NTNU
Norwegian University of
Science and Technology

Operational production planning for multi-plant metal casting using mixed integer linear programming

A case study from the aluminum industry

Marianne Grødem Reine
Thea Sofie Narvestad Ullebust
Håvard Wilsher

Industrial Economics and Technology Management
Submission date: June 2018
Supervisor: Peter Schütz, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Problem Formulation

The aim of this thesis is to generate production plans that allocate a set of aluminum orders (call-offs) to the casting tables at Norsk Hydros' plants. The call-offs are based on long term contracts which specify a tonnage of aluminum products to be delivered every month. However, the alloy, dimensions and specific delivery time are not specified until the customer declares this, typically three weeks before delivery. Meanwhile, a feasible production plan has to be made taking batch production, production capacities and transportation routes into account. This production plan should be cost effective despite call-offs arriving unpredictably and often having to be allocated within short time before the entire realization of orders is known.

The production plan to be developed must account for costs. Most prominently, transportation costs must be considered, but also production costs and storage costs. The production plan should be produced in a rolling horizon manner and one wishes to lock call-offs allocated early in the planning period. This provides some degree of predictability for the casthouses responsible of producing the orders. Efficient computation is also an ambition in the thesis, as the model is supposed to serve as an operational decision support tool. In addition, it should be capable of handling realistically sized demand data from Norsk Hydro.

Preface

This master thesis is the concluding part of our Master of Science in Industrial Economics and Technology Management with a degree specialization in Applied Economics and Operations Management at the Norwegian University of Science and Technology (NTNU).

First and foremost, we would like to thank our academic supervisor Peter Schütz for his dedication and excellent guidance throughout this thesis. We would also like to thank Kjartan Kastet Klyve and Jone Hansen for always helping us when questions arose regarding code-implementation of the model.

The thesis is conducted in collaboration with Norsk Hydro, and we would like to thank them for the opportunity and the proficiency we have experienced. Especially we would like to thank our contact at Norsk Hydro, Ole Reiersen, for improving our understanding of the problem and assisting us when needed.

The thesis is related to previous work done by Truls Flatberg and Kjetil Midthun at SINTEF, and we would like to thank them for extensive answers to all of our questions and for providing valuable input to our project.

Trondheim, June 8th, 2018

Abstract

Significant market shares in the global aluminum industry belong to producers located in low cost countries. Norsk Hydro ASA is a Norwegian firm in this industry, and operate as a Make-to-Order company. They deliver cast products from multiple plants to a wide range of customers. However, the quantities, products and delivery dates of demand is only partly known at any given day.

An operational production planning model is developed, aimed at minimizing costs. The model allocates production to a series of plants, and accounts for demand being a mixture of confirmed orders as well as forecasts. All this is done while taking complications such as batch production and updated demand information into account. The model is a mixed integer linear program, and it is applied with a rolling horizon approach, where demand information is updated daily. The model successfully allocates demand in a way that is consistent with Norsk Hydro's business practices, although a trade-off between optimality gaps and nervousness is observed. The potential of Lagrangian relaxations is tested on certain model constraints, with the subgradient method being used to update Lagrangian multipliers. The approach is successful in providing reduced optimality gaps.

The model is developed with the aluminum industry in mind, but can be applied to any metal industry utilizing casting processes.

Sammendrag

I den globale aluminiumsbransjen tilhører betraktelige markedsandeler produsenter som befinner seg i lavkostland. Norsk Hydro ASA er en aktør i denne bransjen, og driftes etter Make-to-Order-prinsipper. De leverer støpte produkter fra flere aluminiumsverk til et bredt utvalg kunder. Detaljer vedrørende etterspørsel vites bare delvis på en gitt dag, ettersom mengder, produkter og leveringsdatoer sjelden er spesifisert lenge i forveien av en bestilling.

En modell til operasjonell produksjonsplanlegging blir utviklet, den ønsker å minimere kostnader. Modellen allokerer produksjon til en rekke aluminiumsverk, og tar høyde for etterspørsel som er en blanding av bekreftede bestillinger eller prognoser. Dette gjøres samtidig som at hensyn tas til batch-produksjon, stadig oppdatert etterspørsel og andre kompliserende forhold. Modellen er et blandet heltallsproblem, og er anvendt med en rullende horisont, hvor informasjon om etterspørsel oppdateres daglig. Den lykkes i å allokere ordre etter Norsk Hydro's forretningspraksis, men en avveiing mellom optimalitetsgap og "nervousness" blir observert. Potensialet i å Lagrange-relaksere enkelte restriksjoner blir testet, mens subgradientmetoden anvendes for å oppdatere Lagrange-multiplikatorer. Denne tilnærmingen lykkes i å redusere optimalitetsgap.

Modellen er utviklet for bruk i aluminiumsbransjen, men kan med modifikasjoner blir anvendt i andre metallbransjer som benytter støpning.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Aluminum is a material with increasing demand in the global market. It is increasingly preferred in uses ranging from vehicle components to construction to product packaging. It has favorable properties for a wide selection of applications, owing to its low weight and high specific strength when compared to metals such as steel. In addition, the material is resistant to corrosion and has a high degree of quality retention when recycled. The recycling process consumes only 5% of the energy that was originally used to produce the aluminum. These factors cause aluminum to be regarded as a renewable resource.

During the last four decades, the majority of the world's aluminum manufacture has been shifted from western Europe and Northern America, to low cost countries (Nappi, 2013). A clear example of this is that about 60% of current aluminum production is in China according to the International Aluminium Institute (2017). Meanwhile, products from aluminum manufacturing are widely regarded as commodities and traded on global exchanges such as the London Metal Exchange. This encourages manufacturers to constantly be conscious of costs in order to remain competitive.

Norsk Hydro ASA, hereby referred to as Norsk Hydro, is a Norwegian aluminum manufacturer. They had an estimated market share of about 4% in 2013, and practice a corporate strategy implementing a vertically integrated supply chain. They have a global presence, being involved in plants on several continents. These factors, combined with the aforementioned challenge of competing with manufacturers in low cost countries, provide the company with strong incentives to optimize production in order to increase efficiency while lowering overhead costs.

The company currently utilizes decision support tools for production planning, but these are primarily applied on decisions on strategic and tactical levels. Operational planning is often done manually and decentralized at the individual plants. Furthermore, operational decisions are often based on tactical planning which depends on forecasted data, while actual demand often differs from these forecasts.

The sum of these imperfections imply that the operational planning has room for improvement.

In this thesis, an operational production planning model is developed. The model is a mixed integer linear program (MILP), and intends to allocate demand for aluminum products to different plants while minimizing the sum of costs. It provides planning on a daily time resolution that is not currently available to Norsk Hydro from any other planning tools. It also considers demand that is present both as confirmed requests from customers and as predicted requests.

The model is implemented using a rolling horizon approach. The results are discussed with regard to computational efficiency and the patterns observed in solution variables. The concept of nervousness is also discussed, presenting a trade-off between predictability and objective function value in the implemented production plans. In addition, Lagrangian relaxation of different constraints is attempted in order to decrease optimality gaps and improve runtimes. The Lagrangian relaxation is combined with a greedy heuristic to provide feasible results.

A computational study reveals that solving the model in a rolling horizon approach results in sizable and varying optimality gaps. A trade-off between optimality gap and nervousness is also observed. One variant of the Lagrangian relaxation shows promise, obtaining lower optimality gaps in shorter runtimes.

The thesis has the following structure: Relevant background information is provided in Chapter 2 before a literature review is presented in Chapter 3. This is followed by a formal problem description in Chapter 4, and a model formulation in Chapter 5. Suggested solution approaches are described in Chapter 6, along with details on how they are implemented. Input data values, assumptions forming them and procedures for simulating data are all presented as part of the case study description in Chapter 7. A computational study of model results is presented in Chapter 8, before finishing with concluding remarks and suggestions for future research in Chapter 9.

Chapter 2

Background

This chapter provides the reader with relevant background information for the rest of the thesis. Section 2.1 gives an introduction to global aluminum manufacturing and a more detailed explanation of the levels in the aluminum value chain. In Section 2.2 the aluminum manufacturer Norsk Hydro is presented. Section 2.3 revolves around production planning at Norsk Hydro, especially at the tactical and operational level. Potential issues with current practice are identified, providing a foundation for the general problem this thesis is concerned with.

2.1 The Global Aluminum Industry

This section provides a general overview of the aluminum industry. The contemporary market conditions are dicussed in subsection 2.1.1, while a breakdown of the value chain is given in subsection 2.1.2.

2.1.1 Market Conditions

Aluminum is one of the most abundant elements in the earth's crust, accounting for 8.2% of its composition measured in weight. It has many favorable properties such as high specific strength, durability and efficient recycling. This makes it an increasingly preferred material for example in vehicles, one of many sectors contributing to increased year-on-year growth of aluminum demand globally over the last couple of years. Figure 2.1 shows a more detailed breakdown of the business sectors consuming aluminum.

Aluminum products are produced on all of the world's continents, and the majority of global aluminum supply is provided by low cost countries such as China (International Aluminium Institute, 2017). This creates significant incentives for

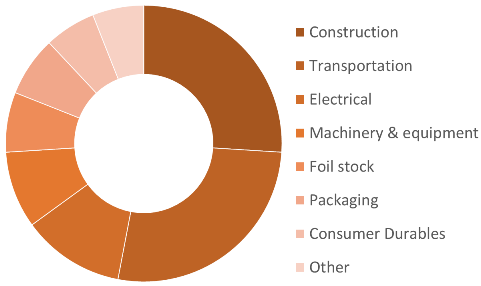


Figure 2.1: A distribution of sectors using aluminum (Fog, 2016).

producers in developed countries with higher cost structures to maintain competitive advantage.

2.1.2 Value Chain

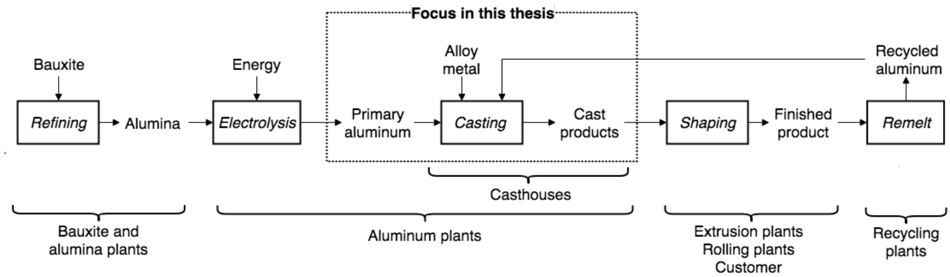


Figure 2.2: Value chain for aluminum.

Figure 2.2 shows the main levels in the aluminum value chain. The value chain for aluminum starts by extracting the raw material bauxite, which is a type of soil commonly found along the equator. Bauxite is then separated into several components, one of them called alumina.

Alumina, in powder form, is transported to the aluminum plants where aluminum is extracted from it. This is achieved through the process of electrolysis which separates oxygen and aluminum in the alumina. The process results in pure, liquid aluminum, more commonly known as *primary aluminum*. The alloy is determined by adding other metals to the liquid. Aluminum manufacturers may also choose to mix the liquid with recycled aluminum. Primary aluminum is further transported to a casthouse located at the aluminum plant to create *cast products*. Each casthouse can utilize a number of casting tables, which determine the geometric dimensions of the cast product. The most common categories of cast products are *extrusion ingots*, *foundry alloys* and *sheet ingots* and are illustrated in Figure 2.3.

Figure 2.3: Main cast products. From left: extrusion ingots, sheet ingots and foundry alloys.

Cast products go on to become a wide variety of end products in a *shaping process*. Extrusion ingots are extruded through a die to become profiles, such as window frames, in extrusion plants. Sheet ingots are rectangular and are usually rolled into foil or plates at a rolling plant. Foundry alloys are reheated into the molten state before being recast into shapes such as engine cylinders. The transformation from cast products to end products is often done at the location of the customer purchasing the cast products.

Aluminum has the property of retaining its quality when it is recycled. It is therefore common to recycle aluminum products that are no longer in use. It is also a less energy intensive process than creating primary aluminum from electrolysis. Electrolysis of alumina requires 13 kWh per tonne of energy to manufacture one kilo of primary aluminum, while recycled aluminum consumes only 5% of this energy.

2.2 Norsk Hydro

Norsk Hydro is a Norwegian manufacturer of aluminum products. They have approximately 35 000 employees in over 40 countries, and serve about 30 000 customers world wide. They are one of the world's largest producers of aluminum, with an annual output of about two million tonnes of aluminum. The company operates with an integrated supply chain, as they own bauxite mines, electrolysis cells, casthouses and recycling facilities. Furthermore, one of its daughter companies operates several extrusion plants. This strategy entails that Norsk Hydro sells finished aluminum products as well as the cast products mentioned in 2.1.

2.2.1 Plants and Locations

There are four main categories of plants in Norsk Hydro's portfolio:

- Bauxite and alumina plants
- Aluminum plants
- Remelt and recycle plants

- Extrusion plants

This thesis focuses primarily on aluminum plants, more specifically the casthouses associated with these.

Norsk Hydro's demand for bauxite and alumina is primarily satisfied by self owned mines and refinement plants in Brazil. The aluminum plants receiving the alumina are either completely owned by Norsk Hydro or operate as joint ventures. They are mostly located in Europe, although a major plant is located in Qatar, along with there being some involvement in aluminum plants in the US and Australia. The largest of the plants are Sunndal (Norway) and Quatalum (Qatar), each with an output of about 200 000 tonnes of cast products annually. In this thesis the aluminum plants in Figure 2.4 are considered.

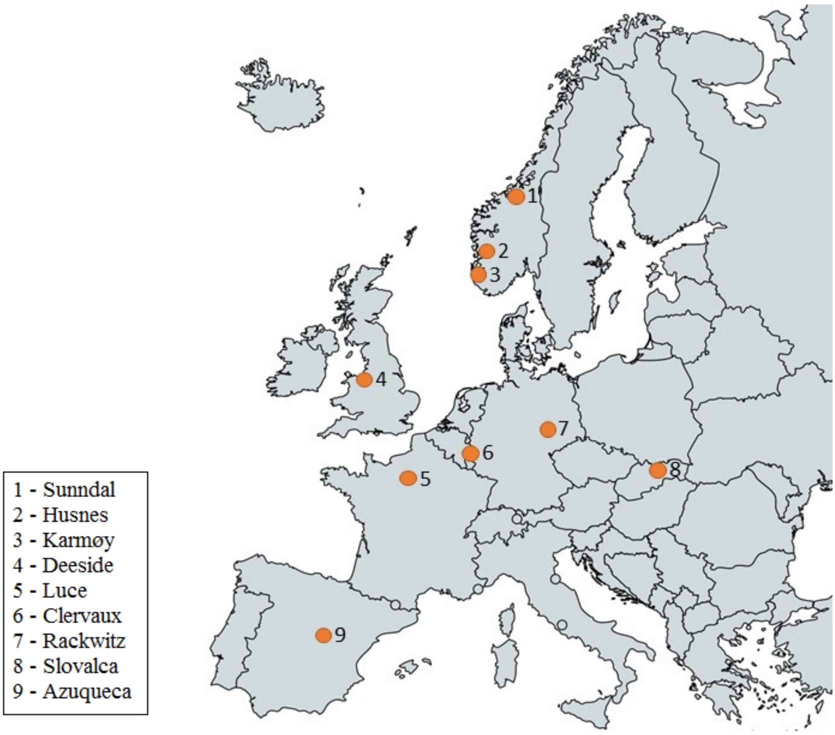


Figure 2.4: Norsk Hydro's locations of aluminum plants in Europe.

All aluminum plants rely on one or more electrolysis cells that create primary aluminum. The liquid aluminum is further transported to the casthouse located in another part of the plant. It can be supplemented with liquid aluminum from recycled aluminum products. At each casthouse, cast products are created at casting tables. One casting table is related to one product dimension.

2.2.2 Products

Norsk Hydro produces and ships both cast and end products to customers, where cast products have not yet gone through the shaping process while end products have. This thesis considers cast products, more precisely extrusion ingots.

A distinct extrusion ingot product is determined by its alloy and dimension. Given the dimension of the product, a specific casting table must be utilized in the casting process. Each casthouse has at most one casting table qualified to cast a given dimension. A casthouse can only produce a selection of the different extrusion ingot dimensions, something that is determined by the casting tables available at the casthouse.

Norsk Hydro has a huge extrusion ingot portfolio, consisting of different alloy and dimension combinations. With 2 650 possible alloys and 14 dimensions, there theoretically exists 37 100 product combinations. However, some combinations are never produced, and Norsk Hydro estimates that they have among 10 000 different extrusion ingot products. Furthermore, Norsk Hydro's customers can request new customized products. Thus, the extrusion ingot portfolio is not a finite set and continues to grow over time.

Production at all of Norsk Hydro's aluminum plants is batch based. This means that when production of a specific product is initiated at least one batch of the product is produced.

2.2.3 Demand and Long Term Contracts

Norsk Hydro operates as a make-to-order (MTO) company. This implies that they do not produce anything unless specifically requested by a customer, and storage of products is avoided. MTO companies often customize products for each customer order.

All customers of Norsk Hydro have a long term contract with the company. These contracts specify a required amount of tonnes of products to be delivered in regular intervals for a time into the future. For instance, a customer can agree to have 500 tonnes of products delivered every month for the next 6 months. The type of cast product (extrusion ingot, sheet ingot or foundry alloy) is specified in the long term contract, but the quantity requested along with a certain alloy, dimension and delivery time are specified when the customer makes a *call-off*. The quantity requested in a call-off is required by Norsk Hydro to be a multiple of a certain *order batch* quantity. A call-off on average arrives at Norsk Hydro two weeks before delivery, and three days before at the latest. Norsk Hydro strive to achieve every requested delivery time, but may insist on postponing it due to practical constraints. In addition, Norsk Hydro always produce the entire call-off at the same casthouse.

Long term contracts bring some degree of certainty in future demand; total demand tonnage from each customer every month is known. Still, Norsk Hydro does not know when call-offs arrive during the month and do not know the specific products to be requested.

2.3 Production Planning

Production planning decisions in Norsk Hydro are divided into different stages where higher stage decisions set the foundation for the lower stages.

Strategic, long term decisions consider for instance plant size and location, as well as investments in machinery. Tactical decisions are made at Norsk Hydro's main office in Oslo, and include sales and operation plans based on forecasts and production plans for each aluminum plant. These decisions are normally taken every two to three months. Operational decisions, with typically day-to-day levels of detail, are made autonomously by representatives at each plant and is mainly based on experience. Norsk Hydro's application of decision support tools in production planning is thus mainly focused on strategic and tactical decisions in the current state.

Further in this section, the tactical and operational decision levels in production planning is presented in further detail in Section 2.3.1, while Section 2.3.2 considers the challenges related to these to levels.

2.3.1 Tactical and Operational Planning

At the tactical level, Norsk Hydro currently utilize CASPAH, an operations research based tool for tactical production planning, used centralized at Norsk Hydro. The model takes demand, costs, capacities and other relevant limitations into account, creating a production plan that allocates demand to the different plants for specific time periods. Customer demand over the planning horizon is estimated using a combination of forecasts provided by Norsk Hydro's marketing department as well as actual call-offs made by customers. The plan generated by CASPAH is capable of providing decision support detailing what casting tables are set to produce which specified products in a given week.

At the operational level, the Production Planner at the individual plants sets up weekly production plans based the tactical production plans generated by CASPAH. If solutions from CASPAH are inadequate for the operational production plan, the Production Planner mainly takes its decisions based on previous experience.

2.3.2 Challenges with the Current Production Planning Process

Production plans made by CASPAH are prone to certain challenges when transitions are made to more detailed production planning at the operational level. Firstly, the tool does not take batch production into account. Secondly, the tool also allows for call-offs to be split among casthouses, something that is never done in practice as individual call-offs are always produced entirely in a single casthouse.

A third challenge arises when centralized planners in Norsk Hydro attempt to emulate production plans made by CASPAH based on received call-offs. CASPAH production plans are heavily based on forecasts, and these forecasts are subject to errors. These production plans may no longer be feasible or optimal when the actual call-offs are realized differently from what was forecasted. If this is the case, the centralized planners work manually and attempt to use their experience to decide what allocations are best. This approach makes it challenging to ensure a level of global optimality in allocations. Furthermore, allocations tend to cause uneven capacity utilization at the different casthouses, sometimes resulting in that the casthouse that is obviously best for producing a certain call-off has its capacity filled for the period in question and cannot be used.

A more operationally oriented planning tool is required, able to account for the MTO production philosophy, demand that is a combination of confirmed call-offs and forecasts, as well as exploiting new information about the arrival of new call-off demand, all while taking factors such as batch production into consideration. Such a tool should also have reasonable computing times, making it possible to deal with received call-offs on a daily bases. Pushing optimality gaps as low as possible is also an advantage, given the tight margins one faces when operating in a global commodity market.

Chapter 3

Literature Review

In this chapter literature from three different aspects of this thesis is reviewed. Firstly hierarchical production planning is discussed in Section 3.1, followed by production planning in Section 3.2. Section 3.3 considers literature on Lagrangian relaxation.

3.1 Hierarchical Production Planning

The concept of hierarchical planning is applied extensively in the literature of production planning. Here, a commonly applied framework is explained, before a review of different implementations is presented. Lastly, a description of comprehensive hierarchical production planning systems is given, called advanced planning systems.

3.1.1 Framework

In hierarchical production planning studied by Bitran and Tirupati (1993), production planning is divided into three levels of decision making: *strategic*, *tactical* and *operational*. This division is based on the framework by Anthony (1965), and following is a description of the different levels. It should be clarified that the decisions taken on the different levels is dependent on the industry considered, and what follows is a generalization based on Bitran and Tirupati (1993).

Strategic decisions concern long term investments and acquisition of equipment. Because of the long term investments, the frequency of such decisions is low and usually taken by the top management. Such decisions consider production on a highly aggregated level and with a high degree of uncertainty.

Tactical decisions consider the utilization and allocation of the acquired resources. These decisions have a medium planning horizon and also an intermediate level of production aggregation. Tactical decisions are constrained by the investments done at the strategic level, often in form of a budget.

Operational decisions are concerned with the daily operation of each plant or facility. Decisions at this level are highly detailed and have a short planning horizon. This implies a low degree of uncertainty and the decisions are taken at low levels in the organization with little or no influence from the top management. The outcome of the operational decisions provides feedback necessary to evaluate the decisions at the higher levels as illustrated in Figure 3.1.

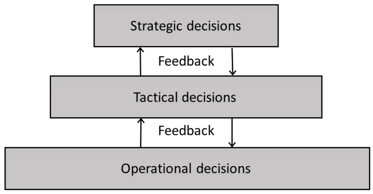


Figure 3.1: Relationship between the hierarchical decision stages.

3.1.2 Implementations of Hierarchical Planning

Even though the hierarchical framework is widely applied, the interaction of the different levels can vary a lot. Schütz et al. (2011) discusses a facility location problem for the meat industry. The location of facilities are considered strategic decisions imposing hard constraints on lower levels. In contrast, in Torabi et al. (2010) a make-to-stock TV manufacturer is studied where the aggregated production plans impose soft constraints on lower levels of the planning process. Thus, the optimal solutions from a certain planning stage can be completely or only partly decisive for the next stage.

Omar and Teo (2007) studies all stages in a three-stage hierarchical planning process from the chemical and pharmaceutical industry. These levels include aggregate production plans with setups, disaggregation by optimizing the number of batches to be produced and finally a job sequencing optimization model. This results in a comprehensive study including all three levels, which can be advantageous from an overall perspective. The interaction between stages may be as important as the optimization of a planning stage itself, and this will be discussed further in relation to advanced planning systems.

3.1.3 Advanced Planning Systems

Hierarchical production planning simplifies a planning problem by breaking it down into smaller subproblems. However, this approach has been criticized for creating challenges concerning optimization of each subproblem, instead of optimization of the system as a whole (Dempster et al., 1981). Enterprise resource planning (ERP) systems were introduced in the early 1990s and intend to be an overall planning tool supposed to ensure better integration of all the different planning operations within an enterprise (Umble et al., 2003).

During the 1990s the concept of supply chains were introduced to producing firms (Stadtler et al., 2015). The firm should focus on what they are best at, the *core activities*, and outsource everything else. The production process thus involves many cooperating independent units, and an optimization of each of these units may be a suboptimal solution for the supply chain as a whole. This enforces planning tools to take this aspect into account and different software companies offer Advanced Planning Systems (APS) which aim at optimizing the production for the entire supply chain.

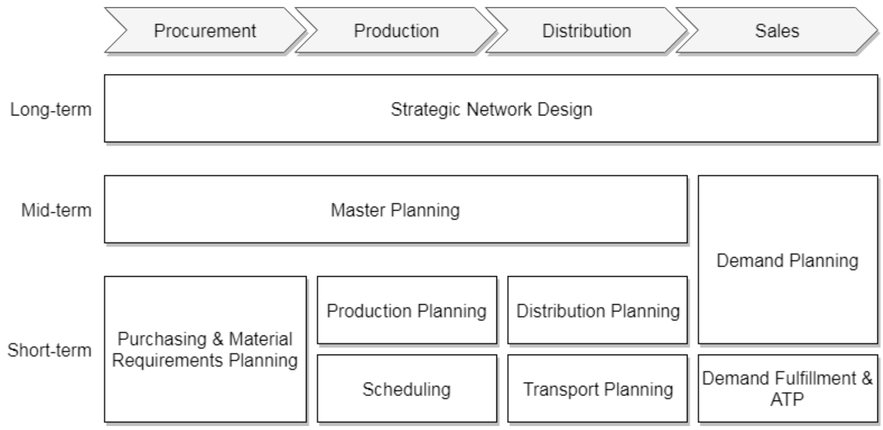


Figure 3.2: Supply chain management matrix (Stadtler et al., 2015).

Figure 3.2 gives an overview of different modules that often are included in APS. The different stages in the hierarchical framework are seen vertically in the matrix, while the supply chain units are found horizontally. Production planning and scheduling is considered short term planning. As mentioned in Section 3.1.2, Omar and Teo (2007) considered these two modules in one single study.

3.2 Production Planning

In this section different approaches to modeling certain features are presented and compared. Section 3.2.1 discusses different feature related to the case of MTO firms. Further, Section 3.2.2 presents the concept of rolling horizon in production planning, while Section 3.2.3 discusses different aspects of dynamic lot sizing problems.

3.2.1 Production Planning in Make-to-Order Firms

MTO firms are characterized by only manufacturing confirmed orders. Consequently, safety stocks are generally avoided, with the aim of not tying up capital in excessive inventories. Meanwhile, there has been a growing trend of utilizing this manufacturing philosophy, especially among companies that deliver customized or diverse product portfolios (Stevenson et al., 2005). The reviewed literature presents several approaches to production planning for firms practicing the MTO philosophy.

Guo et al. (2013) use a three step method to approach MTO production planning. Firstly, orders are allocated based on a memetic optimization process. Uncertainty is accounted for by exposing the solution to Monte Carlo simulation before the third step involves a pruning heuristic. Barut and Sridharan (2004) also apply simulation to account for demand uncertainty, and the solution approach revolves around the use of dynamic programming. Spengler et al. (2007) and Neureuther et al. (2004) both use rolling horizon planning in their approaches as a way of compensating for uncertainty in demand. While Spengler et al. (2007) uses dynamic programming to find solutions and values orders from customers based on giving scores in previous model iterations, Neureuther et al. (2004) is more focused on finding right levels of employment rather than allocating orders, using a MILP to find solutions. Both these sources are especially relevant to this thesis as they are case studies from metal manufacturing, albeit not for casting processes. Finally, Mestry et al. (2011) use a branch and price approach to evaluate orders for a gear manufacturer with a customized product portfolio. Uncertainty in demand is not considered in this source.

Norsk Hydro's problem described in this thesis involves a single value chain level, the casting process, but multiple facilities performing this task. This is contrary to any of the sources reviewed. Guo et al. (2013) is the only source describing a multi-facility problem, while Neureuther et al. (2004), Mestry et al. (2011) and Barut and Sridharan (2004) are those describing single value chain levels.

3.2.2 Rolling Horizon Planning

In production planning, one has to decide upon a planning horizon which can be finite or rolling. Production planning methods with finite horizons create a plan for 1,...,T periods. This plan is then implemented and a new plan for the time periods $1+T$,...,$2T$ is created and implemented, and so on. The plans are independent of each other, with the exception of that the finite state in one plan is the initial state in the next plan. In reality, plans with a finite horizon are often implemented only for the first immediate periods as new information causes the plan to be updated. Changes in forecast or actual demands is typical events that enforce replanning.

Rolling horizon planning is a approach that exploits this practice. A plan for 1,...,T periods is generated, but only the decisions for the first ΔT periods are implemented. Further, the horizon is *rolled* forward with ΔT time periods creating a plan for $1 + \Delta T$,..., $T + \Delta T$. This means that the period $1 + \Delta T$,..., T is replanned. As a result, rolling horizon planning allows decisions in the future to be postponed until new information arrives. In addition, implemented plans can be based on actual demand rather than forecasted demand, creating more realistic production plans. Figure 3.3 illustrates the difference between finite and rolling horizon planning. It is worth noting that with a finite horizon in part a, 3 production plans are generated for 15 time periods, while for the rolling horizon in part b, 15 plans are generated for 15 time periods.

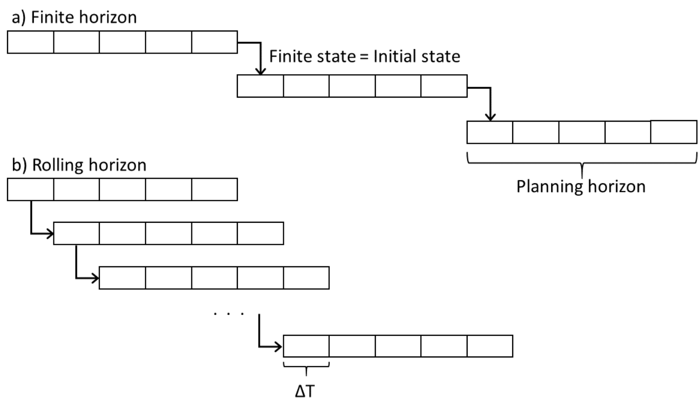


Figure 3.3: Difference between a finite horizon and a rolling horizon.

Baker (1977)'s early studies of rolling lot sizing schedules indicate that the rolling horizon is rather efficient and better in most cases compared to a finite horizon. Moreover, he concludes that the planning horizon, T, most likely is a crucial parameter when using a rolling horizon approach. The same was found in Baker and Peterson (1979). In addition, they found that the cost structure is a significant factor in the performance of the model.

Since Baker (1977) and Baker and Peterson (1979), production planning models using rolling horizon procedures have received considerable attention in research. To name a few, Chung and Krajewski (1987), Sridharan et al. (1987), Sridharan and Berry (1990), Campbell (1992) and Neureuther et al. (2004) have all studied this feature in master production schedules (MPS). On a more operational level Stadtler (2000), Tiacci and Saetta (2012) have studied lot size models with a rolling horizon, while Blackburn and Millen (1982), Simpson (1999) focused on models within material requirement planning (MRP).

As a consequence of rolling horizon planning, decisions from one plan may differ sufficiently from the previous plan. For companies, frequent changes to decisions are usually not desired as they want some degree of consistency and predictability over time. Dependencies between decisions are also an issue as changes from one decision tool might impact other decision tools. In a rolling horizon setting, inconveniences related to frequent changes to decisions is referred to as *nervousness*.

Freezing parts of the plan, meaning that a decision taken can not be altered when the model is rolled forward, is a common approach to reduce nervousness. While freezing part of the plan reduces changes, it might increase other factors such as inventory and production costs. Sridharan and Berry (1990) study nervousness and freezing decisions in MPS. Frequent changes to the MPS induce changes to lower hierarchy decisions like the MRP. It is therefore necessary to generate schedules with a low degree of nervousness. They also study the freezing length, concluding that the freezing length should be at least 50 % of the planning horizon in order to reduce nervousness.

The rolling horizon planning approach can have modifications in order to better fit the problem domain. Interesting features are found in Campbell (1992), Tiacci and Saetta (2012), As'ad and Demirli (2010).

By freezing parts of the decisions, the model can be divided into different phases where changes are either allowed or not allowed for a period when rolled forward. Campbell (1992) has taken this approach a step further, dividing the planning period into three phases; a frozen phase, a phase allowing changes to the lot size and a phase allowing changes to both lot size and location. This way bigger changes are allowed in the most distant time periods.

Tiacci and Saetta (2012)'s generalized lot size model generates production plans for the whole planning period. As only the first period is actually implemented, they only decide the production plan sequence problem (SP) for the immediate period. They argue that it is a waste of time to calculate sequences for other periods as they will be subject to change when the model is rolled forward. By only solving the SP for a small part of the problem, computational time per iteration is reduced.

As'ad and Demirli (2010) consider a rolling horizon MPS for the steel industry. Here, variables for later periods are relaxed while periods in the near future are not. This reduces computational time.

3.2.3 Lot Sizing Problems

Lot sizing problems (LSP) aim at determining the optimal time and level of production where demand is given or has a certain distribution (Jans and Degraeve, 2008). The basic concept of an LSP is illustrated in Figure 3.4. The nodes illustrate time periods i while x_i, s_i and D_i denotes production, storage and demand in time period i respectively.

In dynamic lot sizing problems (DLSP) the demand over a given period is known, but varies within this period (Lee et al., 2001). This is often the case for producers who write contracts with their customers. The contracts may specify an amount to be delivered to the customer during a given time period, but the exact time the goods may be requested is unknown to the producer.

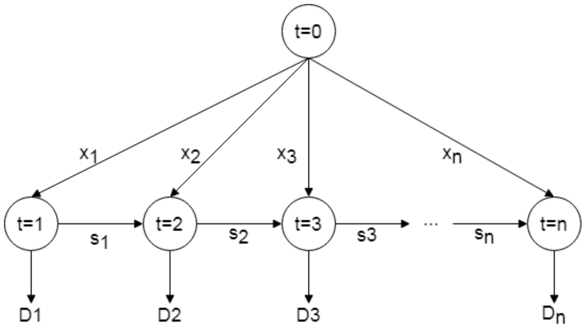


Figure 3.4: Illustration of the single-item uncapacitated LSP (Jans and Degraeve, 2008).

The planning horizon of the LSPs can be either finite or infinite. A finite horizon is usually applied when demand is being dynamic, while an infinite planning horizon applies when demand is being static (Karimi et al., 2003). When data is uncertain a rolling horizon approach should be applied (Karimi et al., 2003). An overview of additional characteristics of the LSPs discussed by Karimi et al. (2003) can be found in Table 3.1. The table indicates how a given characteristic affects the complexity of the problem and which characteristics appear in the problem of this thesis. It is emphasized by Karimi et al. (2003) that resource constraints are an important aspect of modeling LSPs and have a great impact on the complexity of the problem .

Table 3.1: Characteristics of the LSP based on Karimi et al. (2003).

Feature	Options	Complexity	Our problem
Production Levels	Single-level		✓
	Multi-level	↑	
Products	Single-item		
	Multi-item	↑	✓
Capacity	Uncapacitated		
	Capacitated	↑	✓
Demand	Static		
	Dynamic	↑	✓
	Probabilistic	↑↑	

Apart from the features listed in Table 3.1, LSPs with various extensions have been investigated. Jans and Degraeve (2008) review LSP that include characteristics of specific production processes and a rolling horizon approach. Further, where uncapacitated problems in general are easier to solve than capacitated LSPs, capacitated LSPs arise a question of backlogging. Lee et al. (2001) formulate a DLSP including backlogging and time windows. They observe that including the possibility of backlogging show crucial increase in complexity. Moreover, Van Vyve (2006) applies extended formulations to tighten the capacitated LSP with backlogging.

Extensions of the LSP problem provide a more realistic model but at the expense of increased complexity. Bitran and Yanasse (1982) show that a single-item capacitated LSP (CLSP), is NP-hard, and Chen and Thizy (1990) show that a multi-item CLSP is strongly NP-hard. Efficient solution methods for LSP problems are thus of great importance. Previous research has often focused on finding appropriate heuristics to solve these problems.

3.3 Lagrangian Relaxation

Lagrangian relaxation is a solution method used to solve large optimization problems, including integer problems, by exploiting the problem structure. In the method, a set of constraints from the original problem are relaxed and added to the objective function. Infeasible solutions, with respect to the relaxed constraints, are penalized in the objective function using Lagrangian multipliers. The solution of the Lagrangian relaxation will give optimistic bounds (Lundgren et al., 2010).

3.3.1 Theory

The Lagrangian relaxation is explained in detail for a minimization problem with $\geq$-constraints based on Lundgren et al. (2010).

A minimization problem is given as

$$\begin{aligned} & \min f(\mathbf{x}) & \\ \text{subject to} \qquad & g_i(\mathbf{x}) \geq b_i, & i = 1, ..., m \end{aligned}$$

By relaxing the constraints and adding them to the objective function, the *Lagrangian function* is given as

$$L(\mathbf{x}, \boldsymbol{\pi}) = f(\mathbf{x}) + \sum_{i=1}^{m} \pi_i (b_i - g_i(\mathbf{x}))$$

Usually, only a set of constraints are relaxed and it is common to relax constraints that in some way complicate the structure of the problem. $b_i - g_i(\mathbf{x})$ is the slack for constraint i. Given $\geq$-constraints, positive slack implies an infeasible solution, while a negative slack implies a feasible solution. The opposite is correct for $\leq$-constraints. In Lagrangian relaxation, infeasible solutions are penalized in the objective function through the Lagrangian multiplier. For minimization problems, this is achieved by adding a positive value to the objective function. Therefore, the

Lagrangian multiplier must be positive ($\pi_i \geq 0$) if the sign for the relaxed term is positive.

The Lagrangian function should be minimized with respect to $\mathbf{x}$ and maximized with respect to $\boldsymbol{\pi}$ written as:

$$\max_{\pi \geq 0} \min_{x} L(\mathbf{x}, \boldsymbol{\pi})$$

Further, the *Lagrangian dual function* is defined as:

$$h(\boldsymbol{\pi}) = \min_{x} L(\mathbf{x}, \boldsymbol{\pi})$$

For a given $\boldsymbol{\pi'}$, the Lagrangian dual function becomes

$$h(\boldsymbol{\pi'}) = \min_{x} L(\mathbf{x}, \boldsymbol{\pi'}),$$

and is solved by minimizing with respect to $\mathbf{x}$. This problem is referred to as the *Lagrangian subproblem*. The Lagrangian relaxation can be solved by iteratively searching for feasible solutions while evaluating the multipliers, $\boldsymbol{\pi}$.

When the original problem is convex the Lagrangian relaxation guarantees to find the optimal $\mathbf{x}^*$ and $\boldsymbol{\pi}^*$ when $h(\boldsymbol{\pi}^*) = f(\mathbf{x}^*)$. For non-convex problems, like integer problems, a *duality gap* may occur as it is only known that $h(\boldsymbol{\pi}^*) < f(\mathbf{x}^*)$. Meaning it is not necessarily possible to verify that the optimal solution is found.

3.3.2 General Solution Strategy

Following is the general solution strategy for Lagrangian relaxation presented in steps. k denotes the iteration number while the lower bound and upper bound are denoted LBD, and UBD respectively.

Step 1: Initialization. Set k = 0, LBD = $-\infty$, UBD = $+\infty$. Choose initial values of $\boldsymbol{\pi}^{(0)}$.

Step 2: Solve the Lagrangian subproblem for a given $\boldsymbol{\pi}^{(k)}$. This gives a solution $\boldsymbol{x^{(k)}}$ and an optimistic bound $h(\boldsymbol{\pi^{(k)}})$.

- Update optimistic bound, LBD = $\max\left\{\text{LBD}, h(\boldsymbol{\pi}^{(k)})\right\}$
- If solution $\mathbf{x}^{(k)}$ is feasible, update pessimistic bound, UBD = $\min\left\{\text{UBD}, f(\mathbf{x}^{(k)})\right\}$, else try to find a feasible solution using some heuristic and then update UBD.

Step 3: Check stop criterion.

Step 4: With a search direction $\boldsymbol{d}^{(\boldsymbol{k})}$ and steplength $t^{(k)}$, update the Lagrangian multipliers $\boldsymbol{\pi}^{(k+1)} = \boldsymbol{\pi}^{(k)} + t^{(k)}\boldsymbol{d}^{(k)}$.

Step 5: Set k = k + 1. Go to Step 2.

The initial Lagrangian multiplier is often set $\boldsymbol{\pi}^{(0)} = 0$. This can be considered as ignoring the relaxed constraints in the first iteration.

As pointed out by Lundgren et al. (2010), the difficulty in the solution strategy above lays in updating the Lagrangian multiplier in Step 4. The goal is to find multipliers that converge against the optimal multiplier quickly.

Chapter 4

Problem Description

The problem presented below revolves around how to create an operational production plan. One seeks to produce total demand over the planning horizon while minimizing total costs. Two demand types, actual demand and forecasted demand, should be considered when creating the plan. Both production for actual and forecasted demand should be produced in batches and both demand types should be accounted for in capacity limits. Further, the production plan should be updated every day based on the arrival of new demand information. Section 4.1 explains the problem of creating a production plan for one planning horizon, while Section 4.2 explains the problem where new, updated production plans are generated based on the arrival of new demand information.

4.1 Production Planning Problem

A production plan should be created for a set of distinct days, creating a planning horizon. A week is regarded as the period between a Monday and the following Sunday, and not strictly as seven days.

Norsk Hydro has a set of products that are described by an alloy and a dimension. Furthermore, Norsk Hydro have a set of plants where each plant has one casthouse. Each casthouse has a set of different casting tables that are either compatible or non-compatible to produce a certain product dimension. A each casthouse, there is at most one casting table able to produce a specific product dimension. Each casthouse has an upper capacity limit which is limited by available metal flow at the plant each day. Each casting table also has an upper capacity limit, but this is limited by the daily rate at which they can cast products. The sum of all products produced for all customers at a given day at a casting table can not exceed the related capacity limit. In addition, the sum of production at all casting tables located at a given casthouse can not exceed its casthouse capacity limit.

The production plan should account for both actual and forecasted demand in the given planning horizon. Actual demand is demand requested by customers through a *call-off*. A customer will place at most one call-off of a specific product with a specific delivery day. A call-off can therefore be described as a demand quantity with a unique combination of customer, product and delivery day. Forecasted demand is a prediction of call-offs that are assumed to be delivered each week, but have not been confirmed by the customers through a call-off at the time the production plan is to be generated. Forecasted demand can therefore be described as a demand quantity with a unique combination of customer, product and assumed delivery week.

All call-offs must be allocated to at most one casting table, and the entire call-off demand must be produced at the assigned casting table. It is however possible to produce call-off demand over several days. Forecasted demand must be produced the same week as it is assumed to be delivered to the given customer. It is not necessary to produce the whole forecasted demand for a unique combination of customer, product and week at the same casting table. Production of call-off and forecasted demand both use casting table and casthouse capacity. If total demand is higher than total capacity, some demand can not be produced. In such a case, forecasted demand should be discarded before call-off demand is rejected. To simplify the problem, it is assumed that production can be delivered to customers the same day it is produced.

Products are produced in *production batches* and sold in *order batches*. Production must be a multiple of a production batch while the quantity requested in an call-off must be a multiple of an order batch. Production batches and order batches are not equal in volume. Due to this, it must be possible to store products made at a casting table from one day to another. Storage is also necessary in order to satisfy call-off demand when production capacity is full at delivery day. The storage level should be measured at the end of each day.

The objective is to minimize the total costs related to implementing production plans. There is a transportation cost associated with delivery from a casting table to a customer. There is also a production cost associated with production of a product at a casting table. Further, there is a holding cost associated with storing products from one day to another. Call-offs that are not allocated to any casting tables, and are rejected, cause a penalty cost. The same is true for forecasted demand that is discarded in the production plan.

4.2 Production Planning Problem with Updated Demand Information

Section 4.1 explains the problem revolving around how to create a production plan for one planning horizon. In reality, only the production plan for the first day in a planning horizon is implemented as daily changes in demand may leave the remaining production plan sub-optimal. Dynamic demand can be handled better by updating the production plan each day.

Each day a number of new call-offs arrive and should be accounted for in the production plan. From previous production plans, this call-off demand has been represented as forecasted demand with an estimated delivery day. Thus, the associated forecasted demand must be reduced with the quantity of the call-off demand. Forecasted demand does not need to be accurately estimated. As a result, if there exists forecasted demand at the end of the week, the remaining demand is added to the next weeks forecasted demand. The production plan is further updated with the new demand information. In addition, initial storage must be updated to reflect the correct state.

Chapter 5

Mathematical Model

In this chapter the mathematical model for the production planning problem described in section 4.1 is presented. This model corresponds to a production plan created for one iteration in the rolling horizon planning approach. The model considers which call-offs to accept and which to reject. Accepted call-offs are allocated to a specific casting table. In addition, the model utilizes available capacity by considering forecasted demand. As a result, a complete production plan is created. Forecasted demand can be discarded if it is higher than the total capacity. If a call-off is rejected or forecasted demand is discarded, this is penalized in the objective function. Section 5.1 introduces the indices, sets, parameters and variables used in the model. In Section 5.2, the model's objective function and constraints are presented and explained in detail. The complete model is given in Section 5.3.

5.1 Notation

Table 5.1: Indices.

Indices		
c	-	Customer
m	-	Casting table
k	-	Casthouse
p	-	Product
t	-	Day
w	-	Week

Table 5.2: Sets.

Sets		
$\mathcal{C}$	-	Set of customers
$\mathcal{K}$	-	Set of casthouses
$\mathcal{M}$	-	Set of casting tables
$\mathcal{M}^k$	-	Subset of casting tables associated with casthouse k
$\mathcal{M}^p$	-	Subset of casting tables associated with product p
$\mathcal{P}$	-	Set of products
$\mathcal{P}^m$	-	Subset of products associated with casting table m
$\mathcal{T}$	-	Set of days in planning horizon
$\mathcal{T}^w$	-	Subset of remaining days in week w
$\mathcal{W}$	-	Set of weeks in planning horizon

Table 5.3: Parameters

Parameters.		
B	-	Production batch in tonnes
C^P_{mp}	-	Production cost per tonne when producing product p at casting table m
C^T_{cm}	-	Transportation cost per tonne when delivering from casting table m to customer c
C^H	-	Holding cost per tonne per day
C^D	-	Decline cost per tonne when rejecting call-off demand
C^F	-	Penalty cost per tonne when discarding forecasted demand
D_{cpt}	-	Call-off demand in tonnes from customer c, of product p with delivery at day t
K_m	-	Daily capacity in tonnes at casting table m
K_k	-	Daily capacity in tonnes at casthouse k
P_{cpw}	-	Forecasted demand in tonnes for customer c of product p assumed to arrive in week w
s_{mp0}	-	Initial storage level in tonnes at casting table m of product p

Table 5.4: Variables.

Variables		
x_{mpt}	-	Number of batches to be produced at casting table m of product p at day t
y_{mpt}	-	Number of batches forecasted to be produced at casting table m of product p at day t
s_{mpt}	-	Quantity in tonnes stored at casting table m of product p at the end of day t
z_{pt}	-	Quantity in tonnes of discarded forecasted demand of product p at day t
δ_{cmpt}	-	Binary variable being 1 if customer c receives its call-off of product p with delivery at day t from casting table m, 0 otherwise

5.2 Production Planning Model

Objective Function

$$
\begin{aligned}
\min & \sum_{c\in\mathcal{C}} \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^T_{cm} D_{cpt} \delta_{cmpt} \\
& + \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^P_{mp} B \left(x_{mpt} + y_{mpt}\right) \\
& + \sum_{c\in\mathcal{C}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^D D_{cpt} \left(1 - \sum_{m\in\mathcal{M}} \delta_{cmpt}\right) \\
& + \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^H s_{mpt} \\
& + \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^F z_{pt}
\end{aligned}
\tag{5.2.1}
$$

The objective function presented in Equation (5.2.1) is divided into five terms. The first term calculates the transportation cost for delivering a call-off to a customer. The next term calculates the production cost for producing products. The third term is a penalty cost that incurs if a call-off is rejected. The holding cost is calculated in the fourth term and in the fifth term a penalty cost is added if forecasted demand is discarded.

Constraints

$$
s_{mpt-1} + Bx_{mpt} - \sum_{c\in\mathcal{C}} \left(D_{cpt}\delta_{cmpt}\right) - s_{mpt} = 0 \quad m \in \mathcal{M}, p \in \mathcal{P}^m, t \in \mathcal{T} \tag{5.2.2}
$$

The inventory balance for each day is presented in Constraint (5.2.2). This constraint also ensures that accepted call-off demand is satisfied either from new production, Bx_{mpt}, or by inventory, s_{mpt-1}. Further, δ_{cmpt} ensures that the demand is only included at the casting table that is assigned to produce the given call-off. Inventory and production are not assigned to a specific customer, and thus accepted demand is aggregated for all customers.

$$
\sum_{m\in\mathcal{M}} \delta_{cmpt} \leq 1 \qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \tag{5.2.3}
$$

A call-off can either be rejected ($\delta_{cmpt} = 0$) or accepted ($\delta_{cmpt} = 1$) at a casting table. In Constraints (5.2.3), call-offs are assigned to at most one casting table. As a result, the whole demand related to a given call-off must be produced at the same casting table. It is possible to reject call-offs as Constraints (5.2.3) are $\leq$ -constraints. This is not a desired outcome and incurs a penalty cost in the objective function.

$$\sum_{m \in \mathcal{M}^p} \sum_{t \in \mathcal{T}^w} B y_{mpt} + \sum_{t \in \mathcal{T}^w} z_{pt} \geq \sum_{c \in \mathcal{C}} P_{cpw} \qquad p \in \mathcal{P}, w \in \mathcal{W} \tag{5.2.4}$$

The forecasted demand should be accounted for in the production plan in order to pre-allocate, or save, capacity for call-offs assumed to arrive in later in time. Constraints (5.2.4) allocates the weekly forecasted demand to casting tables at a given day. The production day must be in the same week as the forecasted demand is assumed to be delivered to customer, i.e. $\mathcal{T}^w$. Furthermore, z_{pt} enables forecasted demand to be discarded. This decision incurs a penalty cost in the objective function.

$$\sum_{p \in \mathcal{P}} B\left(x_{mpt} + y_{mpt}\right) \leq K_m \qquad m \in \mathcal{M}, t \in \mathcal{T} \tag{5.2.5}$$

$$\sum_{m \in \mathcal{M}^k} \sum_{p \in \mathcal{P}} B\left(x_{mpt} + y_{mpt}\right) \leq K_k \qquad k \in \mathcal{K}, t \in \mathcal{T} \tag{5.2.6}$$

Constraints (5.2.5) and (5.2.6) handle upper capacity limits. Constraints (5.2.5) ensure that call-off production, x_{mpt}, and forecasted production, y_{mpt}, for each day does not exceed the daily capacity of the casting table. Further, Constraints (5.2.6) ensure that the sum of call-off and forecast related production at all casting tables related to a casthouse does not exceed the daily capacity of the casthouse.

$$x_{mpt}, y_{mpt} \in \mathbb{Z}^+ \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{5.2.7}$$

$$s_{mpt} \geq 0 \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{5.2.8}$$

$$z_{pt} \geq 0 \qquad p \in \mathcal{P}, t \in \mathcal{T} \tag{5.2.9}$$

$$\delta_{cmpt} \in \{0, 1\} \qquad c \in \mathcal{C}, m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{5.2.10}$$

Finally, Constraints (5.2.7)-(5.2.10) are non-negativity constraints. In addition, Constraints (5.2.7) state that x_{mpt} and y_{mpt} are integer variables. Constraints (5.2.10) state that δ_{cmpt} are binary variables.

5.3 Complete Model

$$
\begin{aligned}
\min \; & \sum_{c\in\mathcal{C}} \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^T_{cm} D_{cpt} \delta_{cmpt} \\
& + \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^P_{mp} B \left(x_{mpt} + y_{mpt} \right) \\
& + \sum_{c\in\mathcal{C}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^D D_{cpt} \left(1 - \sum_{m\in\mathcal{M}} \delta_{cmpt} \right) \\
& + \sum_{m\in\mathcal{M}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^H s_{mpt} \\
& + \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^F z_{pt}
\end{aligned}
$$

subject to

$$
\begin{aligned}
s_{mpt-1} + Bx_{mpt} - \sum_{c\in\mathcal{C}} \left(D_{cpt}\delta_{cmpt} \right) - s_{mpt} &= 0 && m \in \mathcal{M}, p \in \mathcal{P}^m, t \in \mathcal{T} \\
\sum_{m\in\mathcal{M}} \delta_{cmpt} &\leq 1 && c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \\
\sum_{m\in\mathcal{M}^p} \sum_{t\in\mathcal{T}^w} By_{mpt} + \sum_{t\in\mathcal{T}^w} z_{pt} &\geq \sum_{c\in\mathcal{C}} P_{cpw} && p \in \mathcal{P}, w \in \mathcal{W} \\
\sum_{p\in\mathcal{P}} B \left(x_{mpt} + y_{mpt} \right) &\leq K_m && m \in \mathcal{M}, t \in \mathcal{T} \\
\sum_{m\in\mathcal{M}^k} \sum_{p\in\mathcal{P}} B \left(x_{mpt} + y_{mpt} \right) &\leq K_k && k \in \mathcal{K}, t \in \mathcal{T} \\
x_{mpt}, y_{mpt} &\in \mathbb{Z}^+ && m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \\
s_{mpt} &\geq 0 && m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \\
z_{pt} &\geq 0 && p \in \mathcal{P}, t \in \mathcal{T} \\
\delta_{cmpt} &\in \{0, 1\} && c \in \mathcal{C}, m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T}
\end{aligned}
$$

Chapter 6

Solution Approach

This chapter explains how a rolling horizon approach and a Lagrangian relaxation is applied for the mathematical model presented in Chapter 5. Section 6.1 explains how new demand information is dealt with in the rolling horizon approach followed by a discussion on which variables to lock in order to deal with the concept of nervousness. A flow diagram illustrating the rolling horizon approach is also presented.

A similar model to the one presented in this thesis was studied in Reine et al. (2017). In this work, the assumption of long computational time was confirmed. There is a need of a method that can reduce the computational time and reduce the optimality gap. Therefore the solution method, Lagrangian relaxation, is applied and described in Section 6.2.

6.1 Rolling Horizon Approach

The daily arrival of new call-offs and updated forecasts motivates the use of a rolling horizon approach when the model is implemented. This way, the model can apply the most accurate and recent information for demand. Another benefit from the rolling horizon approach is the ability to alter decisions in the future when more information is available. In this model, a rolling horizon approach with $T = 35$ and $\Delta\mathrm{T} = 1$ is implemented. This indicates that each day a production plan for the next 35 days is created, and the production plan for the first day is implemented. The concept of implementing the results from the first day and shifting the planning horizon one day forward is referred to as a *roll*.

6.1.1 New Demand Information

For each day, the first day in the production plan from the previous roll should be implemented. Further, the model should be rolled forward and create an updated production plan with new demand information.

For each roll R, new call-offs, D^{R*}_{cpt}, arrive. This demand should be added to the call-off demand, D^{R}_{cpt}, that arrived in previous rolls. This is calculated as shown in Equation (6.1.1). For every combination of customer, product and day (cpt) at most one of the parameters D^{R-1}_{cpt} and D^{R*}_{cpt} is positive.

$$D^{R-1}_{cpt} + D^{R*}_{cpt} = D^{R}_{cpt} \qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.1.1}$$

Forecasted demand is an estimate of when call-offs will arrive along with its assumed demand quantity. The forecasted demand must be reduced to reflect the updated demand information. As long as the first day is not Monday, the forecasted demand is updated as shown in Equation (6.1.2).

$$\max\left\{0, P^{R-1}_{cpw} - \sum_{t \in \mathcal{T}^w} D^{R*}_{cpt}\right\} = P^{R}_{cpw} \qquad c \in \mathcal{C}, p \in \mathcal{P}, w \in \mathcal{W} \tag{6.1.2}$$

In the case when the first day is Monday, the forecasted demand that has not been estimated correctly from the previous week must be added to the next week. The calculations for this case is given in Equation (6.1.3).

$$\max\left\{0, P^{R-1}_{cpw-1} + P^{R-1}_{cpw} - \sum_{t \in \mathcal{T}^w} D^{R*}_{cpt}\right\} = P^{R}_{cpw} \qquad c \in \mathcal{C}, p \in \mathcal{P}, w \in \mathcal{W} \tag{6.1.3}$$

Call-offs that are rejected, $D^{R'}_{cpt}$, when running the model, should not be considered in future rolls. It therefore needs to be removed from the call-off demand for the next roll. This is shown in Equation (6.1.4). While rejected call-off demand is not considered in future rolls, discarded forecasted demand is not removed.

$$D^{R}_{cpt} - D^{R'}_{cpt} = D^{R+1}_{cpt} \qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.1.4}$$

6.1.2 Locking Decisions

Section 3.2.2 discusses *nervousness*, a concept that arises in rolling horizon approach models when decisions are altered or re-optimized over time. An issue related to high nervousness in planning systems is the dependency between planning systems. Changes in one planning system will propagate to other systems if they are tightly coupled. The model in this thesis creates operational production plans, a low-level hierarchical planning system exploiting few decisions from higher-level systems. Consequently, altering decisions in this model will have minimal effect on other planning systems used by Norsk Hydro. Nervousness created from this model and its impact on other planning systems is thus considered of low importance.

A more important issue related to nervousness in this model is the low degree of predictability both Norsk Hydro and its customers get if production plans change from day to day. By freezing, hereby referred to as locking, decisions, it is possible to reduce changes and increase the predictability.

Frequent changes to the production plan are not regarded as a positive feature. This is especially true for the first few days in the production plan as these are soon to be implemented. Nevertheless, there exists a trade-off between stable, meaning non-changing, production plans and more accurate demand information. This implies that it would be beneficial to lock production variables if demand is certain.

The first couple of days in the production plan mostly consists of call-off demand, which is certain demand. This suggests that production variables related to call-off demand (x_{mpt}) for these days should be locked. In contrast, the last days in the production plan is mostly based on forecasted demand and are less likely to reflect the true demand and may even give incorrect information. In addition, the decisions are further into the future. Production related to forecasted demand (y_{mpt}) should thus not be locked as this demand is estimated and may contain forecast errors. This implies that discarding variables related to forecasted demand (z_{pt}) should not be locked either.

It is important for Norsk Hydro to give a quick response to whether a customer's incoming call-off is accepted or rejected, a decision related to the δ_{cmpt}-variables. When a call-off is accepted ($\sum_{m \in \mathcal{M}} \delta_{cmpt} = 1$), all updated production plans generated in the future should also accept this call-off as anything else is considered bad customer service. Norsk Hydro may however change the location of the call-off as customers usually are indifferent to where their demand is produced.

From the discussion above it becomes clear that both the call-off production variables, x_{mpt}, and call-off allocation variables, δ_{cmpt}, should be locked.

In order to account for production locked in the previous rolls, lower bounds are created for x_{mpt}-variables. The calculation is shown in Equation (6.1.5), where X_{mpt}^{R-1} is the locked production from the previous roll. The upper bounds are only calculated for a subset of $\mathcal{T}$, here represented as $\mathcal{T}^L$, containing the first n days of the planning horizon. This is to add flexibility by enable re-allocation of call-off demand which possibly will give better solutions.

$$X_{mpt}^{R-1} \leq x_{mpt}^{R} \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T}^L \tag{6.1.5}$$

Further, if a call-off is accepted in one roll, the call-off related to δ_{cmpt} must for all future rolls also be accepted. It should however be possible to change the casting table producing the call-off. By locking δ_{cmpt} variables to either 1 or 0, this will not be possible. Thus, a new constraint must be added to the mathematical model to enable the change of production location. This feature is presented in Constraint (6.1.6). Here, the parameter Δ_{cpt} is one if a call-off is accepted in previous rolls, and zero otherwise.

$$\sum_{m \in \mathcal{M}} \delta_{cmpt} \geq \Delta_{cpt} \qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T}, \tag{6.1.6}$$

where Δ_{cpt} is calculated as shown in Equation (6.1.7).

$$\max \left\{ \Delta_{cpt}^{R-1}, \sum_{m \in \mathcal{M}} \delta_{cmpt}^{R-1} \right\} = \Delta_{cpt}^{R} \qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.1.7}$$

6.1.3 Procedure

For each roll, R, five parameters must be updated to reflect the new demand as well as the state of the system. These are D_{cpt}^{R}, P_{cpw}^{R}, X_{mpt}^{R}, Δ_{cpt}^{R} and s_{mp0}^{R}. Initial storage, s_{mp0}^{R}, is calculated as shown in Equation (6.1.8).

$$s_{mp1}^{R-1} = s_{mp0}^{R} \qquad m \in \mathcal{M}, p \in \mathcal{P} \tag{6.1.8}$$

The rolling horizon approach for the specific problem in this thesis is illustrated with a flow diagram in Figure 6.1.

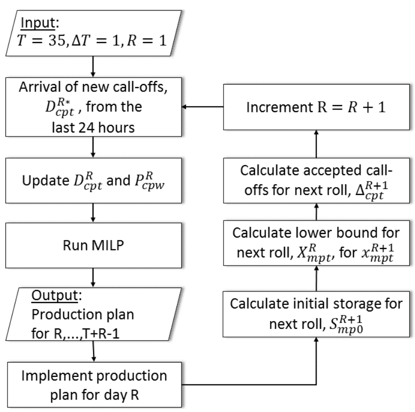


Figure 6.1: Flow diagram of the rolling horizon approach.

6.2 Lagrangian Relaxation

In the preliminary study of Reine et al. (2017) the model ran for 2 hours without reaching optimality. Lagrangian relaxation has been proposed as a solution method with the potential to reach optimality, or at least reduce the optimality gap, for integer problems. Problems may have an underlying block angular structure that is destroyed by one or more sets of constraints. By relaxing these constraints the problem can be divided into distinct subproblems that can be solved separately. These subproblems may be easier to solve, or their structure can make it possible to solve them near optimality by a simple heuristic.

Identifying the complicating constraints to be relaxed is not always obvious before testing. Thus, in this section the implementation of two different Lagrangian relaxations is presented. In Section 6.2.1 a relaxation of the capacity constraints is presented, and Section 6.2.2 presents a relaxation of the constraints which ensure that a call-off is produced only once at one, single casting table. These are further referred to as *capacity relaxation* and *casting table relaxation* respectively.

6.2.1 Capacity Relaxation

As mentioned in Section 3.2.3, Karimi et al. (2003) claim that the capacity constraints have great impact on the complexity of an LSP. As our problem is similar to a LSP, it is expected that a relaxation of these constraints will make the problem easier to solve. The problem of this thesis have two sets of capacity constraints:

$$\sum_{p \in \mathcal{P}} B(x_{mpt} + y_{mpt}) \leq K_m \qquad m \in \mathcal{M}, t \in \mathcal{T} \tag{5.2.5}$$

$$\sum_{m \in \mathcal{M}^k} \sum_{p \in \mathcal{P}} B(x_{mpt} + y_{mpt}) \leq K_k \qquad k \in \mathcal{K}, t \in \mathcal{T} \tag{5.2.6}$$

The Lagrangian function is denoted L, and each of the relaxed constraints have an associated Lagrangian multiplier, $\pi^i \geq 0$. This multiplier is considered constant when solving the Lagrangian subproblem. The Lagrangian relaxation of the capacity constraints is given in (6.2.1) - (6.2.8).

$$
\begin{aligned}
L((\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{z},\boldsymbol{\delta}),\boldsymbol{\pi}) = & \sum_{c\in\mathcal{C}}\sum_{m\in\mathcal{M}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^T_{cm} D_{cpt}\delta_{cmpt} \\
& + \sum_{m\in\mathcal{M}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^P_{mp} B\left(x_{mpt}+y_{mpt}\right) \\
& + \sum_{c\in\mathcal{C}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^D D_{cpt}\left(1-\sum_{m\in\mathcal{M}}\delta_{cmpt}\right) \\
& + \sum_{m\in\mathcal{M}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^H s_{mpt} \\
& + \sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^F z_{pt} \\
& + \sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} \pi^1_{mt}\left(\sum_{p\in\mathcal{P}} B(x_{mpt}+y_{mpt}) - K_m\right) \\
& + \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} \pi^2_{kt}\left(\sum_{m\in\mathcal{M}^k}\sum_{p\in\mathcal{P}} B(x_{mpt}+y_{mpt}) - K_k\right)
\end{aligned} \tag{6.2.1}
$$

subject to

$$
s_{mpt-1} + Bx_{mpt} - \sum_{c\in\mathcal{C}}(D_{cpt}\delta_{cmpt}) = s_{mpt} \qquad m\in\mathcal{M}, p\in\mathcal{P}^m, t\in\mathcal{T} \tag{6.2.2}
$$

$$
\sum_{m\in\mathcal{M}}\delta_{cmpt} \le 1 \qquad c\in\mathcal{C}, p\in\mathcal{P}, t\in\mathcal{T} \tag{6.2.3}
$$

$$
\sum_{m\in\mathcal{M}^p}\sum_{t\in\mathcal{T}^w} By_{mpt} + \sum_{t\in\mathcal{T}^w} z_{pt} \ge \sum_{c\in\mathcal{C}} P_{cpw} \qquad p\in\mathcal{P}, w\in\mathcal{W} \tag{6.2.4}
$$

$$
x_{mpt}, y_{mpt} \in \mathbb{Z}^+ \qquad m\in\mathcal{M}, p\in\mathcal{P}, t\in\mathcal{T} \tag{6.2.5}
$$

$$
s_{mpt} \ge 0 \qquad m\in\mathcal{M}, p\in\mathcal{P}, t\in\mathcal{T} \tag{6.2.6}
$$

$$
z_{pt} \ge 0 \qquad p\in\mathcal{P}, t\in\mathcal{T} \tag{6.2.7}
$$

$$
\delta_{cmpt} \in \{0,1\} \qquad c\in\mathcal{C}, m\in\mathcal{M}, p\in\mathcal{P}, t\in\mathcal{T} \tag{6.2.8}
$$

For a chosen multiplier, $\boldsymbol{\pi}'$, some terms in the Lagrangian function are constant and can be removed in further analyses. The problem can then be separated into a subproblem for every product. Following is the Lagrangian function, L_p, for each of these subproblems:

$$\begin{aligned}
L_p((\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{z},\boldsymbol{\delta}),\boldsymbol{\pi}) = &\sum_{c\in\mathcal{C}}\sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} C^T_{cm} D_{cpt}\delta_{cmpt} \\
&+ \sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} C^P_{mp} B\left(x_{mpt}+y_{mpt}\right) \\
&- \sum_{c\in\mathcal{C}}\sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} C^D D_{cpt}\delta_{cmpt} \\
&+ \sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} C^H s_{mpt} \\
&+ \sum_{t\in\mathcal{T}} C^F z_{pt} \\
&+ \sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} \pi^1_{mt} B(x_{mpt}+y_{mpt}) \\
&+ \sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}^k}\sum_{t\in\mathcal{T}} \pi^2_{kt} B(x_{mpt}+y_{mpt})
\end{aligned} \tag{6.2.9}$$

The sum of the objective functions of all the subproblems and the constant terms will be the value of the Lagrangian function of the complete problem. This function is given in Equation (6.2.10).

$$L = \sum_{p\in\mathcal{P}} L_p + \sum_{c\in\mathcal{C}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} C^D D_{cpt} - \sum_{m\in\mathcal{M}}\sum_{t\in\mathcal{T}} \pi^1_{mt} K_m - \sum_{k\in\mathcal{K}}\sum_{t\in\mathcal{T}} \pi^2_{kt} K_k \tag{6.2.10}$$

From this formulation, it is possible to derive the Lagrangian dual problem written as in Section 3.3:

$$\max_{\boldsymbol{\pi}\geq 0} h(\boldsymbol{\pi}),$$

Let $\boldsymbol{V}$ be a vector containing all the variables in the problem i.e. $\boldsymbol{V} = (\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{z},\boldsymbol{\delta})$. Then the Lagrangian dual function $h(\boldsymbol{\pi})$ is given as in Equation (6.2.1).

$$h(\boldsymbol{\pi}) = \min_{\mathbf{V}\geq 0} L(\mathbf{V},\boldsymbol{\pi})$$

6.2.2 Casting Table Relaxation

For the capacity relaxation it was shown that the problem could be separated into a subproblem per product. In the preliminary study of this thesis, Reine et al. (2017), a relaxation of Constraint (5.2.3) was proposed.

$$\sum_{m\in\mathcal{M}} \delta_{cmpt} \leq 1 \qquad\qquad c\in\mathcal{C}, p\in\mathcal{P}, t\in\mathcal{T} \tag{5.2.3}$$

The model formulation of Reine et al. (2017) is attached in Appendix A. In that model, the constraint corresponding to Constraint (5.2.3) is a sum over casthouses instead of casting tables, and a relaxation would separate the problem into one subproblem per casthouse. However, with the model modifications of this thesis, a relaxation of Constraint (5.2.3) is no longer sufficient for separation into smaller subproblems. Both Constraint (5.2.4) and (5.2.6) are complicating for separation into one subproblem per casting table.

Despite this property it may still be of interest to examine the effect of relaxing Constraint (5.2.3). The extent of this constraint can make its relaxation efficient because of the reduction of constraints. For comparison, the capacity relaxation will imply a reduction of $|\mathcal{M}| \cdot |\mathcal{T}| + |\mathcal{K}| \cdot |\mathcal{T}| = 2\ 030$ constraints, while relaxing Constraint (5.2.3) will give a reduction of $|\mathcal{C}| \cdot |\mathcal{P}| \cdot |\mathcal{T}| = 2\ 720\ 340$ constraints.

The Lagrangian relaxation of Constraint (5.2.3) is given in (6.2.11) - (6.2.19).

$$
\begin{aligned}
L((\mathbf{x}, \mathbf{y}, \mathbf{s}, \mathbf{z}, \boldsymbol{\delta}), \boldsymbol{\pi}) = & \sum_{c \in \mathcal{C}} \sum_{m \in \mathcal{M}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} C^T_{cm} D_{cpt} \delta_{cmpt} \\
& + \sum_{m \in \mathcal{M}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} C^P_{mp} B \left(x_{mpt} + y_{mpt} \right) \\
& + \sum_{c \in \mathcal{C}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} C^D D_{cpt} \left(1 - \sum_{m \in \mathcal{M}} \delta_{cmpt} \right) \\
& + \sum_{m \in \mathcal{M}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} C^H s_{mpt} \\
& + \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} C^F z_{pt} \\
& + \sum_{c \in \mathcal{C}} \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \pi_{cpt} \left(\sum_{m \in \mathcal{M}} \delta_{cmpt} - 1 \right)
\end{aligned}
\tag{6.2.11}
$$

subject to

$$
s_{mpt-1} + B x_{mpt} - \sum_{c \in \mathcal{C}} \left(D_{cpt} \delta_{cmpt} \right) = s_{mpt} \qquad m \in \mathcal{M}, p \in \mathcal{P}^m, t \in \mathcal{T} \tag{6.2.12}
$$

$$
\sum_{m \in \mathcal{M}^p} \sum_{t \in \mathcal{T}^w} B y_{mpt} + \sum_{t \in \mathcal{T}^w} z_{pt} \geq \sum_{c \in \mathcal{C}} P_{cpw} \qquad p \in \mathcal{P}, w \in \mathcal{W} \tag{6.2.13}
$$

$$
\sum_{p \in \mathcal{P}} B(x_{mpt} + y_{mpt}) \leq K_m \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.14}
$$

$$
\sum_{m \in \mathcal{M}^k} \sum_{p \in \mathcal{P}} B(x_{mpt} + y_{mpt}) \leq K_k \qquad k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.15}
$$

$$x_{mpt}, y_{mpt} \in \mathbb{Z}^+ \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.16}$$

$$s_{mpt} \geq 0 \qquad m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.17}$$

$$z_{pt} \geq 0 \qquad p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.18}$$

$$\delta_{cmpt} \in \{0, 1\} \qquad c \in \mathcal{C}, m \in \mathcal{M}, p \in \mathcal{P}, t \in \mathcal{T} \tag{6.2.19}$$

The dual problem for this relaxation will be noted in the same way as for the capacity relaxation.

6.2.3 Subgradient Method

In this section the implementation of the subgradient method for updating the Lagrangian multipliers is described. The explanation of the method is based on Lundgren et al. (2010).

When studying the Lagrangian function it can be argued that the values of the Lagrangian multipliers will determine whether the solution will be feasible in the original problem. Further, the gradient of the Lagrangian dual function will give the ascent search direction for the next iteration. Lundgren et al. (2010) argues that the dual function is piecewise linear and hence not differentiable. Thus, the generalization of the gradient, the subgradient, is applied as the new search direction. The subgradient does not guarantee that the search direction is in ascent direction, and thus the step length must be set carefully to ensure convergence.

In the iterative process of the subgradient method each new optimal solution of the dual function, $h(\boldsymbol{\pi})$, provides a subgradient. This is calculated by inserting the optimal solution into the relaxed constraints. If the multipliers are defined as being positive, the subgradient for the general minimization problem described in Section 3.3.1 is then in iteration k, given as $b_i - g_i(\boldsymbol{x}^{(k)})$.

The selection of step length, $t^{(k)}$, has to guarantee convergence and should be chosen such that $t^{(k)} \to 0$ and $\sum_{i=1}^{k} t^{(k)} \to \infty$, when $k \to \infty$. The formula in Equation (6.2.20) is commonly used to calculate the step length in subgradient optimization.

$$t^{(k)} = \frac{\lambda(z^* - h(\boldsymbol{\pi}^{(k)}))}{||g(\bar{\boldsymbol{x}})||^2} \tag{6.2.20}$$

$||g(\bar{\boldsymbol{x}})||^2$ is the squared length of the subgradient of $h(\boldsymbol{\pi})$. z^* is the optimal value of the objective function. Since this value is unknown it is replaced by the best known feasible solution, the UBD, as an approximation. λ is a constant in the interval $(0, 2]$ and is often initially set to 2 and decreased after a number of iterations, p, where the solution of the the dual function has not improved. Following are the additional steps to include subgradient optimization into the general solution strategy given in Section 3.3.1.

Extension of the General Solution Strategy

Step 1: Choose initial values of $\lambda^{(0)} \in (0, 2], \epsilon, k_{\max}$ and p.

Step 2: Follow the strategy given in Section 3.3.1 until step 3.

Step 3: Check convergence criteria. If (UBD − LBD)/LBD $\leq \epsilon$ or $k = k_{\max} \rightarrow$ STOP and let $\bar{\boldsymbol{x}}$ be the optimal solution.

Step 4: Compute the subgradient $g(\bar{\boldsymbol{x}})$.

Step 5: Determine the step length as in Equation (6.2.20). Update $\lambda := \lambda/2$ if LBD has not improved in the last p iterations.

Step 6: Continue the algorithm in Section 3.3.1.

6.2.4 Heuristic Procedure for Finding UBD

Solving the Lagrangian subproblem may not provide solutions that are feasible in the original problem in Chapter 5. Feasible solutions are needed to calculate the UBD. Consequently, a heuristic has to be applied to the solutions of the Lagrangian subproblems to make them feasible and obtain UBD from these solutions.

As will be shown in the computational study in Chapter 8, the casting table relaxation suggested in Section 6.2.2 is superior with regards to computational efficiency when compared to the capacity relaxation suggested in Section 6.2.1. The following section presents the heuristic used to obtain feasible solutions from the Lagrangian subproblem output of the casting table relaxation. The heuristic is combined with a variant of the model in Chapter 5 called the *resolve problem*. The heuristic and the resolve problem together form what is referred to as the *heuristic procedure*. The algorithm for the heuristic procedure is presented in Algorithm 1, where lines 3 to 12 describe the heuristic and line 13 describes how the resolve problem is applied.

Relaxing Constraint (5.2.3) allows a call-off to be produced several times along different casting tables. This is irrational in a practical sense as a call-off only should be satisfied once. In order to make this output feasible for the original problem formulation, one has to ensure that only one casting table produces an accepted call-off.

The heuristic is based on a greedy algorithm which assigns δ_{cmpt}-variables to the casting table that has the lowest transportation costs to a given customer. The heuristic begins by identifying instances where more than one casting table has been assigned to produce the same call-off as shown in the equation below.

$$\sum_{m \in \mathcal{M}} \delta_{cmpt} > 1 \qquad\qquad c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T}$$

When such an instance has been identified, it checks a list containing the cheapest casting table for the given customer. The casting table is found by choosing the lowest transportation cost, C^T_{cm}. This list is denoted C^{BEST}_c in the algorithm. The δ_{cmpt}-variable associated with this casting table remains 1, while the others are set to zero.

When this has been done for all instances where more than one casting table has been assigned to produce a call-off, the δ_{cmpt}-variables set to 1 are included as input in the *resolve problem*. The resolve problem is similar to the problem in Chapter 5, the only distinction being that δ_{cmpt}-variables set to 1 from the heuristic, are locked to this value. When a large amount of the δ_{cmpt}-variables have predetermined values, the resolve problem is often able to achieve a low optimality gap more quickly. The solution from this model is feasible in the problem in Chapter 5 and thus provides a UBD.

Algorithm 1 Heuristic procedure that creates feasible solutions from Lagrangian subproblem output.

Input
Variables from relaxed problem δ_{cmpt}
List with the cheapest casting table for each customer: C^{BEST}_c
Output
Feasible solution
LBD and UBD to resolve problem
for c,p,t$\in \mathcal{C}, \mathcal{P}, \mathcal{T}$ **do**
  if $\sum_{m \in \mathcal{M}} \delta_{cmpt} > 1$ **then**
    for $m \in \mathcal{M}$ **do**
      $\delta_{cmpt} = 0$
      if m $= C^{\text{BEST}}_c$ **then**
        $\delta_{cmpt} = 1$
      end if
    end for
  end if
end for
Run resolve problem with updated δ_{cmpt}-variables, locking δ_{cmpt}-variables that are equal to 1

Chapter 7

Case Study

The mathematical model and solution approach described in Chapter 5 and 6, needs to be tested with realistic data to reflect Norsk Hydro's case. A data set, hereby referred to as *source data set* is provided by Norsk Hydro. This data set contains data regarding a range of different casthouses and their properties as well as historical call-offs for a year.

Section 7.1 describes data that is directly extracted from the source data set. This data is time independent and is the same for all rolls performed with a rolling horizon approach. Demand has been simulated as the demand from the source data set was incomplete in terms of the attributes required for this model. Section 7.2 outlines how demand is generated through queuing theory and simulation. Finally, Section 7.3 summarizes the data to the reader to provide an overall sense of its scope.

7.1 Input Data

Section 7.1.1 explains how most of the sets and parameters for the model are found using the source data set, while 7.1.2 discusses the relation between the different cost parameters used in the model as most have not been readily available in the source data set. The sets and parameters considered in this section are not time dependent and do not change from roll to roll.

7.1.1 General Data

From the historical call-offs in the source data set it is possible to retrieve 153 distinct customers and 508 distinct products. The products are all in the extrusion ingot category, and are a combination of an alloy and one of 14 different geometric dimensions. Both alloy type and dimension for a product is given in the source data set.

The source data set considers all casthouses located in Europe producing extrusion ingots. This adds up to nine different casthouses. The daily production capacities for each plant is dependent on the available flow of liquid metal at the plant. These have values ranging between 50 and 650 tonnes per day.

Each casthouse has a number of casting tables available, and there are 49 casting tables in total over all casthouses. Each casting table is capable of producing one of the 14 geometric dimensions. It is assumed that the products a casting table can produce is entirely based on the products' dimension, and that all casting tables can produce every alloy. The casting tables are also subject to a daily production capacity. This capacity ranges between 50 to 450 tonnes per day and is based on the rate at which casting can be done over the course of a day. From the data it becomes clear that casthouse capacity is usually more restricting than casting table capacity.

One production batch at Norsk Hydro is 50 tonnes. In addition, one order batch is 25 tonnes. The initial storage level is set to zero for all products at all casting tables.

The planning horizon is set to 35 days, meaning that the model will generate a production plan for 35 days (5 weeks). The model always considers a week to be from Monday to Sunday. It is however possible to create a production plan from e.g. Tuesday in week 1 to Tuesday in week 6. In this case the number of weeks in consideration is 6 weeks. The weeks that forecasted demand belong to may for this reason be any amount of days between one and seven for the weeks at the beginning and at the end of the planning horizon. This depends on the first day in the planning horizon.

7.1.2 Cost Structure

The mathematical model presented in Chapter 5 uses five types of cost parameters. Of these, values on transport costs are the only ones available in the source data set. The remaining costs are for this reason determined using reasonable assumptions, putting emphasis on the appropriate relationship between different costs in order to obtain desirable model output.

The transportation costs given in the source data set are between each casthouse and customer. This is extended to become a transport cost between each casting table and customer based on the casthouse associated to each casting table. The values range between 10-100 EUR per tonne. Some casthouses are also unable to offer transport to certain customers. In this case a route between the customer and given casting tables is not created in the implementation of the model.

In practice, Norsk Hydro use transport costs as a tie breaker when deciding upon the manual allocation of a call-off to a plant. Decisions are not influenced by production costs, and production costs should therefore be set uniform and are chosen to be lower than transportation costs.

$$C^P_{mp} \leq min\left\{C^T_{cm}\right\}$$

As mentioned in Chapter 2, Norsk Hydro operate as an MTO firm. A sizable holding cost should be set to ensure that the model does not suggests production for storage unless it is to satisfy call-offs in the reasonably near future. If capacity is fully utilized on delivery day, it should be possible to store production related to call-off demand that is produced before the delivery day. This is consistent with Norsk Hydro's objective to accept as many call-offs as possible. Meanwhile, the assumption is made that it is preferable to allocate a call-off to a casthouse with spare capacity rather than create storage. Reine et al. (2017) suggested that to achieve these goals for a model that was similar to the one of this thesis, the storage must be higher than the highest sum of production and transportation costs.

$$\max\left\{C^T_{cm} + C^P_{mp}\right\} < C^H$$

If call-off demand exceeds a capacity, either on casthouse level or casting table level, the model should first consider producing call-offs in advance to satisfy demand rather than decline the order altogether. This entails that the decline cost should be higher than holding cost multiplied with the maximum number of days in advance of delivery a call-off can be produced (n).

$$n \cdot C^H < C^D$$

It is reasonable to assume that call-off demand, which is certain demand, is more beneficial to allocate in the production plan than forecasted demand. This is because call-off demand is regarded as certain income, while forecasts cannot be regarded similarly in a practical setting. It is desirable for the model to first discard forecasted demand before declining call-offs when capacity is being fully utilized. This implies that the cost of declining call-offs should be higher than the cost of discarding forecasted demand.

$$C^F < C^D$$

In the model objective function, transportation costs and holding costs only incur for production related to call-offs, x_{mpt}, and not for production related to forecasts, y_{mpt}. If the model results are to prioritize call-off related demand, these costs have to be compensated for in the relationship between C^D and C^F. The difference between C^D and C^F has to be greater than the potential transportation and holding costs suffered from rejecting forecast related demand for call-off related demand. Otherwise, the model will find it "cheaper" to reject call-off related demand as the additional costs associated with x_{mpt} are avoided. The resulting relationship between C^F, C^T_{cm}, C^H and C^D is formalized in the following equation.

$$C^F + \max\{C^T_{cm}\} + n \cdot C^H < C^D$$

By regarding the reasoning above, a relationship between all cost parameters is derived:

$$C^P_{mp} \leq C^T_{cm} < C^H < C^F < C^D$$

Based on the relationship above, the cost parameter values are chosen and presented in Table 7.1.

Table 7.1: Values of the cost parameters per tonnes.

Cost		Value
Production cost	C^P_{mp}	10 EUR/tonne
Transportation cost	C^T_{cm}	10-100 EUR/tonne
Holding cost	C^H	200 EUR/tonne
Discarded forecast cost	C^F	400 EUR/tonne
Decline call-off cost	C^D	1000 EUR/tonne

7.2 Demand

The source data set contains historical call-offs for a year. Here, a call-off has the attributes customer, product, delivery day and demand quantity. The time resolution of delivery day is week. In this thesis the call-off demand (D_{cpt}) should have day as time resolution. In addition, there is no data on call-off arrival day, an important attribute when the rolling horizon approach is used on the model. The need to simulate demand data has thus arisen as the data available does not match the format the model in this thesis is dependent on.

In this section it is necessary to separate three terms from each other. These are order demand list (O), call-off demand (D_{cpt}) and forecasted demand (P_{cpw}). Order demand is the full set of all demand simulated. To create order demand it is necessary to determine five attributes; customer(c), product(p), arrival day ($\underline{t}$), delivery day ($\bar{t}$) and demand quantity(q).

Which customer-product combinations to consider when simulating demand is outlined in Section 7.2.1. Section 7.2.2 explains how to determine both arrival and delivery days. Further, Section 7.2.3 considers which quantities to draw in the simulation of demand. The process of simulating demand is presented in Section 7.2.4, while Section 7.2.5 explains how order demand is distinguished as either call-off demand or forecasted demand in a given roll in the rolling horizon approach.

7.2.1 Sampling of Customer-Product Combinations

From Section 7.1.1, the number of distinct customers and products are 153 and 508 respectively. However, data on historical call-offs shows patterns of each customer ordering only a selection of the available products. Furthermore, some customer-product combinations are hardly ever observed and are deemed to have too small sample sizes for simulations. The process of sampling relevant customer-product combinations is summarized in Algorithm 2. Here, the historical call-off data, D^H, is taken as input as well as the threshold sample size, S^{MIN}. The latter is set to twelve occurrences over the course of a year; in other words if a customer requests a certain product rarer than once a month. Thus, combinations of customer and product occurring less than twelve times are discarded for later simulations of orders. After these have been discarded, 542 different customer-product combinations are used to simulate demand.

Algorithm 2 Finding customer-product combinations.

Input
Historical call-off data: D^H
Min sample size: S^{MIN}
Output
Customer-product combinations: S
Count all (cp)-combination in D^H
if Number of (cp)-combination $\geq S^{\text{MIN}}$ **then**
    S += (cp)
end if

7.2.2 Determining Arrival and Delivery Days

When customer-product combinations have been found, the simulation of order arrival days and delivery days can be done for each combination. When determining arrival days and delivery days for orders, basic queuing theory assumptions are made.

1. The frequency between delivery days requested by a customer for a product is exponentially distributed.
2. The days before a delivery day that the call-off arrived at Norsk Hydro is also exponentially distributed.
3. Delivery days are independent of each other and arrival days are independent of each other.

Arrival and delivery days have different exponential distributions with different rate parameters which need to be estimated.

The rate parameter for delivery days are estimated for each customer-product combination as the source data set shows significant differences here. For a given customer-product combination, the rate parameter of its exponential distribution, λ_{cp}^{D}, is determined using the estimator in Equation (7.2.1)

$$\lambda_{cp}^{D} = \frac{1}{\bar{x}_{cp}}, \tag{7.2.1}$$

where $\bar{x}_{cp}$ is the average number of days between delivery days for a specific customer-product combination. Since the source data set separates distance betwen delivery time by weeks, and the model in this thesis uses a time resolution of days, the number of days between delivery times in the source data set are approximated as seven times the number of weeks separating them.

Order arrival days must also be determined and are especially important to know when using the rolling horizon approach. Data regarding arrival days is not available in the source data set, but Norsk Hydro have expressed that call-offs on average are declared 14 days before delivery, and no later than three days before. Consequently, the number of days prior to delivery is assumed to be exponentially distributed with an average arrival of 14 days before delivery, giving λ^{A} a value of $\frac{1}{14}$ as expressed in Equation 7.2.2. Notice that the arrival rate is the same for all customer-product combinations.

$$\lambda^{A} = \frac{1}{14} \tag{7.2.2}$$

7.2.3 Determining Quantities

The order quantity requested in the simulation of demand is randomly drawn based on the discrete probability distribution of order quantities from all of the call-offs in the the source data set. These range between one order batch, i.e. 25 tonnes, to twelve order batches, i.e. 300 tonnes. The number of instances where a certain quantity occurs in the source data set is counted and divided by the total amount of instances for all quantities. This provides the probability of this quantity being drawn from a discrete probability distribution with all the possible quantities. This probability distribution is shown in Table 7.2.

Table 7.2: Probability distribution of different order quantities.

Order quantity	Probability
25	74.09 %
50	16.24 %
75	5.88 %
100	2.37 %
125	0.84 %
150	0.22 %
≥175	0.32 %

7.2.4 Simulating Order Demand

The process of simulating orders is outlined in Algorithm 3. The simulation of orders is done over the scope of 15 weeks, or 105 days. 105 days is chosen as it equals the length of three planning horizons. This provides one whole planning horizon to subject the simulated data to an initialization phase, along with two planning horizons needed to apply a rolling horizon approach. The mentioned initialization phase is described in more detail in Section 8.1.

The simulation process takes these inputs: the simulation horizon (H), customer-product combinations (S), the rate parameters for arrival and delivery days (λ^A and λ^D_{cp}), the minimum days before delivery allowed to place an order (T^{MIN}) and the probability distribution for order quantities (Q).

Algorithm 3 Simulating order demand O.

Input
Set of customer-product combination: S
Simulation horizon: H
Delivery day rate: λ^D_{cp}
Arrival day rate: λ^A
Min days before delivery allowed to place orders: T^{MIN}
Probability distribution for order quantities: Q
Output
Order demand list: O
t = 0
for c,p $\in S$ **do**
  while t $< H$ **do**
    Draw t' from exp(λ^D_{cp})
    $\bar{t} = t + t'$
    Draw t^* from exp(λ^A)
    while $t^* < T^{MIN}$ **do**
      Redraw t^* with exp(λ^A)
    end while
    $\underline{t} = \bar{t} - t^*$
    Draw quantity q from Q
    Add order with attributes c, p, $\underline{t}$, $\bar{t}$ and q to O
    t $= t + \bar{t}$
  end while
end for
for O **do**
  if $\bar{t} - \underline{t} < 1$ **then**
    Remove order from O
  end if
end for

Orders are simulated for each customer-product combination over the simulation horizon. First the days from the current delivery day to the next is drawn and from this the delivery day is calculated. Further, a draw detailing how many days before delivery the order arrives is performed. If this value is less than or equal to T^{MIN}, a redraw is made. The arrival day is calculated by subtracting the draw from the delivery day previously calculated. In addition, a quantity is drawn from the discrete distribution explained in Section 7.2.3. The order now has all its attributes and is added to the order list. Orders are generated for the same customer-product combination until the sum of days between deliveries exceeds the horizon of the simulations.

As arrival day is drawn in the same way as number of days before delivery, it is possible to simulate orders with negative arrival day, which is outside the simulation horizon. In this case, the order should be discarded as it is desirable to start with an empty system.

The number of orders created by the simulation over the scope of 105 days varies for each trial. From ten different simulations, the average number of orders generated was 7 825, with a standard deviation of 61. When considering the total quantity in all orders generated, the average was 275 305 tonnes with a standard deviation of 2 311 tonnes.

7.2.5 Distinguishing Between Call-offs and Forecasts

Algorithm 3 simulates the complete set of orders spanning over 105 days. The planning horizon is however 35 days. Hence, only orders with delivery day during the planning horizon in question should be considered. These orders are either classified as call-off demand or forecasted demand. It is further assumed that every call-off that is to occur is accurately forecasted.

The process of distinguishing between call-off demand and forecasted demand is explained in Algorithm 4. As input, the first and last day of the planning horizon in consideration is given as T^S and T^E respectively, as well as the simulated order demand list, O. All orders are first checked to see if they have delivery day in the given planning horizon. If an order has delivery day in the planning horizon, it is classified as a call-off if arrival day is before the first day in the planning horizon. Otherwise it is classified as a forecast. Since forecasted demand has week as time resolution, modulo calculations is used in order to get the correct week.

Algorithm 4 Simulating call-offs and forecasted demand for the planning horizon of a given roll in the rolling horizon approach.

Input
Order demand list: O
First day in planning horizon: T^S
Last day in planning horizon: T^D
Output
Call-off demand: D_{cpt}
Forecasted demand: P_{cpw}
for O **do**
  if $T^S \leq \bar{t} \leq T^E$ **then**
    if $\underline{t} < T^S$ **then**
      $t = \bar{t}$
      Add call-off to D_{cpt}
    else
      $w = \lceil \bar{t} \mod 7 \rceil$
      Add forecast to P_{cpw}
    end if
  end if
end for

7.3 Overview of Data

Ranges of values discussed in this chapter are summarized in Table 7.3.

Table 7.3: Overview of input parameters and sets, and the range of their values.

Input Parameter or Set		Value Range
Customers	$\mathcal{C}$	153 customers
Casthouses	$\mathcal{K}$	9 casthouses
Casting tables	$\mathcal{M}$	49 casting tables
Casting tables related to casthouse	$\mathcal{M}_{\mathcal{K}}$	1-7 casting tables
Products	$\mathcal{P}$	508 types
Days	$\mathcal{T}$	35 days
Weeks	$\mathcal{W}$	5-6 weeks
Remaining days in week	$\mathcal{T}_{\mathcal{W}}$	1-7 days
Production batch	B	50 tonnes
Order batch	-	25 tonnes
Transportation costs	C^T_{cm}	0 - 100 EUR/tonne
Production costs	C^P_{mp}	10 EUR/tonne
Holding cost	C^H	200 EUR/tonne
Decline cost	C^D	1000 EUR/tonne
Discard cost	C^F	400 EUR/tonne
Call-off demand	D_{cpt}	0-300 tonnes/call-off
Production capacities casthouse	K_k	100-650 tonnes/day
Production capacities casting table	K_m	50-450 tonnes/day
Forecasted demand	P_{cpw}	0-500 tonnes/forecast
Initial storage	s_{mp0}	0 tonnes
Customer-product combinations	S	542 combinations
Total order demand simulated	-	Avg of 275 305 tonnes

Chapter 8

Computational Study

The mathematical model along with various solution approaches has been written in Mosel and implemented in FICO Xpress. Matlab R2015a was used to call the mathematical model in Xpress when run in the case of a rolling horizon approach or Lagrangian relaxation. All programs are run on a computer with 64-bit Windows 7 Enterprise, Intel(R)CoreTMi7-4790S, 3.20 GHz CPU and 16 GB RAM. *Depth first* has been used as branching strategy in all the computations.

This computational study primarily focuses on model output from the rolling horizon approach and Lagrangian relaxations described in Chapter 6. Section 8.1 discusses the use of an initialization period used to obtain adequate initial states for model runs. It also discusses the selection of the data set used in tests throughout this chapter. Section 8.2 presents computational efficiency and model output from the rolling horizon approach. Concepts such as capacity utilization and nervousness are discussed in relation to this. The impact of the Lagrangian relaxations is discussed in Section 8.3. The effects on computational efficiency is the main focus in the section, but impact on model output is also considered.

Some terminology will be defined that is used in the analyses of this chapter. For any problem instance, the best integer solution objective value found is usually referred to as the *objective value*, although sometimes UBD is used instead. Likewise, the best lower bound found for a problem instance is referred to as LBD.

An alternative to the regular objective value must also be defined. The model presented in Chapter 5 includes soft constraints enforced by costs in the objective function. These costs, such as C^D and C^F, have no practical implications, but are included to encourage the model to not discard demand. Consequently, it is of interest to distinguish between the objective value when these artificial costs are considered compared to when they are not considered. The latter is a more accurate reflection of the costs that would occur in a practical setting, and this value is defined as the *adjusted objective value*.

Note that one expects a trade-off between the amount of demand tonnage accepted and the adjusted objective value. Accepting more demand may necessarily lead to a higher adjusted objective value as more production and transport costs accrue. The primary goal should be to accept as much demand as possible, making accepted demand tonnage an important indicator of a good problem solution. Viewing objective value and adjusted objective value in isolation is thus not recommended.

8.1 Initialization Phase and Selection of Simulated Data Set

Given that the model is run using simulated demand data, one expects different results for every simulated data set used. Meanwhile, we wish a certain sense of consistency in the different analyses of this chapter. Consequently, one simulated data set will be chosen as the main subject of analysis in Section 8.2 and Section 8.3.

The selected demand simulation is chosen based on features observed from ten different demand simulations that have been subject to a 35 day initialization phase. This section wishes to illuminate certain features of the initialization phase as well as the conditions for selecting the data set used in the remainder of this chapter.

Section 7.2 explained that demand is simulated to have delivery days spanning a 105 day simulation horizon. The need for an initialization phase arises as demand is categorized as either call-off or forecasted demand based on the first day in a planning horizon and the arrival days of call-offs. No call-offs have arrived before the simulation horizon, leaving the system free of call-off demand at the first day in the simulation horizon. An initialization period thus becomes necessary to achieve a stable rate of call-off arrivals and available call-off demand to the models in this thesis.

The 105 days in the simulation horizon are shown in Figure 8.1 as day -34 to day 70. The initialization phase spans from day -34 up to and including day 0. The model runs that follow in Sections 8.2 and 8.3 start at day 1. The information from the end of the initialization phase is saved to provide initial states for these runs. Figure 8.1 also illustrates how plans for practical implementation are generated with a rolling horizon approach, by implementing the first day of every roll in practice. As seen, the chosen simulation horizon makes it possible to generate an implemented plan for 35 days after the end of the initialization phase, which equals the length of one planning horizon.

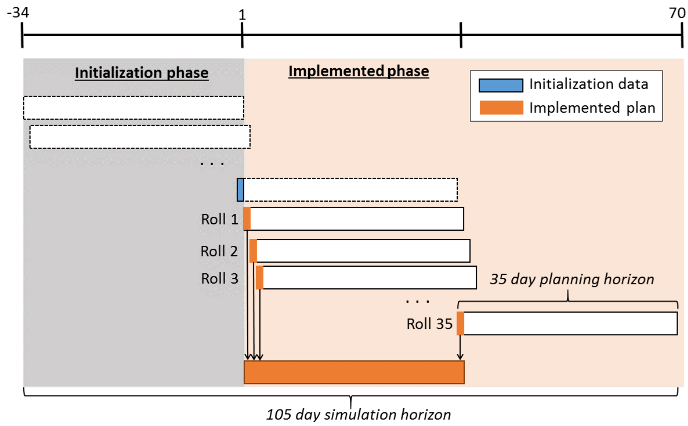


Figure 8.1: Details regarding the rolling horizon approach used in this thesis, including both the initialization phase and implemented phase.

The initialization phase is run using the rolling horizon approach described in Section 6.1. It spans over 35 rolls for each demand simulation. Production and accepted call-offs are locked along the way, with production being locked for the seven first days. This is similar to what is done in most of Section 8.2, which provides analyses on the rolling horizon approach in more detail. The last roll in the initialization phase provides initial states for tests beginning in day 1. This includes initial storage values, s_{mp0}, but also locked variables where applicable.

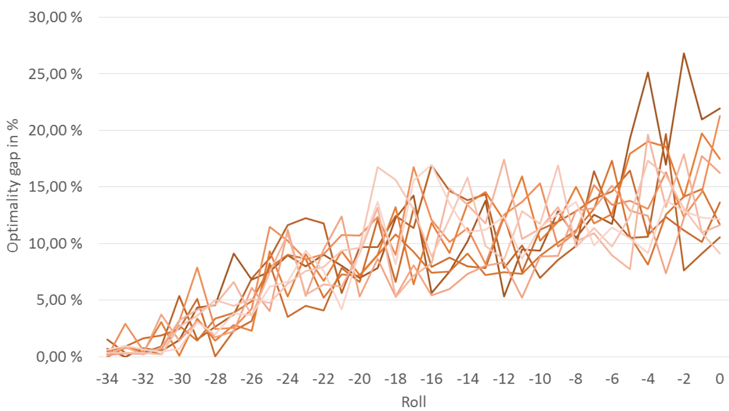


Figure 8.2: The optimality gap for each roll in the initialization phase for ten different simulated data sets.

Each roll has been given a 900 second runtime. The resulting optimality gaps from each roll for ten demand simulations is shown in Figure 8.2. It is observed that optimality gaps are close to zero in the first rolls for each simulated data set. This is a consequence of there being few call-offs to allocate, making the problem easy to solve. The optimality gaps then rise before reaching a plateau of values in the last rolls.

In order to select one demand simulation for further use in this chapter, the average optimality gap for the 35 rolls in the initialization phase has been calculated for the ten simulations. The simulations have then been sorted according to the average optimality gap for these rolls, and one of the data sets closest to the median value for average optimality gap has been chosen as the data set used for testing throughout this chapter. This data set is hereby referred to as the *primary demand simulation*. Another set close to the median value is sometimes used when results from the primary demand simulation are inconclusive. This data set is hereby referred to as the *secondary demand simulation*.

8.2 Rolling Horizon Planning

In this section, the model in Chapter 5 is applied using the rolling horizon approach. We begin by providing information about the parameter and set values used when implementing the rolling horizon approach described in Section 6.1. Subsection 8.2.1 goes on to analyze the features observed from a single roll in the rolling horizon approach. Subsection 8.2.2 provides an analysis of results when the rolling horizon approach is applied to 35 consecutive rolls. Finally, model output is discussed in relation to the concept of nervousness in Subsection 8.2.3.

In addition to the constraints in Chapter 5, Constraints (6.1.6) presented in Chapter 6 is added to the model. For each roll performed, lower bounds on the x_{mpt}-variables are added for those that have locked production from the previous roll. Production is locked for the seven first days in each roll. This entails giving the set $\mathcal{T}^L$ values from one to seven in for every roll, or $\mathcal{T}^L : \{1, ..., 7\}$.

The model is run 35 times, each time creating a production plan with a 35 day planning horizon. Only the first day in every roll is what would be implemented into practical operations from Norsk Hydro's point of view. The reason for choosing to roll 35 times is thus to get implemented production plans for the first 35 days, which equals one planning horizon if a rolling horizon approach wasn't used.

8.2.1 Single Roll Analysis

One roll in the rolling horizon approach is analyzed in this subsection. The purpose is to illustrate its features in isolation, before results from running multiple, consecutive rolls are assessed. The single roll has been run for 7 200 seconds, before output is analyzed in order to verify that the model functions as intended. We wish to examine how the optimality gap develops over time for a single roll. We also wish to examine solution values for the production variables x_{mpt} and y_{mpt}, and how they develop along the horizon of the single roll.

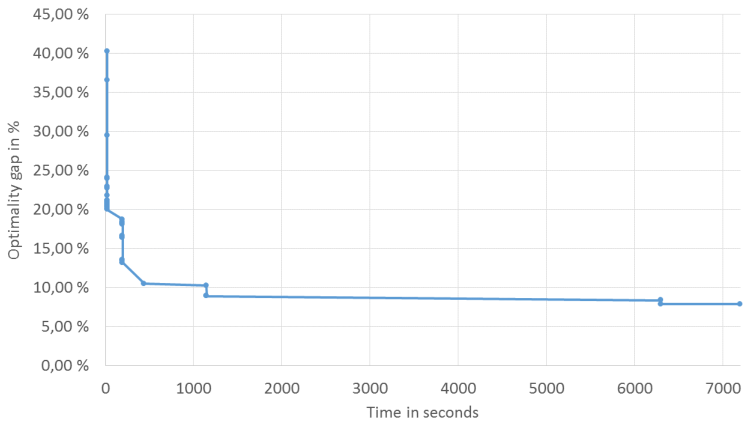


Figure 8.3: Development of the optimality gap with time for the single roll problem.

Figure 8.3 shows the development of optimality gap with time for the single roll. Each data point is registered when the optimality gap is updated. It is worth noting that the rate of improvement of the optimality gap stagnates somewhat from about 1 200 seconds onward. This indicates that a runtime close to this mark is adequate for each roll when multiple, consecutive rolls are applied.

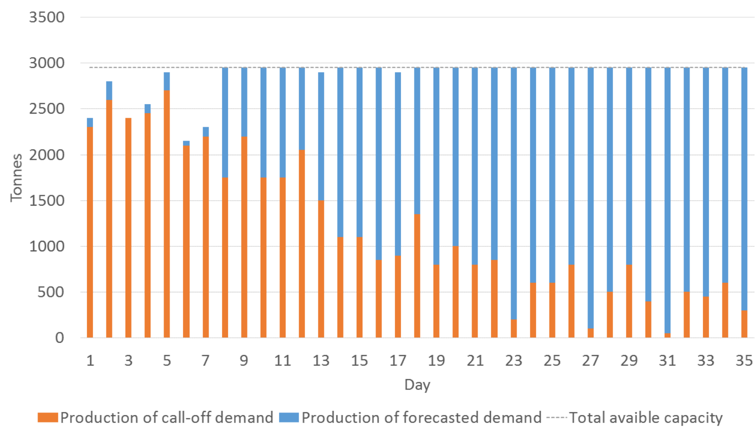


Figure 8.4: Daily sum of production variables along 35 day planning horizon.

The model acts as expected in terms of the two types of production variable. The sum of output values each day is shown in Figure 8.4 for both x_{mpt} and y_{mpt}. The first days of the horizon contains mostly production related to call-offs before the weight shifts towards production related to forecasts late in the horizon.

This tendency can be explained by the composition of the demand on the different days in the horizon, either as related to call-offs or forecasts. The closer the first day in the horizon is to a given day, the more likely it is that call-offs have arrived with that day specified as delivery day. As a result, a higher proportion of the available demand is call-off related early in the horizon compared to late in the horizon. Meanwhile, given that rejecting call-off demand is penalized more in the objective function than rejecting forecasted demand, one expects y_{mpt} to only be allocated if there is spare capacity after x_{mpt} variables have been allocated.

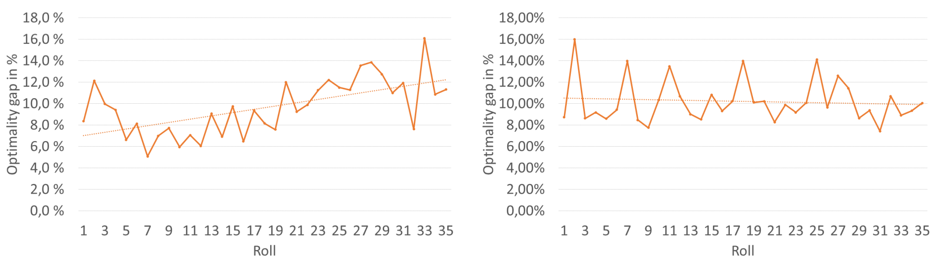
Also worth noting are the effects of forecasts being indexed by week. This explains the spike in y_{mpt} from day 7 to day 8. Because this particular roll started at the beginning of a Monday, it perceives that these two days belong to different weeks given that day 7 is a Sunday and day 8 is the following Monday. In the roll succeeding this one, beginning on a Tuesday, one would expect a spike between day 6 and 7 instead.

8.2.2 Multiple Roll Analysis

In this section, the aim is to observe how the model preforms using a rolling horizon approach with multiple, consecutive rolls. 35 rolls have been run and each roll has a runtime of 2 000 seconds, approximately 30 minutes. Among the output that is given consideration are the values and development of optimality gap and objective value. Solution output such as the degree of accepted and rejected demand is also considered, along with the degree of capacity utilization resulting from the production variables.

Optimality Gap and Objective Value

The optimality gap lies in the interval $[5.08\%, 16.10\%]$ along all 35 rolls, with an average optimality gap of 9.63% and an estimated standard deviation of 2.60%. The optimality gap for all the rolls, as well as a linear trend-line for the optimality gaps, is presented in Figure 8.5a.



(a) Primary demand simulation. **(b)** Secondary demand simulation.

Figure 8.5: The optimality gap for each roll using the a) primary demand simulation and b) secondary demand simulation.

From the figure it is clear that optimality gap for the primary demand simulation varies considerably from roll to roll. The linear trend-line shows that even with high variations from roll to roll, the optimality gap tends to increase with the amount of rolls performed. As an initialization phase has already been performed, one would expect the optimality gap increase to remain fairly stable from roll to roll. Upon inspecting the demand data, the increase may instead be attributed to the primary demand simulation having a large portion of call-offs arriving towards later rolls. To confirm that it actually is a high amount of call-offs causing an increasing optimality gap, new computations using the secondary demand simulation are performed. The optimality gaps from these computations are shown in Figure 8.5b. Clearly the optimality gap does not increase with the amount of rolls performed when using this demand simulation. It might therefore be the case that the optimality gap is quite dependent on the number of call-offs in the specific roll.

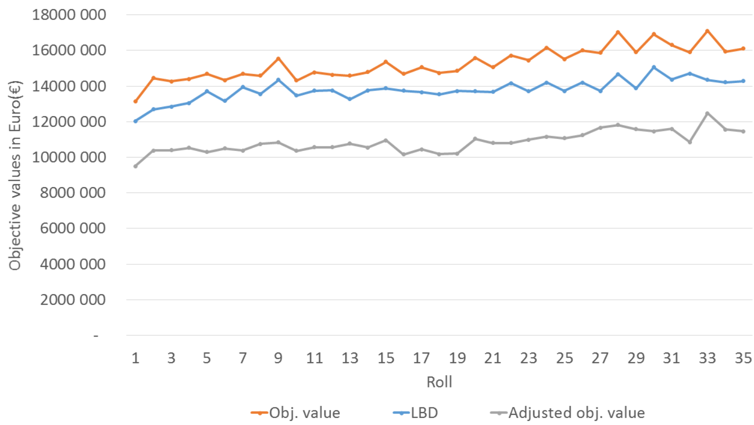


Figure 8.6: The LBD, objective value and adjusted objective value for each roll.

Figure 8.6 shows the best objective value and LBD found in each roll. In addition, the adjusted objective value is included to give a notion of the real costs. As seen from the figure, the LBD lies between 12 million and 15 million EUR, while the objective value lies between 13 million and 17 million EUR. The adjusted objective value is about 70% of the actual objective value, which means that the artificial costs make up the remaining part. Both the LBD and the objective value increases with the number of rolls performed. This observation is expected given that the optimality gap also increases, indicating that the solutions found may be less optimal.

Accepted and Rejected Demand

Figure 8.7 shows both available demand, along with accepted and rejected call-off and forecasted demand in tonnes for each roll. The average and standard deviations for the same data is given in Table 8.1.

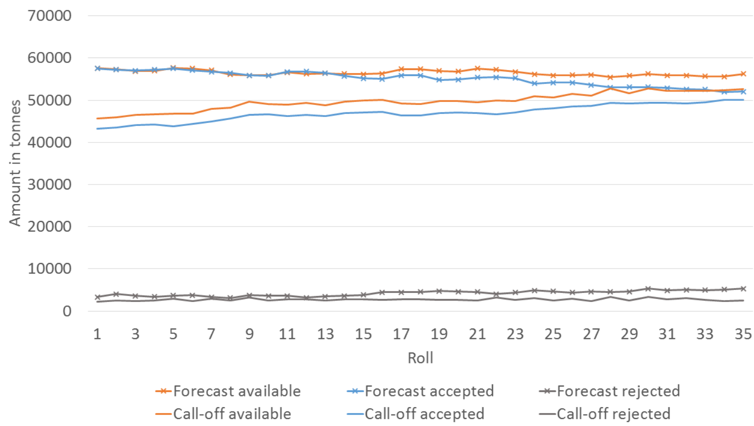


Figure 8.7: Incoming, accepted and rejected call-offs and forecasts for each roll.

Table 8.1: The average and standard deviations for available demand, accepted call-off and forecasted demand, and rejected call-off and forecasted demand for the rolls in question. Measured in tonnes.

Demand type	Average in tonnes	Standard Deviation in tonnes
Available call-offs	49 701	2 028
Available forecasts	56 497	660
Accepted call-offs	46 976	1 932
Accepted forecasts	55 116	1 698
Rejected call-offs	2 725	295
Rejected forecasts	4 224	643

From the data it is possible to see that there's always more forecasted demand than call-off demand, although this difference becomes small towards the final rolls. As pointed out in the optimality gap discussion above, the increase in incoming call-off demand may be a factor to the positive trend-line in optimality gaps. Further, it is clear that amount of accepted call-off demand increases when incoming call-off demand increases. Meanwhile, more forecasted demand is rejected when incoming call-off demand increases, even though incoming forecasted demand does not change much. This indicates that the model prioritizes call-off demand over forecasted demand, as desired.

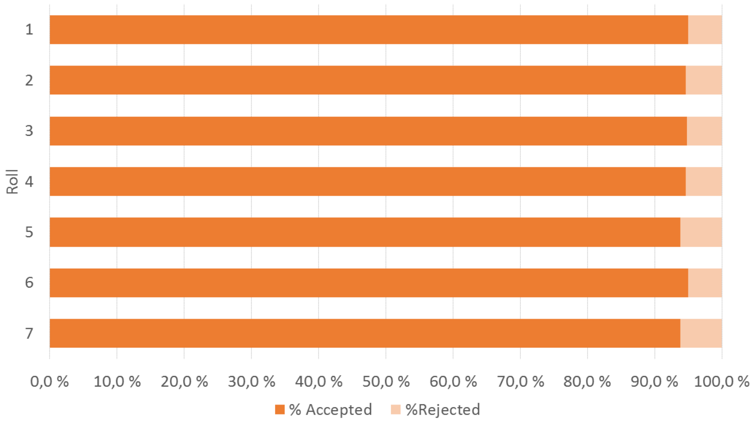


Figure 8.8: Number of accepted and rejected call-offs in percentage of the total number of incoming call-offs for the first seven rolls.

The ratio of the number of accepted and rejected call-offs for the first seven rolls is presented in Figure 8.8. The ratio is very similar for all the other rolls performed. As seen from the figure, about 95% of the incoming call-offs in a roll is accepted, whereas about 5% are rejected.

When considering only the call-offs that are rejected, there are two patterns that seem to appear:

1. Call-offs with quantities of half production batches
2. Call-offs with late arrival day

Table 8.2: Quantities measured in batches, how often they occur in call-offs, and fraction of rejected call-offs belonging to each quantity.

Number of order batches	Fraction of call-offs	Fraction of rejected call-offs
1	74.427 %	88.871 %
2	15.858 %	4.419 %
3	5.912%	4.516 %
4	2.337 %	0.935 %
5	0.825 %	0.839 %
6	0.244 %	0.000 %
≥ 7	0.397 %	0.419 %

Table 8.2 shows the fraction of call-offs with a given quantity in the available call-off demand. It also shows the fraction of rejected call-offs belonging to different quantities. The table makes it clear that call-offs having a quantity equal to one order batch are most frequently rejected. This is as anticipated as this is the most

common quantity requested. The rate at which they are rejected is however more frequent than the rate at which they occur in the set of call-offs entering the model. This indicates that the model is more likely to reject call-offs with quantity equal to one order batch compared to call-offs having quantity equaling a different multiple of order batches.

Recall that order batches and production batches are not equal in volume, one order batch is half of one production batch. One explanation for the trend seen in Table 8.2 may be that call-offs with quantity of half a production batch are in need of storage unless they can be matched with a call-off of the same quantity requesting the same product. Two call-offs requesting one order batch of the same product each can be combined to form one production batch. However, if a call-off cannot be matched with another call-off in this way, a high holding cost incurs until a match appears. This makes the call-off less profitable to accept.

This notion seems more viable when considering call-offs with quantities equal to two order batches, which is the same as one production batch. The percentage of rejected call-offs with this quantity is very low compared to the frequency of this quantity in the data set, which is a trend for any call-off having quantity that is an even number of order batches. It is not necessarily a disadvantage that the model more often rejects call-off quantities of half a production batch compared to whole production batches, as long as capacity is adequately utilized. In addition it is probably more common for the model to reject big call-off quantities, but this is cannot be confirmed, as the data set includes few call-off with these quantities.

Table 8.3: Fraction of rejected call-offs when regarding days prior to delivery they arrived.

Arrival before delivery day	Fraction of rejected call-offs
1 - 7 days	49.359%
8 - 14 days	24.514 %
15 - 21 days	14.731%
22 - 28 days	8.245%
29 days or more	3.151 %

Arrival day of call-offs also has an effect on their likelihood of being accepted. Table 8.3 shows how long in advance of delivery the rejected call-offs arrive. It is clear from the figure that almost 50 % of the rejected call-offs arrived less than one week before delivery. It is therefore safe to assume that call-offs arriving near delivery day are more likely to be rejected. The finding is foreseen, as capacity is locked for the first seven days and accepted call-offs that have arrived early also must be accepted in later rolls, despite more preferable call-offs arriving later.

Capacity utilization

For each roll, the first day in the generated production plan is implemented. It is therefore important that these production plans utilize capacity as much as possible. The casthouse capacity limit is restricted before the sum of casting table capacities at the given casthouse. Thus, it is casthouse capacity which is considered in this context. As all call-offs have arrived a minimum of three days before delivery, and forecasts are accurate, the capacity for the first day in the production plan should be filled by mostly call-off production, x_{mpt}.

Figure 8.9: Capacity utilization at each casthouse for the first day for rolls 1-35.

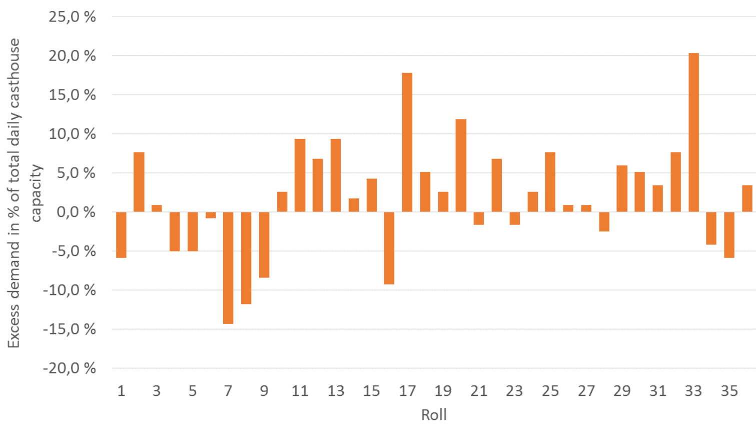


Figure 8.10: Call-off demand with delivery on the first day for each roll. Expressed as an excess percentage of total daily casthouse capacity.

Figure 8.9 shows capacity utilization, as a fraction of available capacity at each casthouse, for the first day in each roll. Here, we see that casthouse 2, 4, 7, 8 and 9 at at all times are fully utilized, while casthouse 1, 3, 5 and 6 are less utilized having an average capacity utilization of 92%, 56%, 57% and 82% respectively. Especially casthouse 3 and 5 are poorly utilized.

One obvious reason for low capacity utilization is low incoming demand. Figure 8.10 shows how much call-off demand that is to be delivered on the first day of rolls 1 to 35. It expresses this as an excess percentage of the total daily casthouse capacity, where the average excess is 1.82%.

Regarding the rolls with the highest and lowest values in Figure 8.10 reveals the extent available demand quantity influences capacity utilization. Roll 33 has call-off demand the first day equaling 20.3% above total capacity, and as seen from Figure 8.9, the roll also has very high capacity utilization. The high utilization of this roll even applies to the casthouses that have lower utilization in general. In contrast, roll 7 has a negative excess demand of −14.4% i.e. demand is lower than capacity. In addition, demand stays low for the next couple of rolls which means that it is less likely that production from the first day in this roll is stored for upcoming delivery days. Based on this, one would assume that roll 7 has one of the lowest capacity utilization the first day, but there are in fact eight rolls with lower utilization. It is also odd that, e.g. roll 25 has the lowest capacity utilization, while its call-off demand is in excess of capacity. Thus, it seems like insufficient capacity utilization is influenced by additional factors to low demand quantity.

To find other possible reasons for low capacity utilization at certain casthouses, the input data is inspected further. Here, two factors seem to contribute to the results:

1. Few routes to customers from the casthouses

2. Low demand of products which are possible to produce at the different casthouses

When considering how many customers each casthouse can deliver to, both casthouse 3 and 5 have a significantly lower number of routes to different customers compared to other casthouses. When inspecting the demand simulated, it is observed that four product dimensions rarely are in demand. At least one of the casting tables available at casthouse 1, 3, 5 and 6 are related to these product dimensions. Hence, it seems that the input data limits the possible degree of capacity utilization in model solutions in several ways. The effect may be enhanced by many customer-product combinations being discarded from simulations due to low sample sizes. This removes diversity from the products that are simulated, isolating high capacity utilization to casthouses that can produce the products that remain.

8.2.3 Nervousness

Until now, production related to call-off demand has been locked for the first seven days in each roll. In addition, all call-offs that are accepted must be accepted in future rolls. This has been done in order to give Norsk Hydro and their customers some degree of predictability, reducing nervousness. Sridharan and Berry (1990) recommends locking variables for at least 50 % of the planning horizon in order to have a positive effect on the nervousness phenomenon. In the following section, the effect of locking bigger fractions of the planning horizon are considered, as only 20 % of the planning horizon has been locked in previous analyses. As before, only the production variables related to call-off demand are locked. As before, runtimes are limited to 2 000 seconds per roll. Three different locking fractions are considered:

1. 20 % of planning horizon. $\mathcal{T}^7 = \mathcal{T}^L : \{1, ..., 7\}$
2. 50 % of planning horizon. $\mathcal{T}^{18} = \mathcal{T}^L : \{1, ..., 18\}$
3. 100 % of planning horizon. $\mathcal{T}^{35} = \mathcal{T}^L : \{1, ..., 35\}$

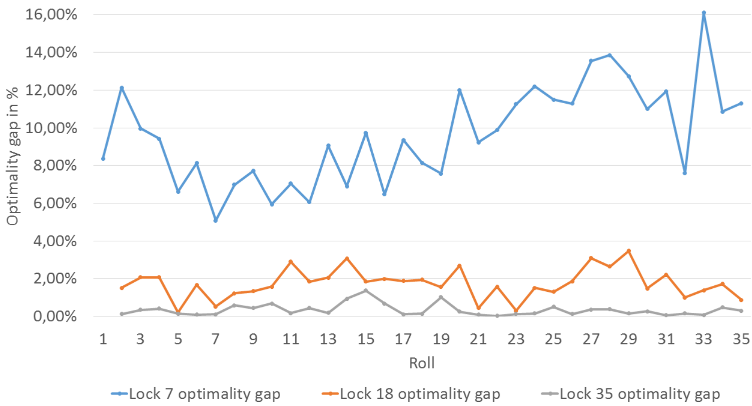


Figure 8.11: The optimality gap for the different lock settings.

Figure 8.11 shows the optimality gap from each roll for the different lock settings. It is obvious from the graph that when L increases, the optimality gap is reduced. This pattern is expected since more of the variables are predetermined when L increases, reducing the complexity of the problem. For $\mathcal{T}^{35}$ the optimality gaps are approximately zero. It is worth noting that even though solutions for $\mathcal{T}^{35}$ are close to or even reach optimality, they still suggest more costly production plans with less accepted demand than the solutions for $\mathcal{T}^7$ and $\mathcal{T}^{18}$.

Sridharan and Berry (1990) points out that a trade-off exists between costs and nervousness in rolling horizon planning. It would therefore be expected that costs are higher when L is high. This notion is confirmed when regarding Figure 8.12.

The lower bound for $\mathcal{T}^7$ is by far the lowest, indicating that this lock setting has the potential to find solutions with the lowest costs. However, the best objective value for $\mathcal{T}^7$ is similar to that for $\mathcal{T}^{18}$. This is probably due to the lower optimality gaps found with $\mathcal{T}^{18}$. As the optimality gap for $\mathcal{T}^{35}$ is approximately zero, its best lower bound and best objective value are very similar, but also much higher than the two other lock settings.

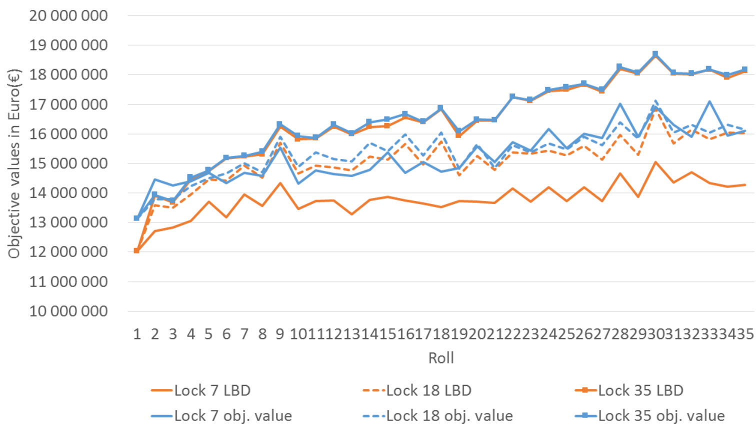
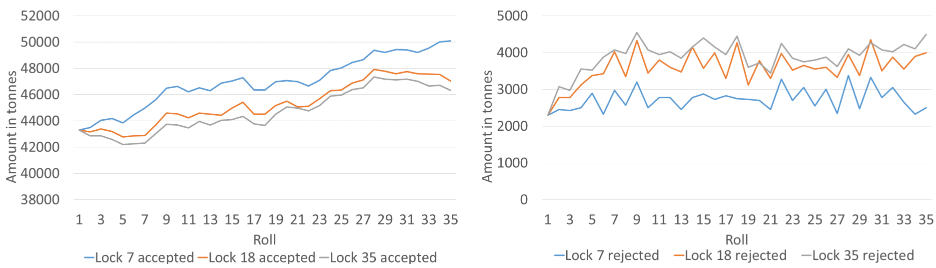


Figure 8.12: The optimality gap for the different lock settings.



(a) Accepted call-off demand. **(b)** Rejected call-off demand.

Figure 8.13: Amount in tonnes of a) accepted call-off demand and b) rejected call-off demand for the different locking-settings.

Figure 8.13 shows how much call-off demand in tonnes that was accepted and rejected in every roll of each production lock setting. As seen from the figure, less call-off demand is accepted and more call-off demand is rejected when $\mathcal{T}^L$ is high. Thus, there exists a trade-off where Norsk Hydro can choose to create production plans with high predictability or production plans that accept more call-off demand. An interesting feature in Figure 8.13 is that the change from $\mathcal{T}^7$ to $\mathcal{T}^{18}$ is larger than that from $\mathcal{T}^{18}$ to $\mathcal{T}^{35}$. In fact, some of the rolls in $\mathcal{T}^{18}$ and $\mathcal{T}^{35}$ have nearly the same values on accepted and rejected call-offs, e.g. at rolls 20 and 21.

To summarize, $\mathcal{T}^7$ accepts more tonnes of call-off demand than the two other lock settings, while having the highest optimality gap. This indicates that if the optimality gap for $\mathcal{T}^7$ is reduced, accepted call-off demand could possibly be even higher. To obtain the full benefits of locking lower fractions of the production in each planning horizon, the trends discussed in this section motivate that methods providing higher computational efficiency should be used.

8.3 Lagrangian Relaxation

The analyses in Section 8.2 made it clear that optimality is difficult to achieve when locking production for few days in the planning horizon of each roll. Given that Norsk Hydro operate in a commodity market, one expects small profit margins for every order. This provides an increased incentive to reduce the optimality gap for each production plan, providing solutions which are cheaper to implement in practice. This section shows the effects of Lagrangian relaxations on the optimality gap.

Lagrangian relaxation has been performed for a single roll on the two models described in Section 6.2.1 and 6.2.2. The implementation of these relaxations exclude some of the features in the rolling horizon approach as neither the production nor the demand variables are locked in any way from the previous roll. Introducing locking of variables, as in the rolling horizon approach, brings more constraints into the problem as shown in Section 6.1. Thus, the simplification of no locked variables is done for testing the potential of Lagrangian relaxation on the underlying problem given in Chapter 5.

The Lagrangian relaxation cannot be compared with the single roll problem from Section 8.2.1 since this problem includes locking of variables. Therefore, an analysis of a single roll problem with no variables locked is performed in Section 8.3.1. This problem is referred to as the *non-relaxed problem*. In Section 8.3.2 preliminary testing for the Lagrangian relaxations is presented. Furthermore, in Section 8.3.3 the computational efficiency of the Lagrangian relaxation is evaluated, and in Section 8.3.4 model output is discussed.

8.3.1 Non-Relaxed Problem

The non-relaxed problem is run for 7 200 seconds. The development of the optimality gap with time is shown in Figure 8.14, where data points have been registered every time the model was able to improve optimality gap. Certain values related to the best solution are also shown in Table 8.4. Among the most obvious and notable features is that the optimality gap is fairly poor after an extensive runtime. Furthermore, the rate of improvement of the optimality gap stagnates significantly from about 3 000 seconds onward.

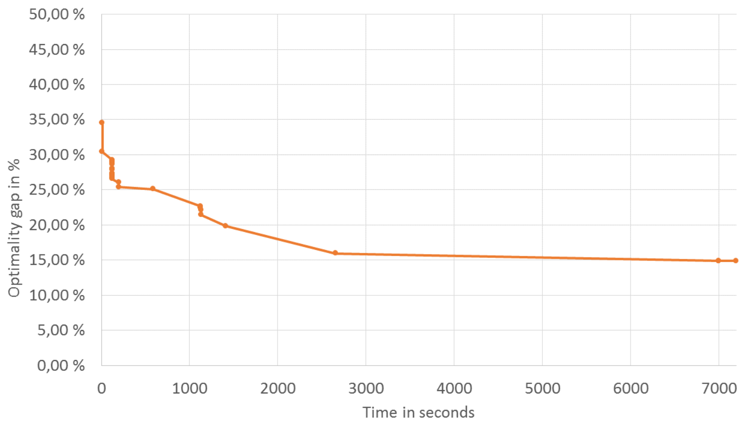


Figure 8.14: Development of optimality gap with time for non-relaxed problem.

Table 8.4: Solution values of non-relaxed problem.

Solution Parameter	Value
Best objective value	12 857 014 EUR
$\text{LBD}_{\text{non-relaxed}}$	10 942 481 EUR
Optimality Gap	14.93%

In terms of model output, the best integer solution to the problem accepts 1 208 out of the 1 343 available call-offs with delivery in the 35 day planning horizon. This results in an objective value of 12 857 014.30 EUR.

8.3.2 Preliminary Testing

This section presents the results of preliminary testing for finding appropriate values for the parameters in the subgradient method. By performing such an analysis, a starting point closer to the optimal solution can be obtained, and the subgradient method will require fewer iterations before convergence. Thus, the consequence should be improved computational efficiency. The parameters to be set based on testing, are the initial values for the multipliers used in different relaxations, π_{cpt}, π^1_{mt}, π^2_{kt}, and also the constant λ.

The values of the Lagrangian multipliers can initially be set to 0, but when examining the Lagrangian function in the two relaxations it may be possible to derive an initial value closer to the optimal one. In the casting table relaxation, the third term in Equation 6.2.11 resembles the sixth term. They can be combined and written as:

$$\sum_{c\in\mathcal{C}}\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}}(C^B D_{cpt} - \pi_{cpt})\left(1 - \sum_{m\in\mathcal{M}} \delta_{cmpt}\right)$$

This relation can give some information on the optimal value of the multipliers, and help choose appropriate initial values. If $C^B D_{cpt} \leq \pi_{cpt}$ it would not be optimal to produce anything as declining a call-off only would incur negative costs. As a result the multiplier should be smaller than $C^B D_{cpt}$.

When $C^B D_{cpt} > \pi_{cpt}$, the term will be negative for $\sum_{m\in\mathcal{M}} \delta_{cmpt} > 1$, and would act as an incentive to accept a call-off on several casting tables. The consequence would be producing duplicates of some highly profitable call-offs, while less profitable call-offs are rejected. This will happen if the Lagrangian multiplier of the more expensive call-offs outweighs the cost of rejecting a call-off. The optimal values of the multipliers are those that can ensure that call-offs should be accepted at maximum one casting table. Initial values in the interval $[0, C^P]$ have been tested, and the most promising initial values are given in Table 8.5 as $\pi_{cpt}^{(0)}$.

For the capacity relaxation this reasoning is not applicable since the last term in Equation 6.2.9 sums over casthouses and thus does not resemble the second term in the equation. Nevertheless, initial values in the interval $[0, C^P]$ have been tested and are set to the values shown in Table 8.5 as $\pi_{mt}^{1(0)}$ and $\pi_{kt}^{2(0)}$.

The value of the constant λ should be in the interval $(0, 2]$ to ensure convergence. However, the testing has shown that after the first iteration, the solution of the subproblem is far from the LBD to the non-relaxed problem, $\text{LBD}_{\text{non-relaxed}}$. Thus, the initial value of λ is set very high in order approach $\text{LBD}_{\text{non-relaxed}}$ more quickly. When approaching the optimal solution, the convergence of the step length is still ensured. After a number of iterations, p, if no better solutions are found, the value of λ is halved. When this happens sufficiently many times, it eventually reaches the interval $(0, 2]$.

The maximum number of iterations, k_{max}, is set in order to deal with a reasonable computing time, and p is set to an appropriate value based on k_{max}. An overview of all the parameter values are found in Table 8.5.

Table 8.5: Values of parameters of the implemented subgradient method.

Parameter	Value
$\pi_{cpt}^{(0)}$	$0.7 \cdot C^B \cdot D_{cpt}$
$\pi_{mt}^{1(0)}$, $\pi_{kt}^{2(0)}$	$0.2 \cdot C^P$
λ	2000
k_{max}	300
p	10

8.3.3 Computational Efficiency

The Lagrangian relaxation is composed of distinct computational elements where the runtime of each element is independent of the others. The sum of these runtimes will determine the computational performance of the Lagrangian relaxation as a whole. These elements are:

- Solving the Lagrangian subproblem.
- Heuristic procedure for finding UBDs from infeasible solutions of the subproblem.
- Time until convergence of the solution of the subproblem, which is determined by the step length when updating the multipliers.

In this subsection, the computational efficiency of the elements is examined.

Solving the Lagrangian Subproblem

Solving the subproblem to optimality or near optimality is crucial for the rest of the procedure, as this solution provides a new subgradient for calculating the step length for the next iteration. It is also important for providing the heuristic with an initial solution that is as close to optimality as possible in order to find a tight UBD. This element is thus considered essential for the rest of the iterative procedure.

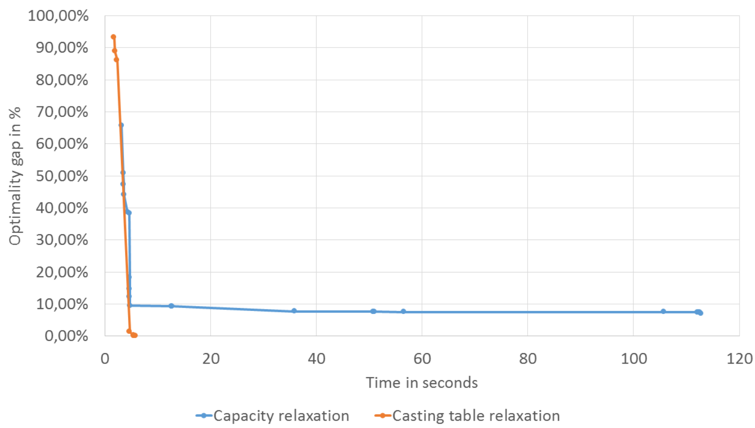


Figure 8.15: Optimality gap for solving the Lagrangian subproblems.

In Figure 8.15 the optimality gap when solving the Lagrangian subproblem for the two relaxations is shown. The casting table relaxation is solved to optimality in 6 seconds, while the capacity relaxation does not reach optimality within a maximum runtime of 120s. The reason for the short runtime of the casting table relaxation can be attributed to the number of constraints relaxed as discussed in Section 6.2.2.

After 120s the solution of the capacity relaxation has reached an optimality gap of 6.89%. Thus, the time for solving the subproblem for the capacity relaxation is at least 20 times larger than the time for solving the subproblem of the casting table relaxation. Because of this, the capacity relaxation is regarded as less efficient in runtime. Given that the goal of the Lagrangian relaxation is to decrease the optimality gap and runtime, the capacity relaxation is not considered in further analyses. The rest of this section studies the computational efficiency of the casting table relaxation.

Heuristic Procedure

The heuristic, as explained in Section 6.2.4, finds feasible solutions for the δ_{cmpt}-variables. From this, the δ_{cmpt}-variables which are equal to one after the heuristic, are locked before the textitresolve problem is run to find a feasible solution for the rest of the variables. Thus, the computational efficiency of finding a feasible solution includes both the time for the heuristic to find which δ_{cmpt}-variables to lock, and the runtime for the resolve problem.

The resolve problem does not solve to optimality. Since the goal of the Lagrangian relaxation is to reduce the computational time and since several iterations are needed to obtain convergence, an appropriate max time for solving the resolve problem has to be set. Upon running the resolve problem, no better solutions are observed after 300 seconds, and this has been set as the max time. In addition, the heuristic procedure, which includes the heuristic and the resolve problem, is only run the first iteration and then every 20th iteration in order to decrease the total runtime of the method. As a result, in 300 iterations the heuristic procedure is run 16 times. The average runtime for the heuristic for these 16 times is 6 seconds and the total runtime for the resolve problem is 430 seconds. This total runtime includes the max time of 300 seconds for solving the problem and 130 seconds for data processing.

Initially, the UBD has been given a high value in accordance with the solution strategy given in Subsection 3.3.2. After the first iteration, the heuristic procedure is run in order to update the UBD. This UBD comes from the solution of the resolve problem and provides a UBD for the entire Lagrangian relaxation. In the definition of the Lagrangian relaxation from Section 3.3.1, the solution of the Lagrangian subproblem provides an LBD, $\text{LBD}_{\text{Lagrange}}$. Meanwhile, the heuristic procedure includes running the non-relaxed problem with some variables locked when the resolve problem is run. The result is a feasible solution to the non-relaxed problem, but also an LBD, $\text{LBD}_{\text{resolve}}$, possibly different from $\text{LBD}_{\text{Lagrange}}$. By definition, $\text{LBD}_{\text{resolve}}$ is not used to update $\text{LBD}_{\text{Lagrange}}$.

It is still of interest to examine the bounds from the resolve problem more thoroughly as this run may have improved computational efficiency given that some variables are locked. In Figure 8.16 the optimality gap as a function of time for the resolve problem in the first iteration is shown. After 12 seconds the optimality gap is 9.55% and after 143 seconds the optimality gap is 7.15%. This means that after only one iteration the resolve problem itself is able to decrease the optimality gap of the non-relaxed problem from 14.93% after 2 hours to 7.15% after 564 seconds (9 minutes and 24 seconds). These 564 seconds is the total runtime of the first iteration of the Lagrangian relaxation including solving and data processing.

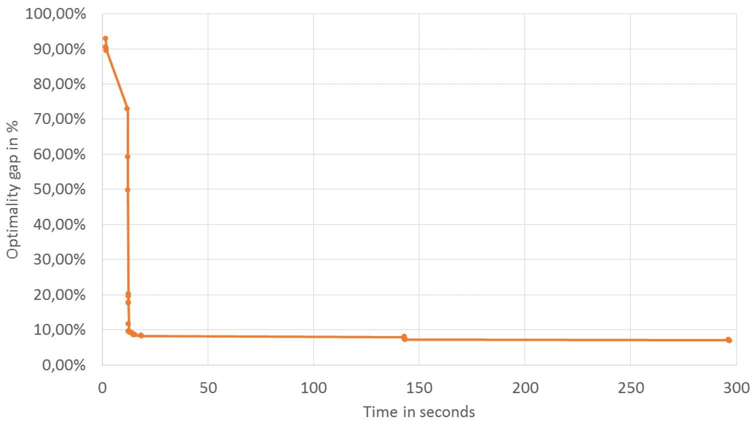


Figure 8.16: Optimality gap of the non-relaxed problem.

An even better optimality gap can be determined by examining the bounds of the non-relaxed problem and the resolve problem together. This is because the non-relaxed problem is similar to the resolve problem, only distinguished by the resolve problem having some δ_{cmpt}-variables locked.

The values for the objective function of the resolve problem after the first iteration is given in Table 8.6. Compared to the values in Table 8.4, $\text{LBD}_{\text{resolve}}$is slightly lower than $\text{LBD}_{\text{non-relaxed}}$, which means that the $\text{LBD}_{\text{non-relaxed}}$ is the highest of the two. However, the best objective value of the resolve problem is significantly lower than for the non-relaxed problem, and thus this solution provides a lower UBD. By combining the lowest UBD, i.e. from the resolve problem, with the highest LBD, i.e. $\text{LBD}_{\text{non-relaxed}}$, an optimality gap of 4.57% is obtained. This results in a 69% reduction in optimality gap after only one iteration of the Lagrangian relaxation. The improvement in best objective value is 12 857 014 EUR - 11 465 907 EUR = 1 391 107 EUR.

Table 8.6: Solution values of the resolve problem.

Solution parameter	Value
Best objective value	11 465 907 EUR
$\text{LBD}_{\text{resolve}}$	10 646 750 EUR
Optimality gap	6.84%

It is important to note that the heuristic procedure is not able to find better solutions in any consecutive iterations. Even though the solution of the Lagrangian subproblem improves, no new UBD is found. This indicates that the heuristic may be too simple and should have included more features. For example the heuristic only considers call-offs that have been accepted at least once in the solution of the subproblem. A more subtle heuristic should also evaluate whether to accept call-offs that were not accepted in the subproblem. Such a change could make it possible for the resolve problem to find better solutions, but would also lead to higher computational time for the heuristic as it in theory would have to consider all the 133 million δ_{cmpt}-variables.

Convergence of the Subproblem

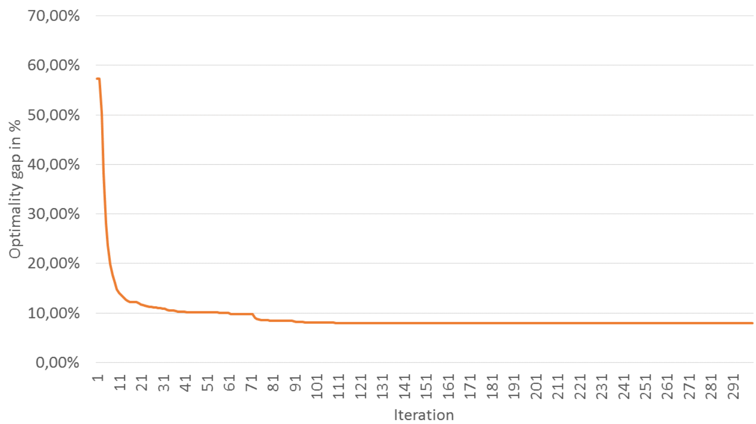


Figure 8.17: Optimality gap for 300 iterations of the Lagrangian relaxation.

The method for Lagrangian relaxation of the casting table constraints has been run for 300 iterations, with a total runtime of 45 979 seconds (12 hours and 46 minutes). The smallest optimality gap, 7.89%, is obtained after 260 iterations which corresponds to 11 hours and 3 minutes. The optimality gap for every iteration is shown in Figure 8.17. For comparison, the optimality gap of the non-relaxed problem after 46 000 seconds is 14.72%. This optimality gap is obtained already after 3 hours and 3 minutes. Thus, the Lagrangian relaxation reduces the optimality gap for the non-relaxed problem, when run for 46 000 seconds, by 46.40%.

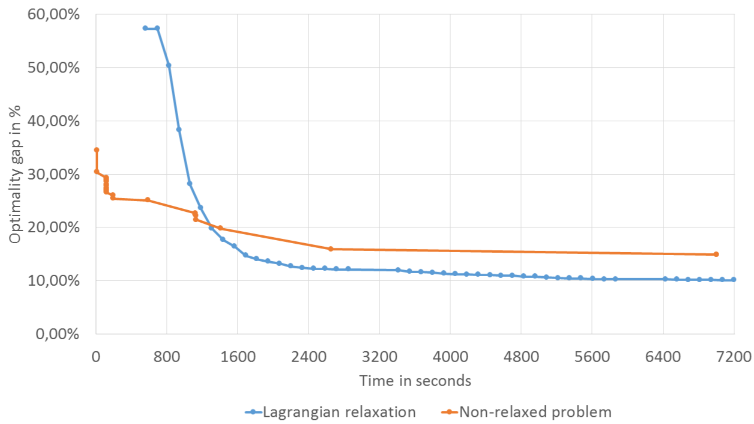


Figure 8.18: Optimality gap for the non-relaxed problem and Lagrangian relaxation.

As can be seen in Figure 8.17 the optimality gap falls rapidly after only a few iterations. It is of interest to analyze this part of the process more closely. In Figure 8.18 the optimality gap for two hours of the Lagrangian relaxation is compared to the non-relaxed problem. After 7 200 seconds (two hours), 46 iterations of the Lagrangian relaxation have been executed. Each point marks the result after an iteration. The parts around 3 000 seconds and 6 000 seconds, where there is a bigger time difference between two iterations, are iteration 20 and 40, when the heuristic procedure has been run. The heuristic procedure is also the reason why the solution for the first iteration occurs at a later time than for the non-relaxed problem.

As can be seen from the figure, the Lagrangian relaxation starts with a higher optimality gap than for the non-relaxed problem, but decreases more rapidly and begins to converge at a smaller optimality gap. The smallest optimality gap for the non-relaxed problem is 14.93% while for the Lagrangian relaxation it is 10.13%. Thus, within two hours, the Lagrangian relaxation reduces the optimality gap for the non-relaxed problem by 32.15%. This is a significant reduction in optimality gap, but it is still larger than the optimality gap obtained from the resolve problem.

Overview

In Table 8.7 an overview of the different cases, with the resulting optimality gap, is presented. The smallest optimality gap is achieved for the case where the $\text{LBD}_{\text{non-relaxed}}$ is combined with the UBD for the resolve problem. However, the runtime for this case is much higher than for the resolve problem in the first iteration of the Lagrangian relaxation. It is also noteworthy that all the results including Lagrangian relaxation, and the resolve problem, have a smaller optimality gap than the non-relaxed problem.

It should be noted that the convergence of the Lagrangian relaxation could be improved with a better heuristic as discussed in Section 8.3.3. When no new UBD is found, the λ is decreased by 50 % every 10th iteration. When no new UBD is found for many iterations, the step length becomes very small and it may look as if the solution converges because the step length is so small. When the heuristic procedure is able to find a new UBD, the step length is reset and it is possible to converge faster and obtain a smaller optimality gap.

Table 8.7: Overview of optimality gap for different computational cases.

Problem case	Runtime in seconds	Optimality gap
Non-relaxed problem	7 200	14.93 %
Non-relaxed problem	46 000	14.03 %
Lagrangian relaxation	7 200	10.13 %
Lagrangian relaxation	45 979	7.89 %
Resolved problem, 1. iteration	564	7.15 %
Combined non-relaxed and resolved	7 764	4.57 %

8.3.4 Comparison of Model Output

It is clear from the results so far, that implementing a Lagrangian relaxation has the potential to improve the computational efficiency. However, it is important to acknowledge that the relaxed problem is different from the non-relaxed problem and may not necessarily converge towards the same solution. Certain combinations of the values of the multipliers can result in solutions that are valid mathematically, but unrealistic from a practical point of view. Therefore, it should be verified that solutions from the Lagrangian relaxation approaches feasible solutions of the non-relaxed problem with increased iterations. In this section the model output of the Lagrangian relaxation is analyzed and compared with the output from the non-relaxed problem.

Figure 8.19 shows the number of accepted call-offs, including duplicates. It also shows the number of unique call-offs accepted. Initially, relatively few unique call-offs are accepted, and since the number of call-offs accepted is much higher, it is clear that call-offs have been accepted on several casting tables simultaneously. After more iterations the number of unique call-offs accepted approaches the number of call-offs accepted. This indicates that the process is getting closer to finding a feasible solution where there are no duplicates of accepted call-offs.

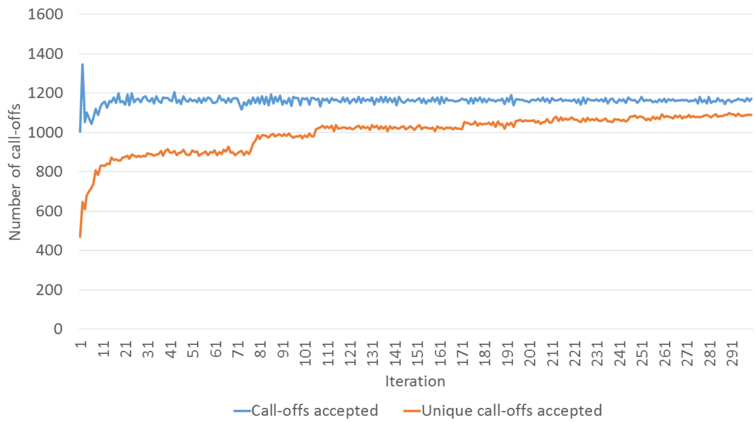


Figure 8.19: Number of call-offs accepted per iteration.

Considering the graph of accepted call-offs, it stabilizes rapidly around 1 200 call-offs, and this happens after only 15-20 iterations. This is close to the number of accepted call-offs from the non-relaxed problem in Section 8.3.1, which accepts 1 208 unique call-offs. When considered together with the discussion on unique call-offs accepted, this is a good indication of the relaxed model behaving as expected in the iterative procedure.

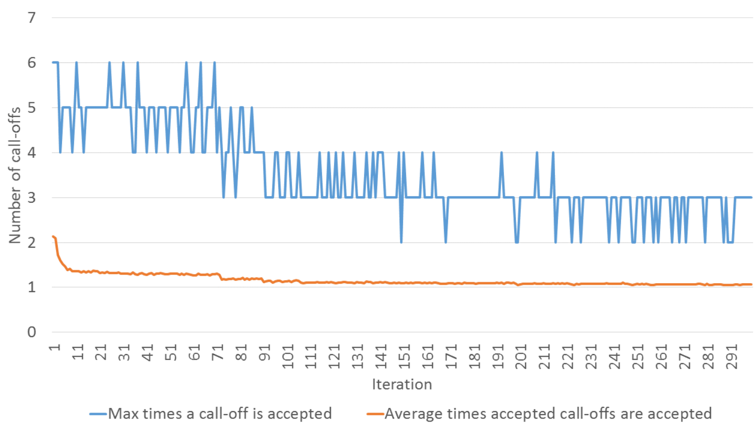


Figure 8.20: Statistics of call-offs accepted per iteration.

In Figure 8.20 the maximum number of times any call-off is accepted is shown, together with the average number of times accepted call-offs have been accepted simultaneously. On the first iteration, an accepted call-off is being accepted simultaneously on 2.1 casting tables on average. This number falls towards one with increased iterations, and after about 200 iterations this number revolves around one.

The maximum number of times a call-off has been accepted is initially six, and then reaches two for the first time after 150 iterations. Together with the average number, this graph should ideally reach one after sufficiently many iterations. These results correspond with the results in Figure 8.19, where the iterative procedure behaves satisfactory but the best solution found is not yet feasible as there are duplicates of accepted call-offs.

In order for the Lagrangian subproblem to find a feasible solution it needs sufficiently many iterations. However in this study, the reason for the subproblem not finding a feasible solution could be the weaknesses in the heuristic procedure discussed in Section 8.3.3. By implementing the suggested change in this procedure, the necessary changes in step lengths could be obtained and a feasible solution of the Lagrangian subproblem could be found.

Chapter 9

Concluding Remarks and Future Research

This thesis has derived a production planning model for the Norwegian aluminum producer Norsk Hydro. Norsk Hydro's problem revolves around allocating demand to their many plants under conditions of daily updated information regarding demand. The demand is either categorized as call-off demand or a forecasted demand, and the model wishes to allocate as much as possible of it while minimizing costs.

Two types of solution approaches have been applied to the problem, a rolling horizon approach and an approach involving Lagrangian relaxation of constraints. The rolling horizon approach has been applied with the purpose of compensating for the dynamic nature of demand information. Meanwhile, two types of Lagrangian relaxations have been tested with the purpose of reducing problem complexity and the following effects on optimality gaps and runtimes.

The scope of values for the input data used has been presented, along with assumptions that were made related to this. The detail level of the data available has made simulating demand data necessary. Simulations have been performed using queuing theory assumptions in order to determine delivery days and arrival days of orders. Orders are later distinguished as either call-off or forecast depending on the rolling horizon iteration, or roll, the model is running. The simulated data has also been exposed to an initialization period to facilitate initial states for various computational studies.

A computational study of the rolling horizon approach has been performed, and both computational efficiency and features of the solution variable output has been discussed. A variation on optimality gap is observed in the consecutive rolls of the approach, and a connection to the amount of call-offs entering the model at each roll is identified. Solution variable output suggests that the model functions as intended; call-off demand is prioritized over forecasted demand, and full capacity utilization is observed on several of the castinghouses. When capacity utilization is limited, the dimensions prevalent in the simulated demand along with unfavourable call-off quantities are seen as contributing factors.

The rolling horizon approach has also been discussed in relation to the concept of nervousness. Nervousness has been investigated by locking production for different times when using the rolling horizon approach. A trade-off is apparent between computational efficiency and performance of the objective function. Locking production for many days in each roll results in higher costs, but solutions close to optimality for this lock setting.

A computational study on the proposed Lagrangian relaxations has shown improvements in the optimality gap and runtime of the model. The so called *casting table relaxation* has been deemed the most efficient, and analyses have primarily been dedicated to this relaxation. The impact of three different computational elements in the Lagrangian solution process has been assessed as they all contribute to the total runtime. Comparisons have been made to the computational efficiency of a non-relaxed variant of the model, and one has shown that a combination of this variant and Lagrangian relaxation gives the best optimality gap.

The analysis of the Lagrangian relaxations has suggested that the solution process can be improved by using a more sophisticated heuristic procedure or by using a different method for updating Lagrangian multipliers between iterations. These are thus natural contenders for future research. More specifically, Lemarechal (1986) suggests a bundling method for updating Lagrangian multipliers. Schütz et al. (2009) use this method to great effect.

An obvious continuation of the solution approaches in this thesis is to combine the Lagrangian relaxations with the rolling horizon approach. The rolling horizon approach was applied by introducing additional constraints to the model in Chapter 5, and the Lagrangian relaxations analyzed did not account for these additions. The computational study of the rolling horizon approach showed that runtimes suffered when production was locked for few days in each roll. Integrating Lagrangian relaxation and rolling horizon can thus contribute to reducing the aforementioned trade-off between nervousness and optimality gaps.

Lastly, an additional subject for potential future research is to extend the model to consider the sequence in which the products are produced. The literature review in this thesis mentioned Tiacci and Saetta (2012), who combine sequencing with a rolling horizon approach. Computational efficiency is preserved by only sequencing the first days of the planning horizons. For Norsk Hydro, preferable sequences for instance have similar alloys following each other. In other cases, certain alloys cannot follow each other without the operators having to clean instruments at the plants in between. The sequencing decisions could be included as part of an operational planning tool, where the call-offs are allocated considering product properties that are advantageous for production sequences, leading to minimal global rescheduling times.

Bibliography

Anthony, R. N. (1965), *Planning and Control Systems: A Framework for Analysis [by]*, Division of Research, Graduate School of Business Administration, Harvard University.

As'ad, R. and Demirli, K. (2010), 'Production scheduling in steel rolling mills with demand substitution: Rolling horizon implementation and approximations', *International Journal of Production Economics* **126**(2), 361–369.

Baker, K. R. (1977), 'An experimental study of the effectiveness of rolling schedules in production planning', *Decision Sciences* **8**(1), 19–27.

Baker, K. R. and Peterson, D. W. (1979), 'An analytic framework for evaluating rolling schedules', *Management Science* **25**(4), 341–351.

Barut, M. and Sridharan, V. (2004), 'Design and evaluation of a dynamic capacity apportionment procedure', *European Journal of Operational Research* **155**(1), 112–133.

Bitran, G. R. and Tirupati, D. (1993), 'Hierarchical production planning', *Handbooks in operations research and management science* **4**, 523–568.

Bitran, G. R. and Yanasse, H. H. (1982), 'Computational complexity of the capacitated lot size problem', *Management Science* **28**(10), 1174–1186.

Blackburn, J. D. and Millen, R. A. (1982), 'The impact of a rolling schedule in a multi-level mrp system', *Journal of Operations Management* **2**(2), 125–135.

Campbell, G. M. (1992), 'Master production scheduling under rolling planning horizons with fixed order intervals', *Decision Sciences* **23**(2), 312–331.

Chen, W.-H. and Thizy, J.-M. (1990), 'Analysis of relaxations for the multi-item capacitated lot-sizing problem', *Annals of operations Research* **26**(1), 29–72.

Chung, C.-H. and Krajewski, L. J. (1987), 'Interfacing aggregate plans and master production schedules via a rolling horizon feedback procedure', *Omega* **15**(5), 401–409.

Dempster, M. A. H., Fisher, M., Jansen, L., Lageweg, B., Lenstra, J. K. and Rinnooy Kan, A. (1981), 'Analytical evaluation of hierarchical planning systems', *Operations Research* **29**(4), 707–716.

Fog, K. (2016), 'Market outlook'. Last accessed 25-04-2018.
URL: *https://www.hydro.com/contentassets/67ae89f8f0bd44a18660ae3b48f13bef/hydro-cmd-2017-market-outlook.pdf*

Guo, Z., Wong, W. K. and Leung, S. (2013), 'A hybrid intelligent model for order allocation planning in make-to-order manufacturing', *Applied Soft Computing* **13**(3), 1376–1390.

International Aluminium Institute (2017), 'Primary aluminium production statistics'. Last accessed 02-05-2018.
URL: *http://www.world-aluminium.org/statistics/*

Jans, R. and Degraeve, Z. (2008), 'Modeling industrial lot sizing problems: a review', *International Journal of Production Research* **46**(6), 1619–1643.

Karimi, B., Ghomi, S. F. and Wilson, J. (2003), 'The capacitated lot sizing problem: a review of models and algorithms', *Omega* **31**(5), 365–378.

Lee, C.-Y., Çetinkaya, S. and Wagelmans, A. P. (2001), 'A dynamic lot-sizing model with demand time windows', *Management Science* **47**(10), 1384–1395.

Lemarechal, C. (1986), Constructing bundle methods for convex optimization, *in* 'North-Holland Mathematics Studies', Vol. 129, Elsevier, pp. 201–240.

Lundgren, J., Rönnquist, M. and Värbrand, P. (2010), *Optimization*, Studentlitteratur AB.

Mestry, S., Damodaran, P. and Chen, C.-S. (2011), 'A branch and price solution approach for order acceptance and capacity planning in make-to-order operations', *European Journal of Operational Research* **211**(3), 480–495.

Nappi, C. (2013), 'The global aluminium industry 40 years from 1972', *World Aluminium* pp. 1–27.

Neureuther, B. D., Polak, G. G. and Sanders, N. R. (2004), 'A hierarchical production plan for a make-to-order steel fabrication plant', *Production Planning & Control* **15**(3), 324–335.

Omar, M. K. and Teo, S. (2007), 'Hierarchical production planning and scheduling in a multi-product, batch process environment', *International Journal of Production Research* **45**(5), 1029–1047.

Reine, M. G., Ullebust, T. and Wilsher, H. (2017), 'An operational production planning tool for casting in metal manufacturing'.

Schütz, P., Tomasgard, A. and Ahmed, S. (2009), 'Supply chain design under uncertainty using sample average approximation and dual decomposition', *European Journal of Operational Research* **199**(2), 409–419.

Schütz, P., Tomasgard, A. and Uggen, K. T. (2011), Using hierarchical planning to exploit supply chain flexibility – An example from the Norwegian meat industry, *in* K. D. Lawrence, R. K. Klimberg and V. Miori, eds, 'The Supply Chain in Manufacturing, Distribution, and Transportation: Modeling, Optimization, and Applications', CRC Press, Boca Raton, FL, chapter 4, pp. 47–68.

Simpson, N. (1999), 'Multiple level production planning in rolling horizon assembly environments', *European Journal of Operational Research* **114**(1), 15–28.

Spengler, T., Rehkopf, S. and Volling, T. (2007), 'Revenue management in make-to-order manufacturing—an application to the iron and steel industry', *OR spectrum* **29**(1), 157–171.

Sridharan, V. and Berry, W. L. (1990), 'Freezing the master production schedule under demand uncertainty', *Decision Sciences* **21**(1), 97–120.

Sridharan, V., Berry, W. L. and Udayabhanu, V. (1987), 'Freezing the master production schedule under rolling planning horizons', *Management Science* **33**(9), 1137–1149.

Stadtler, H. (2000), 'Improved rolling schedules for the dynamic single-level lot-sizing problem', *Management Science* **46**(2), 318–326.

Stadtler, H., Kilger, C. and Meyr, H. (2015), 'Supply chain management and advanced planning'.

Stevenson, M., Hendry, L. C. and Kingsman, B. G. (2005), 'A review of production planning and control: the applicability of key concepts to the make-to-order industry', *International journal of production research* **43**(5), 869–898.

Tiacci, L. and Saetta, S. (2012), 'Demand forecasting, lot sizing and scheduling on a rolling horizon basis', *International Journal of Production Economics* **140**(2), 803–814.

Torabi, S. A., Ebadian, M. and Tanha, R. (2010), 'Fuzzy hierarchical production planning (with a case study)', *Fuzzy Sets and Systems* **161**(11), 1511–1529.

Umble, E. J., Haft, R. R. and Umble, M. M. (2003), 'Enterprise resource planning: Implementation procedures and critical success factors', *European journal of operational research* **146**(2), 241–257.

Van Vyve, M. (2006), 'Linear-programming extended formulations for the single-item lot-sizing problem with backlogging and constant capacity', *Mathematical Programming* **108**(1), 53–77.

Appendix A

Previous Model

The following is the model formulation developed in Reine et al. (2017).

$$
\begin{aligned}
\max\ z = & \sum_{c\in\mathcal{C}} \sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} (P_p D_{cpt} \delta_{ckpt}) \\
& - \sum_{c\in\mathcal{C}} \sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} \left(C^T_{ck} D_{cpt} \delta_{ckpt} + C^F \delta_{ckpt}\right) \\
& - \sum_{k\in\mathcal{K}} \sum_{m\in\mathcal{M}_{\mathcal{K}}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} \left(C^P_{kp} B x_{mpt}\right) \\
& - \sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} C^H s_{kpt} \\
& - \sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} \left(C^A_{kp} y^+_{kpt} + C^B_{kp} y^-_{kpt}\right)
\end{aligned}
$$

subject to

$$
\begin{aligned}
s_{kpt-1} + \sum_{m\in\mathcal{M}_{\mathcal{K}}} B x_{mpt} - \sum_{c\in\mathcal{C}} (D_{cpt} \delta_{ckpt}) = s_{kpt} \qquad & k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \\
\sum_{m\in\mathcal{M}_{\mathcal{K}}} B x_{mpt} + s_{kpt-1} - \sum_{c\in\mathcal{C}} D_{cpt} \delta_{ckpt} \geq 0 \qquad & k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \\
\sum_{k\in\mathcal{K}} \delta_{ckpt} \leq 1 \qquad & c \in \mathcal{C}, p \in \mathcal{P}, t \in \mathcal{T} \\
\sum_{p\in\mathcal{P}} B x_{mpt} \leq K_m \qquad & k \in \mathcal{K}, m \in \mathcal{M}_{\mathcal{K}}, t \in \mathcal{T}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{m \in \mathcal{M}_{\mathcal{K}}} (Bx_{mpt}) + y^-_{kpt} - y^+_{kpt} = \sum_{c \in \mathcal{C}} Z_{ckpt} \qquad & k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \\
Bx_{mpt} \leq K_m Q_{mp} \qquad & k \in \mathcal{K}, m \in \mathcal{M}_{\mathcal{K}}, p \in \mathcal{P}, t \in \mathcal{T} \\
x_{mpt} \in \mathbb{Z}^+ \qquad & k \in \mathcal{K}, m \in \mathcal{M}_{\mathcal{K}}, p \in \mathcal{P}, t \in \mathcal{T} \\
y^+_{kpt}, y^-_{kpt} \geq 0 \qquad & k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \\
s_{kpt} \geq 0 \qquad & k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T} \\
\delta_{ckpt} \in \{0, 1\} \qquad & c \in \mathcal{C}, k \in \mathcal{K}, p \in \mathcal{P}, t \in \mathcal{T}
\end{aligned}
$$