Lowest inventory cost strategy** chooses the delivery profile with lowest inventory related cost for each supplier. This leads to a solution with little or no inventory usage at all.
- **Most efficient pre leg strategy** uses the price per unit ratio for pre leg runs. The delivery profile with the cheapest ratio is selected for each supplier.

As selection strategy a local search algorithm with a taboo list was implemented. The neighborhood was defined using the hamming distance. The hamming distance

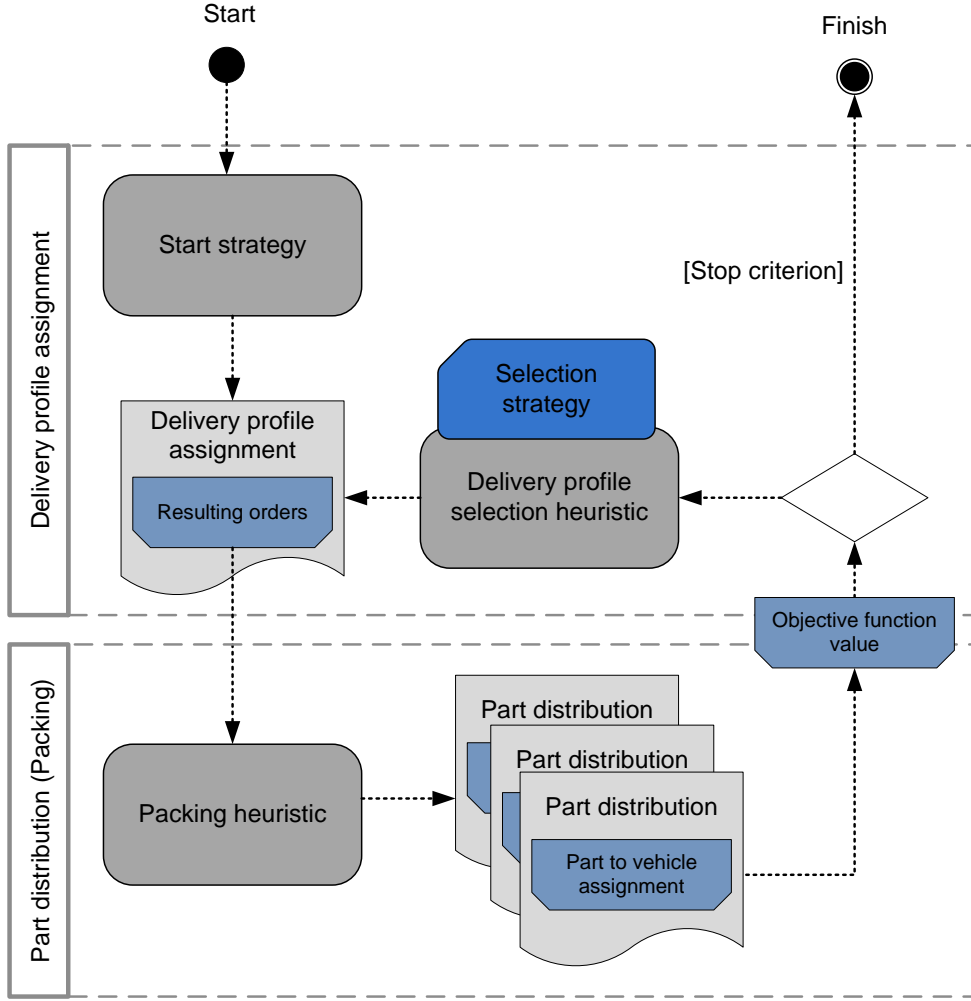


Figure 5.4: Overview of heuristic solution procedure for main leg problem.

$\Delta_{\text{Hamming}}(\vec{X}, \vec{Y})$, which was developed in Hamming [1950] and was originally designated to describe the distance between two binary signals, gives back the number of changed elements between two vectors \vec{X} and \vec{Y} of equal length N . It is defined as

$$\Delta_{\text{Hamming}}(\vec{X}, \vec{Y}) = \sum_{n=1}^N \sum_{x_n \neq y_n} 1$$

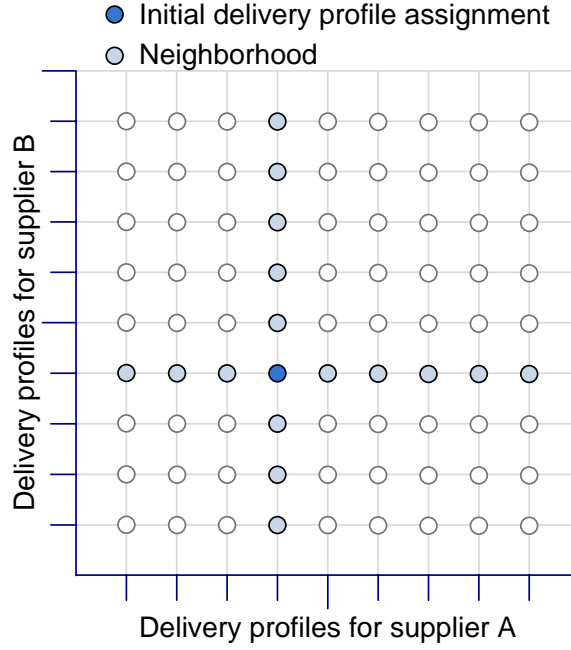


Figure 5.5: Hamming neighborhood for a hamming distance of one and two suppliers.

So for each mismatch of two elements' (x_n, y_n) being in the same position within the vectors but having a different value, the hamming distance is increased by one. Applied to the given problem setting the solution of the assignment problem may be written as a vector \vec{A} of size $|S|$ consisting of one element per supplier $s \in S$. The value of the element at position s represents the delivery profile assigned to supplier s . Using this notation the hamming distance $\Delta_{\text{Hamming}}(\vec{A}, \vec{B})$ between two delivery profile assignment vectors \vec{A} and \vec{B} represents the number of suppliers with a different delivery profile assigned in \vec{A} and \vec{B} . If the hamming distance for the local search is now constrained to $\Delta_{\text{Hamming}}(\vec{A}, \vec{B}) = 1$, then only one suppliers delivery profile may be changed from one solution to another. For the implemented local search a hamming distance of 1 was selected. Figure 5.5 shows the neighborhood for a hamming distance of 1 for the two suppliers case. The size of the neighborhood can be computed as

$$\prod_{s \in S} |DP_s|$$

with $DP_s \subseteq DP$ being the delivery profiles available for supplier $s \in S$. The local

search procedure starts with the delivery profile assignment given by the starting strategy. An *improvement cycle* is then started. An *improvement cycle* tries to find a solution in the neighborhood which leads to an improvement. To pick the best solution from the neighborhood each supplier is tested once in a *test step*. In a *test step* each available delivery profile is once assigned to the supplier and thereafter evaluated by the packing heuristic, whereas all other suppliers delivery profile assignments remain fixed. Executing a test step for each supplier leads to an evaluation of the complete neighborhood. So in each *improvement cycle* the whole neighborhood is evaluated. If there is an assignment within the neighborhood that has a lower objective function value than the assignment selected at the beginning of the *improvement cycle*, it will be used as a starting solution for the next *improvement cycle*. After a successful *improvement cycle* has taken place the supplier whose delivery profile was reassigned will be fixed for the next iteration, thus no *test step* will regularly be executed for that supplier. Given this fixation technique the number of required *test steps* reduces from one *improvement cycle* to another. If no improvement was possible, the algorithm relaxes the fixation constraint, so that all suppliers can be targeted in a *test step*, because otherwise it could get stuck in a local optima. If the *improvement cycle* does not provide a better solution either, it may be said that a local optima was found because

$$\vec{A}\vec{B} : \Delta_{\text{Hamming}}(\vec{A}, \vec{B}) = 1 \wedge \text{Objective}(\vec{A}) > \text{Objective}(\vec{B})$$

holds true and is the condition for a local optimum being found. In this case the algorithm can either be started again with another starting strategy or the solution can be accepted.

5.5.2 A genetic algorithm

Genetic algorithms are primal heuristic algorithms inspired by biologic principles. The basic idea behind a genetic algorithm is to imitate the natural process of evolution that consists of reproduction, mutation and natural selection. In genetic algorithms each solution is therefore seen as an individual that participates in the evolutionary process. Each individual is represented by its genome. If two individuals are used for reproduction, the genome of the offspring will be composed of genome fragments from both parents. Mutations can be applied by randomly changing a part of the genome. The process of natural selection is imitated by discarding the solutions of lower quality

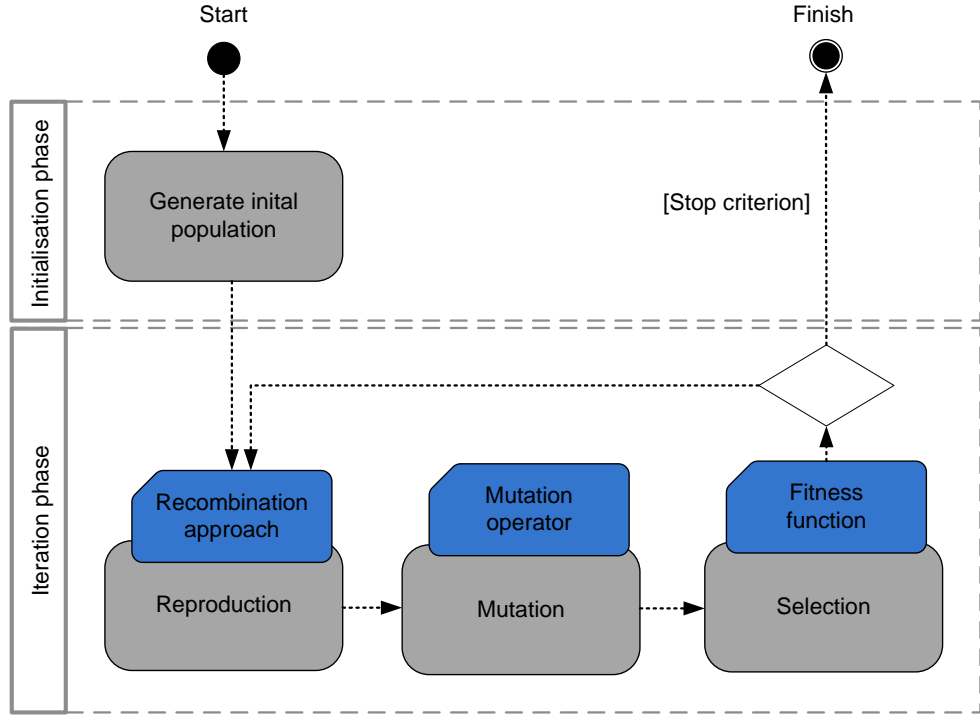


Figure 5.6: Overview of the genetic algorithm.

in the selection process. An overview of the solution procedure can be seen in Figure 5.6. At the beginning an initial starting population will be generated. Thereafter the evolutionary process of reproduction, mutation and selection is repeated iteratively. In the reproduction step parents are chosen and used to reproduce offspring, which will then be added to the population. All individuals in the population are then exposed to random mutations. Finally, the best individuals are selected and all other individuals are removed from the population. To evaluate which individuals are the best ones a fitness function f is used that projects an individual i to a scalar value $f(i)$. For the given problem setting the vector of delivery profile assignments may be used as a genome. An individual is accordingly represented by a delivery profile assignment vector pc . Each entry in the vector represents the assignment of one specific delivery profile dp to a supplier s . In the mathematical model formulations the variable $pc_{s,dp}$ would be set to one accordingly. Figure 5.7 depicts the representation of two individuals, a possible offspring and a possible mutation of the offspring. To generate the initial population

	Supplier 1	Supplier 2	Supplier 3	Supplier 4	Supplier 5
Individual A	W10001	W00100	R2	W10101	W10000
Individual B	W10101	W10000	W10101	W00100	R3
Offspring (A+B)	W10101	W00100	R2	W10101	R3
Mutation (Offspring)	W10101	W01010	R2	W10101	R3

Figure 5.7: Problem specific example for crossover and mutation operators.

each supplier will have a random delivery profile assigned. Alternatively, the starting strategies described in Section 5.5 can be used to create the initial population, or a combination of both randomly generated individuals and individuals that result from the starting strategies can be used.

In the reproduction phase two parent individuals are picked randomly from the population. As the position of a supplier in the vector pc does not correspond to any semantic property in the underlying problem setting, each supplier can be treated separately from all other suppliers. Therefore, it can be decided on the parent whose assignment will be used individually for each supplier. Thus for each supplier it is decided randomly whether the assignment from the first or second parent individual will be inherited by the offspring. The chances are thus equal for both parents to pass over their assignment. This behavior can be seen in Figure 5.7, where for suppliers two, three and four the delivery profiles from individual A are passed to the offspring, whereas for suppliers one and five the delivery profiles from individual B are passed to the offspring. To increase the diversity of the individuals in the population randomly generated solutions may be inserted with a certain probability. In this case, instead of recombining two existing individuals, an individual with random assignments for each supplier will be added to the population. This strategy has proven capable of increasing the diversity of the population and therefore avoiding getting stuck in a local optimum. After the reproduction phase the individuals will suffer a random mutation

with a certain probability. Mutation in this case means that for each supplier a probability holds that the assignment will be changed to another random assignment. Thus each position in the vector pc is flipped to a random profile with a certain probability. In the example given in Figure 5.7 the offspring is mutated by changing the delivery profile from supplier two to the random delivery profile 'W01010' instead of the original 'W00100'. When it comes down to the selection process the packing heuristic described in Section 5.3.4 is deployed to evaluate the economic impact of a delivery profile assignment vector and is thus used as a fitness function. After the solutions have been evaluated by the packing heuristic the best solutions remain in the population and all other solutions are discarded.

5.6 Consideration of demand uncertainty

In the preceding section, the deterministic demand case was discussed. As described in Section 2.2.3, this might not reflect the situation in practice. In fact there are multiple sources of demand uncertainty, ranging from changes in sales forecasting to continuous re-planning of production sequences. This leads in time to the establishment of a distribution different from that previously planned for. When assigning a delivery profile to a supplier a tactical decision is made that usually covers a period of about three months in time. Even though the MRP system adapts the orders depending according to the actual demand situation, a delivery profile assignment may produce a completely different outcome from that which was expected to do during the planning phase. A delivery profile assignment that was planned to be optimal for the whole planning horizon may turn out to be a poor solution given the demands finally established. In multiple applications stochastic programming has shown to be a good method of incorporating uncertainties into the planning process (see Wallace and Ziemba [2005] for an overview of successful applications). The solution of a stochastic programming model is not necessarily optimal for the one possible scenario of the future, but instead provides a high solution quality for a set of alternative scenarios of the future. This ability to provide a high solution quality for multiple scenarios will be called robustness in the following. The basic idea behind stochastic programming is to assume that a certain set of parameters is not deterministic, but rather follows a probability function, and to incorporate this knowledge when making a decision. Instead of choosing optimal decision variable values according to a deterministic expectation

of this parameter set, the goal is to choose decision values such that the the expected value of the objective function under consideration of the probability function is optimized. For the given problem setting a two-stage stochastic program is an appropriate choice. In the first stage a delivery profile assignment is made. This is a decision that has to be made in advance and cannot be changed afterwards. In the second stage the demands are realized with the passage of time. It could be argued that due to the nature of a rolling horizon there should be multiple stages deployed as uncertainty does not reveal itself at once but over time and leaves the possibility to respond. But as the focus of this part of the work is on the selection of delivery profiles, this does not hold true. A delivery profile has to be selected in advance and in following planning cycles only the delivery schedules will be adjusted following that rule. This does not leave any room for optimization purposes, as it will be executed by the MRP systems predefined rules. Thus there would not be any additional information or improvement to the solution quality if a multi-stage stochastic program were deployed. Thus this work will provide a two-stage formulation of a stochastic program for the selection of cost-minimal and robust delivery profiles. A two-stage stochastic program basically has two sets of variables. One set represents the decisions to be made at the first stage and thus are not dependent of the realized scenario. The other set considers decisions to be made at the second stage and thus depends on the realized scenario. In the given problem setting the assignment of a delivery profile to a supplier is a first-stage decision, whereas all decisions on the transport of parts including vehicles packing and tariff selection are second-stage decisions as they depend on actual demand which will be realized over time. Due to this clear separation between the delivery profile assignment problem on the one hand and the part distribution problem on the other hand, the previously described decomposition approach can also be used for the two stage stochastic program. Some modifications have to be made to deal with the presence of multiple demand scenarios. First, preprocessing has to be adopted, as the parameters depending on the suppliers' delivery profile assignment also depend upon the demand scenario. Whereas for the deterministic case each delivery profile assignment has to be evaluated once, multiple scenarios have to be evaluated for each delivery profile assignment in the stochastic case. In addition, the main leg model has to be adapted such that it covers a packing solution for multiple scenarios. In the following, the necessary adoptions and the resulting solution approach will be given.

Former parameter	Current parameter	Explanation
$C_{s,dp}^{Inventory}$	$C_{s,dp,z}^{Inventory}$	Depending on the realized demand scenario, different parts inventory uses can result.
$C_{s,dp}^{Preleg}$	$C_{s,dp,z}^{Preleg}$	A different demand situation also influences the freight cost of pre leg runs.
$C_{s,dp}^{Fullload}$	$C_{s,dp,z}^{Fullload}$	A different demand situation also influences the freight cost of full load runs.
$O_{p,dp,s,t}$	$O_{p,dp,s,t,z}$	A realized demand scenario may lead to another part distribution and thus results in different parts and quantities remaining for the main leg run.

Table 5.1: Preprocessing output parameters requiring an additional subscript for the stochastic case.

5.6.1 Preprocessing

Whereas preprocessing for the deterministic case covered exactly one demand scenario, multiple demand scenarios have to be handled in the stochastic case. This leads to an additional subscript for the parameters that result from the preprocessing algorithm as well as an additional loop over all scenarios for the computational steps. Basically, the existing preprocessing algorithm can be reused, but it has to be started once for each scenario $z \in Z$. Table 5.6.1 gives an overview of the parameters that have to be extended by an additional scenario subscript.

5.6.2 Adapted model formulation

In the following the adapted model formulation for the two stage stochastic program will be given. As a reasonable part of the sets and parameters remains the same as in the deterministic case, only additional sets and revised parameters will be presented. For details on the other parameters and sets see Section 5.4.

Additional Sets

$z \in Z$ Set of demand scenarios

Revised parameters

p_z	Probability that scenario z realizes, $\sum_z p_z = 1$.
$C_{s,dp,z}^{Inventory}$	Inventory-related cost factors in scenario z for assignment of delivery profile dp to supplier s .
$C_{s,dp,z}^{Preleg}$	Cost of pre leg transport in scenario z resulting from the choice of delivery profile dp for supplier s .
$C_{s,dp,z}^{Fullload}$	Cost of full load transport in scenario z resulting from the choice of delivery profile dp for supplier s .
$U_{s,dp,r,t,z}^{Profile}$	Resource usage in scenario z resulting from the choice of the delivery profile dp for supplier s in period t .
$O_{p,dp,s,t,z}$	Quantity of part units of part type p which have to be ordered in period t if delivery profile dp is chosen for supplier s and scenario z is realized.

Decision Variables

$pc_{s,dp} \in \{0, 1\}$	Decision, if delivery profile dp is selected for supplier s .
$o_{p,t,v,z} \in N_0^+$	Quantity of ordered part units of part type p in period t delivered in vehicle v when scenario z is realized.
$u_{r,t,z} \in R_0^+$	Usage of resource r in period t if scenario z is realized.
$v_{t,v,z}^{Active} \in \{0, 1\}$	Decision if vehicle v is used in period t and scenario z .
$v_{t,rl,v}^{Level} \in \{0, 1\}$	Decision if rebate level rl is active in period t for vehicle v in scenario z .
$a_{t,lc,v,z} \in N_0^+$	Number of load carriers of type lc in period t delivered in vehicle v in scenario z .

Model formulation

$$\begin{aligned}
 \text{Min} \quad & \sum_{dp \in DP, s \in S, z \in Z} \left(C_{s,dp,z}^{Inventory} + C_{s,dp,z}^{Preleg} + C_{s,dp,z}^{Fullload} \right) \cdot pc_{s,dp} \cdot p_z \\
 & + \sum_{t \in T, v \in V, z \in Z} C_v^{Vehicle} \cdot v_{t,v}^{Active} \cdot p_z + \sum_{t \in T, v \in V, rl \in RL} C_{v,rl}^{Level} \cdot v_{t,rl,v}^{Level} \cdot p_z \\
 & + \sum_{r \in R^L, z \in Z} C_r^{Unit} \cdot p_z \cdot \left(\sum_{dp \in DP, s \in S} U_{s,dp,r}^{Profile} \cdot pc_{s,dp,z} + \sum_{t \in T} u_{r,t,z} \right) \\
 & + \sum_{r \in R^{NL}, z \in Z} C_r^{Unit} \cdot p_z \cdot \left\lceil \frac{\sum_{dp \in DP, s \in S} U_{s,dp,r}^{Profile} \cdot pc_{s,dp} + \sum_{t \in T} u_{r,t}}{S_r^{Unit}} \right\rceil
 \end{aligned} \tag{5.6.1}$$

subject to

$$\sum_{dp \in DP} pc_{s,dp} = 1 \quad \forall s \in S \tag{5.6.2}$$

$$\sum_{s \in S} O_{p,dp,s,t,z} \cdot pc_{s,dp} = \sum_{v \in V} o_{p,t,v,z} \quad \forall t \in T, dp \in DP, p \in P, z \in Z \tag{5.6.3}$$

$$a_{t,lc,v,z} = \sum_{p \in P_{lc}} \left\lceil \frac{o_{p,t,v,z}}{Q_{p,lc}} \right\rceil \quad \forall lc \in LC, v \in V, t \in T, z \in Z \tag{5.6.4}$$

$$\begin{aligned}
 u_{r,t,z} = & \sum_{v \in V} \sum_{p \in P} U_{p,r}^{part} \cdot o_{p,t,v,z} \\
 & + \sum_{v \in V} \sum_{lc \in LC} U_{lc,r,v}^{Carrier} \cdot a_{t,lc,v,z} \quad \forall t \in T, r \in \bigcup_{v \in V} R^v, z \in Z
 \end{aligned} \tag{5.6.5}$$

$$\begin{aligned}
 u_{r,t,z} = & \sum_{v \in V} U_{v,r}^{Vehicle} \cdot v_{t,v,z}^{Active} \\
 & + \sum_{v \in V} \sum_{rl \in RL_v} U_{v,rl,r}^{Level} \cdot v_{t,rl,v,z}^{Level}
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{v \in V} \sum_{p \in P} U_{p,r}^{part} \cdot o_{p,t,v,z} \\
 & + \sum_{v \in V} \sum_{lc \in LC} U_{lc,r,v}^{Carrier} \cdot a_{t,lc,v,z} \\
 & + \sum_{dp \in DP, s \in S} U_{s,dp,r,t,z}^{Profile} \cdot pc_{s,dp} \\
 & \forall t \in T, r \in R \setminus \bigcup_{v \in V} R^v, z \in Z
 \end{aligned} \tag{5.6.6}$$

$$\sum_{rl \in RL^v} v_{t,rl,v,z}^{Level} = v_{t,v,z}^{Active} \quad \forall v \in V, t \in T, z \in Z \tag{5.6.7}$$

$$UB_{rl,v,r} \cdot v_{t,rl,v,z}^{Level} - \epsilon \geq u_{r,t,z} - BigM \cdot (1 - v_{t,rl,v,z}^{Level}) \quad \forall v \in V, t \in T, r \in R^v, rl \in RL^v, z \in Z \tag{5.6.8}$$

$$LB_{rl,v,r} \cdot v_{t,rl,v,z}^{Level} \leq u_{r,t,z} \quad \forall v \in V, t \in T, r \in R^v, rl \in RL^v, z \in Z \tag{5.6.9}$$

$$BigM \cdot v_{t,v,z}^{Active} \geq \sum_{p \in P} o_{p,t,v,z} \quad \forall v \in V, t \in T, z \in Z \tag{5.6.10}$$

$$v_{t,v,z}^{Active} \leq \sum_{p \in P} o_{p,t,v,z} \quad \forall v \in V, t \in T, z \in Z \tag{5.6.11}$$

A set Z is introduced with one entry $z \in Z$ for each scenario to be considered. Each scenario has a probability to be realized, which is given in parameter p_z . When summed up, these probabilities reach 100%, thus $\sum_z p_z = 1$. As described in the previous section, parameters determined in preprocessing also depend on the realized scenario, thus an additional subscript z is added to each of these. The decision variables are divided into first-stage variables and second-stage variables. At the first stage $pc_{s,dp}$ determines the assignment of a delivery profile to a supplier. All other decision variables are second-stage variables as they depend on the realized scenario. Note that in objective function 5.6.1 all terms now consist of weighted sums over all scenarios, where p_z is the weight factor of the cost terms related to the realized scenario. This reflects the expected value of the objective function under consideration of all scenarios and their probability of being realized. As the assignment of delivery profiles is independ-

ent of the realized scenario, Constraint 5.6.2 can be taken over without changes from the deterministic model. Constraint 5.6.3 has to be modified so that the quantity of ordered parts is equal to the quantity of parts remaining for main leg run according to the assigned delivery profile for each realized scenario. This constraint group creates a connection between the decisions at the first and second stage, as the second-stage part distribution variable $o_{p,t,v,z}$ is linked to the first-stage delivery profile variable $pc_{s,dp}$ via the related scenario-dependent parameter $O_{p,dp,s,t,z}$. In Constraint 5.6.4 the number of load carriers located in a vehicle is computed for each scenario. Constraint 5.6.5 and 5.6.6 compute the resource uses for vehicle-related and non vehicle-related resources for each scenario. Constraints 5.6.7, 5.6.8 and 5.6.9 determine the discount levels according to the vehicle's load in a certain scenario. In Constraints 5.6.10 and 5.6.11, the activation of vehicles is handled in respect to the realized scenario.

5.6.3 Modified solution algorithm for the stochastic case

There are only a few solution algorithms that are applicable under general conditions for two-stage stochastic programs and capable of dealing with integer variables at the second stage. In the model presented above the second-stage variables $o_{p,t,v,z}$ and $a_{t,lc,v,z}$ are integer variables. In addition, $v_{t,v,z}^{Active}$ and $v_{t,rl,v}^{Level}$ are binary variables at the second stage. Hence a specific algorithm should be developed to reduce runtime. As can be seen from the model, all vehicle-related decisions are taken at the second stage. This once again allows for a similar approach to be employed to that in the deterministic case. After selecting a certain delivery profile assignment each period can be evaluated using the packing heuristic described previously for each scenario. This increases the overall solution time by the factor $|Z|$, but as this is a linear growth it may be acceptable. The adapted local search algorithm then reads as summarized in Algorithm 3. The genetic algorithm given in Section 5.5.2 can also be used for the stochastic case, if some minor extensions are made. To adapt to the new objective function the fitness function has to be exchanged. In the deterministic case an evaluation of each separate period is conducted by the packing algorithm in order to retrieve a fitness function value. In the stochastic case the evaluation has to be extended over all scenarios and the expected value has to be computed. Therefore the runtime is increased linearly by the number of scenarios that are considered, but no other changes to the algorithm are required.

Algorithm 3: Modified solution procedure for the stochastic case.

```

foreach  $s \in S$  do
  foreach  $dp \in DP$  do
    foreach  $z \in Z$  do
      | Apply Preprocessing;
    end
  end
end
while !StopCondition do
  NextSolution  $\leftarrow$  Strategy.FindBestNeighbor(CurrentSolution);
  NextSolution.ObjectiveValue = 0;
  foreach  $t \in T$  do
    foreach  $z \in Z$  do
      | NextSolution.ObjectiveValue  $+= p_z \cdot \text{SolveSubProblem}(t, z)$ ;
    end
  end
  if NextSolution.ObjectiveValue < CurrentSolution.ObjectiveValue then
    | CurrentSolution  $\leftarrow$  NextSolution;
    | Fix supplier whose delivery profile was changed;
  end
  else
    | Relax all suppliers;
  end
end

```

5.6.4 A simplified model formulation

As the consideration of multiple scenarios at once increases the computational efforts necessary to solve the model, larger instances from practice cannot be solved efficiently in the stochastic case. Therefore a simplified model formulation that is less generic and has a lower complexity has been developed for the stochastic case. Multiple aspects have been removed from the model formulation in order to trim the formulation. Removing these aspects results in less precise results for the expected cost and limits the applicability to a specific kind of tariff system. This may be acceptable though, as a trade-off between solution quality and level of detail in the model has to be made. The removed aspects include the generic treatment of resources that allowed e.g. the consideration of carbon-dioxide emissions and incoming goods department personnel, and

the explicit modeling of load carriers that allowed the ordering of parts in quantities that do not completely fill a load carrier or to use alternative load carriers. In addition the rebate levels have been approximated by a more moderate approach further to reduce the complexity. Given these reductions, the model reads as follows:

Sets

$z \in Z$	Set of scenarios.
$t \in T$	Set of time periods.
$s \in S$	Set of suppliers connected to the consolidation center.
$dp \in DP$	Set of available delivery profiles.
$l \in L$	Set of load unit types. A load unit describes a load carrier completely filled with parts. Each part has a fixed load carrier assigned to it, and parts will always be ordered in multiples of the load carriers fill level.
$v \in V$	Set of vehicles.
$v \in V^F$	Set of filled vehicles.

Parameters

$C_{s,dp,z}^{Choice}$	Expected total cost in pre leg run, full load run and inventory holding cost of scenario z if delivery profile dp is assigned to supplier s .
$C_v^{Vehicle}$	Cost of the usage of vehicle v .
C_v^{Step}	Cost of one weight step in vehicle v .
$O_{l,dp,s,t,z}$	Quantity of load units of type l which have to be ordered in period t if delivery profile dp is chosen for supplier s and scenario z realizes.
U_l^V	Volume used by one load unit of type l .
U_l^W	Weight used by one load unit of type l .
UB_v^V	Upper bound on volume for vehicle v .
UB_v^W	Upper bound on weight for vehicle v .

LB_v^V	Lower bound on volume for vehicle v to be seen as filled.
LB_v^W	Lower bound on weight for vehicle v to be seen as filled.
W_{Step}	Step size for the partially filled vehicle.
P_z	Probability that scenario z realizes.

Decision Variables

$pc_{s,dp} \in \{0,1\}$	Decision if delivery profile dp is selected for supplier s . This is a first-stage decision.
$o_{l,t,v,z} \in N_0^+$	Quantity of ordered load units of type l in period t delivered in vehicle v if scenario z is realized.
$u_{t,v,z}^V \in R_0^+$	Volume usage in vehicle v in period t if scenario z is realized.
$u_{t,v,z}^W \in R_0^+$	Weight usage in vehicle v in period t if scenario z is realized.
$v_{t,v,z}^{Active} \in \{0,1\}$	Decision if vehicle v is used in period t if scenario z is realized.
$v_{t,v,z}^{FullWeight} \in \{0,1\}$	Decision if vehicle v is filled <i>by weight</i> in period t if scenario z is realized.
$v_{t,v,z}^{Step} \in N_0^+$	Rebate steps of vehicle v .

Model formulation

$$\text{Min} \quad \sum_{z \in Z} P_z \left(\sum_{dp \in DP, s \in S} C_{s,dp,z}^{Choice} \cdot pc_{s,dp} + \sum_{t \in T, v \in V} C_v^{Vehicle} \cdot v_{t,v,z}^{Active} + C_v^{Step} \cdot v_{t,v,z}^{Step} \right) \quad (5.6.12)$$

subject to

$$\sum_{dp \in DP} pc_{s,dp} = 1 \quad \forall s \in S \quad (5.6.13)$$

$$\sum_{s \in S} O_{l,dp,s,t,z} \cdot pc_{s,dp} = \sum_{v \in V} o_{l,t,v,z} \quad \forall t \in T, dp \in DP, l \in L, z \in Z \quad (5.6.14)$$

$$u_{t,v,z}^V = \sum_{l \in L} U_l^V \cdot o_{l,t,sc,v} \quad \forall t \in T, v \in V, z \in Z \quad (5.6.15)$$

$$u_{t,v,z}^W = \sum_{l \in L} U_l^W \cdot o_{l,t,sc,v} \quad \forall t \in T, v \in V, z \in Z \quad (5.6.16)$$

$$u_{t,v,z}^V \leq v_{t,sc,v}^{Active} \cdot UB^V \quad \forall t \in T, v \in V, z \in Z \quad (5.6.17)$$

$$u_{t,v,z}^W \leq v_{t,sc,v}^{Active} \cdot UB^W \quad \forall t \in T, v \in V, z \in Z \quad (5.6.18)$$

$$u_{t,v,z}^V \geq LB^V \cdot v_{t,v,z}^{Active} - v_{t,v,z}^{FullWeight} \cdot UB^V \quad \forall t \in T, v \in V, z \in Z \quad (5.6.19)$$

$$u_{t,v,z}^W \geq LB^W \cdot v_{t,v,z}^{Active} - (1 - v_{t,v,z}^{FullWeight}) \cdot UB^W \quad \forall t \in T, v \in V, z \in Z \quad (5.6.20)$$

$$v_{t,v,z}^{Step} \geq u_{t,v,z}^W \cdot \frac{1}{W_{Step}} \quad \forall t \in T, v \in V, z \in Z \quad (5.6.21)$$

In the objective function 5.6.12 the sum of delivery profile choice cost, fixed cost charges for the use of vehicles and cost for weight steps in the partially filled vehicle are summed up for each scenario. This value is summed up and multiplied with the respective scenarios' probability to obtain the expected total cost. Equation 5.6.13 ensures that each supplier has exactly one delivery profile assigned. In constraint set 5.6.14 it is ensured that each part ordered in accordance to the active delivery profile selection will be delivered in a main leg run vehicle. Vehicles use is computed according to the vehicle's load in constraint set 5.6.15 for the volume and constraint set 5.6.16 for the weight respectively. Constraint sets 5.6.17 and 5.6.18 limit the volume and weight in one vehicle according to the upper bounds for volume and weight. As a vehicle has to be completely filled before another vehicle can be started, constraint sets 5.6.19 and 5.6.20 ensure that each used vehicle is either filled by volume or by weight. If a vehicle is filled by weight, constraint set 5.6.19 will be deactivated because

the the variable $v_{t,v,z}^{FullWeight}$ will be set to 1 and the right hand side becomes negative, as LB^V is always smaller than UB^V . Lastly, constraint set 5.6.21 can be used to compute weight steps of linear size for the partially filled vehicle.

5.7 Scenario generation

The input scenarios used for the stochastic program are essential to the outcome of the model. They should represent the possible future as well as possible. As the future is hardly predictable finding good scenarios becomes a task which may be even more difficult than solving the model itself. A special difficulty is raised by the possibility of net dependent demands' occurring only in certain periods due to lot sizing effects. Traditional time-series based forecasting approaches therefore have poor results when applied to generate demand scenarios. In this work, two approaches will be presented that use different techniques to overcome this problem. One approach uses observed occurrence probabilities to modify the current demand forecast time-slice wise according to change probabilities. The other approach relies on an observed frequency of demand to model future outcomes.

In both cases structured data are required to derive the required probability distributions. As the collected data will be referred to in the following, its structure may be explained shortly. For each part an individual set of forecasts and realized demands has to be collected. A forecast has a period assigned in which it was created and consists of multiple part forecasts which contain forecast data for a single part. Each part forecast consists of multiple forecast entries. A forecast entry represents an expected demand for a certain period with a given quantity. If no demand is forecast for a specific period, so the demand quantity is zero, no forecast entry will be created for that period. In the following two approaches to creating scenarios based on historical data will be presented.

5.7.1 A forecast deviation oriented scenario generation approach

The underlying idea of the procedure presented in this section is that every possible outcome of the future is a deviation from the forecast that is available at the day of planning. thus it may be a valid approach to measuring occurrence probability of changes and using them to build deviation scenarios based on the forecast that is available at the day of planning. Following this assumption every change to a forecast

may be projected onto five basic events depicted in Figure 5.8 which affect demand entries within the forecast:

1. **No change.** A demand entry does not change.
2. **Increase / Decrease.** A demand entry's quantity increases or decreases. A demand entry has been made in the previous forecast, but its quantity is now increased or decreased by a certain amount. The demand entry's quantity does not fall to zero.
3. **Remove.** A demand entry is removed. A demand entry has been made beforehand, but is now removed, so its quantity falls to zero.
4. **Add.** A demand entry is added. No demand entry has been made for a certain period in advance, but it is now added.
5. **Shifting.** A demand entry is shifted forwards or backwards from one period to another. This may be seen as a removing one demand entry and adding the same entry in another period. It can be very hard if not impossible to distinguish between shifted demand entries and demand entries which are added and removed. Therefore this point will not be considered in the following as it can be modeled by the operations add and remove.

The occurrence probabilities of these events can now be estimated by comparing a set of forecasts to the realized demands or by comparing the forecasts with each other. These two comparisons would result in different statements. If forecasts are compared with the realized demands, the total changes occurring may be measured, but no change history will be available. If forecasts are compared with each other, it is possible to create a change history that allows not only the creation of scenarios for the demands that will finally be realized but also how the forecasts may change as time passes. Whereas this is not necessary to find demand scenarios for the given two-stage stochastic program, it can be useful for other purposes which consider the planning process in a rolling horizon environment.

An important aspect when using this concept is to consider that changes to the forecast may have other probability distributions depending on the distance between the period in which the forecast was created and the period in which a demand entry is situated. To account for this issue change probability distributions can be measured

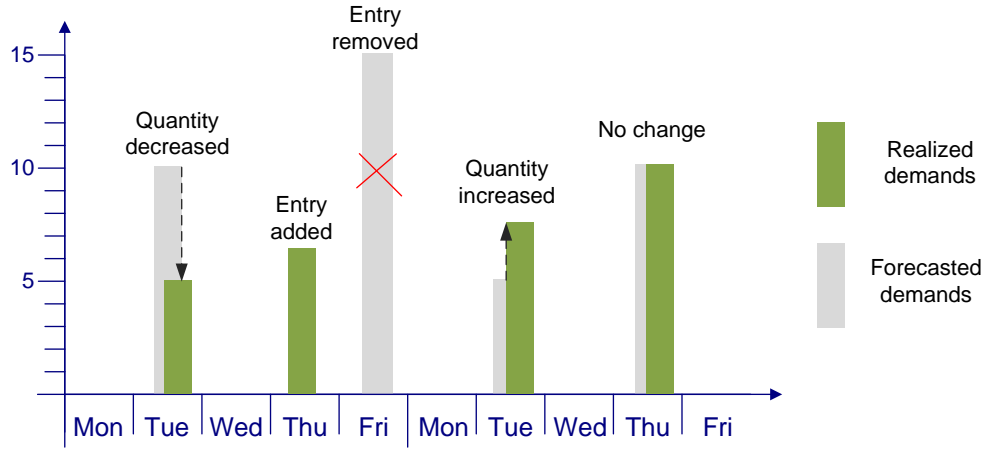


Figure 5.8: Possible changes to a demand forecast.

for different time-slices of the potential planning horizon each with a specific distance from the period in which the forecast was created. This information can then be used to create deviations for each time-slice individually based on the distance of the time-slice to the creation date of the forecast.

It is not sufficient to measure only the occurrence probability of a change event. This may hold true for a removed forecast entry, but when it comes down to an increase or decrease in quantity the amplitude of the change has also to be measured. After the change event distribution information has been gathered forecast scenarios may be created. A summary of the procedure is given in Figure 5.9. In a first step the current forecast is cloned. The planning horizon is then disjoint into multiple time-slices, e.g. into weeks. For each time-slice, multiple random time-slice scenarios will be created based on the given change event probability information. This step is called time-slice scenario generation. During this time-slice scenario generation each period is walked through individually. If there is a demand for the period, one out of four different deviation events may take place. The demand quantity may either be increased or decreased or set to zero, thus the demand will be removed, or no change may take place. Note that no change is also seen as an event. If there is no demand in the period, it may be that a demand is added or that no change takes place. To determine which event will be triggered a random number between zero and one is used and then mapped to the probabilities of the different events. After all periods within the

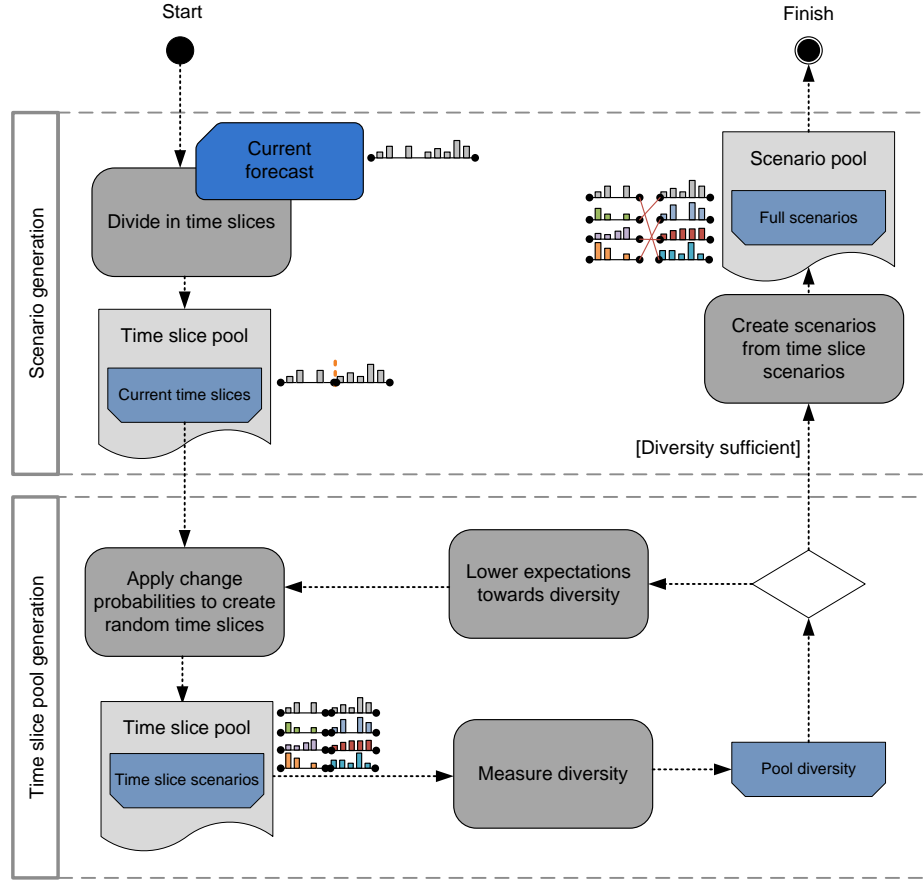


Figure 5.9: Overview of the forecast deviation oriented scenario generation procedure.

time-slice have been walked through it is added to a time-slice scenario pool. The time-slice scenario pool is then evaluated in terms of diversity. Diversity in this case means that the generated time-slice scenarios spread widely throughout the available scenario space. For this purpose different distance measures can be used. Therefore an interface allows plugging in its own distance measures if required. If the distance between the different time-slice scenarios in the pool is too low, additional time-slice scenarios may be created. To avoid a situation where no valid time-slice scenario pool is available due to too high expectations of the diversity of the pool a method inspired by the simulated annealing meta-heuristic is used. In simulated annealing the probability of getting stuck within a local optimum is reduced by introducing a probability

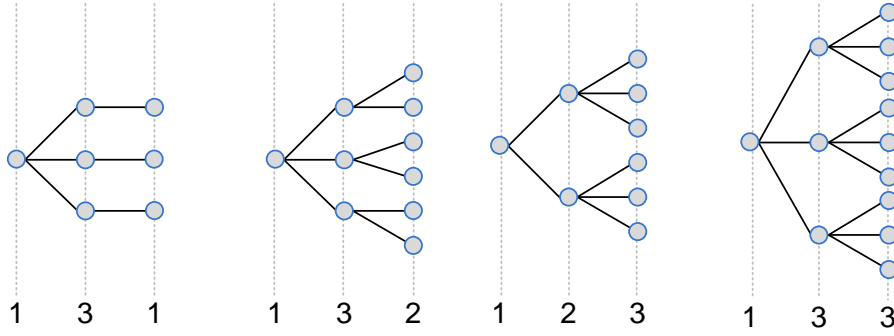


Figure 5.10: Different scenario trees resulting from multiple time-slice scenario pool sizes on each level.

to accept a worse solution. This probability is an analogue of the temperature in a metallurgic cooling process (see Glover and Kochenberger [2003], p. 288 for details). Whereas the probability is reduced over time in simulated annealing, which is referred to as cooling, it is necessary to increase the probability that a time-slice scenario pool will be accepted in this case. Therefore a threshold on the distance measure will be lowered if the pool is invalid until a valid pool is found. During this process the time-slice scenario with the lowest distance to the other time-slice scenarios in the pool will be selected to be thrown out of the pool.

After a valid time-slice scenario pool has been created the next time-slice will be considered and the procedure will be repeated until there is a time-slice scenario pool for all time-slices within the forecast. The size of the pool can be adapted individually for each time-slice. This allows the creation of scenario trees which fit the actual purpose, as depicted in Figure 5.10. When all time-slice scenario pools are filled scenarios have to be constructed out of them through combination. A full combination of all time-slice scenarios from each scenario pool $n \in N$ would lead to a number of

$$|S| = \prod_{n \in N} \text{Poolsize}_n$$

scenarios. Therefore a scenario is constructed by picking a random candidate out of each time-slice scenario pool until the desired number of scenarios is reached.

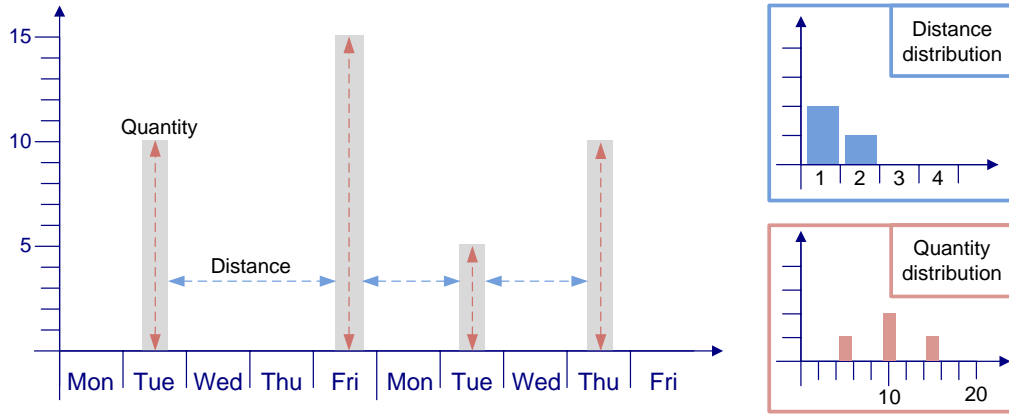


Figure 5.11: Demand entry quantity and distance between two demand entries as derived properties of the demand situation.

5.7.2 A demand distribution oriented scenario generation approach

Unlike the previously described approach the approach presented in the following does not consider a demand scenario to be a deviation from the forecast available on the day of planning, but rather considers it to be an independent demand realization that occurs by chance. Therefore not the changes between forecasts and realized demands but the demands properties will be used as underlying information. Properties of the demand pattern can be measured by two aspects that have been shown to be important for delivery profile selection: the demand entry's quantity and the distance between two demand entries. Figure 5.11 depicts these two measures and their distributions which are collected for scenario generation. In combination the two measures represent the demand pattern of a part and allow us to model demand scenarios for different parts. A part that is used regularly with a low fluctuation in quantity, e.g. tires or screws, will have a low distance between demands and a quantity distribution that deviates closely around an average value. On the other hand, a part that will flow through a lot-production environment before it finally ends up in an assembly line may have a higher distance between demand entries and may underly a high fluctuation in quantity. A summary of the scenario generation procedure is given in Algorithm 4. In a first step an empty scenario is created. A period is then selected randomly in between the first period and a period within the half of the average distance for the current

part. This will reflect the probability that an entry has already been placed before the planning horizon. Alternatively, the last order placed in the MRP system could be used as a starting point for the following steps. When the period has been selected, a demand with a random quantity according to the given quantity distribution will be placed in that period. Afterwards a next period is selected by shifting forward for a random number of period according to the distribution of distances between demands. The procedure is then repeated until the end of the planning horizon has been met.

Algorithm 4: Algorithm to generate scenarios based on demand patterns.

```

 $t_{\text{Current}} \leftarrow t_0 + \text{Random}(0, \frac{\mu_{\text{Distance}}}{2});$ 
while  $t_{\text{Current}} < t_{\text{Max}}$  do
     $d_t \leftarrow \text{NormalDistribution}(\mu_{\text{Quantity}}, \sigma_{\text{Quantity}});$ 
     $t_{\text{Current}} \leftarrow t_{\text{Current}} + \text{NormalDistribution}(\mu_{\text{Distance}}, \sigma_{\text{Distance}});$ 
end

```

5.7.3 Scenario reduction

Either method may lead to a tremendous number of scenarios which cannot be handled at once due to computational problems. Therefore scenario-reduction techniques have been deployed. These allow selection of a subset from a set of scenarios such that the scenarios in the subset form a distribution which is as close as possible to the original scenario distribution. Given a distribution P represented by scenario set S , whereof each scenario $s \in S$ has the probability p_s , the task is to find a new distribution Q with scenario set $Z = S \setminus J$ with probabilities q_z for each scenario $z \in Z$, such that the distance $D(P, Q)$ is as low as possible. Therefore a set of scenarios $J \subset S$ should be removed from S such that $D(P, Q) = \sum_{j \in J} p_j \min_{z \in Z} d(j, z)$ with $d(j, z)$ being the distance between scenario j and scenario z is minimized. Identifying scenario set J^* to be removed from the original scenario set and defining new probabilities q_z for all remaining scenarios is the task of scenario reduction algorithms. Two promising heuristic algorithms for this purpose, the *simultaneous backward reduction* algorithm and the *fast forward reduction* algorithm, have been developed in Dupacová et al. [2003] and Heitsch and Römisch [2003]. The basic idea behind those approaches is that the

optimal distribution Q^* is given by

$$q_z^* = p_z + \sum_{j \in J_z} p_j \quad \forall z \in Z$$

where J_z describes the scenarios $j \in J$ which are closer to z than to any other remaining scenario according to the distance function $d(j, z)$. Both of the scenario reduction techniques presented in the following have been integrated into the scenario generation process. First, a large set of possible scenarios is generated by one of the methods presented in section 5.7.1 and section 5.7.2. The scenarios are then reduced using one of the given scenario-reduction techniques to limit the number of scenarios such that it can be handled by the optimization algorithm.

Fast forward reduction starts with no scenarios selected, thus $J^0 = S$ and $Z = \emptyset$. It can be formalized as in algorithm 5. Before going into iterations the scenario $u^* \in J$ with the lowest sum of weighted distances to all other scenarios $k \in J$ is moved to Z . Afterwards, in each iteration a scenario j is moved from J to Z until the desired number of scenarios has been selected. The scenario j to be moved to the set of remaining scenarios is the one which represents the set of scenarios to delete at best. A representation factor $rf_{k,u}$ is computed for all scenarios $u \in J$ to determine how well the scenario $u \in J$ would represent $k \in J$ if it were moved from J to the set of remaining scenarios Z . The cumulative representation factor rf_u being the sum of $rf_{u,k}$ over all $k \in J$ is then computed and used as selection criterion. The lower rf_u , the better the scenario u represents the remaining scenarios to delete J if it is switched over to Z . Therefore, the scenario u^* with the minimum value of rf_u is selected and moved from J to Z . These steps are repeated until the desired number of scenarios has been selected. Thereafter, for each removed scenario $j \in J$, it is identified which remaining scenario $k^* \in Z$ has the lowest distance to scenario $j \in J$. The probability of the remaining scenario $k^* \in Z$ is then increased by the probability of the scenario to be removed.

Simultaneous backward reduction goes the opposite way. It starts with all scenarios selected, thus $J^0 = \emptyset$ and adds the scenario with the lowest distance to all other scenarios to J in each iteration. The idea behind this is to remove the scenario that can be represented best by the remaining scenarios. In analogy to the fast forward

Algorithm 5: Fast forward reduction algorithm. Notation is adopted from Heitsch and Römisch [2003].

```
 $J \leftarrow S ;$ 
 $Z \leftarrow \emptyset ;$ 
 $u^* \leftarrow \min_{u \in J} (\text{sum}_{k \in J \setminus \{u\}} p_k \cdot d(k, u));$ 
 $J \leftarrow J \setminus u^*;$ 
 $Z \leftarrow Z \cup u^*;$ 
while  $|J| > |J^*|$  do
  foreach  $u \in J$  do
    foreach  $k \in J \setminus \{u\}$  do
       $rf_{k,u} \leftarrow p_k \cdot \min_{z \in Z \cup \{u\}} (d(k, z));$ 
    end
     $rf_u \leftarrow \sum_{k \in J \setminus \{u\}} rf_{k,u};$ 
  end
   $u^* \leftarrow \operatorname{argmin}_{u \in J} (rf_u);$ 
   $J \leftarrow J \setminus u^*;$ 
   $Z \leftarrow Z \cup u^*;$ 
end
foreach  $j \in J$  do
   $k^* \leftarrow \operatorname{argmin}_{k \in J} (rf_{k,j});$ 
   $p_{k^*} \leftarrow p_{k^*} + p_j;$ 
end
```

algorithm the representation factor rf_u is used to select the scenario to be removed. The obvious difference in its use is that in the backward reduction algorithm the scenario u^* which is represented best by the remaining scenarios Z is to be removed, therefore being added to J . After the selection algorithm the same probability distribution rule is applied again to redistribute the probabilities of the removed scenarios onto the remaining scenarios.

Algorithm 6: Simultaneous backward reduction algorithm. Notation is adopted from Heitsch and Römisich [2003].

```

 $J \leftarrow \emptyset$  ;
 $Z \leftarrow S$  ;
 $u^* \leftarrow \min_{u \in Z} p_u \cdot \min_{k \in Z \setminus u} (d(k, u))$ ;
 $J \leftarrow J \cup u^*$ ;
 $Z \leftarrow Z \setminus u^*$ ;
while  $|J| < |J^*|$  do
    foreach  $u \in Z$  do
        foreach  $k \in J \cup \{u\}$  do
             $rf_{k,u} \leftarrow p_k \cdot \min_{z \in Z \cup \{u\}} (d(k, z))$ ;
        end
         $rf_u \leftarrow \sum_{k \in J \cup \{u\}} rf_{k,u}$ ;
    end
     $u^* \leftarrow \operatorname{argmin}_{u \in J} (rf_u)$ ;
     $J \leftarrow J \cup u^*$ ;
     $Z \leftarrow Z \setminus u^*$ ;
end
foreach  $j \in J$  do
     $k^* \leftarrow \operatorname{argmin}_{k \in J} (rf_{k,j})$ ;
     $p_{k^*} \leftarrow p_{k^*} + p_j$ ;
end

```

6 An Evaluation Framework for Delivery Schedule Generation Approaches

One goal of this thesis is to assess the impact of delivery profiles that have been selected by the planning approach presented in Section 5 in respect of their outcome in a rolling horizon application scenario under special consideration of the given demand forecast uncertainty. The delivery profiles as a delivery schedule generation approach should not only be analyzed in respect of their individual outcome, but also be compared with the MRP system's behavior and a state-of-the-art algorithmic delivery schedule generation approach under fair conditions. As was seen in the literature review (see section 3.2 for details), a simulation approach seems to be the most promising technique to be employed in order to analyze the delivery schedule generation approaches' behavior in a rolling horizon environment. Hence a simulation-based evaluation framework was developed in order to examine the outcome of the different delivery schedule generation approaches in a rolling horizon environment under consideration of demand forecast uncertainty. To evaluate the outcome of the simulated behavior it is necessary to provide a set of key figures that reflect the objectives of the problem setting. These performance indicators can then be used to compare the different delivery schedule generation approaches. In addition to the simulation approach itself, a set of performance indicators to quantify the realized cost and the stability of the delivery schedules was developed and included into the evaluation framework. The remainder of this chapter will at first depict the simulation approach itself and in so doing describe the simulation process and the underlying architecture. The selected performance indicators and their computation will then be sketched subsequently.

6.1 Simulation Approach

The primary goal of the simulation approach is to examine the outcome of different delivery schedule generation approaches for the operational order lot sizing problem in a rolling horizon application. As noted in Section 2.2.2 the operational order lot-

sizing problem is integrated into a higher level planning process. Being only a part of a whole process decisions made at this decision stage are influenced by inputs from previous stages and act as outputs for following stages at the same time. This means in turn that a subset of the planning process described in Section 2.2.2 is of relevance for the considered problem setting. This leads to the necessity to divide the planning steps into those which should be simulated and those whose results should be seen as input for the simulation. In this work the line is drawn after the computation of gross dependent demands has finished, which means that the gross dependent demands act as input for the simulation process. This choice allows us to cut off the master production scheduling planning problem as well as the bill of materials explosion. The former can especially be heavily influenced by manual intervention or other planning algorithms, leading to an additional source of freedom. To account for these aspects multiple assumptions would have to be made on the behavior of both the planning system and the planners. By laying the cut after these steps we do not have to differentiate between self-made planning uncertainty created by manual intervention or a nervous master production scheduling on the one hand and demand uncertainty on the other. Therefore uncertainty is considered to reveal itself purely in form of deviation in gross dependent demand predictions varying from one planning cycle to another. Whereas it makes sense to cut off these planning steps, it is important to take the net dependent demand computation based on the safety inventory parameters into consideration. Delivery schedule generation can have a significant impact on inventory levels. Especially if investments in cycle stock are made in order to secure benefits in respect of freight cost, stock levels can undergo heavy fluctuations. This in turn can lead to completely different behavior when determining the net dependent demands which are thereafter being used to create delivery schedules. Not to consider this bidirectional influence would leave strengths and weaknesses of the different approaches in respect to inventory management undiscovered. Another factor of uncertainty that might have been considered is the unreliability of supply. It may be that suppliers cannot deliver goods even though orders have been placed in time, e.g. due to machine breakdown or traffic jam. But, as this work is especially focused on demand forecast uncertainty, supply unreliability has been completely disregarded. This decision has been made to neglect interferences between the two sources of uncertainty so that it can be identified more precisely how the different planning approaches tackle the demand forecast

uncertainty. Hence the gross dependent demand prediction seems to be the most valid interface between the simulation environment and the reality, as it does not require the simulation to be based mainly on assumed data, and at the same time it allows us to step deep enough into the planning process to reveal the impact of a certain planning method on the system's overall behavior. According to Mönch [2007] simulation environments for benchmarking of planning methods consist of two main parts, a base system and a control system. The base system models the considered environment and the natural flow within the environment. The control system gathers information from the base system and incorporates this information to make decisions. These decisions then influence the happenings in the base system. When this concept is adapted to the given problem setting, the base system covers the transportation network, the considered plants inventory and the demand forecasts. The control system covers the MRP system's components necessary to control the flow of goods. The control system can be seen as bipartite, with one part consisting of components which reflect MRP system logic for inventory balancing and net dependent demand calculation, whereas the generation of delivery schedules in particular is carried out by the examined delivery schedule generation approach. The MRP system logic part behaves predictably and is necessary to trigger the delivery schedule generation approach and process information from the base system, but is not directly an object of investigation in this thesis. Section 6.1.2 describes the base system and its components. A detailed description of the MRP system logic is given in Section 6.1.3. The delivery schedule generation approaches employed to create delivery schedules may consider various aspects from the problem setting. Therefore it was identified which parts of both the base system and the MRP system logic state must be submitted to the planning approaches due to their possible influence on the decision-making. To provide the possibility to adapt multiple delivery schedule generation approaches for an examination an interface was developed that covers all properties of the problem setting which may be relevant to the different delivery schedule generation approaches. The interface description can be found in Section 6.1.4.

6.1.1 Representation of time

Due to the course-grained nature of the planning process time is modeled as discrete time steps with a size of one day each. As production facilities do not necessarily oper-

ate twenty-four-seven, there can be days on which facilities are closed and no operations are possible. To account for this fact most facilities employ a factory calendar. This calendar can be used to determine the set of days to consider. In the following a time step will be described as a period. In industry applications delivery schedules may be updated less frequently than in each period. It may be, for example, that the delivery schedule is computed and sent out only once or twice a week. To account for this aspect planning periods as a subset of all periods have been introduced. Whereas the state of the base system can change from period to period, response to these changes can only take place in planning periods. In the remainder the timespan between two planning periods will be described as a planning cycle.

6.1.2 Representation of the underlying base system

The base system consists of its entities and their properties on the one hand and the state of these entities on the other hand. The former part remains unchanged over the whole time horizon whereas the latter changes as time passes. To give a brief understanding of the aspects that were considered in the base system, its entities will first be described. It will then be pointed out what information about the state of the base system can be derived.

Entities in the base system

The partition of the supply network considered in the base system begins at the suppliers outgoing goods department and ends at the warehouse in the incoming goods department of the unloader. Within this partition parts which are packed in load carriers are transported by vehicles along the predefined routes for pre leg, main leg and full load runs are finally stored in the warehouses from where they are taken to satisfy production demands. Given this short summary, the entities within the base system can be divided into those describing the network structure and those that are transported within the network. Both types of entity are to be persisted, e.g. in a relational database or a file based format, so that they can be recalled whenever necessary. The entities in the base system are modeled according to the structure depicted in Figure 6.1 and will shortly be described in the following. An area forwarding network is modeled to consist of a consolidation center to which multiple suppliers are connected. Each supplier is identified by its supplier code and has a set of offered parts and a pre

leg run and a full load run relation assigned. Warehouses with a capacity limit and the capability to store certain items are included to model the inventory within the unloader's plant. The consolidation center has a main leg run relation assigned to it. Pre leg, full load and main leg runs relations each have a tariff structure assigned to it. The tariff structure defines the used vehicle and one or multiple discount levels as well as a variable price per weight or volume unit. Each discount level has lower and upper bounds on weight and volume as well as a fixed price. The vehicles available for use have a capacity given by weight and volume. To account for the fact that parts are packed into load carriers the item entity was introduced. An item represents the combination of a part with a load carrier and a supplier and has a maximum fill level associated with it. Items from different suppliers have to be distinguished as each supplier may have its own packing regulations. A part has a weight and a price associated with it and can be identified via its part code. The load carrier has both weight and volume associated and is identified by its load carrier code. In addition to the entities described above safety parameter configurations have been included into the master data. These parameters strongly influence the MRP system's behavior during the simulations phase. It was therefore decided to allow a configuration of these parameters from the outside. Multiple parameter scenario sets can be persisted and can later be selected for the simulation runs. This allows for what-if-analysis considering different parameter settings and their influence on the different planning approaches or the overall system.

States of the base system

The state of the base system in a certain period is defined by the past demand situation, requested and fulfilled orders and the resulting stock level on the one hand and the demand forecast information available in the period on the other. The past demand situation and the fulfilled orders determine the current state of inventory. This information on the state of the base system is collected during the simulation run as it allows drawing of conclusions on the outcome of the planning approaches. From the stock level multiple performance indicators may be computed, e.g. the service level towards production or inventory cost. From the fulfilled orders in a certain period it can be determined what load has been delivered and therefore the freight cost in that period can be computed. Whereas the stock level and the fulfilled orders represent the

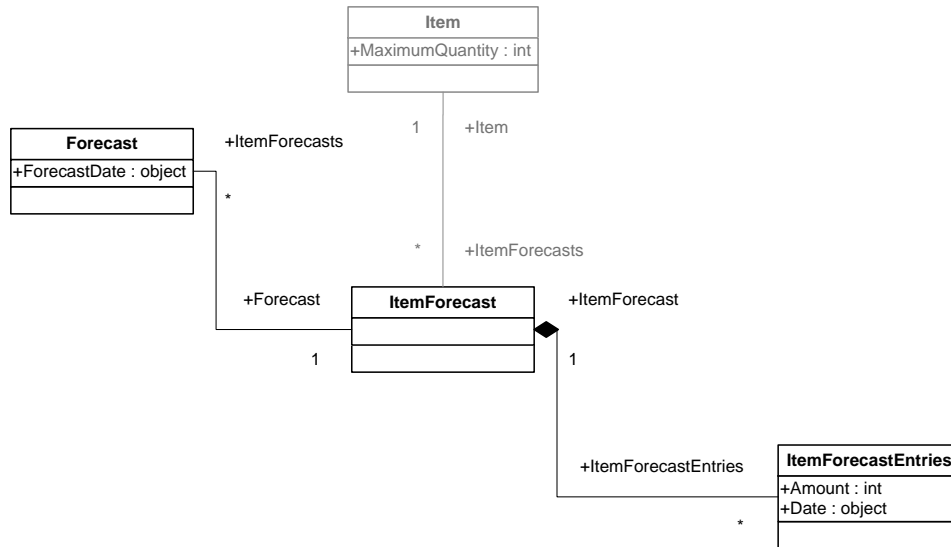


Figure 6.2: Entities of the forecast data, the plant transaction data.

6.1.3 MRP system logic

The MRP system logic derives necessary information from the base system and prepares them for use in the delivery schedule generation approach. Thereafter the results are transformed and send out virtually to the suppliers. The MRP system logic is activated in each period and executes two basic functions. As in a real MRP system a balance of incoming and outgoing materials is created to compute the stock level in the inventory management system. Thereafter a net demand computation takes place. In this process the stock level, requested orders and the expected future demands are combined to identify which quantities are missing in which period to fulfill the expected future demands. The result of the net demand computation is then handed over to the delivery schedule generation approach. The approach is then used to create a delivery schedule. Afterwards the created delivery schedule is again interpreted by the MRP logic. Especially when deriving the net demands from the state of the base system, multiple parameters have to be considered. These parameters can also be found in

typical MRP systems and can have a significant influence on the results of the net demand computation. These parameters will be briefly explained in the following.

- **Lot Ceiling** allows forcing net demand quantities to be ceiled to a certain value. When a ceiling takes place the difference between the initially required amount and the ceiled value is charged off against following demand entries. It allows a sort of small-scale lot-sizing for each item individually. This parameter is often used to assure that only filled load carriers are ordered.
- **Safety lead time** is a parameter to determine safety inventory. The parameter safety lead time gives back the number of periods a part should be ordered earlier than it was originally demanded for. If e.g. the safety lead time has a value of two, each gross demand will be shifted two days towards the planning period in the net demand calculation. In doing so a safety buffer is built up dynamically, as the stock level follows the demand level. On the other hand, orders are fixed earlier than they would have to be and thus are created based on an earlier gross demand forecast.
- **Safety stock quantity** is another parameter to determine safety inventory. The parameter safety stock quantity determines the level of the classical safety stock. The net demand computation will try to assure that the stock level never falls below this threshold value. If it does or is predicted to do so, a replenishment order will be triggered immediately.

6.1.4 Interface between MRP system logic and delivery schedule generation approaches

The delivery schedule generation approaches base their decision-making upon the state of the base system and its general properties. Whereas the state of the base system changes as time passes, the general properties remain unaltered. Therefore it was decided to include three bundles of information as depicted in Figure 6.3. The first bundle consists of the general properties of the base system, including information on the freight network, its tariff structures, part and load carrier data as well as information on warehousing possibilities and related cost. It is passed over from the simulation approach to the delivery schedule generation approach in a first step, the initialization. The second bundle contains the state of the base system which is defined by the actual

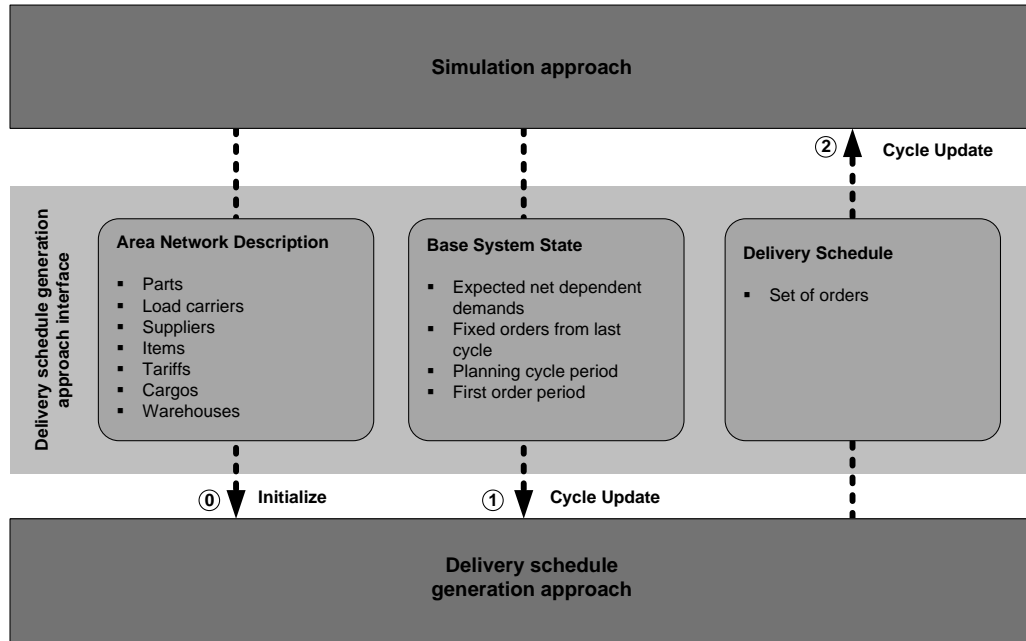


Figure 6.3: Interface between the simulation approach and the delivery schedule generation approaches.

net demand situation and previously placed orders that have already been fixed and hence must be considered but not modified when creating a delivery schedule. As mentioned above, the general properties of the base system remain unaltered and thus need be transmitted only once to the delivery schedule generation approach. By contrast, the state of the base system has to be transferred each time a delivery schedule has to be created. Therefore it is repeatedly passed to the delivery schedule generation approach in each planning cycle. Lastly, the third bundle contains the delivery schedule that has been generated. It is passed back from the delivery schedule generation approach to the simulation approach after each repetition of a planning cycle.

6.1.5 Simulation procedure

The simulation procedure consists of three phases, namely initialization phase, core simulation phase and abandonment phase. In the initialization phase the base system is brought into an initial state. During the core simulation phase planning cycles are iteratively walked through in interaction between the base system and the control

system. The information used to evaluate the different delivery schedule generation approaches is collected during this phase. In the abandonment phase the base system fades out without getting additional input from the control system.

Initialization phase

The initialization phase initiates the state variables for inventory levels, forecasted demand and requested orders. It is thereby assumed that the demands within the replenishment lead time on the period of the first planning cycle have been forecast perfectly during the previous cycles. Even though this is unrealistic, it provides a deterministic startup behavior that does not favor one method over another. Depending on the safety parameter settings and the given initial demands, an inventory level is computed that reflects the desired inventory in the first period. All demands that are forecast within the first forecast and should already be in inventory according to the given safety parameters are summed up and form the initial inventory value. The demands that should have triggered an order that would arrive during the replenishment lead time cause fixed orders to be added with an arrival time within the replenishment lead time.

The core simulation phase

In the following a single planning cycle within the core simulation phase will be discussed in detail. An overview of the core simulation procedure is given in Figure 6.4. The planning cycle starts with the release of a gross dependent demand forecast in the base system. The forecast will then be interpreted by the inventory management system, which belongs to the MRP system logic part of the control system. At first, an update of the current stock level in the base system is computed. This is done by charging up inventory from the last planning period $s_{p,k}^R$ and fixed orders $o_{p,t}^R$ in between both planning periods k and k' against the realized demands $d_{p,t}^R$ during this period range. For each period the stock level $s_{p,t}^R$ for a part p in period t is computed as follows:

$$s_{p,t}^R = s_{p,t-1} + o_{p,t}^R - d_{p,t}^R \quad \forall p \in P, t = k \dots k'$$

Net dependent demands are then derived by offsetting the gross dependent demands against current stock level and fixed orders from the last planning cycle. Fixed orders

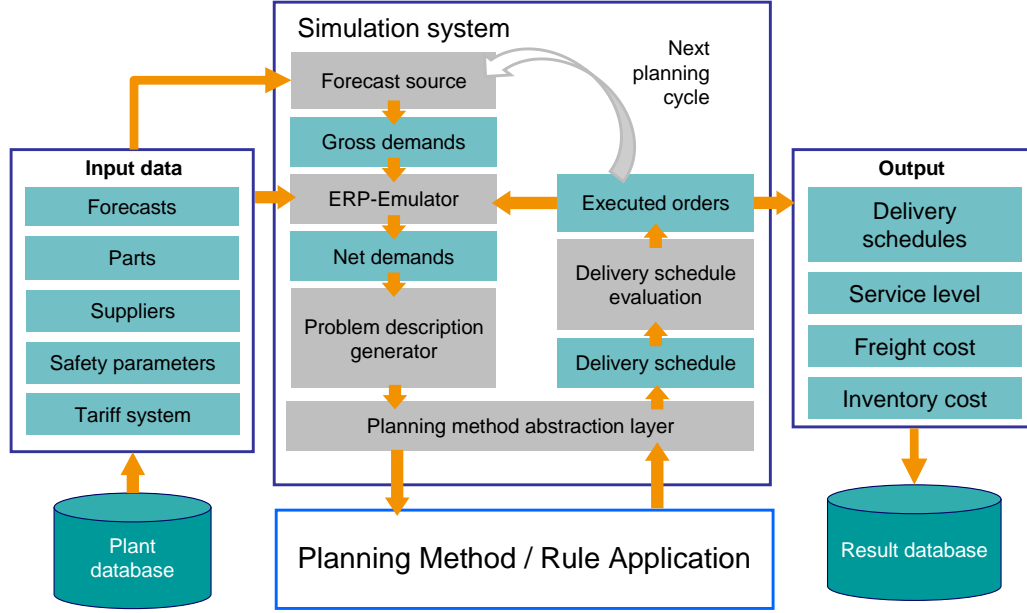


Figure 6.4: Overview of the simulation procedure.

are the set of orders which have their requested arrival period after or equal to the current planning period but before the end of the replenishment lead time. Other orders do not have to be considered as they can still be altered or deleted. When offsetting gross dependent demands against future stock and fixed orders, each period is treated sequentially as shown in Algorithm 7. At first, an expected stock level $s_{p,t}^E$ is computed by charging up planned orders $o_{p,t}^P$ against forecasted demands $d_{p,t}^F$. Then, by applying the safety stock quantity SSQ_p and safety lead time SLT_p parameters to the forecasted demands, the desired stock level $s_{p,t}^D$ is calculated. The difference between the expected stock level and the desired stock level is the net dependent demand $d_{p,t}^N$. If the lot ceiling option is activated, the net dependent demands $d_{p,t}^F$ are ceiled so that a multiple of the given lot L is acquired. In doing so the occurring excess $e_{p,t}$ in one period will be subtracted from the following net demand before it is ceiled. Multiple net demands may thus be drawn together. If there is an excess at the end of the delivery schedule and no further demands are forecast, the excess will be ceiled and ordered on the end of the planning horizon. The resulting net dependent demands are thereafter transmitted to the employed delivery schedule generation approach using the predefined interface. The delivery schedule generation approach then provides a delivery schedule

Algorithm 7: Calculation of net dependent demands including lot ceiling, safety stock quantity and safety lead time.

```

foreach  $p \in P$  do
  foreach  $t \in k' \dots k$  do
     $s_{p,t}^R \leftarrow s_{p,t-1} + o_{p,t}^R - d_{p,t}^R;$ 
  end
  foreach  $t \in k + RLT_p \dots T$  do
     $s_{p,t}^E \leftarrow s_{p,t-1}^E + o_{p,t-1} - d_{p,t-1};$ 
     $s_{p,t}^D \leftarrow SSQ_p + \sum_{t'=t}^{t+SLT_p} d_{p,t'};$ 
     $d_{p,t}^N \leftarrow s_{p,t}^D - s_{p,t}^E;$ 
    if Lotceiling then
       $e_{p,t} \leftarrow \left\lceil \frac{d_{p,t}^N - e_{p,t-1}}{L} \right\rceil \cdot L - d_{p,t}^N;$ 
       $d_{p,t}^N \leftarrow \left\lceil \frac{d_{p,t}^N - e_{p,t-1}}{L} \right\rceil \cdot L;$ 
    end
  end
end

```

which will be interpreted in the next planning cycle. This delivery schedule is evaluated and the orders which should be executed are derived from it. Thereafter the control system gives back control to the base system and awaits the next planning cycle to be initiated by the release of a gross dependent demand forecast.

Treatment of escalation processes

It may be that during the net dependent demand computation described above, an order is placed within the supplier's replenishment lead time. In practice such a situation may be resolved using various alternatives, usually defined in an escalation process to be launched in case of an apprehended future stock-out. Unfortunately the outcome of this escalation process is not as clearly defined as the process itself. It depends strongly on external factors, e.g. the production scheduling at the supplier's facility, the importance of the buyer as a customer to the supplier and thus the supplier's will to deliver, despite the desired delivery time's conflicting with his replenishment lead time, among other factors. As those escalation processes are likely to appear in everyday business, these had to be considered somehow. But as it is not possible to predict their

outcome in advance, it was avoided to assume a possible outcome. In this simulation procedure, escalation processes will always lead to a delivery on the first period after the replenishment lead time. In so doing, the worst case scenario is reflected. Using this behavior does not prefer one delivery schedule generation approach to another, but rather added another evaluation criteria, as it can be measured which delivery schedule generation approach tends to produce how many escalation processes.

The abandonment phase

The abandonment phase is included to avoid censored data. The term censored data refers to the effect that the state at end of the simulation horizon results from the perspective of ongoing operations, whereas these operations will not be considered in the simulation itself. Therefore it may be that a delivery schedule generation approach makes investments within the simulation period that aim to reduce cost in periods after the end of the simulation. Due to this issue it may be that one approach seems preferable to another even if this would not be the case in a real rolling horizon application (see Blackstone et al. [1982]). Awareness of this issue demands the abandonment phase, in which the base system slowly fades out without further interaction with the delivery schedule generation approaches. During this phase cycle inventory is slowly released but not built up any more. It may be that even with an abandonment phase, some parts have been ordered due to a misleading forecast and have not been taken out of the warehouse until the end of the simulation horizon. There is a high probability that these parts will cause additional inventory cost in the time after the simulation horizon or even have to be scrapped. Therefore it was decided to account scrap risk cost for the excess inventory at the end of the abandonment phase. A parts stock is considered to be excess inventory if at the end of the abandonment phase the forecast demand is not sufficient to consume the available stock of the part.

6.2 Performance indicators

The research questions to answer, given the selected performance criteria, are twofold, as described in section 2.4. First, it has to be answered whether the proposed cost advantage holds true in a rolling horizon planning environment. Second, it is of interest if the stability of the generated delivery schedules can be improved when deploying delivery profiles. To answer the first question the costs that are realized during the

simulation have to be assessed. How this is done will be described in the first part of this section. In the second part it will be described how the stability of the generated delivery schedules can be assessed and a new measure to do so will be introduced.

6.2.1 Assessing the realized cost

A planning approach uses a cost function $c(x)$ to determine the cost associated with a solution x . This cost function is used to decide whether a solution is better than another. The value of the cost function is the objective value of the initial planning solution. These costs will be referred to as *expected cost*. When applied in a rolling horizon environment uncertainty comes into play and it may be that the *realized cost* differs from the expected cost. Therefore we distinguish between *expected costs* and *realized costs*. The *realized costs* are the ones that really matter in an industrial application. They can be determined after the simulation process by computing freight cost for each realized order and inventory holding cost for the materials in inventory based on the base system state information gathered during the simulation procedure. Assessing the realized cost means assessing the economic outcome of the employed delivery schedule generation approach in a rolling horizon application. The realized cost C consists of two main components, freight cost $C^{Freight}$ and inventory holding cost $C^{Inventory}$.

$$C = C^{Freight} + C^{Inventory}$$

The freight cost part can be further divided into pre leg run cost C^{Preleg} , full load run cost $C^{Fullload}$ and main leg run cost $C^{Mainleg}$. To compute these cost factors it is necessary to make an ex post analysis of the orders that have been realized during the simulation run. For each period t considered during the simulation a determination of the freight cost associated with the realized orders takes place. Unfortunately these freight values cannot be inferred directly from the order situation in all cases. If the order's weight or volume exceeds one vehicle's capacity it is necessary to determine which share of the goods will be transported in which vehicle. Otherwise the correct discount level cannot be identified. The assignment of load to vehicles is not deterministic as it depends on the LSP's behavior and preferences. Therefore an assumption has to be made on how the goods will be distributed on the different vehicles. One viable approach is to imply that the cheapest possible distribution is always chosen. In this case the mathematical model described in section 5.3.3 can be used to obtain

the freight cost values. The inventory holding cost $C^{Inventory}$ consists of two parts, the cost for warehouse slot usage C^{Slot} and cost for interest on capital commitment $C^{Interest}$. The latter can be computed for each period by multiplying the stock level $s_{p,t}$ of part p in period t with its price p_p and the interest rate divided by the number of periods per year.

$$C^{Interest} = \sum_{p \in P, t \in T} s_{p,t} \cdot p_p \cdot \frac{\text{Interest rate}}{\text{Periods per year}}$$

The cost for warehouse slot usage can be computed for each period by multiplying the number of load carriers lc in stock by the slot cost per year C_{lc}^{Slot} for the specific load carrier lc divided by the amount of periods per year, so the formula reads as

$$C^{Slot} = \sum_{p \in P, t \in T} \left\lceil \frac{s_{p,t}}{Q_{p,lc}} \right\rceil \cdot C_{lc}^{Slot}$$

where $Q_{p,lc}$ is the number of parts of type p that fit into a load carrier of type lc .

Comparison of the realized cost with the planned costs can be used to generate two insights. First, it can be seen how much the cost could be reduced if everything goes as expected in the beginning. Second, the difference between the planned cost determined in advance and the realized cost can be analyzed. In addition to planned and realized cost, it can be determined how a perfect solution would look under post ex conditions. This value will be described as *post-ex-solution value* in the following. The difference between the post ex solution value and the realized cost will be called the *Value of Perfect Information* (VPI). It is the maximum value a planner would pay to get perfect information about the future. The term is inspired by the Expected Value of Perfect Information (EVPI) from the stochastic programming literature (see e.g. Avriel and Williams [1970]), where it describes the difference between a wait-and-see solution and a recourse program solution.

6.2.2 Assessing the stability of the generated delivery schedules

To examine the stability of the delivery schedules produced by the different planning approaches a sequence of delivery schedules generated during the simulation phase has to be analyzed. It has thus to be checked how strongly two consecutive delivery schedules differ, and to which degree these differences may have a negative effects on

the supplier's delivery performance and production planning efficiency. To measure the different variations between the delivery schedules each generated delivery schedule is compared with the previously generated one and the measures are computed. After all delivery schedules have been compared with their predecessor, averages and standard deviations of the measures are calculated. In addition to the quantity increases and decreases, time shifting and additional as well as removed orders have to be considered. As it is hardly possible to determine whether an order has been added or shifted in time, it is almost impossible to compute a measure which reflects changes in time. A sequence of delivery schedules can suffer from different sources of instability. First, the quantity of an order can be increased or decreased from one schedule to another. This will be referred to as variation in quantity. In this case the first delivery schedule contains an order of q_1 product units and the next delivery schedule contains an order on the same day with another quantity $q_2 \neq q_1$. Second, there may be a variation in time. An order can be placed on day t_1 in the first schedule and then be moved to another day $t_2 \neq t_1$. Both cases can have a negative impact on the production plan of the supplier. In addition, orders can be added to the delivery schedule or removed from the delivery schedule. Not all variations have the same influence on the suppliers planning capabilities. Changes following shortly after the day the delivery schedule has been sent out will probably require the supplier to start a major short-term rescheduling, while a change in four to twelve weeks from the day the delivery schedule has been sent out will probably not have greatly affect the supplier. Therefore the focus of the new developed measures lies on the short-term variance within the delivery schedules. A period of three weeks or fifteen working days beginning after the replenishment lead time (frozen zone) is used to analyze the stability of the delivery schedules. As can be seen from the summary in Section 3.3 and the discussion in Section 4.2.2, there are a number of figures to measure the stability or instability of a sequence of delivery schedules, but none of them is capable of covering all important aspects at once. These aspects include differentiation between changes in time and quantity of orders, distinguishing between underestimations and overestimations and considering the temporal distance between the date of the release of the delivery schedule and the date of the change within the schedule. As most existing figures are limited either to observe a change in quantity for a certain period or a change in time for a certain order in the delivery schedule, some changes may be considered more critical than they are, whereas other

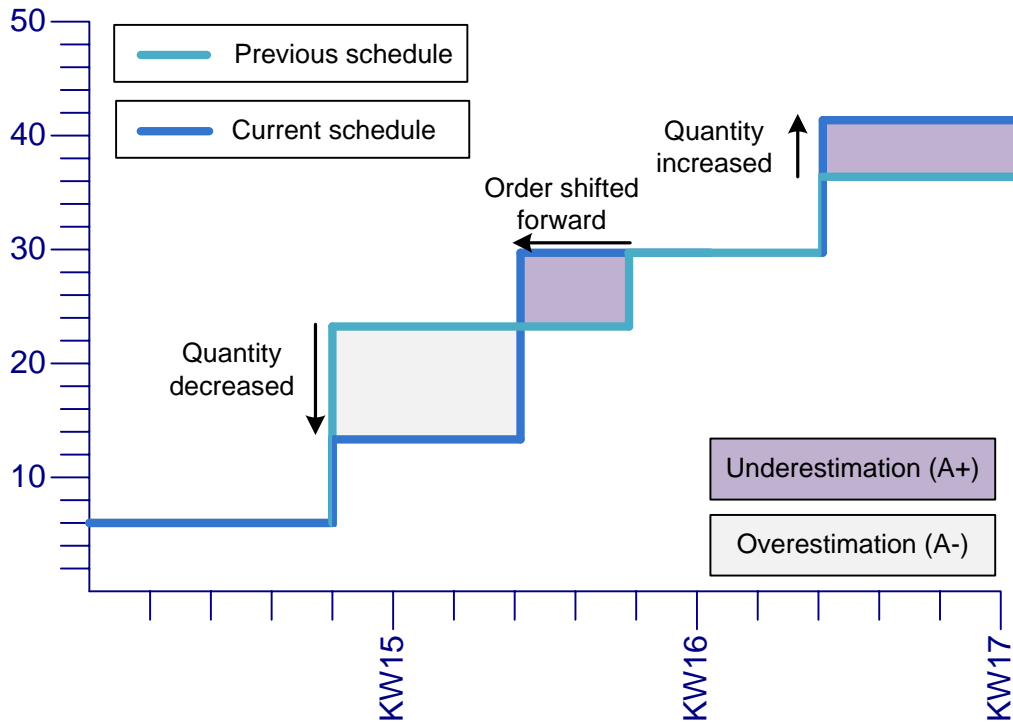


Figure 6.5: The area between two cumulative quantities used as a figure to describe the changes between two delivery schedules.

changes may be seen as less critical than they are. Consider for example a schedule in which an order of n product units is scheduled for period t whereas all other periods have an order amount of zero product units. Then, in the next schedule, the order of n units is moved from period t to period $t + 1$. In quantity based figures this would result in the conclusion that a change of 100 per cent took place in period t and another change of 100 per cent took place in $t + 1$. This may seem rather critical, whereas this change would only cause inventory holding cost for one period and could not be a reason for an escalation process. To account for this issue a new figure has been developed. The cumulative quantities of two consecutive schedules are compared and the areas between the two cumulative quantities are calculated as shown in 6.5. The light gray area A^- between the two cumulative quantities occurs if an overestimation of demand took place in the previous schedule and therefore too many parts have been ordered, or the parts have been ordered too early. In both cases additional inventory

holding costs may occur but no escalation process will be triggered. The dark gray area A^+ on the other hand indicates that demand has been underestimated in the previous schedule and thus orders had to be shifted forwards, be increased in quantity or added to the schedule. These actions may in turn lead to an escalation process, but will not cause additional inventory holding cost. To allow for a comparison between multiple parts with different total quantities both A^+ and A^- areas are set in relation to the total area under the cumulative quantity of the first delivery schedule to gain a relative value that reflects both time and quantity variances. As reflected before, only the first three weeks after the frozen zone are of special importance for the comparison, thus the areas will be computed for this timespan only.

7 A case study from an automotive company

In this chapter a case study from the automotive industry will be presented. The case study followed multiple aims to accomplish a detailed analysis of the presented delivery schedule generation approaches and their behavior when applied to a problem setting from practice. The remainder of this chapter is organized in analogy to the subjects of analysis, which can be summarized as follows:

- **Experimental design** is explained to provide an overview of the case study and how it was conducted.
- **Analysis of selected delivery profiles** for each delivery profile based configuration is made to compare the selection strategies of the different delivery profile based approaches.
- **Algorithmic performance** of the presented planning techniques for delivery profile selection is analyzed in respect to runtime and solution quality. The heuristic approaches are benchmarked against exact solution techniques, and it will be determined whether better results can be expected if both techniques are combined.
- **Realized costs** are computed to provide understanding of the behavior of the different planning techniques applied in a rolling horizon environment. The focus is on the comparison of the planning techniques' expected outcome based on upfront forecasts on the one hand and the realized outcome when applied in a rolling horizon on the other hand.
- **Expected costs** are measured for each planning technique. They provide details on how much the situation could be improved if the forecast holds true. In combination with the realized cost it can be stated how well the planning method was capable of predicting the realized outcome, giving a hint on the reliability of a planning methods plan.

- The **value of perfect information** is derived from the problem setting by comparing a post-ex determination of the optimal solution to the planning result derived based on the initial forecast.
- **Delivery schedule stability** is measured for the different planning techniques to test the hypothesis that delivery profiles help to increase the stability of the delivery schedules.
- **Inventory behavior** is analyzed in respect of safety levels and excess inventory at the end of the planning period to determine the pros and cons of the additional inventory built up depending on the delivery schedule generation method that was deployed.

Before providing details of the different result subjects the experimental design of the case study will be briefly explained.

7.1 Experimental Design

Over a period of six months data from over 3600 parts delivered by more than 330 suppliers distributed to 25 areas all over Europe were collected. Aside from the necessary master data such as part data, load carrier data and tariff structures, movement data, in the form of gross demand forecasts released twice a week, were collected. The data origins from a component plant integrated into an international automotive supply chain. The collected data were separated into two disjunct sets, a training set and a test set, both of three month length. As shown in Figure 7.1 the data from the training set were used to train the delivery profile selection approaches as far as this was necessary. As the stochastic planning methods require historical data to create scenarios based on underlying distributions, the training set was used to derive information on the parts demand behavior and forecast quality. Scenarios were generated for the test set according to the gathered distributions and occurrence probabilities. Delivery profiles were then determined based both on the information gathered from the training set and the first forecast from the test set. This corresponds to the information which is available on the first planning cycle within the test set and therefore reflects a real-world planning application. In a second step the delivery profiles selected by the approaches developed in this thesis are passed over to the simulation approach.

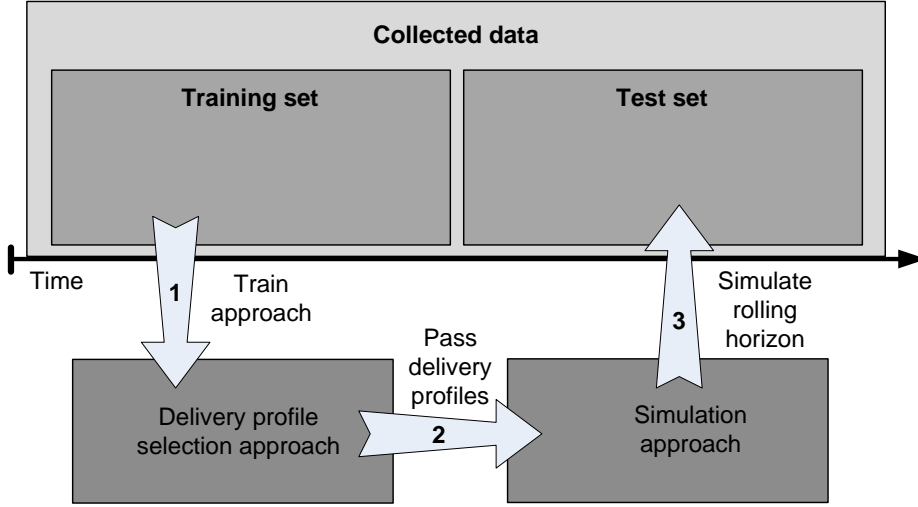


Figure 7.1: Data separation into training set and test set.

Thereafter the simulation approach is employed to simulate a rolling horizon planning situation based on the data from the test set. For each delivery profile assignment vector that has been generated by one of the delivery profile selection approaches one simulation run is conducted. In addition, one simulation run for the default MRP behavior and one simulation run for the state-of-the-art method from Kempkes and Koberstein [2010] are conducted to allow a comparison not only between the different delivery profile selection approaches, but also a comparison with alternative delivery schedule generation approaches.

7.1.1 Description of the examined areas

To give a general impression of the different areas that have been considered in the case study Table 7.1.1 lists the most relevant properties of the different areas. The area instances are divided into four groups, very easy, easy, medium and hard, depending on their size and complexity. The first column is the *instance* number, which will be used to identify a certain area in each different test result table. The next columns reflect the size of each area. In the second column the number of *suppliers* that are assigned to the area is given. In the *parts* column the number of parts in that are delivered through that area is given. The *load carriers* column gives back the number of disjunct load carriers used in that area. As multiple parts may be delivered in equal

load carriers, the number can be significantly smaller than the number of parts in a certain area. In addition to the areas size figures on the number of *demands per period*, the *average item density* and the share of items with a density *above vehicle density* are included. The number of demands per period indicates the density of operations within the area. It is computed by counting number of orders within the considered period range and dividing this value by the number of periods within the period range. The average item density can be computed by dividing a filled load carrier's weight by its volume and indicates how well the goods may fit into a vehicle. In the *above vehicle density* column the share of items in the area that have a density higher than an vehicles empty cargo space is given. One notable aspect is that in all areas the *average item density* is higher than a vehicle's average load density, which is around 0.28. As can be seen from the *above vehicle density* column, most instances also have a significantly higher share of items with a density above a vehicle's average density. This in turn means that the transported goods are mostly high density goods, and that vehicles will more often be filled by weight rather than by volume. Therefore an intelligent packing algorithm may exploit this density distribution to its advance and place the low-density items in the same vehicle as the high-density items.

7.1.2 Considered delivery profiles

The set of delivery profiles that is to be considered during optimization is an important input given by the human planner. On the one hand the set of delivery profiles determines the complexity of the overall solution process. The more delivery profiles are used, the higher becomes the computational effort that is necessary to evaluate them and make the optimal assignments to the different suppliers. If a lesser number of delivery profiles is used, the freedom of choice is reduced, and therefore the solution space is decreased. Thus a trade-off has to be found such that the provided delivery profiles offer substantial solution space while at the same time the complexity remains at a reasonable level. In practice, the considered delivery profiles are often predefined by the options supported by the employed MRP system. The set of delivery profiles that has been used in this case study is shown in Table 7.1.2. In addition to week-based delivery profiles, using a calendar with five working days per week, frequency-based delivery profiles are introduced in order to allow a comparison between frequency-based profiles and week-based delivery profiles. This delivery profile set provides enough free-

Instance	Problem size			Properties		
	Suppliers	Parts	Load Carriers	Demands per period	Average item density	Above vehicle density
1	4	81	6	22.44	0.40	97%
2	5	44	7	10.59	0.81	100%
3	5	84	7	26.17	0.51	91%
4	6	42	10	8.79	0.71	78%
5	6	68	10	13.68	0.49	73%
6	6	85	8	15.14	0.96	91%
7	6	114	17	36.45	0.49	87%
8	7	93	7	15.21	0.70	94%
9	8	33	7	4.01	0.71	75%
10	8	94	11	13.17	0.52	79%
11	9	43	9	16.58	0.50	88%
12	10	53	11	6.77	0.70	70%
13	11	69	13	21.72	0.64	56%
14	11	72	13	12.51	1.00	60%
15	11	129	8	29.10	0.56	72%
16	13	188	13	52.30	0.99	94%
17	14	174	17	40.29	1.14	82%
18	15	83	7	15.98	1.02	83%
19	19	258	17	66.66	1.14	76%
20	21	257	12	71.26	0.75	78%
21	25	352	15	89.10	1.28	91%
22	26	373	18	67.85	1.12	87%
23	28	156	15	48.81	0.54	74%
24	30	430	13	82.68	0.86	91%
25	34	254	16	74.36	0.90	77%

Table 7.1: Overview of the general properties of the examined areas.

dom of choice for the algorithms without overextending performance requirements and has proven viable in practice.

7.1.3 Testing environment

All tests have been conducted on a computer with an Intel Core 2 Duo central processing unit with 3.33 Gigahertz and 8 Gigabyte random access memory running under a 64 Bit-Edition of the Windows 7 operating system with service pack 1 installed. In addition to the operating system the following software has been used during the test runs:

- **Gurobi 5.00** from Gurobi Optimization Inc., a professional implementation of

Week based delivery profiles						
Delivery profile	Mon	Tue	Wed	Thu	Fri	Frequency
W11111	X	X	X	X	X	Weekly
W10101	X	-	X	-	X	Weekly
W01010	-	X	-	X	-	Weekly
W10000	X	-	-	-	-	Weekly
W00001	-	-	-	-	X	Weekly
W00100	-	-	X	-	-	Weekly
Frequency based delivery profiles						
Delivery profile	Frequency					
R2	Every second day					
R3	Every third day					

Table 7.2: Overview of the delivery profiles considered in the case study.

the Branch & Bound algorithm which was used to solve the given model instances.

- **Microsoft .net Framework 4.0** from Microsoft corporation, a software development framework that was used for the implementation of the heuristic algorithms, the model generation and the simulation approach.
- **Microsoft SQL Server 2008 Express RC 2** from Microsoft corporation, a relational database system deployed to collect and persist the necessary data.
- **Optimization.Framework**, a framework for modeling mathematical programs in .net, developed by the Decision Support & Operations Research Lab at the University of Paderborn.

7.1.4 Considered alternatives

For a company that intends to introduce a new operational order lot-sizing control mechanism different alternatives are available. These will be compared in the case study in order to provide an indication of which alternative may be appropriate. Throughout the case study the following alternatives will be considered:

- The alternative *no delivery profiles* refers to the situation before the introduction of a deterministic planning approach. All orders are just passed on by the MRP

system without any bundling taking place except for a ceiling of the lot sizes to integer multiples of the quantity that fits into a load carrier.

- The deterministic planning approach is represented by the *initial forecast* alternative. Delivery profiles are derived from the forecast given on the last day of the training set.
- The *demand-based scenarios* alternative uses the stochastic programming approach with scenarios that are generated from the demand distribution observed in the training set.
- In the *forecast deviation scenarios* alternative scenarios are generated from the forecast deviations observed in the training set.
- A mixture of both scenario generation methods is used in the *mixed scenarios* alternative. In this case scenarios are generated from former demand distribution and forecast deviations in equal share and are thereafter merged during the scenario reduction process.
- To give a comparison with the current state-of-the-art algorithmic delivery schedule generation approaches an implementation of the model presented in Kempkes [2009] was used to generate delivery schedules. This alternative will be called the *Kempkes* alternative in the following.

7.2 Analysis of the selected delivery profiles

Figure 7.2 displays the distribution of the selected delivery profiles for each delivery profile-based configuration. As can be seen at the first sight slightly more than half of all suppliers have the delivery profile W11111 assigned among all configurations. This means that those suppliers may deliver every day and that the savings are mostly generated through the other half of the suppliers involved. We can distinguish between delivery profiles that allow three deliveries (purple color), two deliveries (light red) or one delivery per week (red). The fewer deliveries per week are allowed, the tighter are the boundaries for the supplier's deliveries and the higher are the expected synergy effects. In most cases these are suppliers with only a little material to be delivered. The *initial forecast* configuration prefers the tightest boundaries on deliveries, whereas

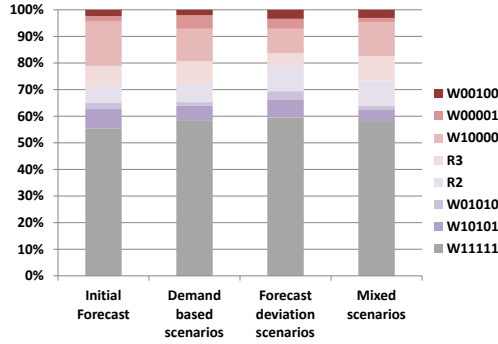


Figure 7.2: Distribution of selected delivery profiles per configuration.

the *forecast deviation scenarios* configuration relies on more frequent deliveries. In particular the *initial forecast* configuration uses single-delivery delivery profiles in 28.6% of all cases, whereas the *forecast deviation scenarios* uses them for only 21.1% of all suppliers. The *demand-based scenarios* configuration with 27.6% and the *mixed scenarios* configuration with 26.5% are settled between these two extreme points. The changes between the different configurations seem small at first sight, but a closer look reveals that various shifts take place. As Figure 7.2 only depicts the total percentage distribution, but not the assignment to a specific supplier, shifts between the different configurations may remain hidden. In fact between 32.0 % and 39.8 % of all delivery profiles are shifted from one configuration to another. The different delivery profiles are shifted with different probabilities. On average only 16.3 % of all suppliers with delivery profile W11111 assigned are shifted from one configuration to another. At the same time 69.5 % of the suppliers with other delivery profiles are shifted. The numbers vary slightly depending on which configurations are compared, but basically it can be said that suppliers with high volumes remain relatively stable in the set of suppliers with delivery profile W11111 assigned, and that the differences between the configurations become apparent on the selection of delivery profiles for the suppliers with less-than-truckload on every day.

7.3 Algorithmic performance

In this section algorithmic performance in terms of required runtime and resulting solution quality of the presented algorithms will be evaluated. The complexity of the planning problem varies from area to area, as some areas are pretty small and cover only four to six suppliers with their respective parts whereas other areas consist of 25 to 34 suppliers. Aside from the pure size in form of suppliers and parts count respectively, the number of orders and their distribution plays an important role. The physical properties of the delivered goods, especially their weight and size may also influence algorithmic performance. To account for these various factors affecting algorithmic performance, each area was tested separately to provide an overall picture. Even though a count of 25 test instances does not allow us to derive a statistical correlation, it may be a sufficient number of experiments to measure the algorithmic performance. The following Section 7.3.1 will first be discuss how the preprocessing algorithm behaves in respect of different problem classes and sizes. Section 7.3.2 then will describe the algorithmic performance of the generic model formulation. In Section 7.3.3 the results for the condensed model formulation can be found. The runtime evaluation concludes with a discussion of the heuristic procedures in Section 7.3.4.

7.3.1 Evaluation of the decomposition approach

Whenever one of the models presented in this work is used to derive delivery profiles it has to be considered that the decomposition approach requires the preprocessing routines as described in Section 5.3 for the deterministic case or Section 5.6.1 for the stochastic case respectively to be executed beforehand. Therefore the algorithmic runtime of these preprocessing routines has to be added to the total runtime of the algorithms and models. As the preprocessing routines consider each combination of supplier, period, delivery profile and scenario separately, their runtime should grow almost linearly in respect of each of these components. Table 7.3 gives an overview of the time required by the preprocessing algorithm to process a certain area instance. In the first column a reference to the area instance is given. The next four columns give back the runtime (in seconds) of the different delivery profile-related alternatives. The column *initial forecast* refers to the runtime for the deterministic case where delivery profiles are determined on the forecast given at the beginning of the planning horizon. The columns *demand-based scenarios*, *forecast deviation scenarios* and *mixed scenarios*

refer to the runtime required for the corresponding stochastic alternatives. In the *average* column the average runtime for the stochastic case is given. The penultimate column shows how many sub-model calls were necessary during the preprocessing procedure. To give a relation to the total amount of processed periods the last column gives the percentage of processed periods in which a sub-model call is necessary. In the last two rows summarizing information on the runtime is given. The penultimate row gives back total runtime required to preprocess all area instances. The last row gives back the ratio between the stochastic and deterministic cases for the three scenario generation approaches.

Instance	Initial Forecast	Demand based scenarios	Forecast based scenarios	Mixed scenarios	Average (stochastic)	# of sub models to solve	% of sub model calls
Very Easy	17	75	260	151	162	774	12.0%
1	17	108	305	211	208	2,068	44.0%
2	10	32	227	41	100	1,502	24.1%
3	12	19	174	61	85	564	7.3%
4	1	2	14	4	7	625	9.7%
5	0	0	0	0	0	0	0.0%
6	5	22	95	34	50	431	9.7%
7	114	551	1,688	1,147	1,129	1,632	19.8%
8	6	11	93	9	38	743	9.0%
9	0	0	0	0	0	0	0.0%
10	0	1	4	1	2	178	1.8%
Easy	1	7	46	16	23	494	3.7%
11	4	22	201	66	96	1,587	14.6%
12	0	3	7	3	4	224	3.4%
13	0	4	8	3	5	220	1.9%
14	0	3	7	3	4	342	5.2%
15	0	2	7	3	4	97	0.6%
Medium	17	87	330	142	187	1149	5.1%
16	37	204	938	405	516	3,350	21.1%
17	46	207	652	252	370	1,570	10.3%
18	0	4	5	14	8	79	0.7%
19	1	13	22	29	21	529	2.4%
20	1	9	32	12	18	218	0.8%
Hard	34	290	736	384	470	3942	12.7%
21	52	229	1,370	483	694	5,965	20.8%
22	80	628	1,205	717	850	5,501	18.3%
23	14	213	515	316	348	2,562	7.9%
24	18	318	465	300	361	3,319	12.2%
25	8	63	126	104	98	2,363	6.4%
Total	426	2,668	8,160	4,218	5,015	6,360	8.36%
Ratio		6.3	19.2	9.9	11.8		

Table 7.3: Time required to preprocess the given area instance with the algorithms described in Section 5.3 for the deterministic case or Section 5.6.1 for the stochastic case with 10 scenarios respectively.

Greater understanding can be gained from the data given in Table 7.3. First of all it has to be mentioned that the runtime required for preprocessing is quite small, but grows rapidly if multiple scenarios are considered. For the deterministic case the total time required to preprocess all 25 area instances is about seven minutes (426 seconds). For the stochastic case this value ranges from 45 minutes (2,668 seconds) for the demand based scenario alternative to two hours and 16 minutes (8,160 seconds) for the forecast deviation scenario alternative. Furthermore, it can be observed that the time required to preprocess an area instance depends not only on the number of suppliers and delivery profiles. Even though it can be seen that instances with a higher number of suppliers take longer to be processed in general, there are some instances that differ exceptionally from this rule, especially areas 7, 16, 17, 18, 19 and 25. Whereas areas 7, 16 and 17 take longer to be preprocessed than could be expected given the number of combinations, areas 17, 18 and 25 can be processed much faster than would be expected, even though area 25 provides the highest number of suppliers. This finding can be explained by the fact that the runtime of the sub-model that is solved for the pre leg cost evaluation accounts for the highest share of runtime in the preprocessing algorithm. As described in Section 5.3.3 runtime can be saved by leaving out the sub-model for periods in which freight does not exceed the capacity of a single vehicle. When an area instance tends to provoke more sub-models to be solved, the linear relation between number of suppliers and runtime is broken. This can be seen when the number of sub models that have been solved in an area instance preprocessing routine with its runtime is considered. It can be stated clearly that the high share of pre leg sub-models to be solved explains the higher runtimes. In addition to the number of sub-model calls the complexity of the sub-models is an important factor in the preprocessing runtime. The more freight is involved in a single period, the higher becomes the complexity of the sub-model as more vehicles and rebate levels have to be included. More complex sub-models take longer to solve. Therefore the preprocessing seems to be more sensitive to the freight volume transported in a certain area than to the number of suppliers involved, as a high freight volume demands more sub-models be solved.

7.3.2 Evaluation of the generic model formulation

The generic formulation of the main model provided in Section 5.4 allows us to model various tariff systems independently of the number of resources involved, the complexity of the rebate level structure and so on. In addition it gives the opportunity to include additional factors like carbon-dioxide, incoming goods personnel or penalties for certain delivery days. This flexibility comes at the cost of runtime. Table 7.4 gives an overview of the runtime and the remaining gap for the deterministic case. All area instances were tested with a time limit of three ours or 10,800 seconds. Before the model solving was started a starting value from the heuristic algorithms was passed over. In the first column the area instance is referenced. The second column shows the runtime of the solver in seconds. The next column gives back the gap between the starting value passed over and the best bound that was found by the solver. The last column gives back the gap between the best integer solution found by the solver and the best bound found by the solver. Even for the deterministic case only the smaller area instances could be solved to optimality. As the starting value is nearly optimal in most cases only small improvements to the primal solution can be made. For the instances from the hard set and instances 16 and 19 from the medium set, no optimal solution values could be obtained within three hours. Thus no improving integer solutions could be found. This may be a hint that the models formulation is not tight enough, as it takes a lot of time to improve the bounds.

7.3.3 Evaluation of the simplified model formulation for the stochastic case

The results from the algorithmic performance tests for the simplified model for the stochastic case are summarized in Table 7.5. In the first column a reference to the area instance is created. The runtime of the solver in seconds is then given in the next column. Thereafter the gap between the best LP relaxation bound and the starting value is given in the *starting value* column. The gap between the best linear programming relaxation bound and the best integer solution is presented in the last column. As can be seen most instances can be solved to optimality easily. In most cases the heuristic algorithms already find the optimal solution, and the solver only needs to prove the optimality. Only for the instances 16 from the medium set and instances from the hard set, could no optimal solution values be derived within the given time limit of three hours. In these cases the solution of the heuristic algorithms could not

Instance	Runtime MIP Gurobi	Gap starting value	Gap MIP Gurobi
Very Easy	923	0.19%	0.01%
1	6970	0.73%	0.00%
2	23	0.26%	0.01%
3	1090	0.01%	0.01%
4	19	0.01%	0.01%
5	41	0.63%	0.00%
6	942	0.18%	0.01%
7	36	0.09%	0.01%
8	53	0.01%	0.01%
9	42	0.00%	0.00%
10	11	0.03%	0.01%
Easy	462	0.24%	0.01%
11	138	0.54%	0.01%
12	725	0.52%	0.01%
13	424	0.01%	0.01%
14	481	0.11%	0.01%
15	544	0.01%	0.01%
Medium	5,039	7.28%	7.21%
16	10,808	9.23%	9.23%
17	847	0.27%	0.01%
18	1,936	0.01%	0.01%
19	10,812	26.80%	26.80%
20	790	0.09%	0.01%
Hard	10,818	13.90%	13.90%
21	10,820	35.99%	35.99%
22	10,820	5.28%	5.28%
23	10,806	7.87%	7.87%
24	10,827	5.13%	5.13%
25	10,816	15.23%	15.23%

Table 7.4: Overview of the runtime for the generic model formulation for the deterministic case.

be improved by the solver, but the optimality could not be proven either.

7.3.4 Evaluation of the heuristic algorithms

In the previous sections a heuristic starting value has been referred to. This heuristic starting value was generated by the algorithms presented in Section 5.5. The detailed data on runtime and solution quality of the heuristic solution approaches is given in Table 7.6. After the first column gives a reference to the area instance, the second column shows the size of the search space available in form of *possible combin-*

Instance		Runtime MIP Gurobi	Gap starting value	Gap MIP Gurobi
Very Easy		332	0.19%	0.00%
1		417	0.73%	0.00%
2		99	0.26%	0.01%
3		92	0.00%	0.00%
4		42	0.00%	0.00%
5		5	0.63%	0.00%
6		71	0.18%	0.00%
7		2,549	0.09%	0.01%
8		21	0.01%	0.01%
9		1	0.00%	0.00%
10		26	0.03%	0.01%
Easy		111	0.13%	0.00%
11		487	0.54%	0.01%
12		8	0.01%	0.00%
13		21	0.01%	0.01%
14		4	0.11%	0.00%
15		33	0.01%	0.01%
Medium		5,221	1.07%	1.01%
16		10,849	4.58%	4.58%
17		402	0.27%	0.01%
18		30	0.00%	0.00%
19		10,845	0.42%	0.42%
20		3,979	0.09%	0.01%
Hard		10,956	3.52%	3.52%
21		10,941	3.54%	3.54%
22		10,989	1.87%	1.87%
23		10,842	2.70%	2.70%
24		10,924	3.51%	3.51%
25		11,086	5.97%	5.97%

Table 7.5: Overview of the runtime for the stochastic case with mixed scenarios for the simplified model formulation.

ations for these algorithms. The *Local search* column refers to the local search method presented in Section 5.5.1. The *genetic algorithm* column gives back the results of the implementation of the genetic algorithm as described in Section 5.5.2. In the *Gurobi* column the results from the simplified stochastic programming formulation presented in Section 5.6.4 are given to allow an evaluation of the solution quality of the heuristic algorithms. The heuristic solution values were close to the optimum or even optimal in most cases. Note that the genetic algorithm performs much better than the local search algorithm. This can be explained by the fact that synergy effects between different de-

Instance	Possible Combinations	Runtime in Seconds		Gap		Best MIP Gap
		Local search	Genetic Algorithm	Local search	Genetic Algorithm	
Very Easy	4,688,448	377	332	0.46%	0.19%	0.00%
1	65,536	310	306	0.73%	0.73%	0.00%
2	390,625	416	320	0.26%	0.26%	0.01%
3	390,625	340	334	0.00%	0.00%	0.00%
4	1,679,616	356	449	0.00%	0.00%	0.00%
5	1,679,616	329	313	0.00%	0.63%	0.00%
6	1,679,616	305	303	0.18%	0.18%	0.00%
7	1,679,616	481	340	1.04%	0.09%	0.01%
8	5,764,801	449	321	1.15%	0.01%	0.01%
9	16,777,216	316	305	0.00%	0.00%	0.00%
10	16,777,216	467	331	1.28%	0.03%	0.01%
Easy	157,224,673	672	342	1.80%	0.13%	0.00%
11	43,046,721	513	346	0.90%	0.54%	0.01%
12	100,000,000	606	321	0.02%	0.01%	0.00%
13	214,358,881	1164	349	2.14%	0.01%	0.01%
14	214,358,881	671	305	0.11%	0.11%	0.00%
15	214,358,881	407	388	5.82%	0.01%	0.01%
Medium	11,932,166,561	712	731	13.91%	1.07%	1.01%
16	815,730,721	978	998	36.40%	4.58%	4.58%
17	1,475,789,056	669	419	1.50%	0.27%	0.01%
18	2,562,890,625	605	355	2.62%	0.00%	0.00%
19	16,983,563,041	572	836	12.64%	0.42%	0.42%
20	37,822,859,361	735	1048	16.41%	0.09%	0.01%
Hard	636,222,171,687	858	1211	34.81%	3.52%	3.52%
21	152,587,890,625	631	1105	51.30%	3.54%	3.54%
22	208,827,064,576	443	895	37.74%	1.87%	1.87%
23	377,801,998,336	963	1115	21.62%	2.70%	2.70%
24	656,100,000,000	454	1450	20.21%	3.51%	3.51%
25	1,785,793,904,896	1800	1489	43.16%	5.97%	5.97%

Table 7.6: Overview on the heuristic runtimes and solution quality for the given area instances.

livery profile assignments have to be achieved in order to improve the solution. If a promising combination was found within a solution, it can be transferred to other solutions during the recombination phase. Thus the principle of genetic evolution fits the problem setting very well. This can also be seen when the solution progress over time is shown in detail. Figure 7.3 shows the solution progress over time for selected area instances. Whereas the genetic algorithm provides a steady improvement with medium sized steps, the local search algorithm swiftly rushes towards a local optimum and gets stuck. To avoid this behavior a backup method has been introduced that restarts the

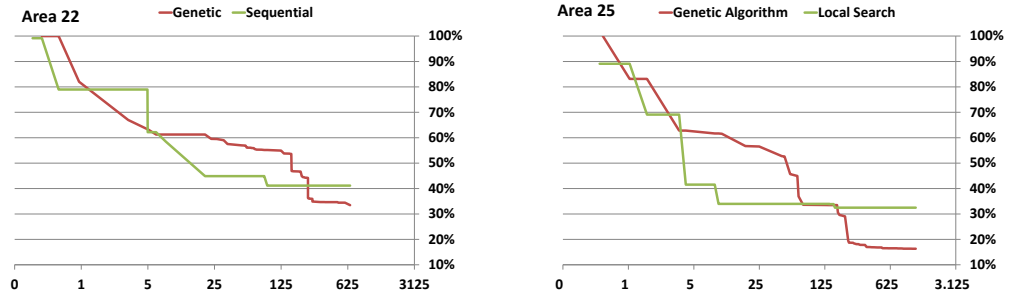


Figure 7.3: Solution progress over time in comparison between local search and genetic algorithm for two area instances.

local search on a random position if no further improvement can be achieved within a reasonable number of tries. But as the probability of finding a better assignment by chance is quite low, only limited success has been achieved. In general it can be stated that the genetic algorithm is clearly preferable to the local search method and that the results yielded by the genetic algorithm are optimal in most cases for the smaller instances. When it comes down to the implementation details of the genetic algorithm it has to be noted that the possibility of adding a random solution, in case no improvement could be made, during the last iterations is of high practical use. In Table 7.3.4 the number of improving solutions that have been found during the solution process of the stochastic area instances for the forecast deviation based scenarios and their source is depicted. In the first column a reference to the considered area instance is given. The *offspring* column gives back how many improving solutions have been found by combining two parents. In the *mutation* column the respective number for the mutation step is given. Finally, the *random* column displays the number of solutions that have been found by the function which adds a random solution after no improvement has been made in the previous iterations. It can be seen that on average the largest share (45.8 %) of the improving solutions is found in the recombination step, followed closely by the mutation step which is responsible for 41.8% of all improvements. Only 12.3 % of all improving solutions have been found by the 'fresh blood' function. When interpreting these numbers it has to be kept in mind that the random solutions will only be added in case that the algorithm could not find an improvement. Thus it can

be shown that it can be prevented from getting stuck in a local optima by providing the possibility to move through the solution space.

Instance	Source of improvement		
	Offspring	Mutation	Random
Very Easy	176	185	145
1	7	8	12
2	11	21	15
3	26	16	18
4	9	28	12
5	18	13	13
6	11	11	11
7	24	19	14
8	24	27	20
9	19	9	16
10	27	33	14
Easy	197	155	60
11	44	29	18
12	38	21	8
13	34	35	12
14	19	20	12
15	62	50	10
Medium	340	285	73
16	57	59	11
17	62	45	21
18	61	37	12
19	84	62	13
20	76	82	16
Hard	573	549	68
21	90	82	10
22	94	118	12
23	125	129	18
24	116	112	16
25	148	108	12
Total Share	1286 45.8%	1174 41.8%	346 12.3%

Table 7.7: Improving solutions found during the genetic algorithm procedure grouped by their source.

7.4 Evaluation of monetary effects

In this section the monetary effects incurred by the different delivery schedule generation approaches will be discussed. The different outcomes expected by the different

planning approaches will first be described, followed by a discussion of the costs that have finally been realized during the simulated application in a rolling horizon. The expectations and the realized outcome will then be set in relation to each other to identify to what degree the expected cost situation reflects the final outcome. An examination of the optimal post-ex solution is then offered to identify a lower bound on the realized cost and to give an impression of how close the different delivery schedule generation approaches get to this lower bound. As the case study was performed on data from practice the results are anonymized such that no conclusions on the company's freight budget can be made. In order to make the results anonymous, so that no detail on the company's business-critical data is given, the results are normalized, which means that each area's realized cost was divided by a common variable. In so doing the cost values are divided by the average total realized cost among all areas for the *no-delivery profiles configuration*. A value of 100 equals the average areas realized total cost for the *no delivery profiles* configuration. Thus the relevance of the depicted areas can still be derived from the data without publishing the real cost value. At the same time it is possible to draw comparisons between one configuration and other configurations.

7.4.1 Expected costs

When it comes down to measuring monetary results case studies in the field of operations research often provide a comparison of an existing and an optimal solution for the same problem instance. Two issues, however, remain unsolved. On the one hand it is not clear whether these theoretic improvements can be achieved. On the other hand a comparison of different models can hardly be achieved, especially if objective functions differ or approximations allow different solutions for both models. Aside from these issues, providing the gap between the optimal solution and the achieved solution can be used to measure the efficiency of heuristic algorithms as these do not necessarily provide an optimal solution. In the given case a comparison of different models is also possible as they provide the same objective function and level of detail for cost modeling. In addition, the objective function value of the optimal solution reveals the different costs expected by the various solution approaches. Considering the application of these approaches in a rolling horizon environment, it is of interest what gap emerges between the expected and the realized costs.

Table 7.8 shows the expected savings of the different configurations. The *Area*

7.4 Evaluation of monetary effects

Area	No Delivery Profiles	Initial Forecast	Demand based scenarios	Forecast deviation scenarios	Mixed scenarios	Kempkes
1	57.34	39.86 (30.5%)	39.86 (1.5%)	58.65 (1.4%)	48.84 (9.6%)	26.61 (53.6%)
2	43.28	41.45 (4.2%)	46.20 (4.4%)	45.90 (2.8%)	45.66 (11.1%)	32.22 (25.6%)
3	29.44	20.57 (30.1%)	20.57 (1.2%)	30.48 (7.0%)	25.65 (9.6%)	7.51 (74.5%)
4	27.85	17.90 (35.7%)	17.90 (1.2%)	28.53 (1.4%)	22.78 (10.4%)	4.30 (84.6%)
5	4.83	3.09 (35.9%)	3.09 (0.7%)	4.05 (2.7%)	3.71 (10.9%)	3.44 (28.8%)
6	44.49	37.51 (15.7%)	37.51 (0.3%)	47.16 (16.3%)	35.62 (11.9%)	18.77 (57.8%)
7	158.12	103.26 (34.7%)	103.26 (2.3%)	247.89 (1.9%)	182.42 (9.9%)	121.76 (23.0%)
8	29.72	18.84 (36.6%)	18.84 (0.2%)	28.87 (0.4%)	20.01 (8.4%)	10.96 (63.1%)
9	8.31	5.65 (32.0%)	5.65 (0.0%)	7.73 (1.0%)	6.36 (8.3%)	3.52 (57.7%)
10	19.96	14.16 (29.0%)	14.16 (0.0%)	21.47 (0.0%)	15.20 (8.3%)	3.96 (80.2%)
11	37.49	31.33 (16.4%)	31.33 (2.4%)	37.64 (3.1%)	34.96 (10.0%)	17.20 (54.1%)
12	40.71	29.89 (26.6%)	29.89 (2.2%)	43.01 (3.8%)	29.74 (12.0%)	4.74 (88.4%)
13	18.51	12.97 (29.9%)	12.97 (1.4%)	22.37 (2.3%)	18.42 (9.9%)	7.10 (61.6%)
14	22.70	19.75 (13.0%)	19.75 (19.9%)	22.65 (14.9%)	21.67 (23.0%)	10.74 (52.7%)
15	19.26	13.97 (27.5%)	13.97 (1.7%)	22.33 (0.9%)	18.82 (9.3%)	8.12 (57.8%)
16	42.28	31.69 (25.0%)	31.69 (1.7%)	40.62 (1.0%)	35.25 (10.0%)	36.90 (12.7%)
17	52.41	49.41 (5.7%)	49.41 (2.9%)	56.08 (0.8%)	52.71 (10.8%)	48.34 (7.8%)
18	25.84	25.63 (0.8%)	25.63 (3.5%)	30.12 (1.3%)	27.42 (10.5%)	6.94 (73.1%)
19	51.22	43.27 (15.5%)	43.27 (0.5%)	58.24 (3.8%)	50.21 (10.9%)	11.79 (77.0%)
20	31.58	25.59 (19.0%)	25.59 (0.0%)	38.88 (0.0%)	32.37 (8.3%)	27.03 (14.4%)
21	88.57	68.41 (22.8%)	68.41 (1.7%)	103.88 (1.7%)	85.33 (9.7%)	78.83 (11.0%)
22	234.66	183.66 (21.7%)	183.66 (2.8%)	255.55 (1.9%)	222.99 (11.0%)	100.47 (57.2%)
23	38.93	37.02 (4.9%)	37.02 (2.2%)	47.92 (3.1%)	37.83 (10.5%)	36.04 (7.4%)
24	98.99	92.12 (6.9%)	92.12 (0.4%)	108.65 (2.6%)	89.15 (10.1%)	83.74 (15.4%)
25	57.96	48.01 (17.2%)	48.01 (1.0%)	61.70 (2.0%)	56.38 (9.4%)	50.62 (12.7%)
No delivery profiles		1284.5	1034.9	1512.3	1358.9	1284.5
Optimal solution		1015.0	1019.8	1470.4	1219.5	761.6
Expected savings		21.0%	1.5%	2.8%	10.3%	40.7%
Minimum		0.8%	0.0%	0.0%	8.3%	7.4%
Maximum		36.6%	19.9%	16.3%	23.0%	88.4%
Average		21.5%	2.2%	3.1%	10.6%	46.1%
Standard deviation		11.1%	3.9%	4.0%	2.8%	27.2%

Table 7.8: Expected costs and expected savings for all configurations.

column refers to the area instance that is considered in the row. The following columns *no delivery profiles*, *initial forecast*, *demand-based scenarios*, *forecast deviation scenarios*, *mixed scenarios* and *Kempkes* each refer to the cost resulting from a certain alternative as described in Section 7.1.4. In each cell two values are given. The first value is that of the normalized expected cost value, whereas the latter value (given in brackets) reflects the expected savings towards the *no delivery profiles* configuration. The last rows gives a summary of the expected savings of all areas. The *no delivery profiles* row gives back what total cost is expected as the *no delivery profiles* configuration is applied to the given demand set that is used by each configuration. The

optimal solution row gives back the total cost expected following the optimal solution for each configuration. The *savings* row indicates the percentage saving that can be expected. These are computed by subtracting the best objective value available from the objective value for the *no delivery profiles* configuration. For the stochastic models it has to be considered that these work with scenarios that have been generated on the given information rather than with the actual forecast, their underlying demand expectations may differ from the demand forecast which is used in the other two configurations. Therefore their absolute expected value may be higher than the absolute expected value for the *no delivery profiles* configuration in the deterministic case. However, their expected absolute value is still lower than the value for the *no delivery profiles* configuration in the stochastic case. The difference between these two is given in the brackets in each cell. It is to be noted that the expected savings vary between the different areas. Accordingly, the last four rows contain the minimum expected saving, the maximum expected saving, the average expected saving and the standard deviation of the savings among all areas. Apparently, the highest savings are expected in the *Kempkes* configuration, which proposes a total saving of 40.7 % over all areas. The proposed saving varies from 7.4% to 88.4% with an average of 46.1% and a standard deviation of 27.2%. Only two out of 25 areas have proposed savings of less than 10 %. On the other hand, savings of more than 50 % are proposed for 15 areas. The superb expected results of the *Kempkes* configuration can easily be explained. Whereas the delivery profile based configurations are limited to ordering points and lot sizes that comply with the given set of delivery profiles, the algorithm behind the *Kempkes* configuration has a much higher degree of freedom. It can place any order in any period as long as it reduces the objective value and is settled before the demand date. The delivery profile based configurations, however, show diverse behavior. The *initial forecast* configuration promises an expected saving of 21.0% in total, with numbers ranging from 0.8% up to 36.5% per area. The average saving is given with 21.5%, the other values are settled around the average with a standard deviation of 11.1%. For the scenario-based configurations the values are not as high. In fact each scenario-based configuration has at least 14 values below 2.0 %. All three configurations provide a small positive total expected saving. However, the savings range from 0% up to only 19.9 % for the *demand-based scenarios* configurations and from 0% to 16.3% for the *forecast deviation scenarios* configuration. The highest expected savings among the

stochastic configurations can be observed in the *mixed scenarios* configuration. The expected savings range goes from 8.3% up to 23.0%, with an average of 10.6% and a standard deviation of only 2.8 %. If the absolute numbers are considered, it can be seen that there are differences between the expectations on future demand volume. The *no delivery profiles* configuration is expected to provide a total cost value of 1284.5. Both deterministic approaches that have made their computations based on the available forecast data, the *initial forecast* configuration and the *Kempkes* configuration, offer significantly lower values with values of 1015.0 and 761.6 respectively. The stochastic approaches, however, with an exception for the *demand-based scenarios* configuration with a value of 1034.9, expect higher values than in the deterministic forecast-based configurations. The *forecast deviation scenarios* estimates the total cost in the *no delivery profiles* configuration to be 1512.3, whereas the *mixed scenarios* configuration expects a value of 1358.9. For the *forecast deviation scenarios* configuration, even the optimal solution values are larger than the expected total cost in the deterministic *no delivery profiles* configuration. This is a first indicator that the forecast tends to underestimate demand in general, as the scenarios are based on the observations on forecast errors from the past. This indication is undermined by Figure 7.4, which shows the expected procurement volumes and the real consumption. The figure is normalized so that the average consumption reflects a value of 100 %. It can be seen that the forecast procurement volume is lower than the real consumption. The demand-based scenarios have the second lowest average value. As these are generated on historic consumption data it could be argued that demand has risen during the training set and this change has not yet been reflected in the forecast that is used to forecast the test set. In addition, it can be seen that the forecast deviation-based scenarios follow patterns similar to those of the forecast, but is higher on average. This shows that the problem of too low forecasts also existed in the test set and leads to the expectation that future forecasts will also be lower than the actual consumption. The mixed scenarios move somewhere between the demand-based and the forecast-based scenarios, as they are combined between the two.

7.4.2 Realized costs

In this section the costs that would finally have been realized when the planning methods are used in a rolling horizon are evaluated. In so doing it can be evaluated how

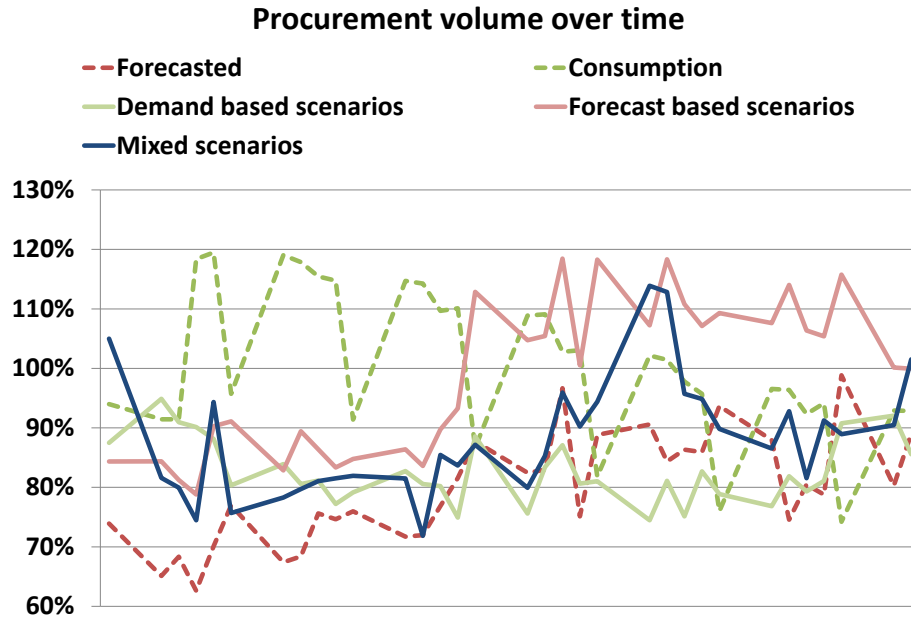


Figure 7.4: Expected procurement volume for the different sources of demand information and real consumption.

the delivery schedule generation approaches perform when uncertainty comes into play and how performance differs from that predicted upfront using the optimal solution objective values. In Table 7.9 an overview of the results from the simulation study is given. The *Area* column refers to the area instance that is considered the row. The following columns *no delivery profiles*, *initial forecast*, *demand-based scenarios*, *forecast deviation scenarios*, *mixed scenarios* and *Kempkes* each refer to the cost resulting from a certain alternative as described in Section 7.1.4. In addition to the absolute normalized value as described above, the *initial forecast*, *demand-based scenarios*, *forecast deviation scenarios* and *mixed scenarios* columns contain a percentage savings value. For the *initial forecast* configuration the percentage value reflects the savings in comparison with the *no delivery profiles* configuration. For the three configurations that have been created based on the stochastic programming approach the percentage value gives back the savings in comparison with the *initial forecast* configuration. The additional savings due to the employment of a stochastic programming approach are

thereby stressed. For the *Kempkes* configuration column percentage savings value is also based on the *initial forecast* configuration. It can therefore be shown how much additional savings may be achieved by employing a sophisticated order lot-sizing algorithm for the delivery schedule generation. The last three rows of the table include a summary of the results. In the *total* row the absolute normalized value over all areas is summed up for each configuration. The *savings* row gives back the percentage savings towards the situation without delivery profiles employed. In the *additional* row, the total additional savings that can be achieved towards the *initial forecast* configuration are given. Three main findings can be derived from these results. First, it can be

Area	No Delivery Profiles	Initial Forecast	Demand based scenarios	Forecast deviation scenarios	Mixed scenarios	Kempkes
1	43.0	20.38 (52.6%)	20.38 (0.0%)	20.38 (0.0%)	20.38 (0.0%)	16.97 (7.9%)
2	72.8	46.02 (36.8%)	46.02 (0.0%)	46.02 (0.0%)	46.02 (0.0%)	44.34 (2.3%)
3	31.2	18.09 (42.1%)	18.03 (0.2%)	18.09 (0.0%)	18.09 (0.0%)	16.73 (4.3%)
4	35.0	19.61 (44.0%)	18.50 (3.2%)	18.51 (3.1%)	18.51 (3.1%)	18.60 (2.9%)
5	13.0	9.31 (28.1%)	8.76 (4.2%)	8.43 (6.8%)	8.76 (4.2%)	8.97 (2.6%)
6	57.3	25.92 (54.8%)	25.92 (0.0%)	25.92 (0.0%)	25.92 (0.0%)	22.50 (6.0%)
7	273.5	237.65 (13.1%)	237.65 (0.0%)	237.65 (0.0%)	237.02 (0.2%)	227.52 (3.7%)
8	32.9	18.49 (43.7%)	18.49 (0.0%)	18.49 (0.0%)	18.49 (0.0%)	15.99 (7.6%)
9	12.5	7.20 (42.3%)	7.20 (0.0%)	7.20 (0.0%)	7.20 (0.0%)	5.89 (10.5%)
10	22.7	14.88 (34.5%)	14.83 (0.2%)	14.88 (0.0%)	14.83 (0.2%)	13.22 (7.3%)
11	43.5	30.25 (30.5%)	30.28 (-0.1%)	30.25 (0.0%)	30.25 (0.0%)	28.62 (3.8%)
12	68.0	33.44 (50.8%)	32.68 (1.1%)	33.54 (-0.1%)	33.44 (0.0%)	27.54 (8.7%)
13	29.6	19.56 (33.9%)	19.58 (-0.1%)	19.56 (0.0%)	19.56 (0.0%)	17.30 (7.6%)
14	63.2	45.78 (27.6%)	45.78 (0.0%)	45.78 (0.0%)	44.56 (1.9%)	44.31 (2.3%)
15	37.0	27.54 (25.5%)	27.57 (-0.1%)	27.55 (0.0%)	27.55 (0.0%)	26.19 (3.7%)
16	44.8	28.15 (37.2%)	28.15 (0.0%)	28.15 (0.0%)	28.15 (0.0%)	26.42 (3.9%)
17	87.2	54.28 (37.8%)	53.06 (1.4%)	53.09 (1.4%)	53.03 (1.4%)	53.53 (0.9%)
18	40.4	20.44 (49.5%)	20.44 (0.0%)	19.81 (1.5%)	20.44 (0.0%)	18.85 (3.9%)
19	106.5	73.27 (31.2%)	73.26 (0.0%)	73.28 (0.0%)	73.28 (0.0%)	70.25 (2.8%)
20	102.1	79.84 (21.8%)	79.85 (0.0%)	79.80 (0.0%)	79.85 (0.0%)	77.21 (2.6%)
21	212.3	200.36 (5.6%)	200.39 (0.0%)	200.36 (0.0%)	200.36 (0.0%)	188.40 (5.6%)
22	472.6	391.81 (17.1%)	391.76 (0.0%)	391.76 (0.0%)	391.76 (0.0%)	379.08 (2.7%)
23	59.5	38.61 (35.1%)	38.61 (0.0%)	38.74 (-0.2%)	38.58 (0.0%)	35.12 (5.9%)
24	401.0	371.82 (7.3%)	359.31 (3.1%)	359.31 (3.1%)	359.31 (3.1%)	345.34 (6.6%)
25	138.5	95.13 (31.3%)	95.97 (-0.6%)	95.27 (-0.1%)	95.45 (-0.2%)	85.33 (7.1%)
Total	2500.00	1927.84	1912.46	1911.84	1910.81	1814.23
Savings		22.89%	23.50%	23.53%	23.57%	27.43%
Additional			15.39	16.00	17.03	113.61
%			0.62%	0.64%	0.68%	4.54%

Table 7.9: Overview of the realized total cost of each area as derived from the simulation study.

seen that intelligently chosen delivery profiles not only provide a theoretical savings potential but also perform well in a rolling horizon application. An overall saving of

22.89 % of freight and inventory-related cost could be achieved for the deterministic case. This value differs largely depending on the considered instance. Whereas some smaller areas can obtain savings of up to 54.8 %, the largest four instances can be improved by only 7.3 % to 17.1 %. A chief reason for this observation is that a large share of goods from these areas is transported via full load runs, because a reasonable share of goods in high volumes is ordered from few suppliers. This leads to multiple full load run vehicles per day even without an assigned delivery profile and therefore does not provoke a need for consolidation. Most suppliers of such size have been removed from the area forwarding network and their goods are delivered via separated freight contracts. Owing to multiple products' life cycles interfering with each other, the set of those high-volume suppliers changes over time. As the identification of those suppliers is made infrequently, some may still be connected to the area forwarding network, as is the case for some areas in the case study. Second, a robust planning approach based on the consideration of multiple scenarios in a stochastic programming approach can further increase the savings. When it comes down to a comparison between the deterministic solution based on the initial forecast and the stochastic programming approaches it can be seen that for most area instances no significant improvement towards the deterministic solution can be achieved. The overall savings can be increased from 22.89 % for the deterministic case up to 23.57 % for the *mixed scenarios* configuration. For 7 out of 25 area instances the savings have even been reduced for the stochastic approach independently of the scenario generation approach. However, these savings reductions are rather small with an average of -0.12 % for the *demand-based scenarios* configuration, -0.07 % for the *forecast deviation scenarios* configuration and -0.04 % for the *mixed scenarios* configuration. On the other hand the area instances with an improvement benefit with an average of 1.23 %, 1.78 % and 1.31 % respectively. If these values are weighted with the area instances relevance, a further improvement of 0.62 % for the *demand-based scenarios* configuration, 0.64 % for the *forecast deviation scenarios* configuration and 0.68 % for the *mixed scenarios* configuration can be observed. Even though these numbers may seem ridiculously small, the absolute savings per year are still quite presentable and account for 15 % to 17 % of an average areas total cost in the *no delivery profiles* configuration. Third, even though the results based on the rule-based delivery schedule generation through delivery profiles provide valuable savings in comparison with the default configuration with no

order lot-sizing optimization methods in place, a state-of-the-art algorithmic approach may still provide higher savings than a rule-based approach. On average the *Kempkes* configuration provides additional savings of about 4.54 % towards the *initial forecast* configuration, which accounts for 113 % of an average areas total cost in the *no delivery profiles* configuration. The overall savings can be raised from 22.89 % for the *initial forecast* configuration or 23.57 % for the *mixed scenarios* configuration up to 27.43 % for the *Kempkes* configuration.

To understand the details of these cost reductions it is necessary to take a closer look at the cost components involved in the case study. Figure 7.5 shows the distribution of the total realized cost on the different cost components. The cost components considered in the case study include cost for full load runs, pre leg runs, main leg runs and inventory holding cost. In the *no delivery profiles* configuration the full load run accounts for 33.65 % of all cost, the pre leg run 15.94 % and the main-leg runs 33.6 % respectively. At the same time inventory holding costs of about 16.8 % occur based on the given safety stock parameters. When delivery profiles are applied based on the *initial forecast* configuration, pre leg run cost are cut by almost 50 % to about 7.41 % of the total cost in the *no delivery profiles* configuration. Main-leg run costs are also reduced by about 50 % down to 15.92 % of the total cost in the *no delivery profiles* configuration. This improvement is achieved by increasing vehicle use by cumulating orders on delivery days. This leads to an increase in both full load run cost and inventory holding cost. The full load run costs are increased as some additional full load runs can be used. Inventory holding costs increase as materials have to be ordered in advance and are then stored in inventory until they are finally used for production. In the stochastic cases for the *demand-based scenarios* configuration, the *forecast deviation scenarios* configuration and the *mixed scenarios* configuration show only little changes. There is a slight tendency to use fewer full load runs and inventory, while at the same time pre leg runs and main leg runs are used more efficiently. This may be caused by a more cautious use of delivery profiles. In the *Kempkes* configuration the inventory holding costs are further increased to lower pre leg run costs and main leg run costs. Thereby the degree of freedom in order making reveals its strength. In comparison with the *no delivery profiles* configuration both pre leg run and main leg run costs are reduced to about one third of the original value. At the same time share of inventory holding cost is further increased to 21.94 % of the total cost in the *no delivery profiles*

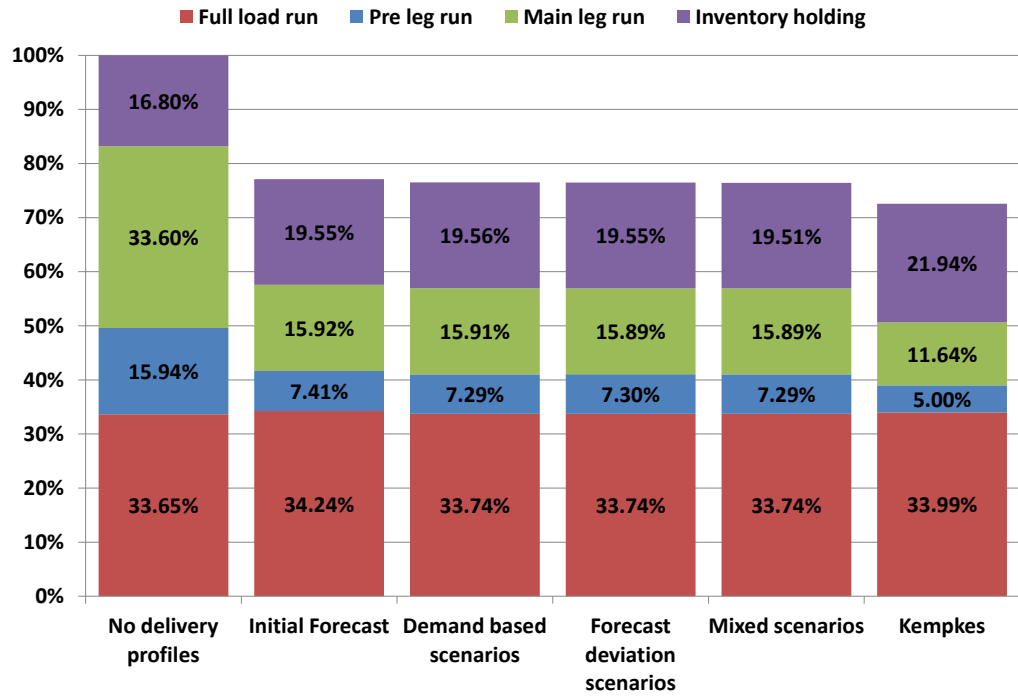


Figure 7.5: Cost components for the realized cost.

configuration.

7.4.3 Value of perfect information

In this Section a comparison is drawn between the theoretically expected costs, the realized outcome when the delivery schedule generation approaches are used in a rolling horizon, and the best possible solution that can be derived post-ex. The result of this examination is twofold. On the one hand the predictability of the outcome of the different configurations can be determined. On the other hand the disruptions caused by forecast errors and their effect on the different configurations can be numbered. To evaluate the predictability of the different configurations the expected savings are compared with the realized savings. It would not make sense to compare the absolute values as both are created upon different demand information. However, the percentage saving may still be comparable, especially if multiple values are taken into considera-

tion. If a correlation between expected and realized savings can be observed, it eases the development of a business case analysis for a possible application to practice, as the savings can be expected to be similar in both theory and practice.

	No Delivery Profiles	Initial Forecast	Demand based scenarios	Forecast deviation scenarios	Mixed scenarios	Kempkes
Expected No delivery profiles	1284.5	1284.5	1034.8554	1512.2505	1358.94081	1284.5
Expected						
- Optimal solution	-	1015.0	1019.8	1470.4	1219.5	761.6
- Savings	-	21.0%	1.5%	2.8%	10.3%	40.7%
Realized						
- Total cost	2500.0	1927.8	1912.5	1911.8	1910.8	1814.2
- Savings	-	22.9%	23.5%	23.5%	23.6%	27.4%
Savings Prediction error						
Total savings		-1.9%	-22.0%	-20.8%	-13.3%	13.3%
Average savings		-11.9%	-31.6%	-30.9%	-23.4%	7.8%

Table 7.10: Savings prediction error of the different configurations.

Table 7.10 shows the predictability of the different configurations outcome. At first sight it can be seen that the expected total values differ significantly from those that have finally been realized. The expected total values vary from about 40% of the value that has finally realized for the *demand-based scenarios* configuration to about 60 % for the *forecast deviation scenarios* configuration. This huge gap can be explained by three factors. First, the underlying expectations of future demand vary from the demand situation finally realized. Second, in the realized total cost, inventory safety parameters are considered, which force a higher overall inventory holding cost value. Third, the expected costs do not consider the rolling horizon effects which may force inefficiencies in form of orders that were not necessary and emergency orders that have to be placed. Aside from the gap in respect of the total values, differences in respect of designated savings can be observed. This gap varies between the different configurations. In the *initial forecast* configuration the total savings prediction error is about -1.9%, which means that savings are estimated to be slightly lower than the realized savings. However, the average savings prediction error is higher with a value of -11.9%. For the stochastic configurations these values are much higher, ranging from

-13.3% total savings prediction error for the *mixed scenarios* configuration to -22.0% for the *demand-based scenarios* configuration. It can be noted that all delivery profile-based configurations underestimate the possible savings. The expected savings from the stochastic approaches are so low that they cannot be considered to be valid for an approximation of the expected savings. This may be the case because the stochastic approaches have to satisfy multiple scenarios at once, thus it is harder to obtain a serious savings value as the distribution of demand among the week days - an important factor for the synergy effects in the main leg run - is different in the different scenarios and thus delivery profiles become less advantageous in comparison with a solution without delivery profiles. The *Kempkes* configuration is the only configuration that actually overestimates the savings. The expected total savings of about 40.7% cannot be held in the rolling horizon. Rather, the total savings are reduced by 13.3% to 27.4% in total. The larger areas are more strongly affected than are the smaller ones, as the average savings prediction error per area with a value of 7.8% is lower than the total savings prediction error.

To measure the disruptions caused by forecast errors and the rolling horizon effects the value of perfect information (VOPI) can be used. The value of perfect information gives back how much the realized costs could have been improved if the uncertainty on demand forecasts had not existed and the real consumption of goods had been known in advance. At the same time the value indicates how much a company would be willing to pay to eradicate the errors from the forecast. This is derived by computing an optimal post-ex solution as a lower bound on the realized cost value. This post-ex solution is then subtracted from the realized cost and the remainder is called value of perfect information. If the value is significantly high, it is a hint that investments in better forecasting systems or more stable production schedules may be profitable. Aside from the perspective on information quality, it can be used to evaluate the quality of the solution derived in the rolling horizon. The optimal post-ex solution is the lower bound for all planning approaches. Therefore the gap between the realized costs and the lower bound gives back how well the planning approach has performed in the rolling horizon, as the VOPI indicates to what degree the planning approach has been affected by the difficulties that occur in a rolling horizon planning scenario.

Table 7.11 compares the optimal savings that could have been achieved with the savings that have been realized to derive the value of perfect information. The table is

Instance	MRP	A priori		Post-ex	
		Best delivery profiles	Kempkes solution	Optimal delivery profiles	Optimal Kempkes Solution
1	43.0	20.38 (47.4%)	16.97 (39.5%)	14.66 (34.1%)	13.53 (31.4%)
2	72.8	46.02 (71.1%)	44.34 (112.4%)	51.72 (71.1%)	50.28 (69.1%)
3	31.2	18.03 (55.4%)	16.73 (95.9%)	17.29 (55.4%)	14.16 (45.3%)
4	35.0	18.50 (51.1%)	18.60 (96.9%)	17.92 (51.1%)	16.01 (45.7%)
5	13.0	8.43 (80.0%)	8.97 (122.9%)	10.36 (80.0%)	10.25 (79.1%)
6	57.3	25.92 (44.1%)	22.50 (97.6%)	25.29 (44.1%)	20.49 (35.7%)
7	273.5	237.02 (52.6%)	227.52 (60.7%)	143.87 (52.6%)	132.39 (48.4%)
8	32.9	18.49 (49.6%)	15.99 (88.2%)	16.31 (49.6%)	15.33 (46.6%)
9	12.5	7.20 (64.1%)	5.89 (110.9%)	7.99 (64.1%)	7.72 (61.9%)
10	22.7	14.83 (60.2%)	13.22 (92.3%)	13.69 (60.2%)	11.38 (50.1%)
11	43.5	30.25 (93.8%)	28.62 (134.9%)	40.83 (93.8%)	39.22 (90.1%)
12	68.0	32.68 (38.9%)	27.54 (80.8%)	26.41 (38.9%)	24.01 (35.3%)
13	29.6	19.56 (75.6%)	17.30 (114.3%)	22.36 (75.6%)	21.89 (74.0%)
14	63.2	44.56 (76.1%)	44.31 (108.0%)	48.13 (76.1%)	45.88 (72.6%)
15	37.0	27.54 (68.9%)	26.19 (92.5%)	25.47 (68.9%)	22.65 (61.3%)
16	44.8	28.15 (48.0%)	26.42 (76.4%)	21.50 (48.0%)	19.56 (43.7%)
17	87.2	53.03 (80.3%)	53.53 (132.1%)	70.07 (80.3%)	58.12 (66.6%)
18	40.4	19.81 (51.2%)	18.85 (104.5%)	20.70 (51.2%)	19.98 (49.4%)
19	106.5	73.26 (67.5%)	70.25 (98.0%)	71.82 (67.5%)	62.54 (58.8%)
20	102.1	79.80 (80.0%)	77.21 (102.3%)	81.64 (80.0%)	76.54 (75.0%)
21	212.3	200.36 (69.5%)	188.40 (73.6%)	147.48 (69.5%)	136.50 (64.3%)
22	472.6	391.76 (46.9%)	379.08 (56.6%)	221.69 (46.9%)	217.99 (46.1%)
23	59.5	38.58 (58.4%)	35.12 (90.1%)	34.75 (58.4%)	31.24 (52.5%)
24	401.0	359.31 (78.4%)	345.34 (87.5%)	314.27 (78.4%)	250.98 (62.6%)
25	138.5	95.13 (58.8%)	85.33 (85.6%)	81.40 (58.8%)	76.86 (55.5%)
Total	2500.00	1908.61	1814.23	1547.61	1395.52
Savings		23.66%	27.43%	38.10%	44.18%
VOPI absolute			94.38	361.01	418.71
% VOPI			3.78%	14.44%	16.75%

Table 7.11: Realized savings in comparison with the optimal savings that could have been achieved if all information had been available at planning time.

divided into two parts. The first part displays the best realized value in the rolling horizon for both the delivery profile based configurations and the *Kempkes* configuration. This part is called *a priori* as the delivery schedule generation approaches have worked with the information that was available at planning time. The second part of the table shows the optimal solutions of a model that has been solved based on the demand that has finally realized. This part is called *post-ex* as the optimal solution has been computed on information that was available after the considered period range. The

savings that could have been realized if all information had been available in advance and are significantly higher than the realized savings. For the delivery profile based configurations a total saving of about 38.10% could have been realized, which means that the savings could have been increased by 14.44% of the realized total costs in the *no delivery profiles* configuration. Under availability of all information the *Kempkes* configuration could have lowered the total costs by 44.18% in total, thereby increasing the savings towards the realized savings by 16.75% of the total realized costs in the *no delivery profiles* configuration. Here the limitation of the delivery profile based approaches to using delivery profiles rather than exploring the whole solution space is obvious. However, the distance between the optimal and achieved values is smaller for the delivery profile based approaches than for the *Kempkes* configuration. This is consistent with the observations on the predictability of the savings that has been discussed above. In addition it shows that delivery profile based approaches are less affected by rolling horizon difficulties than is the *Kempkes* configuration.

7.5 Stability of the generated delivery schedules

As shown in section 2.3.1 the stability of the delivery schedule plays an important role in practice as it determines how efficiently suppliers can react to the changing requirements. It will accordingly be determined if and how well delivery profiles can improve the stability of the delivery schedules in comparison with the standard MRP planning and the algorithmic delivery schedule generation. To measure the stability of the generated delivery schedules all delivery schedules have been collected during each simulation run. These schedules have then been compared and the following measures have been computed separately for each simulation run:

- **MAD** The mean absolute deviation of the observed quantities per period, measured as percentage of average demand.
- **MPE** The mean percentage error of the observed quantities per period, measured as percentage of average demand.
- Q^- The average percentage overestimation of demand. An underestimation of demand will be measured as a zero value.

- A^+ Time and quantity based measure for underestimation of demand, as described in Section 6.2.2.
- Q^+ The average percentage underestimation of demand. An overestimation of demand will be measured as a zero value.
- A^- Time and quantity based measure for overestimation of demand, as described in Section 6.2.2.

The general results of these stability assessments can be seen in the Table 7.12. Each row represents one delivery schedule generation configuration, whereas the columns cover the indicators. It can be seen that the fluctuations are significant. For the *no delivery profiles* configuration the mean absolute deviation is about 74.1 %, which means that on average the quantity-fluctuations follow a bandwidth from +75% to -75% around the targeted value. It has hereby to be considered that a shift in time will always be regarded as two changes with 100 % each, as discussed in Section 3.3. The mean percentage error indicates that the fluctuations are mainly driven by underestimations of demand, thus lower values are submitted at the beginning and are scaled up in later delivery schedules. On average the quantity is 23.8 % below the value in the next forecast. The overestimations in quantity (Q^-) are given with an average of 25.1 %, which reflect themselves in an area of overestimation (A^-) of 24.4 %. Much higher are the values for the underestimation, with 49.0 % for the quantities and 40.5% for the area of underestimation. This high fluctuation can be dampened strongly by employing delivery profile based delivery schedule generation approaches. For the corresponding configurations *initial forecast*, *demand-based scenarios*, *forecast deviation scenarios* and *Mixed scenarios*, the average mean percentage error is lowered to only 5.1 % up to 6.2 %. The overestimation of demand is slightly increased, from 25.1% to around 28 % for Q^- and from 24.4 % to 24.7 % for the area of overestimation. However, the much more critical underestimations are drastically reduced. The average quantity underestimation is lowered from 49 % to around 34 %, and the area of underestimation can be decreased from 40 % to 16 %. Thus it can be said that employing delivery profiles reduces the overall fluctuation in demand especially by lowering the undesired underestimations of demand. Comparison of the different techniques which are based on

delivery profiles shows that the *initial forecast* configuration provides stronger stability improvements than do the three stochastic cases. This holds true for all considered indicators. The numbers for the *Kempkes* configuration also show an improvement towards the *no delivery profiles* configuration. However, in comparison with the delivery profile based configuration the *Kempkes* configuration generates less stable delivery schedules. Interestingly, the *Kempkes* configuration tends more often to underestimate than to overestimate demand. This leads to an increased mean percentage error (17.1 % for the *Kempkes* configuration vs. 5.1% for the *initial forecast* configuration) and increased values for Q^+ and A^+ (41.2 % vs 33.3 % and 26.9 % vs 16.0 % respectively). Though the key figures for the underestimation of demand are worse than in the delivery profile based configurations, the figures for overestimations of demand are even better than in the *no delivery profiles* configuration.

	MAD	MPE	Q^-	A^-	Q^+	A^+
No delivery profiles	74.1%	23.8%	25.1%	24.4%	49.0%	40.5%
Initial forecast	61.5%	5.1%	28.2%	24.6%	33.3%	16.0%
Demand based scenarios	61.6%	5.5%	28.1%	24.5%	33.6%	16.6%
Forecast deviation scenarios	62.1%	6.2%	28.0%	24.2%	34.2%	17.1%
Mixed scenarios	62.4%	5.5%	28.5%	24.7%	34.0%	16.8%
Kempkes	65.3%	17.1%	24.1%	20.9%	41.2%	26.9%
Average	64.5%	10.5%	27.0%	23.9%	37.5%	22.3%

Table 7.12: Overview of key stability indicators for the different configurations.

Not all parts are affected in the same intensity by the positive effects on the stability of the generated delivery profiles. Especially for the delivery profile based approaches it has to be considered that some suppliers have the delivery profile W11111 assigned, which will lead to a delivery schedule that equals the delivery schedule from the *no delivery profiles* configuration. In Table 7.13 only the parts which do not have delivery profile W11111 assigned in the *initial forecast* configuration are considered. For these parts it can be shown that the improvements are much stronger than they are on average. These parts are mainly those with infrequent demands that pose additional trouble for both the suppliers and the OEMs due to their irregular demands. The mean absolute deviation is almost halved from 68.6% in the *no delivery profiles* configuration to 35.1 % in the *initial forecast* configuration. The mean percentage error is reduced to 1.1 % from 30.5 %. This is achieved by a slight reduction of overestimations and a

significant reduction in underestimations. The quantity-based measure Q^- is reduced from 19.1 % down to 17.0 % and the time and quantity based measure A^- is reduced from 19.3 % to 13.9 %. In the case of underestimations the improvement is even more significant, reducing the value for Q^+ from 49.6% to 18.1% and the value for A^+ from 41.6% to 8.9%. It has to be mentioned that the distance between the *initial forecast* configuration and the three stochastic delivery profile selection approaches becomes larger in this analysis. As the base for the given analysis is the set of parts that have not assigned the delivery profile W11111 in the *initial forecast* configuration, it may be that other delivery profiles are assigned to the corresponding suppliers in the other configurations. Therefore these values cannot be seen as representative. However, it is possible to draw a comparison with the *Kempkes* configuration. Again, the pattern of lower overestimations but higher underestimations can be observed for this subset of parts. The Q^- and A^- values are close to zero with 7.5% and 7.0% respectively, whereas the underestimation values Q^+ and A^+ are quite high with 36.8% and 32.4%. The underestimation of part demands in subsequent delivery schedules is the most

	MAD	MPE	Q^-	A^-	Q^+	A^+
No delivery profiles	68.6%	30.5%	19.1%	19.3%	49.6%	41.6%
Initial forecast	35.1%	1.1%	17.0%	13.9%	18.1%	8.9%
Demand based scenarios	49.2%	4.5%	22.3%	19.0%	26.9%	15.4%
Forecast deviation scenarios	53.3%	5.6%	23.9%	18.9%	29.5%	16.7%
Mixed scenarios	51.4%	4.2%	23.6%	19.2%	27.8%	15.3%
Kempkes	44.4%	29.3%	7.5%	7.0%	36.8%	32.4%
Average	50.3%	12.5%	18.9%	16.2%	31.4%	21.7%

Table 7.13: Overview of key stability indicators for parts that do **not** have delivery profile **W11111** assigned in the initial forecast configuration.

important indicator of instability in delivery schedules. Demands that have been increased or shifted forward may pose problems for the supplier's production planning. To provide a more detailed analysis of this key figure the parts have been grouped into categories based on their average underestimation. Figure 7.6 shows the distribution of the parts into the different risk groups. Parts without any underestimations are grouped in the dark green section; parts with an average between 0 % and 10 % underestimation are grouped into the second section marked in a slightly lighter shade of green; parts with an average between 10 % and 20 % in the third section marked with a very light green, and so on. The dark red section covers parts with an average

underestimation of more than 50 %. On the x-axis the different configurations are listed. Each configuration has a bar assigned that gives back the percentage distribution of parts within each risk group. In the *no delivery profiles* configuration the share of parts with an acceptable level of underestimations with values below 20 % is 35.3 %. Another 33.5 % of parts are within 20 % and 50 % underestimation, while 31.2 % of all parts have such instable delivery schedules that underestimations are higher than 50 %. In the *initial forecast* configuration 77.8 % of all parts are in the acceptable region. Only 15.4 % are in the middle groups between 20 % and 50 %, and the group of very instable parts with more than 50 % underestimation is reduced to 6.7 %. The share of parts with huge underestimations from one schedule to another are reduced drastically, more than three out of four parts are acceptable. The limitation of worst parts down to 6.7 % is an especially worthy improvement. The only drawback is that the number of parts with no underestimation at all is slightly reduced from 26.2 % to 20.9 %. However, 54.7 % of all parts are below 10 % average underestimation, which is an improvement of 26.4 %. For all four delivery profile based methods the figures are similar. It can be observed that the configurations that chose delivery profiles based on stochastic models provide slightly smaller enhancements. The *Kempkes* configuration provides better results than the *no delivery profiles* configuration. However, the underestimations from one delivery schedule to another is significantly higher than in the delivery profile based configurations. The share of parts in the acceptable range below 20 % is reduced by 21.4 % to 56.4 %, even though the share of parts with no underestimations is increased from 20.9 % to 25.7 %. More critical than the reduction of acceptable parts is the increased share of parts with more than 50 % uncertainty, which is 17.4 % in the *Kempkes* configuration. The share of these irregular parts is thereby doubled towards the delivery profile based configurations.

In summary it can be stated that all other configurations provide an improvement in stability towards the *no delivery profiles* configuration. The delivery profile based configurations yield the highest improvement in stability, with slight differences between the configurations. The deterministically chosen delivery profiles in the *initial forecast* configuration provide the strongest enhancement but the distance to the stochastic configurations is small. The *Kempkes* configuration provides an improvement towards the *no delivery profiles* configuration, but is considerably less successful in terms of delivery schedule stability than the delivery profile based configurations. In addition it

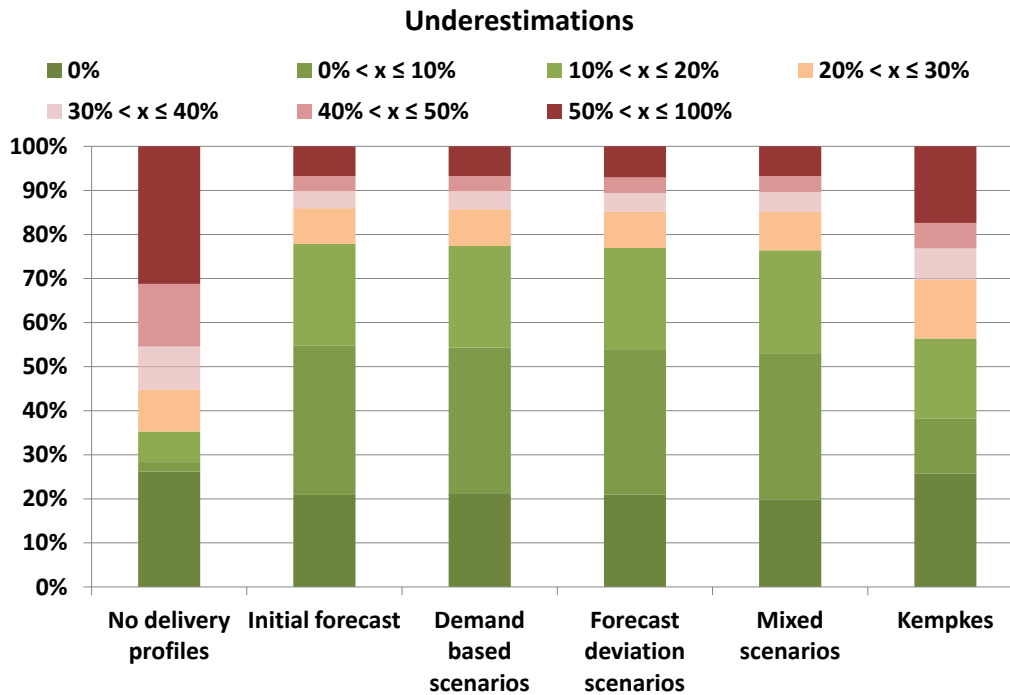


Figure 7.6: Distribution of parts based on the percent degree of underestimation.

can be stated that the *Kempkes* configuration tends greatly to underestimate demands, whereas the delivery profile based configurations have overestimations and underestimations on a similar level.

7.6 Inventory behavior

The level of inventory should be determined by the given safety parameter settings. As the safety lead-time component provides a dynamic inventory level, the desired level cannot exactly be determined beforehand, but is rather a result of the simulation process. Every delivery schedule generation approach discussed in this thesis uses a time-based consolidation approach which requires additional inventory in comparison with the basic MRP scheduling. It was therefore analyzed what effects this additional inventory has had on the inventory performance in total. Two indicators are of special

importance. On the one hand the additional inventory deployed may have a positive effect on the number of necessary escalation processes. Escalation processes should be avoided if possible. Thus the influence of the delivery schedule generation approach on the number of necessary escalation processes is measured and analyzed. On the other hand it may be that due to the uncertain future demand inventory is built up according to a forecast and never used afterwards, leading to the necessity to scrap material in the worst case. Accordingly, the effect of the underlying delivery schedule generation approach on the quantity of excess inventory is discussed in the following.

Figure 7.7 shows the number of escalation processes in relation to the inventory holding costs for each configuration that has been measured in the simulation study. In so doing both the number of escalation processes and the inventory holding costs are normalized so that the number of escalation processes in the *no delivery profiles* configuration equals a value of 100 %. It can be seen that employing delivery profiles independently from the planning configuration significantly improves the total number of escalation processes. The total number of escalation processes is reduced by more than 52.5 % for all four configurations. At the same time the inventory holding costs are increased by about 16 %. It can therefore be said that the additional inventory holding costs are justifiable both in terms of freight cost reduction and increased supply safety. This holds true for all configurations with delivery profiles involved. Slight differences appear between the exact values but the trend is the same. The algorithmic delivery schedule generation approach from the *Kempkes* configuration also deploys additional inventory. Thus the number of escalation processes is also reduced. But, as can be seen, the effect is not as strong as it is for the delivery profile related configurations, even though the additional inventory holding costs are higher. Here a 30.5 % increase in inventory holding costs results in a reduction of only 33.5 % in escalation processes. Thus, it can be argued that the additional inventory is used less efficiently by the algorithmic approach. For the configurations using the delivery profile control rule to generate delivery schedules a one-percent increase in inventory holding costs yields 3.25 % reduction in escalation processes, whereas the *Kempkes* configuration only yields 1.1 %. Of course these numbers cannot be transferred to other contexts, but for the given case study it can be argued that the additional inventory created by the delivery profile based approaches is almost three times as effective as the inventory built up by the algorithmic approach in terms of increasing supply safety.

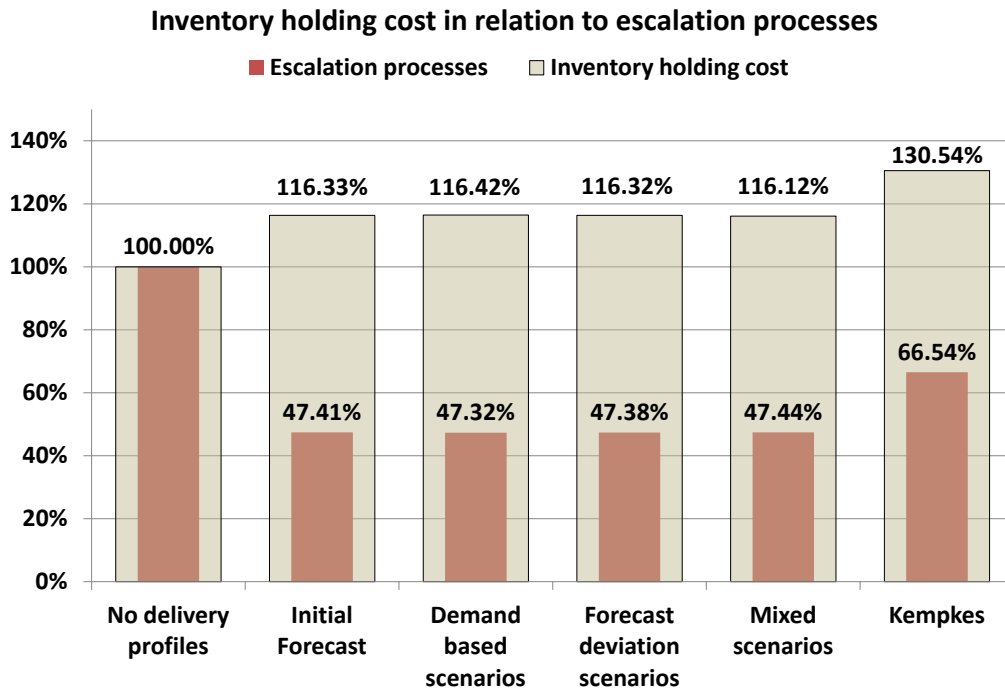


Figure 7.7: Number of escalation processes, percentage of orders without escalation processes and inventory holding costs for each configuration.

Figure 7.8 illustrates the inventory level over time for the *no delivery profiles*, *initial forecast* and the *Kempkes* configurations. The values are normalized on the average inventory in the *no delivery profiles* configuration. The normalized consumption volume is also given. The inventory rises over the whole simulation horizon in all three configurations. A first assumption may be that the consumption volume also rises and causes additional inventory due to the given safety parameters. But as can be seen in the diagram this is not the case. Rather, the consumption volume slowly decreases over time. The increased inventory levels can be explained by two facts. On the one hand, the initial inventory value is an idealized value, as is explained in Section 6.1.5. On the other hand the forecasts contain overestimations of demand in various periods. Over time these overestimations pile up as goods are ordered but not consumed directly or even at all during the whole simulation horizon. These parts provide a certain risk of

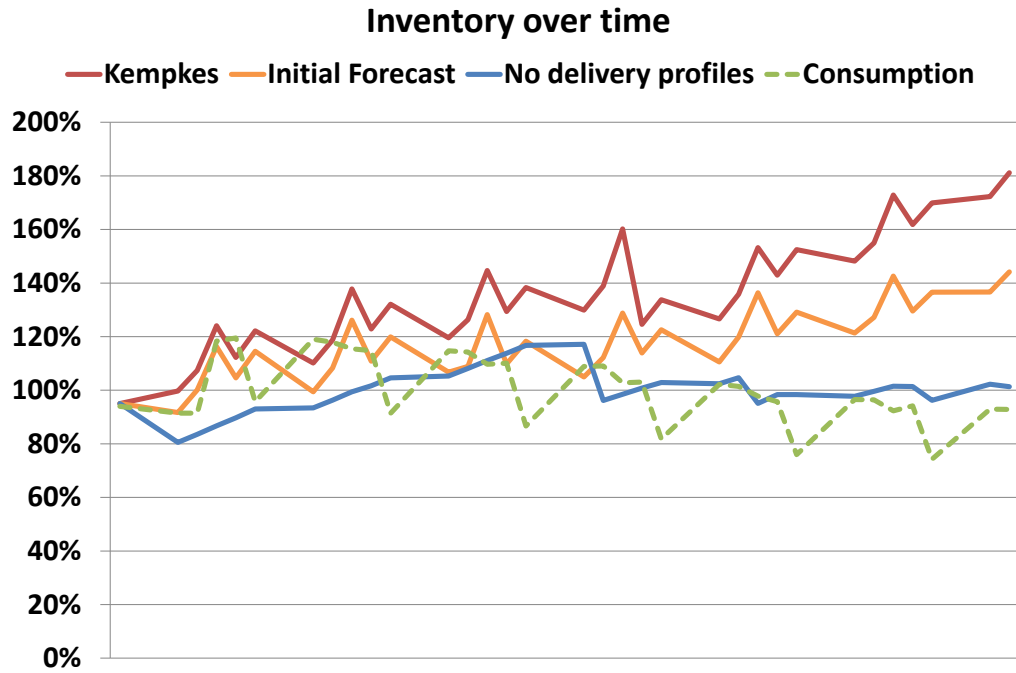


Figure 7.8: Inventory level over time for three selected configurations.

being scrapped in the end, as it is unsure whether or not they will be used in future. For the *no delivery profiles* configuration the effect is the weakest. At the beginning the inventory is about the idealized 95.1 % of the average inventory. At the end of the simulation horizon inventory value rose to 107.8 %. For the *initial forecast* configuration, however, the inventory value rises to 151.7 %. In the *Kempkes* configuration the value is doubled to 190.7 % at the end. It can also be seen that the *Kempkes* configuration and the *initial forecast* configuration produce irregular inventory peaks. These peaks can be explained by the deployment of inventory to achieve freight cost reductions. They are higher in the *Kempkes* configuration than in the *initial forecast* configuration. The extend of the peaks is reflected by the standard deviation of the inventory value per period. The standard deviation for the *no delivery profiles* configuration is the lowest with 8.2 %, thereafter following the *initial forecast* configuration with 13.4 % and the *Kempkes* configuration with a standard deviation of 21.5 %.

	No Delivery Profiles	Initial Forecast	Demand based scenarios	Forecast deviation scenarios	Mixed scenarios	Kempkes
Average days on stock	4.19	5.98	6.02	6.06	6.01	7.13
% of parts with						
· inventory > safety inventory	25.25%	33.92%	34.12%	33.69%	33.84%	39.72%
· days on stock > 1 month	16.80%	19.51%	19.70%	19.31%	19.25%	20.19%
Value of above parts	36.26	39.64	40.09	39.93	39.83	43.23
Difference vs. <i>no delivery profiles</i>		3.38	3.84	3.68	3.57	6.98
Increase in %		9.34%	10.58%	10.14%	9.86%	19.25%

Table 7.14: Figures on end-of-simulation inventory for the different configurations.

Table 7.14 shows the most important figures on end-of-simulation inventory for the different configurations. Each column gives back the numbers for a specific configuration. To collect the figures the inventory at the end of the simulation has been gathered and compared with the forecast given in the last simulation period. The days on stock have then been computed for each part by walking over the days after the simulation range in the forecast and marking the day on which the inventory is going to be exhausted if no additional orders are placed. The number of working days between the last simulation day and the day of inventory exhaustion is the number of days on stock. The desired safety inventory has then been calculated for each part as described in Section 6.1.5 and the maximum quantity of one load carrier has been added. Based on these numbers, the share of parts with more than the desired safety inventory and one load carrier has been computed. If parts inventory could not be exhausted during the first month after the last day of simulation, they have been added to the share of parts with days on stock larger than one month. For these parts the value of inventory has been computed and normalized by the average areas total cost. This value is that of parts that have a risk of being scrapped if they are replaced by a newer version of the part within the next month. The last rows show the difference in these values towards the *no delivery profiles* configuration in absolute normalized values as well as in percentage values. The percentage of parts with more than safety inventory and one load carrier in inventory is about one quarter for the *no delivery profiles* configuration and rises to about one third for the delivery profile based configurations. For the *Kempkes* configuration the value is about 40%, meaning that two out of five parts have more inventory than desired. The number of parts with days on stock above one month varies from 16.8 % in the *no delivery profiles* configuration to about 20 % for all other configurations. Here no significant difference between the delivery profile based

and the *Kempkes* configurations can be found, even though the value is slightly higher. The value of these parts and their inventory, however, is higher than in the delivery profile based configurations. Here the difference towards the *no delivery profiles* configuration is about 3.38 to 3.84 average areas, whereas the *Kempkes* configuration has a doubled additional value at risk of 6.98. In percentage of the value of the inventory at risk, this means an increase from between 9.34% and 10.58% for the delivery profile based configurations up to 19.25% for the *Kempkes* configuration. Thus the value at risk of being scrapped has increased by about 10 % for the delivery profiles and twice as much by about 20 % for the *Kempkes* configuration. Still, these numbers are small in relation to the overall savings achieved by the different configurations.

7.7 Summary of the case study results

The merits gained from the case study are threefold. First, the models and algorithms presented in this work are evaluated in terms of runtime and applicability to problems from practice. Second, information on the predictability of the delivery schedule-generation approaches outcome and their relation to the best possible solutions was gathered. Third, implications for possible implementations in practice can be drawn based on the different configurations' behavior in the given circumstances. The following will first discuss how the algorithmic results could be interpreted. It will then be determined to what degree the delivery schedule generation algorithms are predictable if deployed in practice. Finally, the pros and cons of the different configurations are discussed to provide an overview of the implications for practical applications.

7.7.1 Algorithmic results

The solution algorithms presented in this thesis provided good results. It could be shown that given the decomposition approach it was possible to solve most instances to optimality with only few exceptions. The increase in runtime for the stochastic case could be limited to a linear progression. The two heuristic algorithms performed well; the genetic algorithm especially provides solutions that are close to the optimum in most cases. The density-based packing concept showed its strength, as it is fast and reliable. The heuristic approaches did not exceed the maximum runtime of 30 minutes except in one single case on the largest stochastic model. In the case of the few exceptions in which optimality could not be proven within the limited time the

genetic algorithm still provided the best solution. For this reason the formulation of the mathematical model can be seen as the weak-point in terms of algorithmic quality. The standard mixed-integer solver implementations takes a long time to prove the optimality of optimal solutions, which suggests that the linear programming relaxation of the model could still be improved.

7.7.2 Predictability and quality of applications in a rolling horizon

As far as the predictability of the different approaches outcome is concerned, it can be stated that the *initial forecast* configuration provides the best results. The stochastic delivery profile based configurations, however, provide the worst predictability. Among the stochastic approaches the *mixed scenarios* configuration provides the best prediction quality. Whereas all delivery profile based configurations underestimate the savings, the *Kempkes* configuration overestimates the savings that can be realized. In addition, its predictability is higher on small than on larger areas - which may be a hindrance if applications in practice are considered. The quality of the the application in a rolling horizon can be evaluated from two sides. On the one hand it can be seen that the *Kempkes* configuration would be able to achieve higher savings than the delivery profile based configurations if all information were available at planning time. At the same time the delivery profile based configurations get closer to the optimal value and thus provide a lower value of perfect information. This leads to the gap between the *Kempkes* configuration and the delivery profile based configurations being narrowed in a rolling horizon application.

7.7.3 Implications for applications in practice

As could be seen in the case study each configuration has its strengths and weaknesses. When implementing a delivery schedule generation approach in practice, multiple targets have to be considered and balanced according to companies' preferences and demands. The case study presented in this chapter provides hints on the applicability of the different approaches depending on these preferences and demands. Multiple insights can be drawn from the given case study and are briefly summarized in the following.

Realized cost It has to be noted that from a purely freight cost orientated perspective the *Kempkes* configuration provides the best results. On average the sum of freight cost and inventory holding cost can be reduced by 27.43 % towards the *no delivery profiles* configuration. The distance to the best delivery profile based approach, the *mixed scenarios* configuration, which reduces freight cost by about 23.57 %, is 4.54 % of the initial freight and inventory holding cost. It has to be noted, however, that these values may have to be reevaluated for other relations between freight cost and inventory holding cost. Especially for more expensive parts or higher capital interest rates, it may be that the potential savings are reduced.

Applicability of stochastic approaches Slight cost reductions could be achieved by employing stochastic programming methods to identify more robust delivery profiles. On average 0.65 % additional savings could be generated. It has to be considered, though, that these improvements can only be used in practice if an infrastructure for data collection and preparation is set up, and that the complexity of the involved algorithms is increased.

Stability of delivery schedules The stability of delivery schedules could to an extent be improved both by the delivery profile based configurations and the *Kempkes* configuration. However, the delivery profile based configurations do provide a much higher level of stability than the *Kempkes* configuration. Especially considering the more dangerous underestimations of demand, the stability of the generated delivery schedules of the delivery profile based approaches with an average of 16.6 % for A^+ is far better than in the *Kempkes* configuration with 26.9 %. The share of stable parts is also much higher, reaching 77.1 % on average for the delivery profile based configurations in comparison to only 57.7 % for the *Kempkes* configuration. At the same time the number of absolutely irregular parts is significantly lower with 6.8 % on average towards 17.0 % in the *Kempkes* configuration.

Escalation processes The number of required escalation processes can be reduced by employing a delivery schedule generation approach. For the delivery profile based approaches an average improvement of 52.61 % towards the *no delivery profiles* configuration can be achieved, whereas the *Kempkes* configuration still provides an advantage of 33.46 %. Given that the delivery profile based configurations use less inventory to

achieve these results, it can be said that the additional inventory is used more efficiently. The delivery profile based approaches use a one-percent increase in inventory holding cost to yield 3.25 % reduction in escalation processes, whereas the *Kempkes* configuration only yields 1.1 % per cent of additional inventory piled up.

Inventory behavior The inventory level suffers greater fluctuations if a delivery schedule generation approach is used. The highest fluctuation can be observed for the *Kempkes* configuration, with a standard deviation of 21.5 %, in contrast to 13.4 % for the *initial forecast* configuration and only 8.2 % for the *no delivery profiles* configuration. Aside from the increased fluctuation the tendency of collecting unnecessary parts in inventory also increases. The value of inventory at risk of being scrapped is increased by about 10 % for the delivery profile based configurations, whereas it is increased by about 20% for the *Kempkes* configuration.

8 Summary and Conclusion

In the following the results of this thesis will briefly be summarized to give a retrospective on the contributions in this thesis. An outlook on possible further research will then be given to provide a starting point for researchers that want to extend the work in this area.

8.1 Summary of the achieved contributions

In this thesis methods of selecting cost-minimal delivery profiles for application in area forwarding inbound logistics networks have been developed, and an assessment of their impact on both cost and delivery schedule stability has been conducted. In Chapter 2 the problem setting was identified in its broader context and thereafter depicted in detail. It was pointed out that two major options are available for delivery schedule generation in general, namely algorithmic schedule generation and rule-based schedule generation. While algorithmic approaches repeatedly built new delivery schedules from scratch in each planning iteration, rule-based approaches make a tactical decision to identify a control rule and then apply the control rule to create the delivery schedules. A promising control rule that has been discussed both in the literature and practice is that of so-called delivery profiles, which define a subset of days on which deliveries should take place. Both types of schedule-generation approach propose different advantages in respect of the given goals: cost reduction, delivery schedule stability and avoidance of escalation processes. As could be derived from the literature review in Chapter 3, where the most important literature on the topic has been presented and discussed in respect of its applicability to the given problem setting, no scientific validation of these propositions has yet taken place. To close the gap between the demand defined in the problem setting and the existing literature, multiple contributions have been targeted and will be named in the following.

A deterministic model formulation and efficient solution algorithms A deterministic model formulation and efficient solution algorithms for the selection of cost-

minimal delivery profiles were presented in Chapter 5. A decomposition-based solution algorithm was developed to determine cost-minimal delivery profiles. A mixed-integer model formulation was set up that allows treatment of various forms of tariff system and can be applied to all area forwarding networks following the common industrial practice. To solve the model two heuristic algorithms were implemented, namely a sequential heuristic based on the concept of the hamming-neighborhood and a genetic algorithm that uses the delivery profile assignments as genotypes. These heuristic algorithms share a sophisticated packing heuristic that uses the concept of load and vehicle density to obtain good packing results with high computational efficiency.

Consideration of demand uncertainty The decomposition-based solution algorithm and the model formulation were extended to a stochastic programming formulation in order to cope with the uncertainty involved in the tactical delivery profile selection decisions. The heuristic algorithms were extended so that they can deal with both the stochastic and the deterministic model formulations. In addition, a simplified model formulation was developed for the stochastic case. The simplified model formulation does not provide the same universality in terms of modeling tariff systems, but it is much more efficient and can be applied in cases where computational efficiency is valued more highly than the flexibility of the modeling approach. To support the stochastic methods with plausible scenarios, two scenario-generation approaches were developed, one using historical consumption data to create scenarios, the other using historical data on forecast errors to derive scenarios based on an actual forecast. Both scenario generation approaches can be combined with scenario-reduction techniques to obtain a subset of scenarios that reflect the desired probability distributions.

Development of an evaluation framework To assess the impact of the delivery profile based delivery schedule generation, a simulation environment for planning methods for the operational order lot-sizing problem in area forwarding based inbound logistic networks with complex tariff structures was developed and is described in Chapter 6. A simulation environment was set up that covers the whole process of delivery schedule generation based on gross demand forecasts and current inventory levels. The simulation environment can be equipped with data from practice, and can then simulate a rolling horizon planning environment under realistic conditions. Different delivery schedule generation approaches can be connected over a common interface, allowing

them to be examined under the same conditions and their results to be compared. During the examination data on logistics cost and schedule stability can be gathered. To obtain the latter a new indicator for delivery schedule stability has been developed. Unlike approaches discussed in literature, both time shifts and quantity changes are considered and consolidated in a single value.

Assessment through case study The simulation approach was then used to evaluate the methods developed and compare them with state-of-the-art algorithmic approaches from the literature in a large scale case study in Chapter 7. In the case study, multiple insights into the impact on both cost and delivery schedule stability in respect of an application in a rolling horizon planning environment was derived. It was shown that significant savings can be achieved by deploying the methods to select optimal delivery profiles described in this thesis. These savings are not as high as the savings that can be observed when state-of-the-art algorithmic approaches from literature are used. However, it could be shown that delivery profile based approaches provide advantages in respect of delivery schedule stability, inventory behavior and the number of necessary escalation processes. The pros and cons of the various approaches were widely discussed in the case study. Aside from the economic perspective, the case study provided a proof of concept for the presented solution algorithms. The heuristic approaches worked well and provided close-to-optimal solutions in all cases. The runtime of the heuristic algorithms was very short, which indicates that these are highly efficient. The model formulation, however, still leaves room for improvement, as standard mixed-integer solvers take very long to prove the optimality of a given solution.

8.2 Outlook on further research

To provide an outlook on future research, two major directions have to be looked at. On the one hand, improvements to the models and methods depicted in this thesis could be made, for example, to the model formulations. A formulation that would tighten the linear programming relaxation could, in particular, help to prove optimality in multiple cases. On the other hand, the problem setting could be extended to include more alternative delivery schedule generation approaches. Recently, some automotive manufacturers have relaxed the boundaries on the delivery schedule generation approaches. In the past only approaches that do not shift orders forward but

rather order demands earlier to obtain synergy effects have been accepted. Lot-sizing models for operational order lot-sizing and production applications therefore do not allow dissatisfaction of demand. The main argument is that demand fulfillment is the highest goal. Though this is true, the implications for model development may have been mistaken. In fact safety parameters are used to protect against stock-outs. As the demand that is passed to the delivery schedule generation approaches is the net dependent demand after safety parameter consideration, the safety stock cannot be used to provide flexibility. The idea is that if this barrier were weakened, more stable delivery schedules could be generated. One promising approach in this direction was developed at the Technical University Braunschweig in cooperation with a German automotive manufacturer (see Grunewald [2011] and Grunewald [2012]). Thus it could make sense to extend the comparison to approaches from that direction. Another interesting possibility would be to develop a hybrid approach that combines the highly flexible and effective dynamic lot-sizing approach from Kempkes and Koberstein [2010] with the regularity of delivery profiles that provokes a high schedule stability. Such an approach could, for example, use regularity as an objective, as do approaches from public transport.

Bibliography

- Yasemin Aksoy and S. Selcuk Erenguc. Multi-Item Inventory Models with Co-ordinated Replenishments: A Survey. *International Journal of Operations & Production Management*, 8(1):63–73, 1988.
- Kurt Andler. *Rationalisierung der Fabrikation und optimale Losgröße*. 1929.
- David Arnott and Gemma Dodson. Decision Support Systems Failure. In Frada Burstein and Clyde W. Holsapple, editors, *Handbook on Decision Support Systems 2*, pages 763–790. Springer, 2008.
- M. Avriel and A.C. Williams. The Value of Information and Stochastic Programming. *Operations Research*, 18(5):947–954, 1970.
- S Axsäter. *Inventory control*. International series in operations research & management science. Springer, 2006.
- J Banks. *Handbook of simulation: principles, methodology, advances, applications, and practice*. A Wiley-Interscience publication. Wiley, 1998.
- Robert T Barrett and Lawrence Laforge. A study of Replanning Frequencies in a Material Requirements Planning System. *Computers & Operations Research*, 18(6): 569–578, 1991.
- Karen Lynn Basinger. *Impact of Inaccurate Data on Supply Chain Inventory Performance*. Phd, The Ohio State University, 2006.
- Helmut Becker. *High Noon in the Automotive Industry*. Springer, 2006.
- W.C. Benton. Quantity discount decisions under conditions of multiple items, multiple suppliers and resource limitations. *International Journal of Production Research*, 29(10):1953–1961, 1991.

- W.C. Benton and Seungwook Park. A classification of literature on determining the lot size under quantity discounts. *European Journal Of Operational Research*, 92(2): 219–238, 1996.
- Joseph D Blackburn, Dean H Kropp, and Robert A Millen. A Comparison of Strategies to Dampen Nervousness in MRP Systems. *Management*, 32(4):413–429, 1986.
- J.H. Blackstone, D.T. Phillips, and G.L. Hogg. A state-of-the-art survey of dispatching rules for manufacturing job shop operations. *International Journal of Production Research*, 20(1):27–45, 1982.
- T Blecker, W Kersten, and C Herstatt. *Key factors for successful logistics: services, transportation concepts, IT and management tools*. Operations and Technology Management. Erich Schmidt Verlag, 2007.
- F. F. Boctor, G. Laporte, and J. Renaud. Models and algorithms for the dynamic-demand joint replenishment problem. *International Journal of Production Research*, 42(13):2667–2678, 2004.
- James H Bookbinder and Jin-Yan Tan. Strategies for the Probabilistic Lot-Sizing Problem with Service-Level Constraints. *Management Science*, 34(9):1096–1109, 1988.
- D C Bose. *Inventory Management*. Prentice-Hall Of India Pvt. Ltd., 2006.
- Donald J. Bowersox, David J. Closs, and M. Bixby Cooper. *Supply Chain Logistics Management*. Mc Graw Hill, 2007.
- Paolo Brandimarte. Multi-item capacitated lot-sizing with demand uncertainty. *International Journal of Production Research*, 44(15):2997–3022, 2006.
- Robert L. Bregman and Edward A. Silver. A Modification of the Silver-Meal Heuristic to Handle MRP Purchase Discount Situations. *Journal of the Operational Research Society*, 44(7):717– 723, 1993.
- Lisbeth Buschkühl, Florian Sahling, Stefan Helber, and Horst Tempelmeier. Dynamic capacitated lot-sizing problems: a classification and review of solution approaches. *OR Spectrum*, 32(2):231–261, 2008.

- K Button, H Vega, and P Nijkamp. *A Dictionary of Transport Analysis*. Elgar Original Reference Series. Edward Elgar Pub, 2011.
- Thomas E Callarman and Robert S Hamrin. A Comparison of Dynamic Lot Sizing Rules for Use i n a Single Stage MRP System with Demand Uncertainty. *International Journal of Operations & Production Management*, 4(2):39–48, 1983.
- Amiya K Chakravarty. Joint Inventory Replenishments with Group Discounts Based on Invoice Value. *Management Science*, 30(9):1105–1112, 1984.
- Sohail S Chaudhry, Frank G Forst, and James L Zydiak. Vendor selection with price breaks. *European Journal Of Operational Research*, 70(1):52–66, 1993.
- Ping-shun Chen. *Cost Minimalization in Multi-Commodity, Multi-Mode Generalized Networks with Time Windows*. PhD thesis, Texas A & M University, 2005.
- Chung-Mei Ho Chern. A Multi Product Joint Ordering Model with Dependent Setup Cost. *Management Science*, 20(7):1081–1091, 1974.
- S Y Chou, A Trappey, J Pokojski, and S Smith. *Global Perspective for Competitive Enterprise, Economy and Ecology: Proceedings of the 16th ISPE International Conference on Concurrent Engineering*. Advanced Concurrent Engineering. Springer, 2009.
- Teodor G Crainic and Jean-marc Rousseau. Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research Part B: Methodological*, 20(3):225–242, 1986.
- Teodor Gabriel Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.
- J. J. DeMatteis. An economic lot-sizing technique, I: The part-period algorithm. *IBM Systems Journal*, 7(1):30–38, 1968.
- A Drexl and A Kimms. Lot sizing and scheduling - Survey and extensions. *European Journal Of Operational Research*, 99(2):221–235, 1997.

- J Dupacová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming An approach using probability metrics. *Mathematical Programming, Series A*, 511(95):493–511, 2003.
- B P Dzielinski, C T Baker, and A S Manne. Simulation Tests of Lot Size Programming. *Management Science*, 9(2):229–258, 1963.
- Torsten Fahle, Stefan Schamberger, and Meinolf Sellmann. Symmetry Breaking. In Toby Walsh, editor, *Principles and Practice of Constraint Programming – CP 2001*, volume 2239 of *Lecture Notes in Computer Science*, pages 93–107. Springer Berlin / Heidelberg, 2001.
- Awi Federgruen and Chung-ye Lee. The Dynamic Lot Size Model with Quantity Discount. *Naval Research Logistics*, 37(5):707–713, 1990.
- Bernhard Fleischmann. On the use and misuse of holding cost models. In Paul Stähly and Peter Kischka, editors, *Models, methods and decision support for management*, pages 147–164. Springer, 2001.
- J Forrester. Industrial Dynamics, A major breakthrough for decision makers. *Harvard Business Review*, 36:67–96, 1958.
- John W Fowler and Oliver Rose. Grand Challenges in Modeling Manufacturing Systems. *Simulation*, 80(9):469–476, 2004.
- G Ghiani, G Laporte, and R Musmanno. *Introduction to logistics systems planning and control*. Wiley-Interscience series in systems and optimization. J. Wiley, 2004.
- F. Glover and G.A. Kochenberger. *Handbook of metaheuristics*. International series in operations research & management science. Kluwer Academic Publishers, 2003.
- Suresh K Goyal and Ahmet T Satir. Joint replenishment inventory control : Deterministic and stochastic models. *European Journal Of Operational Research*, 38(1):2–13, 1989.
- H Graf. *Innovative logistics is a vital part of transformable factories in the automotive industry*. Springer, 2006.

- T.; Spengler T. Grunewald, M.; Volling. Towards leveled inventory routing for the automotive industry. *Operations research proceedings 2010 : selected papers of the annual International Conference of the German Operations Research Society (GOR) at Universität der Bundeswehr München*, pages 423–428, 2011.
- T.; Spengler T. Grunewald, M.; Volling. Combining routing and leveled replenishment for the automotive industry under stochastic demand. *Applied Mathematical Optimization and Modelling - APMOD 2012 Extended Abstracts*, 8:181–188, 2012.
- T Gudehus and H Kotzab. *Comprehensive logistics*. Springer, 2009.
- H O Günther and H Meyr. *Supply Chain Planning: Quantitative Decision Support and Advanced Planning Solutions*. Springer, 2009.
- A L Haghani. Formulation and Solution of a Multi-Commodity, Multi-Modal Network Flow Model for Disaster Relief Operations. *Transportation Research Part A: Policy and Practice*, 30(3):231–250, 1996.
- R.W. Hamming. Error detecting and error codes. *The Bell System Technical Journal*, 29(2):147–160, 1950.
- Stephen Hanman. Benchmarking Your Firm’s Performance with Best Practice. *The International Journal of Logistics Management*, 8(2):1–18, 1997.
- F. W. Harris. How Many Parts to Make at Once. *Factory, The Magazine of Management*, 10(2):135–136, 1913.
- Holger Heitsch and Werner Römis. Scenario Reduction Algorithms in Stochastic Programming. *Computational Optimization and Applications*, 24:187–206, 2003.
- Frank Herrmann. SIM-R / 3 : Softwaresystem zur Simulation der Regelung produktionslogistischer Prozesse durch das R / 3-System der SAP AG. *Wirtschaftsinformatik*, 49(2):127–133, 2007.
- J L Heskett, N A Glaskowsky, and R M Ivie. *Business logistics; physical distribution and materials management*. Number Bd. 1. Ronald Press Co., 1973.
- C.-J. Ho and T.C. Ireland. Correlating M RP system nervousness with forecast errors. *International Journal of Production Research*, 36(8):2285–2299, 1998.

- Juhwen Hwang. Improving Supplier â€™ s Performance Using Common Replenishment Epochs in a Vendor-Managed Inventory System. *Operations Research*, (2002):199–206, 2008.
- Robert R. Inman and David J.A. Gonsalvez. Measuring and analysing supply chain schedule stability: a case study in the automotive industry. *Production Planning & Control*, 8(2):194–204, 1997.
- Raf Jans and Zeger Degraeve. Modeling industrial lot sizing problems: a review. *International Journal of Production Research*, 46(6):1619–1643, 2008.
- Thomas Jensen. Measuring and improving planning stability of reorder-point lot-sizing policies. *International Journal of Production Economics*, 30-31:167–178, 1993.
- S.N. Kadipasaoglu and S V Sridharan. Measurement of instability in multi-level MRP systems. *International Journal of Production Research*, 35(3):713–737, 1997.
- S K Kapoor and P Kansal. *Basics of Distribution Management: A Logistical Approach*. Prentice-Hall of India, 2004.
- B Karimi, S.M.T. Fatemi Ghomia, and J.M. Wilson. The capacitated lot sizing problem: a review of models and algorithms. *Omega*, 31(5):365–378, 2003.
- Jens Peter Kempkes. *Kostenoptimale Materialflüsse in der operativen Zulieferungslogistik der Nutzfahrzeugindustrie*. PhD thesis, University of Paderbon, 2009.
- JP Kempkes and Achim Koberstein. A Resource Based Mixed Integer Modelling Approach for Integrated Operational Logistics Planning. *Advanced Manufacturing and Sustainable Logistics*, 46(4):281–294, 2010.
- Florian Klug. *Logistikmanagement in der Automobilindustrie - Grundlagen der Logistik in der Automobilindustrie*. Springer Berlin Heidelberg, 2012.
- J.B. Lasserre, C. Bes, and F. Roubellat. The Stochastic Discrete Dynamic Lot Size Problem : An Open-Loop Solution. *Operations Research*, 33(3):684–689, 1985.
- H. Lee, V. Padmanabhan, and S Whang. The bullwhip effect in supply chains. *Sloan Management Review*, 38(3):93–102, 1997.

- Alain Martel, Moustapha Diaby, and Fayez Boctor. Multiple items procurement under stochastic nonstationary demands. *European Journal of Operational Research*, 87(1):74–92, 1995.
- Lothar März, Tutsch Hansjörg, Auer Stefan, and Sihn Wilfried. Integrated Production Program and Human Resource Allocation Planning of Sequenced Production Lines with Simulated Assessment. In Wilhelm Dangelmaier, Blecken Alexander, Delius Robin, and Klöpfer Stefan, editors, *Advanced Manufacturing and Sustainable Logistics*, pages 408–419. Springer-Verlag Berlin Heidelberg, 2010.
- M Meixell. The impact of setup costs, commonality, and capacity on schedule stability: An exploratory study. *International Journal of Production Economics*, 95(1):95–107, 2005.
- Herbert Meyr. Supply chain planning in the German automotive industry. *OR Spectrum*, 26(4):447–470, 2004.
- Stefan Minner. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81-82(Proceedings of the Eleventh International Symposium on Inventories):265–279, 2003.
- Lars Mönch. Simulation-based benchmarking of production control schemes for complex manufacturing systems. *Control Engineering Practice*, 15(11):1381–1393, 2007.
- Max Müller. *Essentials of inventory management*. AMACOM, 2003.
- I Nyoman Pujawan. Schedule nervousness in a manufacturing system: a case study. *Production Planning & Control*, 15(5):515–524, 2004.
- G Reith-Ahlemeier. *Ressourcenorientierter Bestellmengenplanung und Lieferantenauswahl: Modelle und Algorithmen für Supply Chain Optimierung und E-Commerce*. Books on Demand, 2002.
- P Robinson, a Narayanan, and F Sahin. Coordinated deterministic dynamic demand lot-sizing problem: A review of models and algorithms. *Omega*, 37(1):3–15, 2009.
- Yaakov Roll and Reuven Karni. A Heuristic Algorithm for the Multi-Item Lot-Sizing Problem with Capacity Constraints. *IIE Transactions*, 14(4):249–256, 2011.

- Paul A Rubin, David M. Dilts, and Beth A Barron. Economic Order Quantities with Quantity Discounts: Grandma does it best. *Decision Sciences*, 14(2):270–281, 1983.
- A Rushton, P Croucher, and P Baker. *The handbook of logistics and distribution management*. Kogan Page Series. Kogan Page, 2006.
- B Scholz-Reiter, M Freitag, Ch de Beer, and Th Jagalski. Modelling Dynamics of Autonomous Logistic Processes : Discrete-event versus Continuous Approaches. *CIRP Annals - Manufacturing Technology*, 54(1):413–416, 2005.
- Tim Schöneberg, Achim Koberstein, and Leena Suhl. An optimization model for automated selection of economic and ecologic delivery profiles in area forwarding based inbound logistics networks. *Flexible Services and Manufacturing Journal*, 22(3-4): 214–235, 2011.
- J Shah. *Supply Chain Management: Text and Cases*. Pearson Education, 2009.
- J F Shapiro. *Modeling the supply chain*. Duxbury applied series. Thomson-Brooks/Cole, 2007.
- Edward Silver. Inventory control under a probabilistic time-varying demand pattern. *A I I E Transactions*, 10(4):371–379, 1978.
- Edward A Silver. A Simple Inventory Replenishment Decision Rule for a Linear Trend in Demand. *The Journal of Operational Research Society*, 30(1):71–75, 1979.
- D Simchi-Levi, P Kaminsky, and E Simchi-Levi. *Designing and managing the supply chain: concepts, strategies, and case studies*. Irwin/McGraw-Hill series in operations and decision sciences. McGraw-Hill/Irwin, 2003.
- Donald M Simmons. Optimal Inventory Policies under a Hierarchy of Setup Costs. *Management Science*, 18(10):B591–B599, 1972.
- Herbert A Simon. On the Application of Servomechanism Theory in the Study of Production Control. *Econometrica*, 20(2):247–268, 1952.
- N. C. Simpson. Questioning the relative virtues of dynamic lot sizing rules. *Computers & Operations Research*, 28(9):899–914, 2001.

- C Sox. Dynamic lot sizing with random demand and non-stationary costs. *Operations Research Letters*, 20(4):155–164, 1997.
- Charles R Sox, Peter L Jackson, Alan Bowman, and John A Muckstadt. A review of the stochastic lot scheduling problem. *International Journal of Production Economics*, 62(3):181–200, 1999.
- Sri V Sridharan, William L Berry, and V Udayabhanu. Measuring Master Production Schedule Stability Under Rolling Planning Horizons. *Decision Sciences*, 19(1):147–166, 1988.
- V. Sridharan and R Lawrence Laforge. The Impact of Safety Stock on Schedule Instability, Cost and Service. *Journal of Operations Management*, 8(4):327–347, 1989.
- H Stadtler and C Kilger. *Supply chain management and advanced planning: concepts, models, software, and case studies*. Springer, 2008.
- Hartmut Stadtler. A general quantity discount and supplier selection mixed integer programming model. *OR Spectrum*, 29(4):723–744, 2006.
- Gordon Stewart. Supply Chain Operations Reference Model (SCOR): The first framework for integrated supply-chain management. *Logistics Information Management*, 10(2):62–67, 1997.
- P M Swamidass. *Encyclopedia of production and manufacturing management*. SpringerLINK ebook collection. Kluwer Academic, 2000.
- S Armagan Tarim and Brian G Kingsman. The stochastic dynamic production/inventory lot-sizing problem with service-level constraints. *International Journal of Production Economics*, 88(1):105–119, 2004.
- H Tempelmeier. On the stochastic uncapacitated dynamic single-item lotsizing problem with service level constraints. *European Journal of Operational Research*, 181(1):184–194, 2007.
- Horst Tempelmeier. A Simple Heuristic for Dynamic Order Sizing and Supplier Selection with Time-Varying Data. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, 11(4):499–515, 2002.

- J A Tomlin. Minimum-Cost Multicommodity Network Flows. *Operations Research*, 14(1):45–51, 1966.
- J W Toomey. *Inventory management: principles, concepts and techniques*. Materials management/logistics series. Kluwer Academic Publishers, 2000.
- Paul Valckenaers, Sergio Cavalieri, Bart Saint Germain, Paul Verstraete, Hadeli, Romeo Bandinelli, Sergio Terzi, and Hendrik van Brussel. A benchmarking service for the manufacturing control research community. *Journal of Intelligent Manufacturing*, 17(6):667–679, 2006.
- Tim van Kampen, Dirk Pieter van Donk, and Durk-Jouke van der Zee. Safety stock or safety lead time: coping with unreliability in demand and supply. *International Journal of Production Research*, 48(24):7463–7481, 2010.
- S Viswanathan. Coordinating supply chain inventories through common replenishment epochs. *European Journal Of Operational Research*, 129:277–286, 2001.
- Harvery M. Wagner and Thomson M. Within. Dynamic Version of the Economic Lot Size Model. *Management Science*, 5(1):89 – 96, 1958.
- S W Wallace and W T Ziemba. *Applications of stochastic programming*. MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics, 2005.
- H P Williams. *Model Building in Mathematical Programming, 4th Edition*. Wiley, 4 edition, 1999.
- Jiefeng Xu, Leonard L Lu, and Fred Glover. The deterministic multi-item dynamic lot size problem with joint business volume discount. *Annals of Operations Research*, 96(1-4):317–337, 2000.
- X. Zhao, J.C. Goodale, and T.S. Lee. Lot-sizing rules and freezing the master production schedule in material requirements planning systems under demand uncertainty. *International Journal of Production Economics*, 33(8):2241–2276, 1995.