

UNIVERSITY OF TWENTE.

Optimal capacity planning with
uncertain supply in a whitefish
processing plant

Yelte Kikke

Industrial Engineering and Management

Submission date: June 2018

First Supervisor: Marco Schutten

Second Supervisor: Martijn Mes

University of Twente

Faculty of Behavioral Management and Social Sciences

Department of Industrial Engineering and Business

Information Systems

Preface

Before you lies the thesis “Optimal capacity planning with uncertain supply in a whitefish processing plant”, which is the result of my research at the Norwegian University of Science in Technology in combination with Lerøy Norway Seafoods. I wrote this thesis to finish my master degree in Industrial Engineering and Management at the University of Twente.

Although I had no previous knowledge about the fish industry, I must say that I enjoyed doing research in this field. This industry is coping with many difficulties, since the freshness of the fish is so important. I found it a challenge to come up with solutions to help with these difficulties. Besides this, it was quite an experience visiting the processing plant in Båtsfjord, which is almost at the most northern part of Norway.

The fact that a lot of the literature and data was in Norwegian was a serious challenge at the beginning. Later however I found it quite fun to learn the language in this way and I think I picked up some Norwegian, which I will proof to you in the last paragraphs of this preface.

I want to thank Marco Schutten and Martijn Mes, which are my supervisors from the University of Twente. They gave valuable and detailed feedback on my intermediate work, which helped to improve the quality of my thesis.

Fra NTNU vil jeg takke Jørn Vatn, som var veilederen min fra dette universitet. Diskusjonene vi hadde gjennom møtene hjalp meg med å forstå og finne løsninger på problemene.

Ketil Pettersen, sentralplanlegger hos Lerøy, sørget for dataene jeg trengte for å bygge modellene og forstå situasjonen på fabrikken. Han viste oss også fabrikken i Båtsfjord, som tok hele dagen. Jeg vil gjerne takke ham for dette.

Finally, I hope you have fun reading this report.

Yelte Kikke
Trondheim, 24-04-2018

Executive summary

This research aims to improve the planning of the workhours, which we call the capacity, in the whitefish processing facility of Båtsfjord, which is owned by LNS. LNS has problems planning the capacity, due to uncertainty in the supply of raw fish to the processing plant. LNS cannot catch fresh fish, which is used as raw material, itself due to regulations by the Norwegian government. Hence, the company depends on supply of external fishing vessels, where the catch varies every shipment. The only information that LNS has is which vessels are coming to deliver fish, three days in advance. The research consists of three steps:

1. Finding a method to reduce the uncertainty in the supply of raw material;
2. Creating a mathematical model of the processing plant in Båtsfjord and implementing the method of step 1 in this model;
3. Estimating the gains of the reduction in uncertainty of the supply of raw material.

In the first step we compare three methods to reduce the supply uncertainty. With the requirements that LNS set, we decide that a method where we forecast the fish catch per vessel fits the needs of LNS the best. We use the information about the vessel length, the vessel age, the wind speed from the west, and the visibility to forecast the number of kilogram Cod a vessel delivers to LNS. All the four predictors are significant with a maximum error of 5%. This error gives the probability that a predictor seems to have influence, while this is not the case.

For step 2 we create a mathematical model of the processing plant. We decide to use Stochastic Programming since we consider the processing times deterministic. The age and quality of the raw material dictates which products LNS can produce. Hence, when not all the supply is used at a given day, the possible value of the raw material deteriorates. Therefore, there is a recursion in the calculations, making it too difficult to calculate the gains with exact formulae. We decide to use a simulation method to evaluate the gains of our forecasting method. We use a rolling planning horizon and simulate for each day a realized supply of raw material and for the coming three days a forecasted supply of raw material, using the forecasting method of step 1.

To calculate the gains in step 3 of the research approach, we need a current scenario to compare the results of the simulation model with. Because LNS cannot provide us with detailed data of the planned capacity in the processing plant, we create our own scenario where LNS uses the average catch of a vessel to forecast the supply, we name this scenario the base scenario. With this scenario we show the gains when we reduce the supply uncertainty. For this we create the KPI “profit”, which consists of the generated profit per kilogram raw fish. Also, we compare the results with the use of our forecasting model with actual data of the planned hours in the cutting department of LNS. We create the KPI “productivity”, which consists of the kilogram processed fillet per cutting hour. We call this scenario the current situation.

Next, we create the new scenarios where we use the forecasting method of step 2 of the research. We create three scenarios that can potentially improve the current situation at LNS:

1. Only use the forecasting method to forecast the supply of raw material;
2. Use extra capacity that is flexible against extra costs;
3. Use method 2 in combination with an extra filleting machine.

Table 1 shows the results of the comparisons between the methods we create, and the current situation. We see that in all the cases the lower bound of the confidence intervals yields an improvement. This leads to the conclusion that the forecasting method increases the current situation.

Although the case with an extra filleting machine yields the best profits, we advise LNS to implement the case with only extra capacity. The investment in the extra filleting machine has a 42% probability of yielding less profit than only using overtime. The differences in profits follow a normal distribution. This has the result that the probability distribution is symmetric, making the probability of saving a given number of NOK the same as losing this amount. This results in a high volatility of the returns of the investment. Therefore, we consider the investment too risky for LNS currently, and advice only the use of extra capacity, as this does not require any investment.

Table 1: Comparison of KPIs with current situation (higher is better)

Method number	KPI	Confidence interval 95% Current situation		Confidence interval 95% Base scenario	
		Min	Max	Min	Max
1 (Forecasting method)	Productivity	14%	28%	3%	8%
	Profit	No data		2%	76%
2 (Extra capacity)	Productivity	15%	59%	13%	23%
	Profit	No data		10%	73%
3 (Extra capacity with extra filleting machine)	Productivity	20%	41%	12%	22%
	Profit	No data		18%	90%

Table of Contents

List of used figures	7
List of used tables	8
Abbreviations	9
1 Introduction	11
1.1 Company description	11
1.2 Problem statement	13
1.3 Research questions and methods	14
1.4 Scope of the research	15
2 Current situation LNS: External and internal analysis	17
2.1 Norwegian fish industry	17
2.2 Supply chain of whitefish: General analysis	17
2.3 Analysis of LNS	19
2.4 Conclusions about the processes within LNS	28
3 Theoretical background	31
3.1 Optimization of capacity planning	31
3.2 Reduction of uncertainty of supply	35
4 Reducing uncertainty of the supply to LNS	39
4.1 Choosing best method to reduce uncertainty	39
4.2 Data explanation and sampling	41
4.3 Prediction model: Response- and prediction variables	44
4.4 Causal model: linear regression	46
4.5 Machine learning models	49
4.6 Weather forecast accuracy and impact on prediction model	53
4.7 Selection of best model	58
4.8 Conclusions of the reduction of supply uncertainty	58
5 Using the supply information in the capacity planning	59
5.1 Assumptions about the processes	59
5.2 Choice of the modelling approach	60
5.3 Modelling the processing plant in Båtsfjord	62
5.4 Evaluating the planning model	64
5.5 Conclusions of the plant model and simulation process	69
6 Estimating the gains with the new planning method	71

6.1	Analyzing the output of the model: choosing the variables	71
6.2	Current cost level and deterministic solution.....	72
6.3	Solution with forecasting model.....	74
6.4	Solution with possibility of overtime	75
6.5	Solution with capacity increase of filleting machine	76
6.6	Comparison of all the setups and testing the significance	76
6.7	Choosing the best setup and determination of the expected improvement.....	79
6.8	Conclusions about the gains when using the forecasting method	82
7	Conclusions, recommendations, and discussion	83
7.1	Conclusions	83
7.2	Recommendations	84
7.3	Discussion.....	84
8	Bibliography	87
9	Table of content appendices.....	91

List of used figures

Figure 1.1: LNS and their production facilities (Google Maps, 2017)	11
Figure 1.2: Strategy LNS analyzed with Treacey & Wiersema	12
Figure 1.3: Financial results Norway Seafoods AS (in 10 ³ NOK)	12
Figure 1.4: Ishikawa diagram from problems and causes at LNS	13
Figure 1.5: Schematic overview project LNS.....	15
Figure 2.1: Export statistics Norway 2005 until 2016	17
Figure 2.2: Largest fish export nation in 2012	17
Figure 2.3: Export characteristics from Norwegian fish export	18
Figure 2.4: Information flow for the receipt of raw material	19
Figure 2.5: Raw material bought by Båtsfjord in 2017	21
Figure 2.6: Raw material bought per month by Båtsfjord	21
Figure 2.7: Optimal cut of a raw fish.....	22
Figure 2.8: Primary production processes	26
Figure 2.9: Information flow between processing facility and central planner.....	27
Figure 2.10: Plotted processed raw material against cutting hours in Båtsfjord	28
Figure 3.1: Process of creating the RCCP according to Smunt (1996)	31
Figure 3.2: Deterioration process of raw material.....	32
Figure 3.3: Matrix of problem combinations according to Peidro et al.....	33
Figure 3.4: Schematic visualization of the random forest algorithm.....	37
Figure 3.5: Categorization of the forecasting methods	38
Figure 4.1: Score of each method visualized (blue= idea a, red=idea b, green=idea c)	41
Figure 4.2: Norway and its regions	42
Figure 4.3: Catch per kilogram coastline for the relevant regions	43
Figure 4.4: Data distribution of the response variable	43
Figure 4.5: Distribution of the catch	44
Figure 4.6: Distribution after transformation with natural logarithm	44
Figure 4.7: Star schema of the prediction variables	46
Figure 4.8: Predicted values from the linear regression model.....	49
Figure 4.9: Forecasted values together with observed values from the REP Tree	50
Figure 4.10: Forecasted values together with observed values of the MSP Tree	52
Figure 4.11: Forecasted values together with observed values from the random forest	52
Figure 4.12: Scatterplot observed values and predicted values.....	54
Figure 4.13: Variance of forecast temperature	54
Figure 4.14: Schematic model for Monte Carlo simulation	56
Figure 4.15: Estimated variance for various simulation runs	57
Figure 4.16: Estimated standard deviation based on Monte Carlo simulation	57
Figure 4.17: Comparison of all forecast methods.....	58
Figure 5.1: Recursion in the calculations	61
Figure 5.2: Simulation process schematically explained	67
Figure 5.3: Welch's graphical method for 20 simulation runs of length 500	69
Figure 6.1: Comparison of average productivity in cutting department	77
Figure 6.2: Comparison of average profit per kilogram raw material	78
Figure 6.3: Total NPV compared with probabilities.....	81

List of used tables

Table 2.1: Distribution of the length of the active fishing vessels in Norway	18
Table 2.2: Purchasing prices raw materials	20
Table 2.3: Selling prices of the end products.....	21
Table 2.4: Production matrix with respect to age and quality raw material.....	22
Table 2.5: Possibilities end products for each part.....	22
Table 2.6: Production times and capacity for Cod.....	25
Table 2.7: Production times for Haddock and Pollock.....	25
Table 2.8: Summary statistics productivity of cutting department	28
Table 3.1: Types of uncertainty with solutions	35
Table 3.2: Variables with the used units.....	36
Table 4.1: Date specific prediction variables	45
Table 4.2: Vessel specific prediction variables.....	45
Table 4.3: Performance indicators REP Tree.....	50
Table 4.4: Performance indicators MSP Tree	52
Table 4.5: Performance indicators random forest.....	52
Table 4.6: Expected value and expected standard deviation for VISIBIL and WINDOPT.....	55
Table 5.1: Variables with their units and elements	60
Table 5.2: Scenarios with their probabilities	62
Table 5.3: Indices of the SPP	63
Table 5.4: Variables of the SPP	63
Table 5.5: Parameters of the SPP.....	63
Table 5.6: Clusters for analysis.....	65
Table 5.7: Predictors and their distribution.....	65
Table 5.8: Expected daily shipments in each cluster	66
Table 5.9: Two examples of output simulation model	68
Table 6.1: Current productivity according to data LNS.....	73
Table 6.2: Summary output deterministic situation.....	73
Table 6.3: Summary output deterministic situation.....	74
Table 6.4: Summary output with forecasting model	74
Table 6.5: Summary output with overtime model	75
Table 6.6: Summary output with capacity increase model	76
Table 6.7: Significance table differences in productivity	78
Table 6.8: Significance table differences in profit.....	79

Abbreviations

BoR	Bill of Recourses
CoV	Coefficient of Variance
CPuOF	Capacity Planning using Overall Factors
LNS	Lerøy Norway Seafoods
LPP	Linear Programming Problem
MBoR	Modified Bill of Resources
MILPP	Mixed Integer Linear Programming Problem
MPMP	Multi-period, multi-product
MSP	Multi-stage stochastic programming
NFSO	Norwegian Fishermen's Sales Organization
NOK	Norwegian kroner
NPV	Net Present Value
RAE	Relative absolute error
RCCP	Rough-Cut Capacity Planning
REP	Reduced error pruning
RMSE	Root mean squared error
SDR	Standard Deviation Reduction
SPP	Stochastic Programming Problem
VIF	Variance Inflation Factor
WS	Work station

1 Introduction

This chapter discusses the background of this research. It explains the problem and discusses the research, which follows from the problem. Section 1.1 gives a brief explanation of the company, which is necessary to understand the problem. Section 1.2 continues with the problem statement and Section 1.3 gives the research questions and methods. Section 1.4 finishes this section by giving the scope of the research.

1.1 Company description

Lerøy Norway Seafoods (LNS) is a processor of whitefish, where the most important fish species are Cod, Pollock, and Haddock. LNS is active in the northern part of Norway. The company has 8 processing plants. It has 3 big plants, operating mostly all year, and 5 smaller plants, which are mostly opened in season (Figure 1.1) (Norway Seafoods, 2017). The season starts around the end of January and ends in July. LNS needs raw fish to process, which comes from fishing vessels. It does not own its own fishing vessels. The reason for this is that it is prohibited for a land-based processing company to catch fish. Hence, LNS depends on external fishermen to deliver the raw fish.

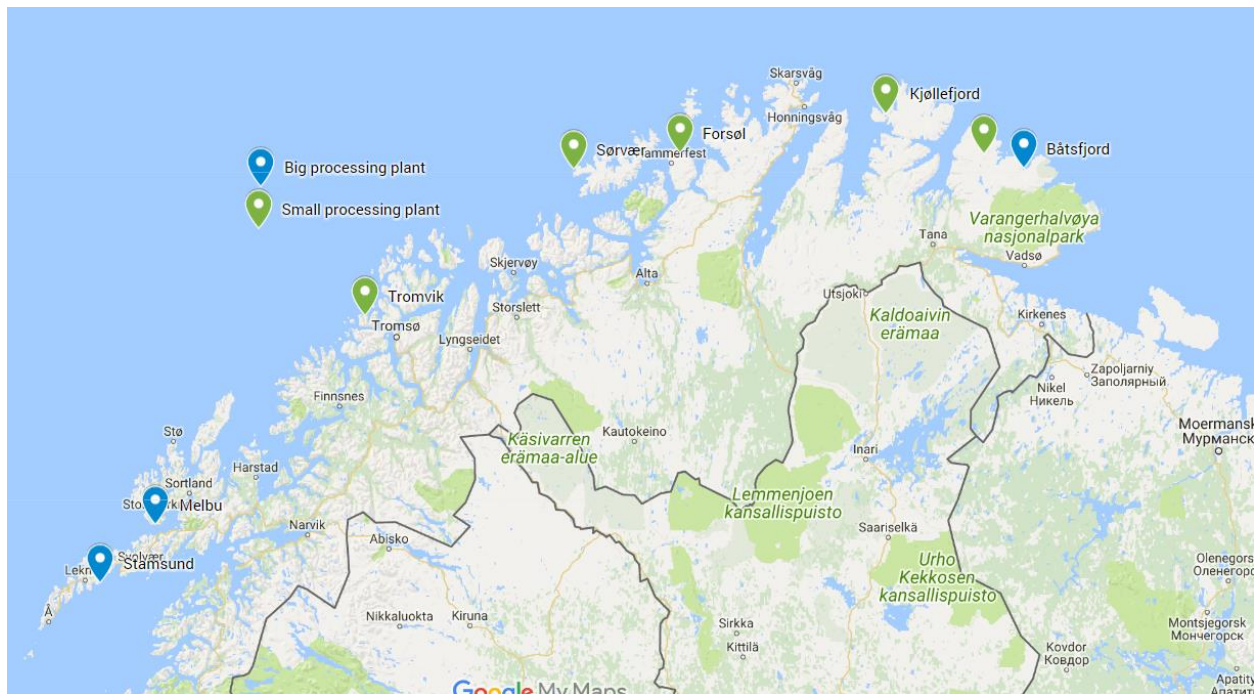


Figure 1.1: LNS and their production facilities (Google Maps, 2017)

The company specializes its core activities on the filleting of freshly caught fish. Its strategy is to keep supply times low, by producing and delivering the filets as fast as possible. At this way LNS is able to deliver fresher products as its competitors. This gives a competitive advantage. LNS supplies to a wide customer base and does not keep close customer relationships.

To determine the competitive strategy of LNS, we use the model from Treacy & Wiersema. The strategy is determined by using three value disciplines (Wiersema, 1992):

1. Operational excellence: when there is a focus on Operational Excellence, a company will try to outperform the competition by having the most efficient and streamlined operation to supply with low costs.
2. Product leadership: a company focusing on product leadership will compete by creating products with higher quality or which are more innovative.
3. Customer intimacy: customer service and attention are the main competition aspect on this discipline. Also, the possibility to deliver a wide range of products, or fast delivery can be seen as a customer intimacy strategy.

When combining the facts of LNS and the knowledge of the model of Treacy & Wiersema, we conclude that LNS focusses on product leadership (Wiersema, 1992). In Figure 1.2 we show the model when we apply it on LNS. Hence, for the company it is important to be able to supply the end products as freshly as possible.

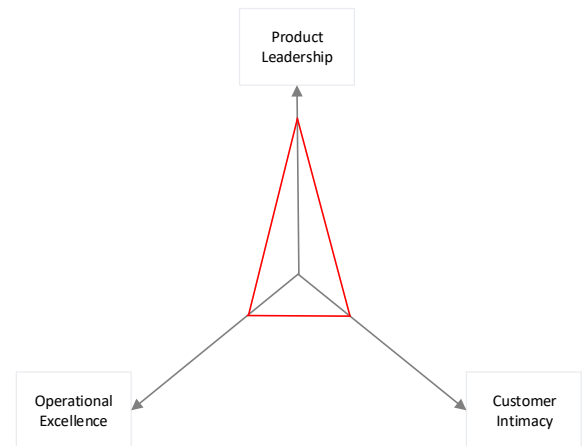


Figure 1.2: Strategy LNS analyzed with Treacy & Wiersema

Financially LNS made losses since 2011. Figure 1.3 summarizes the financial results from 2010 until 2015. (*Konsernregnskapet*, 2015). The raw material costs, which consists of fresh fish, are a big part of the expenses of the company. In 2017 Norway Seafoods AS is bought by Lerøy, after which the financial reports are not specified anymore for Norway Seafoods. Norway Seafoods remains operating autonomously as a business unit within Lerøy, making it Lerøy Norway Seafoods.

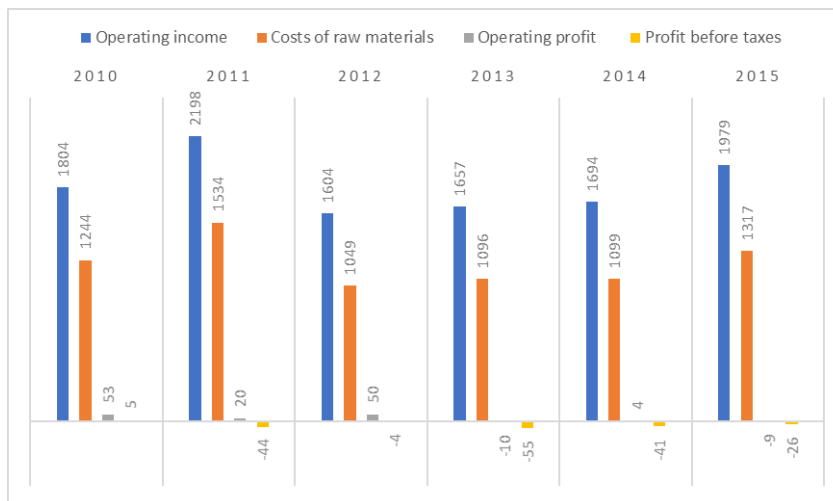


Figure 1.3: Financial results Norway Seafoods AS (in 10^3 NOK)

1.2 Problem statement

Since LNS does not catch fish itself, the company depends on external parties to deliver the (fresh) fish to the processing plants. These external parties are independently operating fishing vessels, where LNS must make agreements with about when to deliver a shipment of fish. LNS knows approximately three days in advance which vessels are going to deliver. Besides this, there is no further information exchanged between a vessel and a processing facility.

The problem that LNS faces, is that the quantity of fish a vessel will deliver is uncertain for LNS until the moment of actual delivery. Since the strategy of LNS is to compete on product quality, LNS must process the fish as soon as possible to be able to deliver the highest quality products. It is unclear for LNS how much fish it will get delivered on a given day, and thus how much demand there will be for capacity. Therefore, the production planning is made under uncertainty. Currently, this planning is not based on any objective measurements, but is made on intuition of one central planner. Also, the supply characteristics of each vessel in terms of quantity are not being tracked or stored in a database. This can result in a human bias in the planning, where systematically too much or too little capacity is planned, based on the objective of the planner.

There are two different cases that cause problems: over- and under capacity. When LNS plans too much capacity, given the amount of raw material, there is overcapacity and LNS pays employees while employees are not able to perform value added activities. When there is less capacity planned than there is capacity needed for the raw material, not all the raw material can be processed. As a result, the fish becomes one day older and it reduces in value. Alternatively, the employees have to work in overtime, resulting in extra costs.

Figure 1.4 shows the Ishikawa diagram corresponding to the problems of LNS.

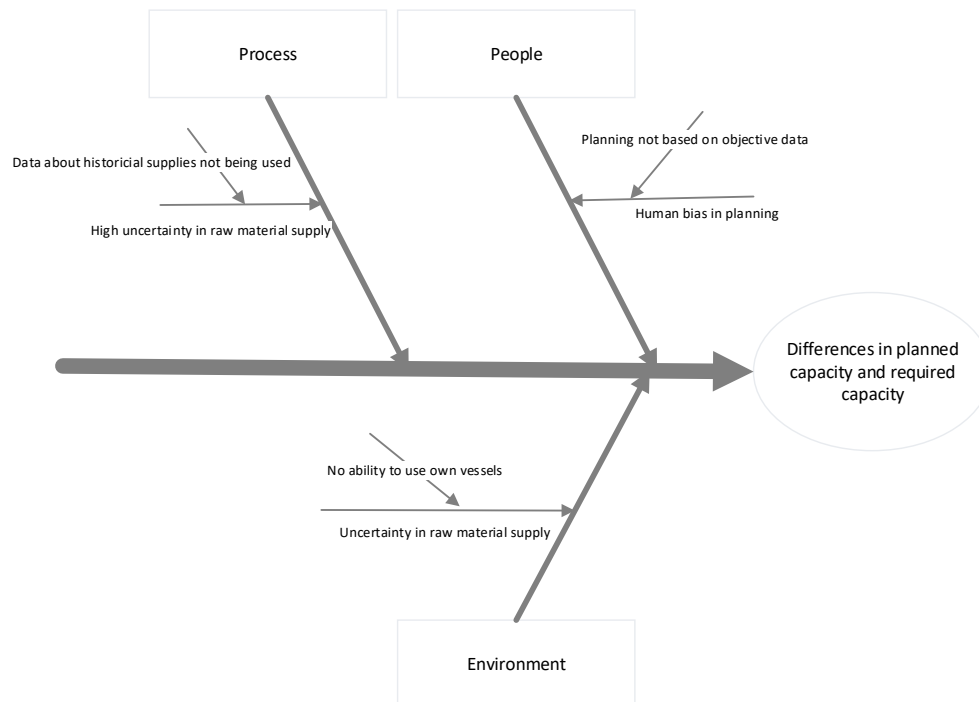


Figure 1.4: Ishikawa diagram for the problems and causes of LNS

1.3 Research questions and methods

Now that we analyzed the problems that LNS faces, we formulate the research starting with the research questions. After the research questions, we discuss the methods that we use to answer these questions.

The main research question this report answers is:

“How can information about the raw material supply be used to improve the capacity planning within the processing facilities of LNS to reduce under- and overcapacity of labor and optimize the use of the raw material”

Since this main research question is too broad to answer in one time, we divide it in 5 sub questions.

1. How are the primary processes organized within LNS that are relevant for the capacity planning?

To improve the current situation, it is important to first analyze the current processes and clarify how the flow of materials and information are organized within the processing plants. Chapter 2 focusses on these aspects.

In this research question we use descriptive research, together with both quantitative and qualitative data that comes from historical data of LNS and a plant visit and interview with employees.

2. Which methods exist to plan capacity with uncertain supply?

We need information to optimize the current capacity planning. Chapter 3 consists of the literature research about planning capacity with uncertain supply, which gives the necessary information about capacity planning with uncertain supply. In this research question we use descriptive research for the literature research.

3. How can the uncertainty about the supply of fish be reduced?

To answer this research question, we need information about the possibilities to reduce the uncertainty in supply. Chapter 3 focuses on this. In Chapter 4 we select the best method for LNS and we develop this method further with respect to the situation of LNS. To answer this research question, we use both descriptive research in the literature research and exploratory research to analyze which methods LNS can use. We use both quantitative and qualitative data. The data consists of catch statistics of vessels in Norway and qualitative information about reduction of uncertainty in supply which we find in the literature.

4. How can the information about the supply be used in the capacity planning and what are the advantages of using this information?

We use the information and method to reduce uncertainty of research question 3 in the capacity planning. We do this in Chapter 5. We use exploratory research to answer this question, in combination with both qualitative and quantitative data.

5. Which savings can LNS expect when using the supply information in the capacity planning?

To determine the effectiveness of the method, we need an objective measurement of the savings for LNS. Chapter 6 focusses on this. This question consists of testing research together with quantitative data. Testing research consists of research where we test if our findings in the other research questions are true.

There are four major stages in the project: analysis of the current situation, reviewing the literature, reducing the uncertainty of supply, implementing and testing the reduction in the processing plant of LNS. Figure 1.5 shows these phases together with the research questions that we answer. The report consists of 6 main chapters. Chapter 7 finishes the report, with the conclusions and recommendations, this chapter is not included in Figure 1.5.

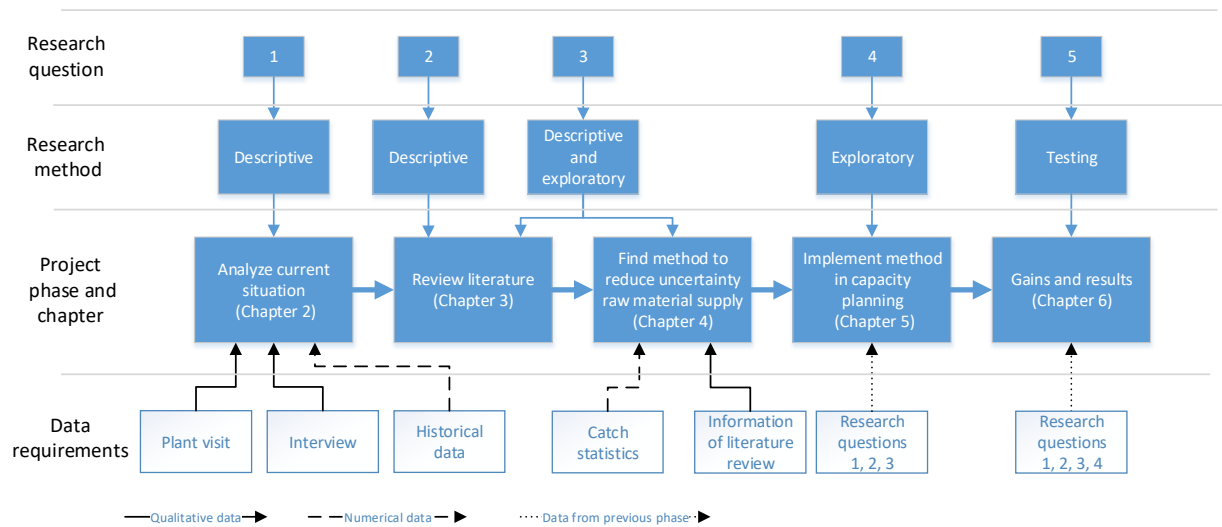


Figure 1.5: Schematic overview project LNS

1.4 Scope of the research

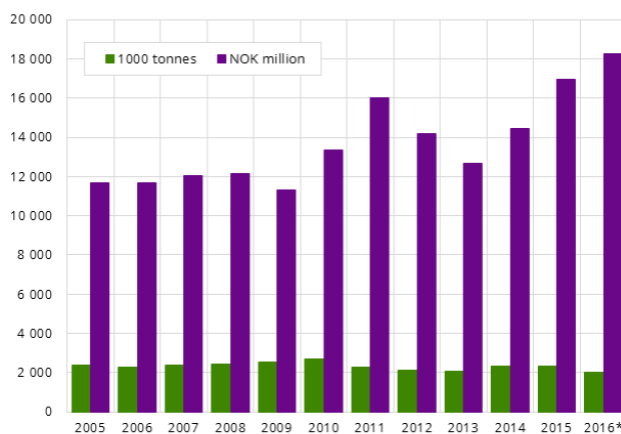
Due to the time limitations, we focus our research on 1 processing plant. We choose a big processing plant, since these are operating the whole year. In this way we can test our method for in-season months and off-season months. We choose to test our method on the processing plant in Båtsfjord, since this processing plant can produce fresh and frozen finished products and has the most problems with the uncertainty in supply according to LNS.

2 Current situation LNS: External and internal analysis

This chapter analyzes the current situation of LNS and the supply chain in which it operates, answering research question 1. Section 2.1 gives some general information about the Norwegian fish industry and the supply chain of whitefish. Section 2.2 continues with a general analysis of the supply chain of whitefish. After this background information, Section 2.3 continues with the current situations specified on LNS. Finally, Section 2.4 finishes this chapter with the conclusions.

2.1 Norwegian fish industry

The fish industry is important for Norway, considering its value and its share in the GDP and export. In 2016 the value of the caught fish was over 18 billion NOK (Statistisk sentralbyrå, 2017). Comparing this to the GDP of Norway in 2016, which was 3118 billion NOK, we see that the fish industry accounts for 0.6% of the GDP. The export of Norway was 1064 billion NOK in 2016 (Statistisk sentralbyrå, 2017). 91.6 billion of this export was as result of seafood export, making it account for 8.6% of the export value (Norwayexports, 2017). Figure 2.1 shows this visually. The total value of the fish catch increased over the last 3 years. Compared to other fish exporting countries, Norway is the second largest exporter when looking at the total value in USD (Figure 2.2). Considering the small population of Norway compared with China, which is the first, and the USA, which is the third largest exporter, it is understandable why the fish industry is an important source of income.



Source: Statistics Norway

Figure 2.1: Export statistics Norway 2005 until 2016

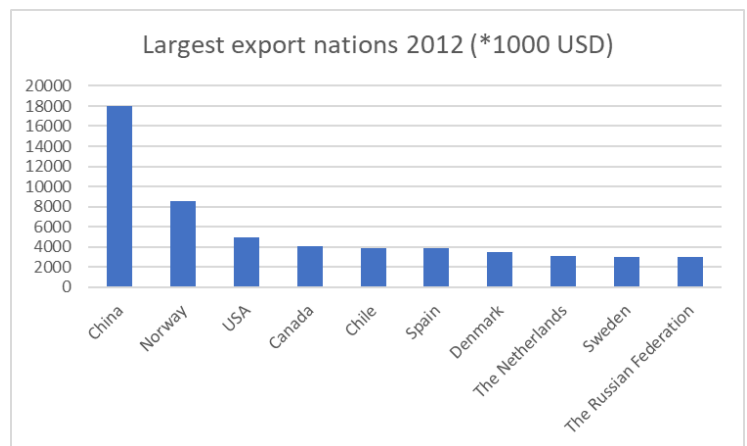


Figure 2.2: Largest fish export nation in 2012

When looking at the fish catch in Norway, Cod is the most important single specie for the Norwegian fish catch, it accounts for 20% of the total fish catch in 2016 (Statistisk sentralbyrå, 2017).

2.2 Supply chain of whitefish: General analysis

It is important to know from which sources LNS gets its supply of raw materials. Therefore, we analyze the supply chain of whitefish in this section. We use the research of Thakur (2017), which is a qualitative research about the whitefish supply chain. In Section 2.2.1 we explain the information and raw material flows in the supply chain. Section 2.2.2 shows the demand characteristic of the export of the Norwegian whitefish industry.

2.2.1 Supply of raw material and information flow

LNS gathers its fish from three sources: seagoing vessels, coastal vessels or internal transport between processing plants. Coastal vessels fish with lines or small nets and return the same day to deliver fish (Thakur, 2017). Coastal vessels deliver the fish fresh, since there is no capacity to freeze the fish onboard. Seagoing vessels take around 3-5 days to deliver the fish. In 2012, there were a total of 5400 vessels that operate the whole year. Table 2.1 shows the distribution of these vessels with respect to the size. As we can observe, the majority of vessels is under 11 meter, and are coastal vessels (*Facts about Fisheries and Aquaculture 2013, 2013*).

Table 2.1: Distribution of the length of the active fishing vessels in Norway

0-11 m	11-14,9m	15-20,9m	21-27,9m	28m and over	Total
4157	702	159	135	247	5400

As Chapter 1 states, it is not allowed for land based processing plants to catch fish with own fishing vessels (Tveterås, 2014). This is done to ensure that one party in the chain cannot have the total power.

Each time a fishing vessel comes to deliver fish, it stores data about the date, location, specie and weight of the supplied fish in the database of the Norwegian Fishermen's Sales Organization (NFSO), which is publicly available.

2.2.2 Demand characteristics of finished products

Raw whitefish can be produced into different end products. Besides the customer requirements with respect to shelf life, the quality and age of the fresh fish also determines the kind of end product that can be produced.

The main products can be classified into three different groups: frozen fillets, fresh filets, and whole fish. Figure 2.3 shows the export characteristics of each product group (Henriksen, 2013). We conclude that the whole fish and the fresh fillets are taking a bigger part in the export last years.

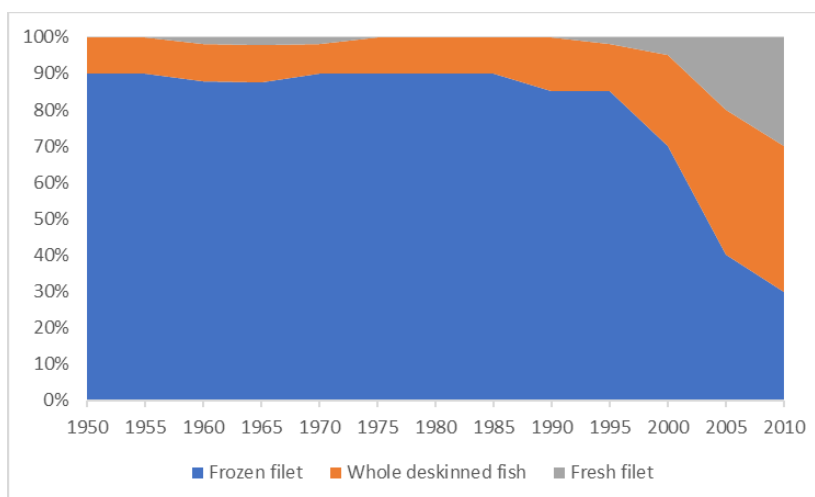


Figure 2.3: Export characteristics from Norwegian fish export

2.3 Analysis of LNS

Now that we explained the global fish market and the supply chain globally, we focus the analysis on LNS internally. This section covers the current situation of LNS, qualitatively and quantitatively. The section starts with Section 2.3.1, which analyzes the supply characteristics and uncertainty in general. Section 2.3.2 focusses on the production matrix and the raw material requirements. Section 2.3.3 continuous with the processing plant in Båtsfjord and explains the production processes. Finally, Section 2.3.4 concludes with the planning processes. We base this section on a plant visit and interview in the processing plant in Båtsfjord.

2.3.1 Purchasing and receiving of raw materials and uncertainty

LNS buys its fish directly from a vessel. As Section 1.2 states, LNS knows approximately three days in advance which vessels come to deliver fish. However, the amount of fish that a vessel on the delivery day is uncertain. Per week, approximately 30 ships deliver fish at each of the processing plants (Pettersen, 2017). Figure 2.4 shows the information and raw material flows between a vessel and a processing plant of LNS. There is no information exchanged about the catch while the vessel is at sea. The process starts with a vessel coming to deliver fish at LNS. The vessel plans its next moment it goes out to sea, which is generally three days later. LNS checks the production schedule to see if it has enough capacity free to receive the shipment. In practice, it seldomly happens that LNS can, or will not accept a shipment, since raw material is scarce.

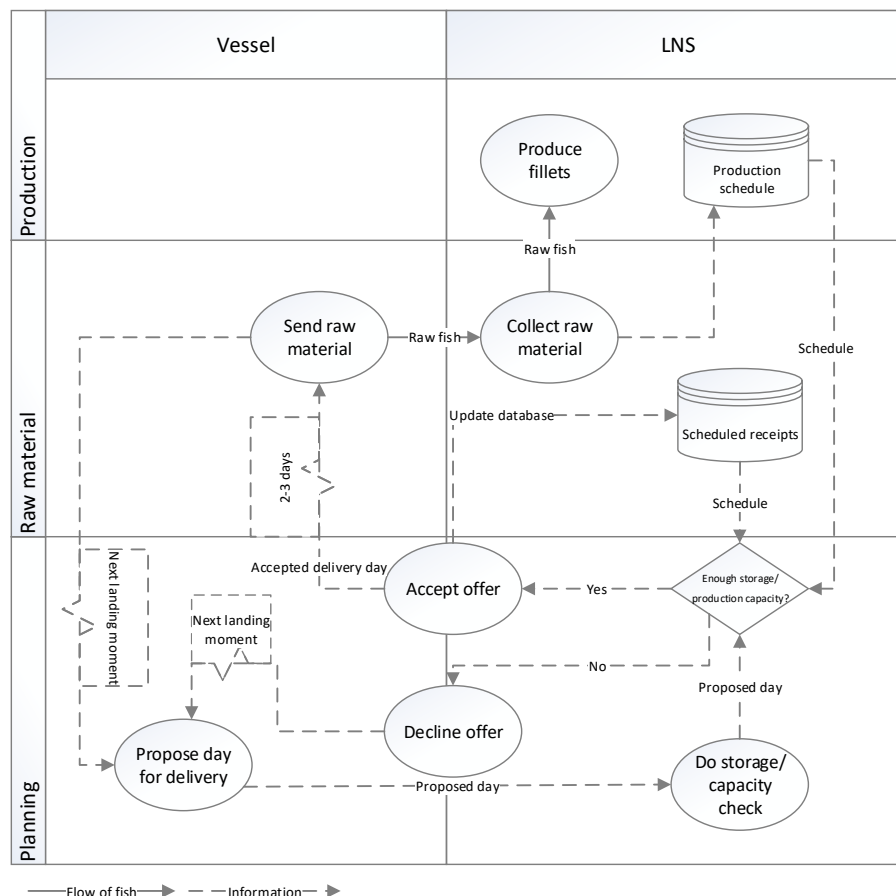


Figure 2.4: Information flow for the receipt of raw material

Fish can be purchased frozen or fresh from the fishing boats. Coastal vessels deliver fresh fish, since this type of vessel does not have freezing capacity. Seagoing vessels deliver slaughtered, frozen fish, since the period these vessels are on sea is too long to deliver the fish fresh. Of all the raw fish purchased by all the facilities, 97% consists of fresh fish and thus 3% is purchased frozen. (LNS (1), 2017). There are two processing plants that process frozen raw material (Melbu and Stamsund). When a fisherman arrives with its shipment to a processing facility LNS checks the fish for damage and weighs each fish. The fish is either delivered with, or without head. The minimum prices for raw fish and the quality control is standardized and set by the government (Rafisklaget, 2017). When fish is delivered without head, the price is converted to fish with head using official conversion ratios (Appendix 1). Since LNS uses fish without head in their production planning, we use this conversation as well. The conversion ratio from fish without head to fish with head is 1.5.

The prices depend, besides the fact if there is damage or not, on the weight of each individual fish. The weight of the fish has a relatively small impact on the price, the difference between the heaviest fish and the lightest fish is approximately 2.5 NOK per kilogram. When the fish is damaged, the prices drops with 4.8 to 5.9 NOK per kilogram.

Table 2.2 shows average weighted price LNS paid to the fisherman of each specie of the three most processed species without head, which are Cod, Haddock, and Pollock. LNS did not buy frozen Pollock in 2017.

Table 2.2: Purchasing prices raw materials

	Cod	Haddock	Pollock
Fresh (NOK per kg)	23.65	14.28	9.14
Frozen (NOK per kg)	29.90	22.22	-
Damaged (NOK per kg)	17.75	9.49	3.91

Since the strategy of LNS is to compete on product quality (see Section 1.2) LNS must process the fresh fish as soon as possible to maintain its freshness and thus be able to sell the end products for a higher price. LNS does not take fish in production which is older than 6 days, counted from the moment it is caught at sea. After this, LNS cannot process the fish anymore, and sells it for a lower price than when it would be taken into production.

In 2017, LNS bought in total 44 million kilograms of raw material for all the processing facilities. 1.45 million comes from seagoing vessels, 37.65 million from coastal vessels and 4.90 million from external suppliers (LNS (2), 2017).

Out of these 44 million kilograms, the processing plant in Båtsfjord bought 16, making it the largest processing plant with respect to raw material purchasing. Figure 2.5 shows the fish species bought by Båtsfjord. Most of this raw material consists of fresh, undamaged Cod (almost 10 million kilogram). The amount of fish that is available fluctuates each month. Figure 2.5 shows the total number of kilogram purchased fish by Båtsfjord. The production plant was 249 days operational in 2017.

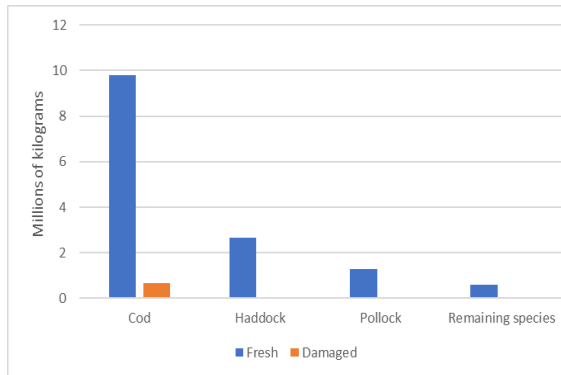


Figure 2.5: Raw material bought by Båtsfjord in 2017

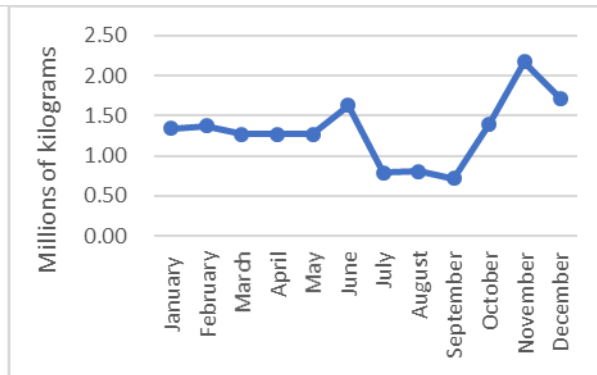


Figure 2.6: Raw material bought per month by Båtsfjord

2.3.2 Production matrix and raw material requirements

After LNS receives the fish, it grades the shipment by quality. The grading process is done visually and by sample, based on the amount of blood spots, the shininess of the skin and the temperature (LNS (3), 2017). Then LNS categorizes the fish into three categories (good, regular, and bad). The quality bad is given when the fish is received damaged (see Section 2.3.1). In 2015, LNS stored the quality of the shipment of 361 vessels the quality of the shipments. Out of the 2597 shipments, 1307 were given the quality good, 1180 the quality regular and 110 shipments had bad quality fish. We need this information in Chapter 5.

Out of the raw material, LNS sells and produces different end products that require each a minimum standard of quality and age of the raw material. Table 2.3 shows the possible products together with the selling prices (LNS (4), 2017). For Pollock, there are two products that have no demand: frozen tail and whole fish older than 5 days. Therefore, the product price of these products is 0.

Table 2.3: Selling prices of the end products

Category	Product number	Product	Selling prices (NOK per kg)		
			Cod	Haddock	Pollock
Processed products	1	Fresh loin	74	64	48
	2	Frozen loin	68	57	47
	3	Frozen portion	62	22	22.50
	4	Fresh tail	54	12	0
	5	Frozen tail	52	10	0
	6	Block	45	19	18.5
Unprocessed fish (whole fish)	7	Good or regular quality fish between 1 and 5 days old	20	12.79	9.75
	8	Bad quality fish between 1 and 5 days old	18	5	4
	9	Whole fish 6 days and older	16	3	0

Table 2.4: Production matrix with respect to age and quality raw material
(L=Fresh Loin, FL=Frozen Loin, FP=Frozen portion, B=Block, T=Tail, FT=Frozen tail, S=Sell without processing)

Quality Age (days)	Good	Regular	Bad
1 until 2	L, FL, FP, B, T, FT	L, FL, FP, B, T, FT	B
3 and 4	FL,FP,B,T, FT	FL,FP,B,T, FT	B
5	FP,B, FT	FP, B, FT	B
6	S	S	S

If the fish is taken into production, the fillets are removed out of the fish first. The weight of the fillet that can be cut out of a fish without head is on average 50.0% for Cod, 60.9% for Haddock and 50.1% for Pollock of the weight of the whole fish (LNS (5), 2016). This is in line with the official Norwegian conversion rates, which convert the weight of fillet products to the original weight of whole fish (see Appendix 1) (Fiskeridir, 2009). Once the fillet is out, it is cut into different end products. Figure 2.7 shows the optimal cutting plan of a fillet, where all parts are used to produce the most expensive product. The percentages show the weight of each part with respect to the total weight of the fillet. These factors are constant, independent of the weight and specie of the fish (LNS (6), 2017).

The quality and the age of the raw fish dictate which parts LNS can use for which product. When the fish gets too old, or the quality is not good enough, LNS cannot produce the highest quality products of Table 2.3. As a result, the cutting plan is different for certain combinations of age and quality of the raw material. The loin can be produced into portion or block. The tail and portion can be used to produce block. Fish older than 5 days is sold without taking it into production. Table 2.4 shows the production matrix of each age/quality combination. For example, when we look at a good quality fish of 5 days old we see that LNS cannot produce loins. Table 2.5 shows which parts of the fish (Figure 2.7) can be used for which products.

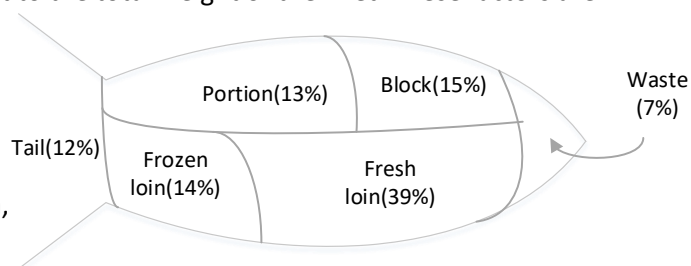


Figure 2.7: Optimal cut of a raw fish

Table 2.5: Possible end products for each part

	Fresh Loin	Frozen loin	Frozen portion	Fresh tail	Frozen tail	Block
Fresh loin	1	1	1	0	0	1
Frozen loin	0	1	1	0	0	1
Frozen portion	0	0	1	0	0	1
Fresh tail	0	0	0	1	1	1
Frozen tail	0	0	0	0	1	1
Block	0	0	0	0	0	1

We give an example to clarify the usage of the raw material: LNS wants to know the optimal cutting plan for regular quality fish of 5 days old. We see in Table 2.4 that LNS can produce Frozen portion, Block and Frozen tail. We thus cannot produce Fresh loin, Frozen loin, and Fresh tail. We search in Table 2.5 the most expensive product that we can produce for this age/quality combination. For both Fresh loin and Frozen loin, this is Frozen portion.

For Fresh tail, this is Frozen tail. This fish then yields a total of 66% Frozen portion, 15% block and 12% frozen tail. Bad quality fish fresher than 6 days always yields 93% block.

2.3.3 Primary production process in the plants, processing times, and capacity

Once the fish is weighted and given a quality grade, LNS decides if and how much fish of each quality and age combination it takes into production that day and how much raw material it sells without processing. If fish is sold without processing, it skips the whole production process, and goes directly to the packing station, where it is packed in a retail pack. We can neglect the production time of whole fish, since the packing process of whole fish is automated and there is sufficient capacity in the plant to pack the whole catch if this is necessary.

When fish is taken into production the fillets must be removed first. We call this process the filleting process. There are 4 lines to fillet the fish. Each individual line consists of two machines: a beheading machine and a filleting machine. There are two lines that are capable of processing fish with a weight of over 2.3 kg. Two employees are needed per production line to feed the fish to the machines. The change over time from one fish specie to another at the filleting line can be neglected, since this time is almost zero.

After the filleting of the fish, the filets need to be cut into the different products of Table 2.3. The cutting of the filets is done manually, and there is place for 26 employees at the cutting station. Since all the products are cut at the same time, the production time at this station is the same for every product. After the cutting the products that must be frozen are taken into a freezing tunnel, where two employees are needed to put the products correctly on the conveyor belt. Next the products are checked if they are free of bones, using an automated X-Ray machine. When there are still bones in the fillet, employees must remove these before the products are checked again. Finally, the products are packed in a retail or wholesaler packing. This requires, in contrast to packing the whole fish, production capacity. Employees must place the products manually in the packages.

Employees are not trained to be multifunctional, they can only perform one task in the factory (Pettersen, 2017). Figure 2.8 visualizes the production process and choices that LNS makes daily, based on the quality of the fish.

The bottleneck in the process is the filleting process. When the machines are set to process big fish (>2.3 kg), it processes 28 fishes per minute and the set up for small fishes has a capacity of 33 fishes per minute. The weight of the fish or the type of specie has no influence on the beheading, filleting or cutting time. Also, the machines can be changed in setup from small to big fish, so each fish can be filleted on its optimal speed. Since each product has a fixed routing within the manufacturing system, this setup can be considered as a product oriented layout (Slack, 2008).

The production times per kilogram fish depend on the weight of the fish. To calculate the production time of the filleting process, we need to know the weight of the average whole fish purchased in Båtsfjord. We give the calculations for Cod. Since some fish is purchased without head, we convert all the purchased fish to whole fish without head, using official conversion ratios (see Appendix 1 for the conversion rates). If the fish is purchased with or without head does not influence the production time since the fish still must be processed on both machines. We need the average weight for fish under 2.3 kilogram and over 2.3 kilogram, since this differentiate on which speed the fish is produced. The average

weight for fish under 2.3 kg is 1.28 kg and for fish over 2.3 kg the average weight is 2.86 kg. 80% of the purchased fish is under 2.3 kilogram.

We now have all the numbers we need to calculate the processing time. We calculate the processing time for small fish and big fish and take the weighted average of both processing times. We determine the processing time using:

$$ptime = \left(P_{w<2.3} * \frac{1}{C_{w<2.3}} * \frac{1}{A_{w<2.3}} + (1 - P_{w<2.3}) * \frac{1}{C_{w\geq 2.3}} * \frac{1}{A_{w\geq 2.3}} \right) * Q$$

Where: $P_{w<2.3}$ = Fraction of purchased fish lighter than 2.3 kg

C_w = Capacity in fishes per minute given that the fish is lighter or heavier than 2.3 kg (2.1a)

A_w = Average weight of a fish given that the fish is lighter or heavier than 2.3 kg

Q = Conversion factor fish without head to fish with head

When we fill in Equation 2.1a with the numbers we calculated in this section we obtain:

$$ptime = \left(0.8 * \frac{1}{33} * \frac{1}{28} + (1 - 0.8) * \frac{1}{28} * \frac{1}{2.3} \right) * 1.5 = 0.220 \text{ minutes per kilogram} \quad (2.1b)$$

Equation 2.1b gives the processing time for a whole fish without the head. To calculate the total capacity of the filleting lines in fillet, given 4 machines and 6.6 operational hours per day, we use:

$$Capacity = \frac{1}{0.0205} * 4 * 60 * 6.6 * 0.50 = 23,956 \text{ kilogram of fillet per day} \quad (2.2)$$

We define 0.50 as the conversion ratio from whole fish without head to fillet. We define 60 as the conversion from minutes to hours and 4 stands for the 4 filleting lines in Båtsfjord.

Next, we determine the capacity at the quality station. We know that on average 85% of all checked fillet parts are free of bones. 15% needs extra operation to make them bone free. Since the X-Ray machine scans the products continuously, there is no maximum capacity given for this work station. Hence, we only consider the production time of removing the bones to determine the capacity of this station, which is 0.0566 minutes per kilogram fillet. There is place for a maximum of 5 employees simultaneously to remove the bones. The capacity is then:

$$capacity \text{ quality station} = \frac{1}{0.0566} * 60 * 6.6 * 5 = 34,982.33 \text{ kilogram of fillet per day} \quad (2.3)$$

Table 2.6: Production times and capacity for Cod

Work station number	Work station	Number of parallel lines or places	Employees needed per line or place	Minute per kilogram Cod per line or place	Capacity in kg filet Cod per day with 6.6 hours daily
1	Filleting	4	2	0.0220	23,951
2	Cutting	26	1	0.350	24,882
3	Freezing	Continuous flow	2	0.0105	>filleting
4	Quality check	Continuous flow	1	0.0566	>filleting
5	Packing	2	2	0.0298	53,154

The production times of the other work stations are given by LNS. Table 2.6 summarizes all the information and we conclude that the filleting line is the bottleneck of the processing plant (LNS (7), 2018).

Next, we need to determine the production time for Haddock and Pollock. We calculate the average weight of a Haddock fish purchased in Båtsfjord to be 0.92 kg and the average weight of a Pollock fish to be 0.90 kg. Haddock and Pollock can be produced on both machine setups. We calculate the production time of the filleting line for the remaining two species in the same way as for Cod. Table 2.7 summarizes the production times of the remaining two species

Table 2.7: Production times for Haddock and Pollock

Work station number	Work station	Minute per kilogram Haddock per line or place	Minute per kilogram Pollock per line or place
1	Filleting	0.0329	0.0334
2	Cutting	0.406	0.502
3	Freezing	0.0216	0.0221
4	Quality check	0.1167	0.1192
5	Packing	0.0614	0.0628

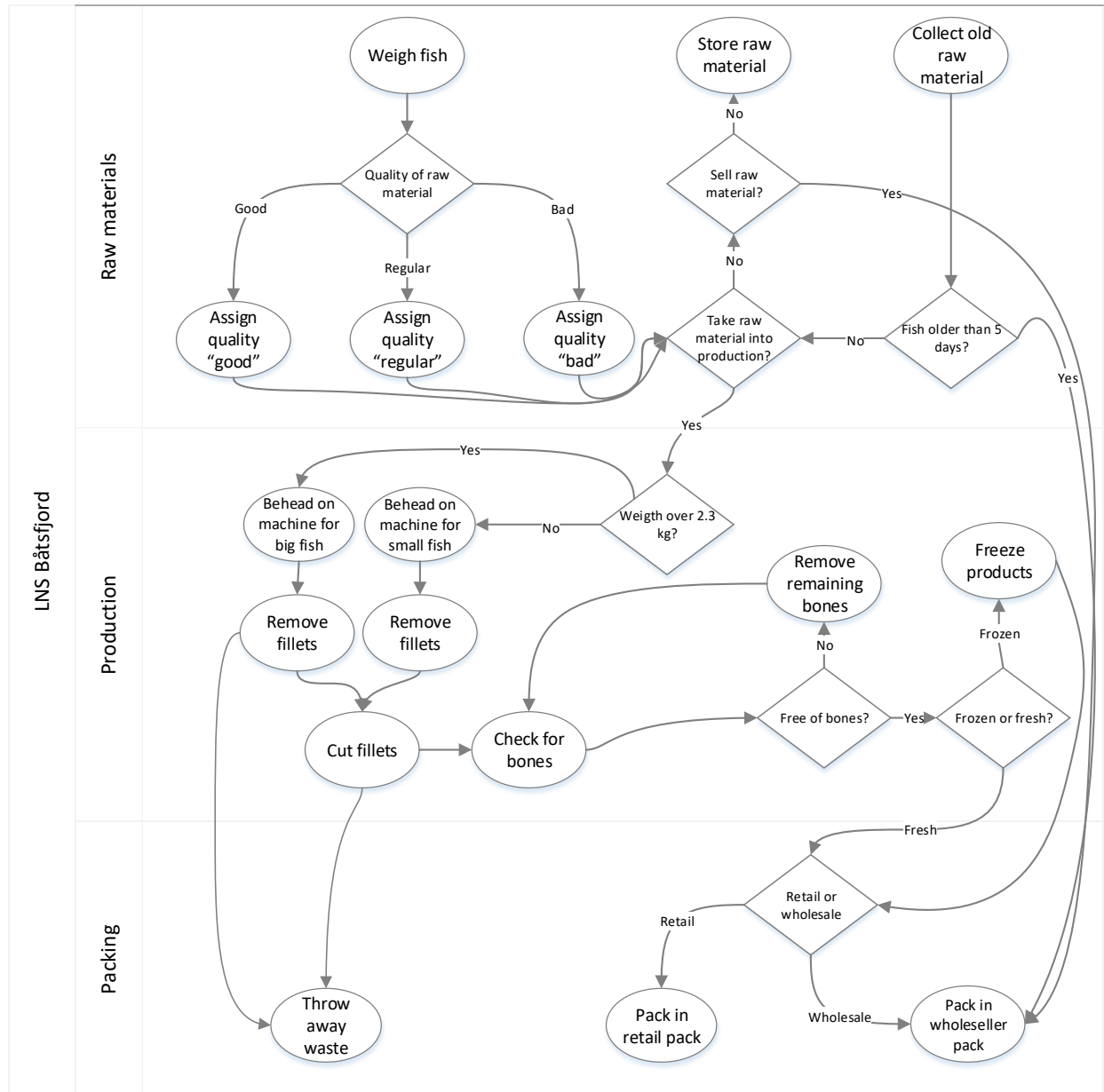


Figure 2.8: Primary production processes

2.3.4 Planning processes and productivity

Since both the quality and the quantity of the supply of raw fish is uncertain, the capacity planning in the plants is made under uncertainty. The maximum age of the raw material that can be used for processing is 6 days and the age of the raw material determines, together with the quality, which products LNS can produce. Because of this the capacity planning for today influences the amount of raw materials in coming days and the value of this raw material. After 6 days, LNS sells the raw material against a lower price than when it would take it into production and incurs a loss.

The planning structure is hybrid, where each processing facility receives a global number of hours it can use each day. After this the local plant manager makes a detailed planning. The global planning, or Rough-Cut Capacity Planning (RCCP) is made by a central planner. The central planner uses information

about the current raw material inventory and amount of shipments each facility receives to determine the number of hours each facility can use on each workstation. Currently the central planner does not make the RCCP in any systematic way or uses an algorithm, but the central planner uses his experience. Since the production process has a production-oriented layout, the products incur in general the same production steps. This leads to a low variability in the production times, making it easy to translate the RCCP to a detailed planning. Hence, we conclude that the most difficult aspect LNS faces is to make the RCCP for the processing plants.

LNS has an agreement with its workers that they always must know their work schedule for three days in advance. For example, at the beginning of Monday, the work schedule must be made for Monday until Wednesday. LNS has flexible workers, which means there is no minimum number of hours a worker must be planned in.

Figure 2.9 shows the information flow between the processing plants and the central planner, where processing plant X stands for a random plant within LNS. The local plant manager updates the central database of the plant with the production plan and the inventory planning of the raw material. He then transmits this information through a conference call with the central planner after which the central planner makes the RCCP for the plant. The central planner sends the RCCP to the central database of the local plant. The RCCP consists of the number of hours that a local plant can use on each workstation, in the coming three days. The local plant manager transforms this RCCP to a detailed production plan and plans in the employees the coming three days.

The number of hours used by each processing facility is recorded. This is split out in total hours and cutting hours. The total hours include overhead hours for management and the other working stations. On average 38% of the total hours consisted of cutting hours (LNS (8), 2016).

To analyze if the current usage of hours is efficient, we plot the processed raw material of a certain day against the used capacity of this day. Since the total hours have a fixed character for the overhead, we use the cutting hours as planned capacity indicator. Figure 2.10 shows the result. The trendline shows the expected number of cutting hours, given an amount of processed raw material. Since the processing times have little variation due to the production-oriented layout, we expect little variability in the plotted values, given that the planned capacity accurately matches the needed capacity. We see in Figure 2.10 this is not the case, since the data points show big fluctuations compared with the trendline.

Table 2.8 shows the summary statistics of the productivity of the plant (kilogram raw material processed per cutting hour). We observe that there are indeed big fluctuations, given the large range of the confidence interval.

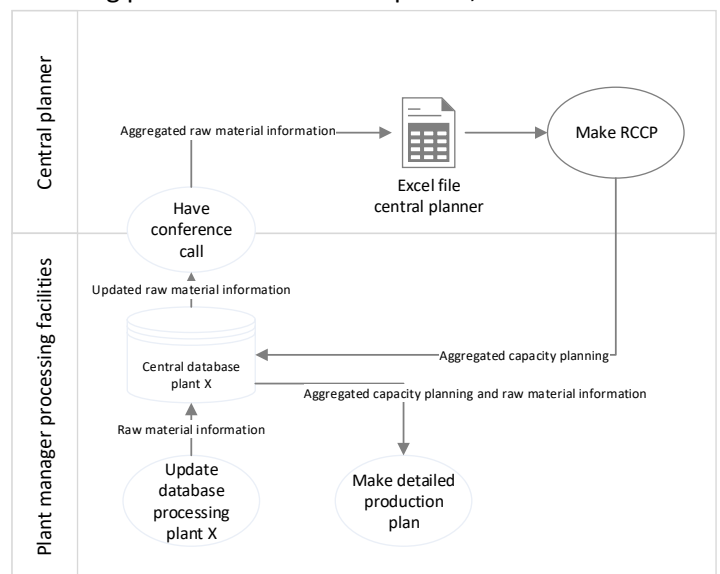


Figure 2.9: Information flow between processing facility and central planner

Table 2.8: Summary statistics productivity of cutting department

Variable	N	Mean kilograms per cutting hour	SE Mean	StDev	Confidence interval 95%	
					Lower bound	Upper bound
Productivity	402	180.87	2.33	46.73	86.65	303.84

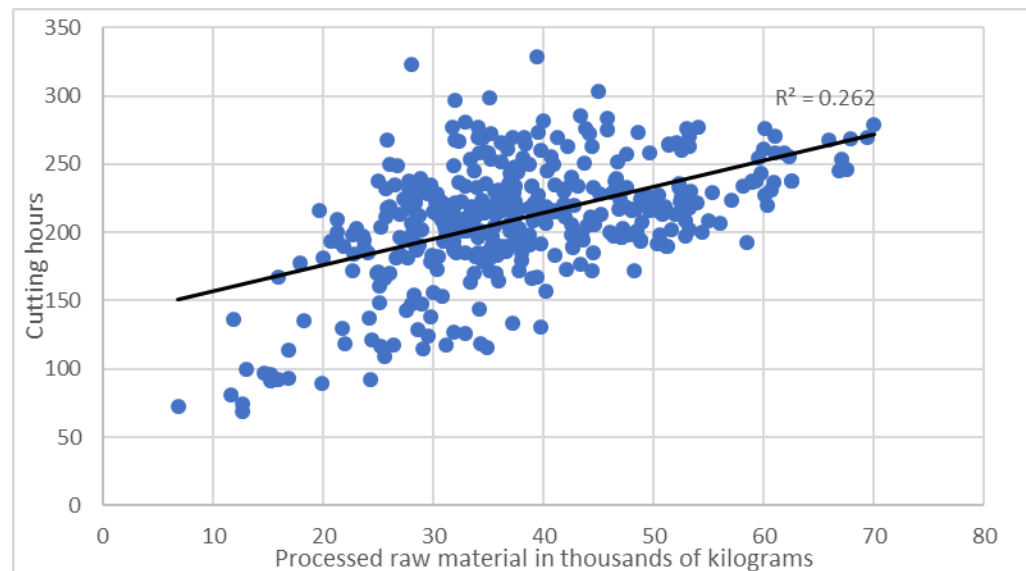


Figure 2.10: Plotted processed raw material against cutting hours in Båtsfjord

2.4 Conclusions about the processes within LNS

Fish is an important product for the Norwegian industry, considering that 8.6% of the total export consists of fish. Fresh products have more value than unprocessed or frozen products. This makes it important for LNS to produce the fish as soon as possible. As a result, we conclude that matching the supply and demand from capacity is important for LNS.

LNS only has information about when a specific fishing vessel comes to deliver raw fish. Besides this there is no information available in advance about the amount and quality of raw material. Since the capacity planning is based on these facts, LNS has difficulties planning the daily capacity in the plants.

The quality and age of the raw material determine which products can be produced with the raw material. Although it is theoretically possible to produce block products from parts that can be produced in more expensive products, LNS always produces the most expensive product possible. Hence, the production matrix is constant for each individual age/quality combination of the raw material given there is enough freezing capacity available.

When there is insufficient capacity to produce all the raw material, the remaining inventory becomes one day older and reduces in value since less expensive products can be produced. When there is too much capacity planned, money is wasted on excess salaries for the employees. Hence, the profit that LNS generates out of the raw material is a measure for the quality of the capacity planning.

The production process is relatively simple, since the bottleneck of the process is the beheading machine that has a deterministic processing time. Since the capacity of the fillet products in the production plant is thus limited to the beheading machine we can conclude that we can plan the capacity based on this machine. The required capacity of the other processing steps (cutting, freezing, quality check, and packing) depend on the planned capacity of the beheading machine. As the production has a production oriented lay-out and the fact that the bottleneck has a deterministic processing time, the employee's productivity, given that there is available work, is relatively constant. Therefore, we expect the processed kilograms of fillet per cutting hour to be constant when the planning is accurate. Hence, the differences in the number of kilograms processed raw material per cutting hour (Section 2.3.3) each day must be caused by planning errors. The difference between the highest observed productivity (304 kilogram per cutting hour) and the lowest (86), is more than factor 3.5. Therefore, we conclude that the current RCCP is not optimal and leads to overcapacity, reducing the productivity of the processing plant.

3 Theoretical background

In this chapter we analyze the literature and find the necessary literature to use during the research. This chapter consists of 2 sections. Section 3.1 focusses on capacity planning with uncertainty, gathering the knowledge to answer research question 3. Section 3.2 gathers the theory about the reduction of uncertainty within the supply. This sections answers research question 2.

3.1 Optimization of capacity planning

Since the problem LNS faces copes with an optimization problem with uncertain supply, we need theory on how the literature solves these problems. To understand the problem Section 3.2.1 first focusses on the capacity planning in the deterministic case. Section 3.2.2 continues with the stochastic case.

3.1.1 Capacity planning in the deterministic case

Smunt (1996) describes the process of making a RCCP in a deterministic situation on a high aggregation level. He argues that the capacity planning should be based on the Master Production Schedule (MPS), which consists of demand forecasts, desired inventory levels, the production plan, constraints and customer orders. Figure 3.1 shows the process according to Smunt (1996). The goal of the RCCP is to transform the MPS into capacity needs for the most important work stations (WS), for example the bottlenecks of the production process.

The actual determination of capacity is not given by Smunt (1996). Figure 3.1 only states that an appropriate technique should be used. Guide JR. et al. (1997) investigate several techniques. They base their research on a remanufacturing environment. The research distinguishes three techniques to determine the capacity requirements:

1. Bill of Resources (BoR):

The BoR is the simplest technique for capacity planning and can be applied when all products go through the same work stations.

$$C_{ik} = \sum_{j=1}^r [M_{jk} L_{ij}]$$

Where: C_{ik} = Capacity required at WS i in time period k (3.1)

M_{jk} = MPS quantity product j in time period k

L_{ij} = Production time product j in WS i

2. Modified Bill of Resource (MBoR):

This method is comparable with the BoR but adds the probability that a product needs operation in a work station, making it suitable for job shops. Also, the MPS is adjusted by multiplying the original MPS quantity with the Material Recoverability Rate (MRR). The MRR gives the percentage of total incoming raw material that is qualified for processing.

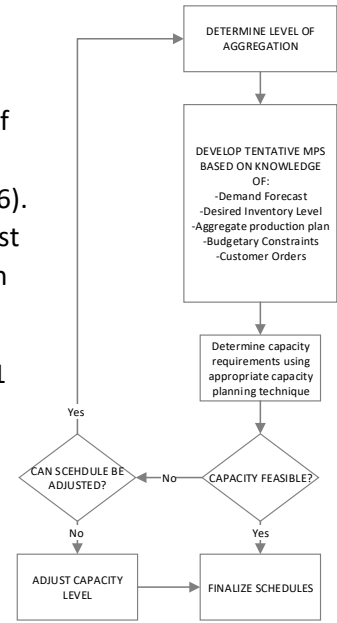


Figure 3.1: Process of creating the RCCP according to Smunt (1996)

$$C_{ik} = \sum_{j=1}^r [M_{jk}^* L_{ij} O_{ij}]$$

Where: M_{jk}^* = adjusted MPS quantity (3.2)

O_{ij} = Occurance factor product j in WS i

3. Capacity Planning using Overall Factors (CPuOF):

This method takes the utilization into consideration. To do so the utilization of each work station should be estimated, based on historical data. The capacity of each work station is then calculated with:

$$C_{ik} = \left[\sum_{j=1}^r M_{jk} S_{ij} \right] \frac{1}{P_i}$$

Where: S_{ij} = standard time required for product j in work station i (3.3)

P_i = Utilization rate of work centre i $0 < P_i \leq 1$

In all proposed methods for capacity planning the MPS is needed. Hence, we need a method to determine which products to produce on a given day and make a RCCP based on this. Begen et al. (2003) do research for a processing plant for red fish in Canada. They develop a mathematical model to support the short time (operational) production planning for the processing plants once the catch is known. This model takes the quality of the inventory in consideration. Figure 3.2 shows the quality deterioration model.

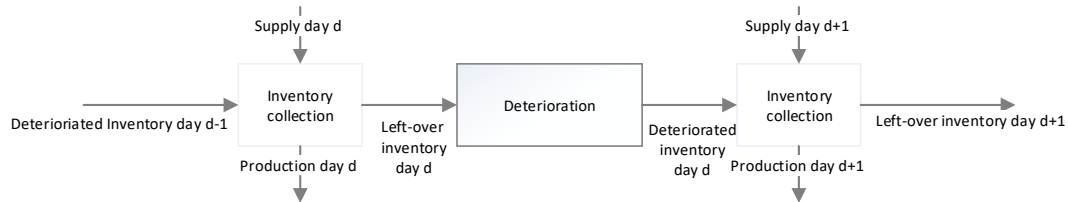


Figure 3.2: Deterioration process of raw material

The mathematical model that Bergen et al. (2003) create consists of a Linear Programming model which goal function optimizes the total profit of the processing plant. Bergen et al. (2003) assume that the product mix (which quality and which specie is needed for each end product) and the profits of each end product are known and deterministic. Further they include the recovery rates for each specie. The recovery rate is a number in the interval [0,1] that gives the fraction of the weight of the fish that can be used to produce a certain end product. For example, if a cod of 5 kg yields a loin filet of 1 kg, the recovery rate for this product is 0.2.

The LP uses an objective function, which maximizes the total profit. The LP maximizes the profit by taking the sum over the produced raw materials of all different quality and species. The goal function multiplies all quality/specie combinations with the profit coefficient and recovery rate to obtain the total profit.

We omit the 12 constraints of the LP program. They ensure, together with the decision variables, that there is sufficient capacity, raw material and that the product matrix is respected. The LP does not take customer demand characteristics into consideration, as it assumes there is always sufficient demand for each product. Also, the labor costs are not specifically included in the goal function but are taken into consideration in the profit coefficient. This means that the marginal labor costs per end product have to be known. The objective function maximizes the weighted average of the profit and the recovery rate of each end product.

Jensson (1988) also proposes a planning method for daily fish production in a processing firm. He investigates the case when a processing firm produces frozen end products and thus takes the freezing capacity into consideration. He uses a 5-day planning horizon. Other differences with the model of Begen et al. (2003) is that the daily production is bounded by an interval. The lower bound gives the accumulated quantity of fish, from day 1 to T, that must be processed, or else the fish will be damaged. The upper bound is the accumulated received quantity of fish from day 1 to T. Hence, in this model all the raw material has to be processed. Also does the model include an upper and lower bound for each finished product if the product is bounded by market constraints. The goal function of the model of Jensson (1988) is different when we compare it with the model of Bergen et al. (2003). The first model has the capability to vary the personnel costs with the time as well as with the department. Again, we omit the constraints, which ensure that the freezing capacity is respected, all the raw material is being processed and that there is enough capacity on all departments.

3.1.2 Capacity planning in the uncertain case

To position the problem which LNS faces we use the research from Peidro et al. (1998). They classify the source of production uncertainty in three types: demand-, process- and supply uncertainty. Especially the third uncertainty is important for LNS, since the supply of fish is uncertain. Next, there are three types of problem: strategic, tactical and operational problems. We conclude that LNS copes with operational problems which are caused by supply uncertainty. In Figure 3.3 we show the framework of the problems with respect to the two classification methods. We place LNS in the combination of operational problem type with supply uncertainty, which is the under right quadrant.

Problem type	Strategic	Find optimal location for distribution centers	Optimize location, inventory, and allocation decision under random parameters	Find optimal location for processing plants under uncertain supply
	Tactical	Determine optimal reorder point in all echelons in supply chain	Determine optimal replenishment policy with uncertain factors such as breakdowns	Develop a control mechanism for the allocation of orders with uncertain suppliers
	Operational	Optimize inventory levels in retail shops	Optimal planning in a SC considering uncertain factors such as breakdowns	Plan operational capacity in the future with uncertain supply
		Demand	Process	Supply
		Source of uncertainty		

Figure 3.3: Matrix of problem combinations according to Peidro et al.

When looking at a similar problem formulation compared to LNS, one can look at an emergency department in a hospital. Here the capacity planning also must be made under uncertain supply. For example, Rosetti et al (1999), solve this kind of problem using discrete event simulation. By analyzing data of 17 weeks they find an arrival distribution for the patients for each hour, which is used as input for the simulation model. Next to this they analyze the transport routes within the facilities since patients can move between various departments within the hospital. Using this approach, they manage to improve the current situation significantly, reducing the waiting time of patients.

Begen et al. (2003) use stochastic programming to solve a case of uncertain supply of red fish at a Canadian fish processor. They use the model that we explained in Section 3.1.1 and adapt it to cope with uncertain supply. To cope with the uncertainty, they propose to use a forecasting technique to forecast the catch size each day, and define possible outcomes each day. Each possible outcome, defined as a catch size, should then be given a probability.

Next, the stochastic version of the model sums over all the different possible scenarios multiplied with their probability. The model maximizes the expected value of the profit based on the probabilities of the scenarios. The other decision variables and the restrictions assure again that the capacity restrictions are respected and that it is not possible to produce more than there is raw material available.

Kazemi Zanjani (2008) present a stochastic model of a wood producer, who has to deal with uncertainty in the dimensions of wood logs he gets supplied as well as uncertainty in demand for finished products. This situation can be compared with the problem LNS faces, since the quality of raw fish that LNS receives is also uncertain. Kazemi Zanjani (2008) proposes a multi-stage stochastic programming (MSP) model to solve this multi-period, multi-product (MPMP) producing planning problem.

In this model all the different possible scenarios of demand are given a probability. $P(n)$ is then the probability of scenario demand scenario n . The decision variable gives the number of times a process is run. Since the supply of logs is stochastic, the yield of finished products by running a scenario is uncertain. The goal function first calculates the raw material costs over all different end products. The second part of the equation calculates the holding costs of the inventory and the backorder costs. Then the goal function minimizes the expected raw material, inventory and backorder costs over all possible scenarios.

We observe that there are different approaches in solving the planning problems with uncertainty in the literature. Allen et al. (2015) describe that the complexity of the problems determines the best solving approach. They argue that complex problems with many different decision to make can best be solved using discrete event simulation. When problems are less complex, analytical models provide better solutions. To classify the complexity of a problem, Ullman (2007) classifies the algorithms that are constructed to solve a problem into the maximum running time that is needed to solve the problem to optimality. This is known as the “Big-O” notation. When a problem can be solved to optimality in polynomial time ($O(n^x)$, $x \in \mathbb{N}$), an analytical model has the preference. When a problem cannot be solved in polynomial time we call this problem NP-hard, resulting in a longer running time of the algorithm compared with algorithms with polynomial time. When the problem becomes too big, the running time explodes, and one must use heuristics or simulation.

3.2 Reduction of uncertainty of supply

There are several ways listed in the literature to reduce the uncertainty of supply in a manufacturing system. Vorst et al. (1998) investigate the uncertainty in a food supply chain. They use a supply chain of chilled salads with a producer, one central distribution center, and 100 retail outlets. Since a food supply chain also has to cope with the problems that come with perishable goods, we can compare this situation can with the situation of LNS. In this reasearch the uncertainty is split out in different sources, which each require a different approach to cope with this type of uncertainty. The source of the uncertainty is split up in three types: order forecast horizon, input data and decision process. The order forecast horizon refers to the uncertainty in the time between placing and receiving an order. Input data refers to the uncertainty of available data that is available to base a decision upon. The decision process copes with the uncertainty with respect to human behavior

Table 3.1 shows the three sources together with general solutions. The order forecast horizon problems can be solved by using Electronic Data Interchange (EDI) to acquire more accurate information about the lead times. Next decision support systems can be used, such as Computer Assisted Ordering in the retail shops or automated production planning systems can be used.

For the input data general solutions are to share more information in the supply chain, or to create new information by combining different data sources to gain new insights.

The solutions for the decision process problems are based on changing the processes to create standardization, or to eliminate human interference to lower the chances of human bias.

Hameri et al. (2010) investigate the supply chain of fish in Iceland, where they deal with the questions how to match the fluctuations in demand with the supply of raw materials and how to increase the speed of the material flow in the whole supply chain. Also they observe in this research that the supply of fish is subject to fluctuations due to natural causes. There are several solutions proposed to solve this problem:

1. Share more information in the first two echelons in the supply chain. That means increase the information flow between the supplier of raw material and the first processor.
2. Start producing fish at fish farms. By producing fish in a controlled manner, fish can be supplied with almost no uncertainty. There are a few downsides to this: it requires big investments to feed the fish and diseases are common within the fish farms, reducing the quality of the fish.

Since it could also be possible for LNS to create new data, we gather theory for this approach.

Table 3.1: Types of uncertainty with solutions

Type of uncertainty	General solutions
Order forecast horizon	Use Electronic Data Interchange
	Use decision support systems
	Decrease process throughput time
Input data	Create new information flows in different stages in the supply chain
	Use real-time management systems
	Create new information
	Use information systems to exchange information
Decision process	Eliminate decisions or redesign procedure
	Co-ordinate procedures in supply chain
	Eliminate or reduce human influences

To use the stochastic programming models explained in Section 3.1.2, we need forecasts for the raw material supply. Therefore, it is necessary to predict the fish catch. Begen et al. (2003) propose the use of forecasting to predict the catch size and thus the raw material supply. Stergiou et al. (1997) distinguish two categories of forecasting methods:

1. Time series methods: These methods use historical observations to predict future outcomes. Examples of time series methods are the moving average method, exponential smoothing and trend estimation
2. Causal methods: These methods use underlying factors that explain the outcome of a variable (the response variable) with the help of one or more factor(s) (predictors). Regression analysis is an example of a causal method.

Next to these two methods, Witten et al. (2005) propose a third relevant category:

3. Machine learning methods: Machine learning consists of making categorizing or predictions based on an algorithm and training data. Popular machine learning methods are decision and categorization trees, which make a prediction based on a set of rules. These classification rules are based on the training data.

Bergen et al. (2008) try to forecast the catch using time series models such as simple exponential smoothing, moving average models and mean models, but none of these gave significant results. They propose the use of causal models.

Wikstrom (2015) researches how to predict fishing success on recreational level. In this report, he gathers weather variables from public databases and puts them in a model, together with the response variable. The response variable in this case is the weight and the amount of fish caught by recreational fishers. Since coastal vessels also use lines to catch fish and LNS gather its fish mostly from line fishing vessels, this method is potentially interesting for LNS. Table 3.2 summarizes the variables.

Wikstrom (2005) uses a linear regression model and two machine learning techniques. To rate the models by their accuracy he uses the relative absolute error (RAE), which is calculated using:

$$E(a) = \frac{\sum_{i=1}^n |y'_i - y_i|}{n}$$

Where:

a = Expected accuracy of prediction model (3.4)

y'_i = prediction value observation i

y_i = observed value observation i

n = number of observations in dataset

Table 3.2: Variables with the used units

Variable	Unit
Weight	Kg
Cloud coverage	%
Humidity	%
Precipitation	Mm
Air temperature	°C
Water temperature	°C
Visibility	km
Wind direction	°
Wind speed	Km/h
Moon phase	-
Day of year	days

Wikstrom (2005) uses, next to the regression model, machine learning techniques. In machine learning a prediction variable is known as an *attribute*. Wikstrom (2005) uses the following techniques:

1. **Multilayer perceptron:** A multilayer perceptron is an artificial neural network, which is commonly used for the prediction of values with a non-linear classification. It consists of nodes (neurons) that are connected to each other. Input nodes are used to input an attribute from the data set, after which this value flows through the network using connection equations between the nodes. Finally, an output node gives the predicted value.
2. **Random forest:** The random forest uses prediction trees to predict a value. These trees are generated based on training data and predict a value with the attributes of the dataset based on different choices in the tree. From the training data, randomly n attributes are drawn from which a prediction tree is made. This is done x times to make x random trees. Then the final prediction is the average of all x predictions of all the trees. An example of a random forest with two random trees is given in Figure 3.4. The values are fictional.

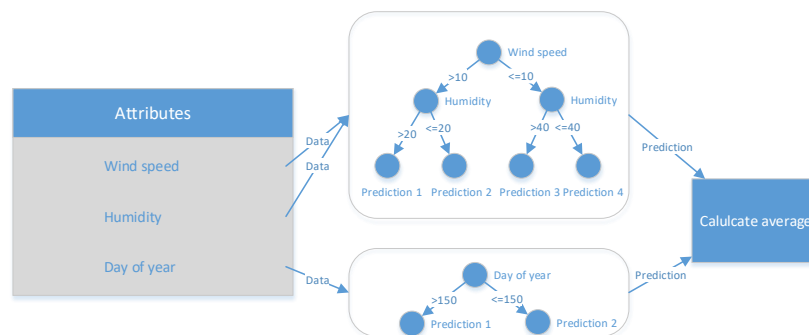


Figure 3.4: Schematic visualization of the random forest algorithm

The random forest performs well on large datasets and requires little computational time. Since it relies on more than one prediction tree it has the potential to have a low absolute error. The disadvantage is that the random forest has a high risk of overfitting the training data, which means the algorithm performs well on the training data but is inaccurate on new data.

3. **Random tree:** The random tree is comparable with the random forest, but instead of taking the average of multiple trees, only one tree is constructed, and the prediction of this tree is used. The random tree is easy to interpret, and the outcomes are easy to implement in models. Disadvantages is the fact that the prediction is based on one tree. Therefore, it is on average less accurate than the prediction of the random forest.

Wikstrom (2005) evaluates the three methods based on the prediction error. He puts the linear regression model as base model and finds that the multilayer perceptron algorithm performs worst, increasing the RAE with 9%. The random forest decreases the RAE with 10% and performs best.

According to Witten (2005) there are two other effective methods that can be used to predict numerical data:

1. **The M5P Tree:** This algorithm generates a decision tree just like the random tree. The difference between the two algorithms is that the M5P Tree predicts values in the leaves with a regression

model, where the random tree takes the average of all the values reaching a leaf. The M5P Tree performs well in situations where there is a linear dependency between the response variable and the predictors.

2. The Reduced Error Pruning (REP) Tree: The REP Tree also starts building a full decision tree based on all available attributes. This in contrast with the random tree, where attributes are selected randomly. The initial tree can be, based on the number of attributes, relatively big compared with the other algorithms. After constructing the full tree, the algorithm examines all nodes and checks the total prediction error in the whole tree if this node is *pruned*. Pruning means that all nodes and leaves under the selected node are removed. The tree is pruned, using a greedy heuristic, on the node where the forecasting error reduces the most. This is done until the forecasting error does not reduce any further. The REP Tree has the advantage that it is a strong algorithm when attributes combined explain the response variable, but each of the individual attributes not.

Witten (2005) states that there is not one model which outperforms the others in general, this depends on the kind of data and the attributes. We present the summary of all the methods together with the preferred data dependency in Figure 3.5.

Category	Method	Preferred dependency
Time series methods	Moving average	Dependency between the observations
	Exponential smoothing	
	Trend estimation	
Causal methods	Regression analysis	Dependent on external factors
Machine learning methods	M5P Tree	Lineary dependent on external factors
	Random Tree	No specific requirements
	Random forest	
	Multilayer perceptron	Non-linear dependent on external factors
	Multilayer perceptron	Combinations of attributes are significant

Figure 3.5: Categorization of the forecasting methods

4 Reducing uncertainty of the supply to LNS

In Chapter 3 we observe that we need a method to estimate the supply to optimize the capacity planning. Hence, the first part of improving the planning of LNS is to find a proper method to reduce the uncertainty of the supply of raw fish to the processing plants. This chapter starts with Section 4.1, where we define the alternative methods and requirements of LNS and choose the best method to reduce the supply uncertainty. Section 4.2 gives the characteristics of the vessel fleet of the Norwegian fishing vessels and we choose a proper sample to research based on the fleet characteristics. Section 4.3 explains the data that we use to reduce the supply uncertainty. Sections 4.4 to 4.7 develop the best method that we chose in Section 4.1. Section 4.8 finishes with the conclusions about this chapter. This section answers research question 3.

4.1 Choosing best method to reduce uncertainty

To analyze what the best method for LNS is, we need to define when we consider a method to be effective. To do so, we need the requirements for LNS. The only certain information LNS currently has is the number of vessels coming to deliver raw material. This information becomes known approximately 3 days in advance. Section 4.1.1 starts by defining the requirements of LNS, so we can determine when a method is effective. This section then obtains three different possibilities to reduce the uncertainty based on the literature research of Chapter 3. Section 4.1.2 compares the defined methods and choose the best one for LNS based on the requirements.

4.1.1 Obtaining methods for reducing uncertainty

First, we need to define how to measure the effectiveness of a method. We choose the measurements in consultation with LNS, where we agreed on three dimensions that measure the effectiveness. These three dimensions comply with what is mentioned in the literature. The dimensions are:

1. **Earliness:** The earlier LNS receives the information needed to make the production planning, the more this information is worth. This is because LNS then can plan employees in an early stage, and does not need to invest more in flexibility, such as flexible employees.
2. **Accuracy:** It is obvious that data that reflects the reality better is worth more. When LNS obtains accurate data about the supply, the models mentioned in Section 3.1.2 provide more accurate calculations and the capacity planning reflects the reality better.
3. **Financial impact:** The financial impacts of acquiring and using the data or method to reduce the uncertainty are also important. We define the financial impact as the total of the investment costs and the potential to save costs for LNS. As mentioned in Section 1.1, LNS makes financial losses and can therefore not invest a large capital.

Section 3.2 gives methods to reduce the uncertainty. We test these methods on the three dimensions and compared them with each other. Then we select and work the best method further out. We do not take the methods that are not feasible due to external constraints, such as catching fish with own vessels, into consideration.

We base the possible methods on the literature research (Section 3.2). We can derive three methods out of the literature (a, b and c), which we base on two general solutions (1 and 2). We derive these general solutions from Table 3.1 of Section 3.2.

1. Increase information sharing between echelons in the supply chain:
 - a. Create a real-time information flow between the fishing vessels and the processing plants, where information is exchanged about the catch size and the quality of the fish.
 - b. Buy fish from fish farms. When LNS buys fish from fish farms, it can make contracts about what to receive daily. Fish farmers can accurately predict when fish is ready for production (Hameri, 2010). In this way LNS has total control over the information that is needed to support the capacity planning. This method is tested in an Icelandic whitefish supply chain (Hameri, 2010). In 2015 Norway produced 1.4 million tons of fish on fish farms (Statistisk sentralbyrå, 2016). Hence, there is sufficient production to buy all the fish LNS needs.
2. Create new information:
 - c. LNS can use the information they have available (e.g. about the fishing vessel) to make an estimation about the delivery it can expect. This method has already been executed on recreational level (Wikstöm, 2015).

4.1.2 Ranking the methods and choosing the best fit for LNS

We need to analyze the presented methods based on their advantages and disadvantages and give a score on each dimension so that we can make an objective comparison, and we can choose the best method.

1. Earliness of the information:

The method to buy fish from fish farms (method b) scores best on this dimension, since the fish can be monitored constantly, and it can be accurately predicted when a fish is ready for processing (Vorst, 1998). The estimation based on known information (method c) scores second on this dimension, since it is known three days in advance which vessel will deliver the fish. The information exchange between supplier and processor (method a) scores worst, since the information is known the moment the vessel will catch the fish. Since LNS obtains most of its raw fish from coastal vessel, which are only fishing for one day, the information is known at its most one day in advance.

2. Accuracy of the information:

On this dimension, method b scores the best since the information about the supply can be most accurately predicted and thus the fish farmers can supply exactly the amount LNS requests. Second is method a since the information LNS obtains is based on the real catch. Method c scores last on this aspect because the information is based on an estimate and can have deviations from the real supply.

3. Financial impact of the information

Method c scores the best on financial impact. The information can be acquired internally, and the method can be developed without external expertise. Method a scores second. Some meetings with the fisherman are required and it could be necessary to invest in information systems. Method b scores worst. This is because operation and feeding the fish at the farms is expensive and labor intensive (Hameri, 2010). Next to this, diseases are common at fish farms and the quality of the fish is worse than caught fish at sea, resulting in inferior product quality, reducing the possible profits. This would interfere with the strategy of LNS to compete on product quality (Section 1.1).

Figure 4.1 visualizes the rankings on each dimension. Closer to the origin means an idea performs worse on this dimension. After discussing the ideas with LNS, they indicated that the costs are an important aspect, since the company has been making financial losses over the last years. Hence, LNS does not have the resources and permission from the stakeholders to make big investments. Since method c does not require any significant investment, we use this idea to reduce the supply uncertainty.

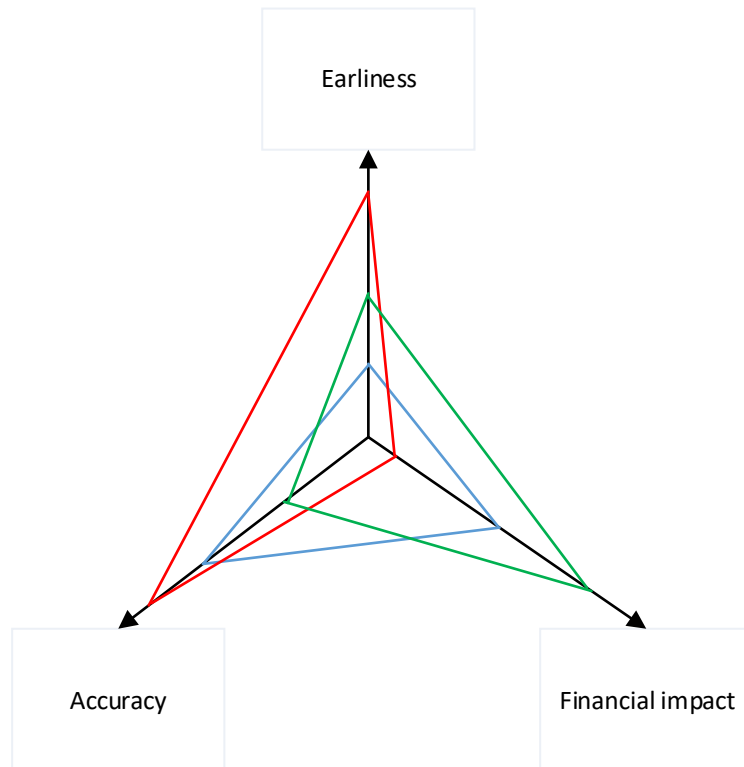


Figure 4.1: Score of each method visualized (blue= idea a, red=idea b, green=idea c)

As Chapter 3 shows, it is possible to forecast the fish catch of each individual vessel, based on the vessel information and weather characteristics, such as wind speed and temperature (Begen, 2003) (Wikstöm, 2015). Since it has been tried already to use time dependent forecasting techniques, such as exponential smoothing and moving average, we use causal models and machine learning models.

4.2 Data explanation and sampling

A causal model must explain the data with external variables (see Chapter 3). Therefore, we need to choose a sample on which we base our forecasting model. We try to include variables that are based on weather data because research suggest that it is possible that weather variables explain the catch size (Stergiou, 1997) (Wikstöm, 2015). Since the weather varies over the location, we need to select a region where we base the model on.

This section chooses this region and explain the data that we use to build the models. In Section 4.2.1 we explain the facts about the statistics of each region in Norway after which we choose the region and sample in Section 4.2.2.

4.2.1 Data explanation

To forecast the amount of fish caught by a given vessel on a given day we first need a response variable. Since LNS always has to buy the complete shipment of a fisherman, the forecast model should predict the total catch of fish per vessel, per shipment. The Norwegian Fishermen's Sales Organization (NFSO) stores the weight of all individual shipments per specie in a public database. The data is split out per region of Norway (see Figure 4.2) The database has data available from 01-01-2017 until 31-12-2017. This is important as there is a main fishing season, which starts in January and lasts until July (see Section 1.1). Therefore, we need to include a whole year of data to be able to analyze if the season influences the catch. There is no information about the quality of the caught fish stored in the public database.

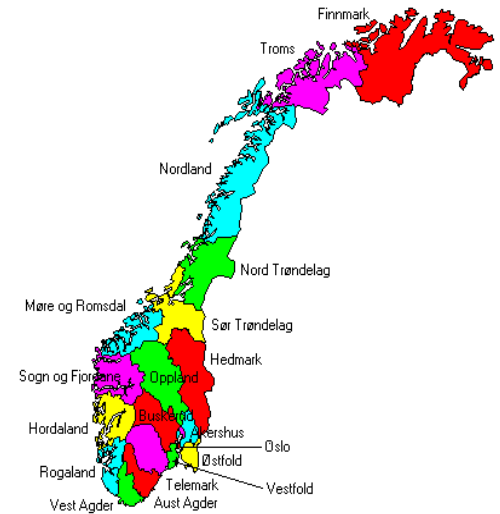


Figure 4.2: Norway and its regions

We give the data of the response variable the following notation:

$$Y_{ijkl} = \text{Tonnes of fish caught of species } i, \text{ on day } j, \text{ with vessel } k, \text{ in region } l \quad (4.1)$$

Since we use weather variables in the models, we prefer to make the region where the fish is caught as small as possible so that the weather information is as accurate as possible. The exact location where the fish is caught is not stored in the database. The only information that is known is the location where the fish is delivered. Due to the limited time we cannot forecast the catch for all regions. We select the best region with respect to size in the next section and base our models upon this region.

4.2.2 Statistics of each region

The data of the fish catch is distinguished into five regions in Norway. From each region the public information about the catch of individual vessels is known. It is only known that the fish is caught in this region, but not in which exact location. To keep the weather information accurate, we want to have a small location where the fish can be caught. However, it is also important to have as much as data available, since the more data there is available, the more potential there is to create an accurate forecasting model (Witten, 2005).

The possible surface where the fish can be caught is determined by how many kilometers a vessel ships from the coastline into the sea, multiplied with the length of the coastline. Since we do not know the first distance, we consider the number of kilometers of coastline of a region the measurement of how big the catching surface is. Hence, we want to maximize the amount of caught kilograms fish per kilometer of coastline. To measure the coastline, we choose the coastline of the mainland, excluding the islands. Figure 4.3 shows the total catch per region per kilometer coastline (Norges råfisklag, 2017) (Statistisk sentralbyrå, 2013). We conclude that the region Troms has the highest catch per kilometer coastline. Hence, we choose this region to base our forecasting model on.

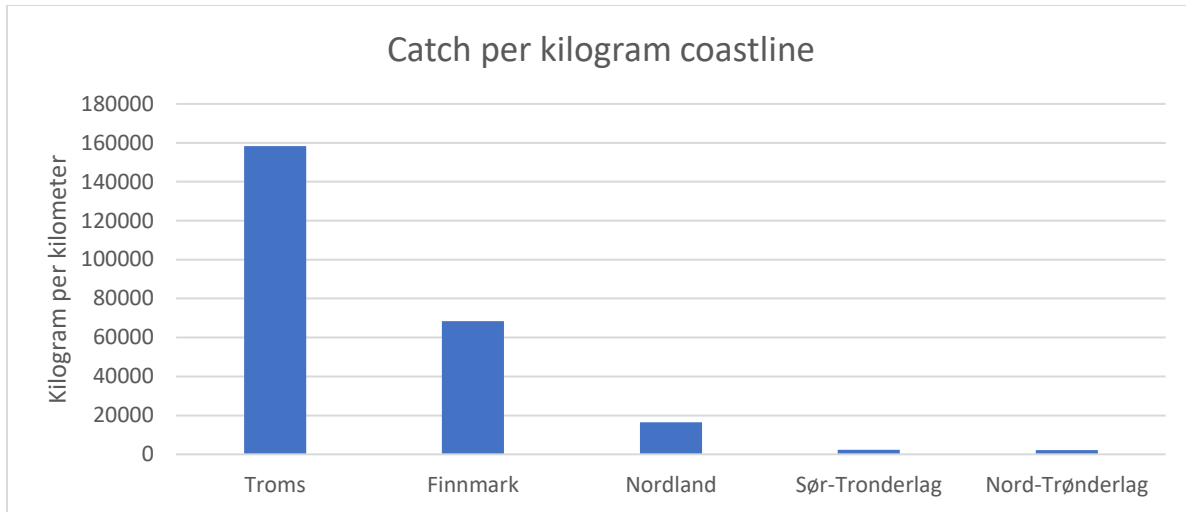


Figure 4.3: Catch per kilogram coastline for the relevant regions

4.2.3 Choice of the sample

Finally, we need a sample out of the region Troms to base our forecasting model on. To do so, we first need to know how many individual data points we need. A general rule that is used is that one needs a minimum of 10 data points per covariate (predictor) (Harrell, 2001). In total, we have a maximum of 14 different covariates (see Section 4.3.2), which means that we need more than 140 usable data points for a region.

For two species (Pollock and Haddock), 58.6% of all the observations are smaller than 1 ton. These small values are the result of bycatch. Bycatch happens when a fisherman catches a different fish specie than the specie he intends to fish on. This bycatch data makes the data unreliable since we do not want to predict bycatch. When removing these observations, we have less than 140 data points left for each specie. For Cod, only 1.8% of the catch data is smaller than 1 ton (see Figure 4.4). We still have 236 data points for Cod, which is sufficient to use linear regression. Hence, we base the model on the catch of Cod in the Troms region.

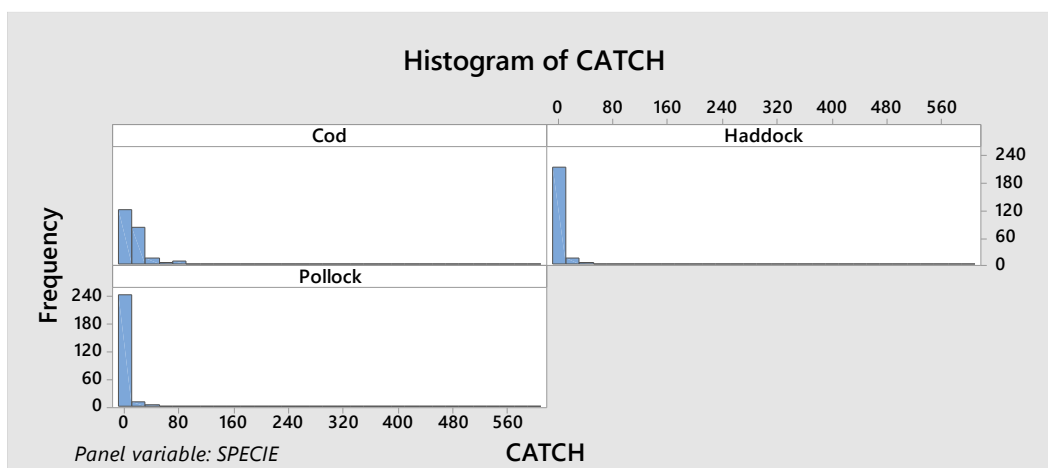


Figure 4.4: Data distribution of the response variable

4.3 Prediction model: Response- and prediction variables

To make a proper model, it is important to analyze how the dependent variable is distributed and if there are outliers. Also, we present the prediction variables (independent variables) here, together with the data scheme to show how the variables relate to each other. In Section 4.3.1 and 4.3.2 we explain the response variable and the prediction variables respectively.

4.3.1 Response variable: Distribution and transformation

As stated in Section 4.2.3, we base the models on the Cod catch in the region Troms in 2017.

We define the response variable as follows:

$$Y_{jk} = \text{Tonnes Cod caught on day } j, \text{ with vessel } k \quad (4.2)$$

To apply causal models such as linear regression we need the errors of the forecasts to be normally distributed, since otherwise the model is unreliable. To achieve this, it is desirable to have the response variable distributed as a normal distribution (Walpole, 2012).

When looking at the histogram of the data and applying a normality test (Kolmogorov-Smirnov test), we conclude that the data is not normally distributed (see Figure 4.5 and Appendix 2).

We try to normalize the data using a Box-Cox transformation. This is commonly used to try to normalize not-normal distributed data (Sakia, 1992). In Appendix 3 we proof that taking the natural logarithm of the data is the best transformation.

When making a histogram from the transformed data, we indeed conclude that the data is more normally distributed (Figure 4.6). The QQ-plots and the value of the test statistic of the Kolmogorov-Smirnov test confirm this (Appendix 2). However, the p-value of the test is still smaller than 0.05, which means there is still enough evidence to reject the hypothesis that the data is normally distributed. This means that we have to analyze the forecasting errors carefully, which we explain in Section 4.4. We call the transformed data *LNCATCH*.

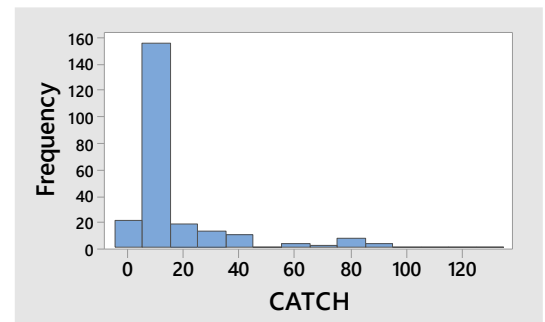


Figure 4.5: Distribution of the catch

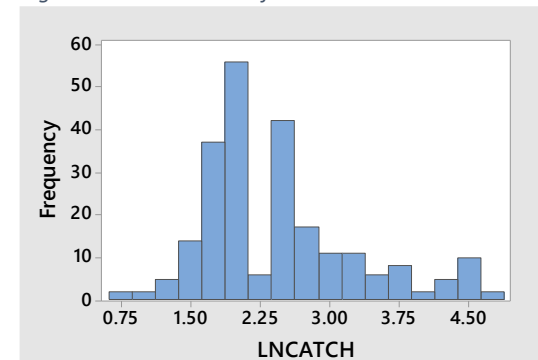


Figure 4.6: Distribution after transformation with natural logarithm

4.3.2 Prediction variables: data schema

To predict the catch of a vessel, we need independent variables. We can categorize the independent variable into two categories: date specific, and vessel specific. Table 4.1 and Table 4.2 show the variables together with the interval of possible values they can obtain, the measurement, and the acronym we use in this report.

Table 4.1: Date specific prediction variables

Date specific	Acronym	Measurement	Interval/set
Average temperature	AVGTEMP	°C	$(-273.15, \infty)$
Rain	RAIN	millimeter	$[0, \infty)$
Average wind	AVGWIND	Kilometers hour ⁻¹	$[0, \infty)$
Wind direction	WINDDIR	Degrees	$[0, 360)$
Visibility	VISIBIL	Kilometers	$\{0, 1, \dots, 13, 14\}$
Humidity	HUMID	Percentage	$[0, 100]$
Pressure	PRESSURE	Pascal	$[0, \infty)$
Sun hours	SUNH	Hours day ⁻¹	$[0, 24]$
Moon phase	MOON	-	$[0, 29.53]$

Table 4.2: Vessel specific prediction variables

Vessel specific	Acronym	Measurement	Interval/set
Vessel length	LENGTH	Meters	$(0, \infty)$
Vessel width	WIDTH	Meters	$(0, \infty)$
Vessel age	AGE	Years	$(0, \infty)$
Material	MATERIAL	-	$\{\text{Steel, Aluminium, Plastic, Composite}\}$
Power	HP	Horsepower	$(0, \infty)$

The moon phase is the only variable we did not extract directly from the weather database. We calculate it using:

$$\begin{aligned}
 MOON_i &= (date_i - date_m) \bmod(29.530588853) \\
 \text{Where : } MOON_i &= \text{moonphase on day } i \\
 date_m &= \text{date of a know new moon}
 \end{aligned}
 \tag{4.3}$$

Figure 4.7 shows a star schema of the data to visualize how the variables are connected in our database.

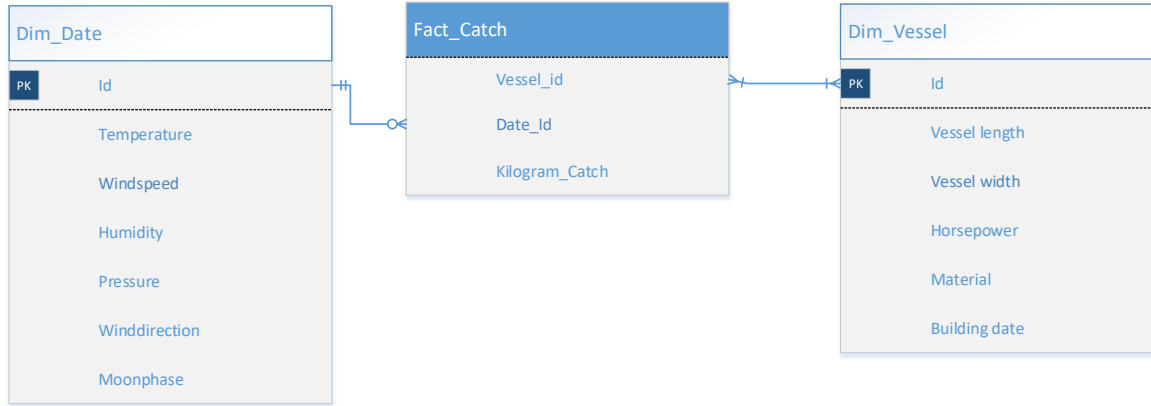


Figure 4.7: Star schema of the prediction variables

4.4 Causal model: linear regression

In this section we construct a linear regression model to forecast the catch observations, using one or more prediction variables. As shown in Section 4.3.1, the response variable is, after the best possible transformation, still not normally distributed according to the Kolmogorov-Smirnov test. A normal distributed response variable is desired when using linear regression (Larsen, 2012). However, it is more important that the errors of the regression model ($\hat{Y}_i - Y_i$) have a normal distribution, as this is one of the requirements for a regression model to be reliable (Li, 2012). In Section 4.4.1 we make a first model where we include all the variables, after which we transform some independent variables and construct a new model in Section 4.4.2. For all significance levels we choose 0.95. Hence for a predication variable to be significant, the maximum Type I error is 0.05.

4.4.1 Regression with all independent variables: a first impression

To get a first impression of the significance of the variables, we construct a model where we include all the independent variables. First, we analyze if there is a significant correlation between the prediction variables. If this is the case, we should pay attention when interpreting the regression model, since it may look like a prediction is significant, while this could be due to multicollinearity (Larsen, 2012). Appendix 4 shows the correlation that have a p-value smaller than 0.05

Especially when a vessel specific variable is correlated with a weather specific variable we have to pay attention, since it could happen that prediction variables are predicting each other, instead of the response variable. When this happens, it seems like a prediction variable has a significant explanation power with respect to the response variable, while it is only explaining another prediction variable. This phenomenon is known as multicollinearity (Walpole, 2016). We observe that AGE is correlated with SUNH and WIDTH with AVGWIND. Hence, we need a measurement for the multicollinearity in the model.

To check for multicollinearity, we use the Variance Inflation Factor (VIF), which is calculated in a linear regression model with m as the number of prediction variables as:

$$VIF_i = \frac{1}{1 - R_i^2}$$

$$R_i^2 = R^2 \text{ in } X_i = \sum_{\substack{k=1 \\ k \neq i}}^m \alpha_k X_k + c_0$$

Where : VIF_i = Variance Inflation Factor of prediction variable i

α_k = regression coefficient of prediction variable k (4.4)

X_i = prediction variable i

R_i^2 = coefficient of determination of prediction variable i

c_0 = intercept value (constant)

Hence, we check the exploratory power of each prediction variable by constructing a linear regression model without the response variable. When the VIF equals 1, it means there is no multicollinearity. A general rule for the upper bound is that the VIF should be lower than 6 (O'Brien, 2007).

When we construct a linear regression model with all the predication variables as shown in Section 4.3.2 we observe that the model has an adjusted R^2 of 65%. The adjusted R^2 corrects the regular R^2 when there are too many prediction variables included in the model, something we call overfitting. An adjusted R^2 of 65% means that 65% of the variation of the response variable can be explained by the prediction variables. The p-value is equal to 0.000, which means the model has enough exploratory power to be significant. However, we can see that there are some variables with an increased VIF, such as the length (VIF=6.59) and the width (VIF=7.51). This means there is multicollinearity in the model. Appendix 5 shows the summary of this model.

4.4.2 Transforming independent variables

Currently, the variables may not be optimally transformed to explain the catch properly. Therefore, we look for transformations that can improve this. For this, it is necessary to acquire information about how the different weather factors influence the catch size.

When looking at the literature review, we see that there is a research where full moon with a margin of three days, has a positive effect on the catch size (Courtney, 1996). Besides this, there is a common believe in the fishery that offshore wind is most promising for sea fishing and onshore wind the worst (Britishseafishing, 2016).

Hence, we transform the moon phase and the windspeed combined with the wind direction to match the findings as stated before.

We transform the wind direction to be 1 when there is onshore wind, and 0 when there is offshore wind. Then, we multiply this value with the windspeed. We do this using:

$$WINDOPT_i = \begin{cases} AVGWIND * \frac{180 - (270 - WINDDIR)}{180} & \text{if } 90 < WINDDIR < 270 \\ AVGWIND * \frac{180 - (WINDDIR - 270)}{180} & \text{if } 270 \leq WINDDIR \leq 360 \\ AVGWIND * (0.5 + \frac{WINDIR}{180}) & \text{if } 0 \leq WINDDIR \leq 90 \end{cases} \quad (4.5)$$

For the moon phase, the data is similarly transformed. We want the transformed value to be at its maximum (1) at full moon and decreasing to its minimum (0) when the moon phases reaches new moon. We transform the moon phase using:

$$MOONOPT_i = \begin{cases} \frac{(MOON - \frac{29.53058853}{2})}{\frac{29.53058853}{2}} & \text{if } MOON > \frac{29.53058853}{2} \\ \frac{(\frac{29.53058853}{2} - MOON)}{\frac{29.53058853}{2}} & \text{if } MOON < \frac{29.53058853}{2} \end{cases} \quad (4.6)$$

Now we create a new model with only the predictors that had a p-value smaller than 0.10 and adding the two newly constructed predictors. Also, we remove WIDTH from the model, since LENGTH and WIDTH have multicollinearity, given the high VIF.

Model 1 shows the output of this model. The transformation with the wind direction and wind speed is indeed significant. The moon phase transformation is highly insignificant, and we remove it from the model. This is in contrast with what we saw in the literature, where the moon phase had a significant positive influence on the catch size.

In the final model (Model 2), all used predictors are significant, and the model has an adjusted R^2 score of 62.48%. All the VIF values are around 1, which indicate there is no, or very limited multicollinearity in the model. We also see that the standard deviation of the forecasting error (S) is equal to 0.514.

Model Summary

S	R ²	R ² (adj)	R ² (pred)
0.514116	63.28%	62.48%	60.88%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.016	0.223	13.55	0.000	
LENGTH	0.10389	0.00567	18.32	0.000	1.02
VISIBIL	0.0327	0.0197	1.66	0.098	1.17
AGE	-0.01033	0.00237	-4.36	0.000	1.01
WINDOPT	0.02250	0.00866	2.60	0.010	1.13
MOONOPT	-0.110	0.119	-0.92	0.357	1.15

Model 1: Linear regression with transformed variables

Model Summary

S	R ²	R ² (adj)	R ² (pred)
0.513953	63.15%	62.51%	61.12%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.856	0.205	4.18	0.000	
LENGTH	0.10406	0.00567	18.37	0.000	1.02
VISIBIL	0.0376	0.0189	1.99	0.048	1.09
AGE	-0.01046	0.00236	-4.43	0.000	1.01
WINDOPT	0.02113	0.00853	2.48	0.014	1.10

Model 2: Final model with all significant predictors

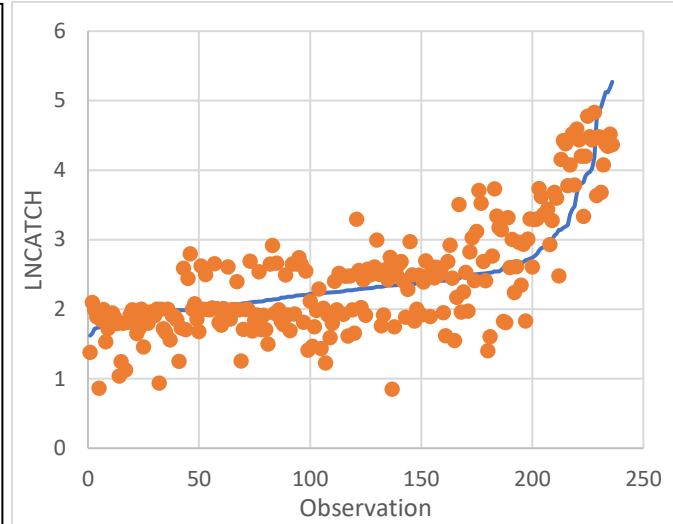


Figure 4.8: Predicted values from the linear regression model

As we stated in Section 4.4 it is important that the prediction errors are normally distributed. Hence, we test the residuals for normality using a Chi-squared test. The p-value of the test statistic is equal to 0.987. Hence, there is no evidence to believe the distribution of the residuals is not normally distributed. We conclude the regression model, based on this dataset, is reliable. Appendix 6 shows the calculation of the test statistic and the test. Figure 4.8 shows the predicted values, which are plotted in blue, when sorted from small to large. The orange dots show the corresponding predicted value. When we look at this figure, one can doubt the effectiveness of this model, since there are some large deviations from the predicted value. However, we have to keep in mind that these deviations are smaller with the regression model than in the case where we do not use a forecasting model and use the expected catch as forecasting method. We proof this in Chapter 6, where we observe the differences when we use a forecasting model and when we use the expected value of the catch size.

4.5 Machine learning models

Now that we have a base model, and a way to measure the accuracy of this model (the RMSE, or S), we try to improve this model using machine learning techniques. We use the most common used techniques that are applicable to regression models with a continuous response variable: the REP tree, the random forest, and the M5P Tree. We omit the Multilayer Perceptron algorithm as literature tells us this gives the worst results when applied to the forecasting of fishing success (see Section 3.2). Section 3.2 gives a brief overview and comparison of these algorithms.

To not lose observations while building the models based on training data, we use 10-fold cross validation. This means the algorithm constructs a model 10 times using 90% as training data and 10% as test data. The accuracy is then determined by taking the average accuracy of these 10 models. We use the standard deviation of the forecasting error (S) as measurement for the accuracy of the models. The accuracy of the model is estimated using:

$$S = \frac{1}{10} * \sum_{i=1}^{10} \sqrt{\left(\frac{1}{n} \sum_{j \in i} (y_{ij} - \hat{y}_{ij})^2 \right)}$$

Where : y_{ij} = observation j in iteration i

\hat{y}_{ij} = predicted value of observation j in iteration i

(4.7)

In Section 4.5.1 we build the Reduced Error Pruning (REP) Tree, after which we continue with the M5P Tree in section 4.5.2. In Section 4.5.3 we finish with the random forest algorithm. The attributes we use in these algorithms are the same as the prediction variables used in Model 2 of Section 4.3.2, where we built the linear regression model.

4.5.1 REP Tree

Since it is possible that the weather attributes interact with each other we need an algorithm that can cope with this. The Reduced Error Pruning (REP) Tree has this ability. Because the REP Tree is pruned after construction it is optimized in size, making it easier to use in the capacity planning (Belouch, 2017).

To measure the error in a decision tree with numerical variables, we use:

$$S = \frac{1}{\sum_{c=1}^k n_c} \sum_{c=1}^k \sum_{i \in l_c} (y_i - \hat{y}_{l_c})^2$$

Where: l_c = leaf with number c

y_i = observed value of observation i

\hat{y}_{l_c} = predicted value of observation i in leaf c

n_c = number of observations falling in leaf c

$$\hat{y}_{l_c} = \frac{1}{n_c} \sum_{i \in l_c} y_i$$
(4.8)

Appendix 7 shows the algorithm to construct this tree. Table 4.3 shows the summary of this method. We can see that the standard deviation of the error has a value of 0.525, which is slightly worse than the linear regression model that has an S value of 0.514.

Table 4.3: Performance indicators REP Tree

Mean absolute error	0.4092
Root mean squared error (S)	0.525
Total Number of Instances	236

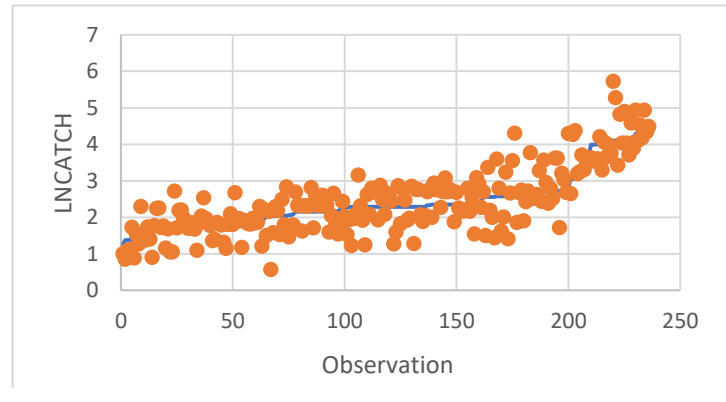


Figure 4.9: Forecasted values together with observed values from the REP Tree

In Chapter 5 we explain how we evaluate the production planning when including the forecasting model. To do so we need the errors of the REP Tree to follow a statistical distribution. Hence, we test the errors with a Chi-squared test to check for normality. The p-value of this test is equal to 0.18 (Appendix 6). Hence, there is no reason to assume the distribution of the errors is not normal. Figure 4.9 plots the predicted values together with the observed values. Appendix 8 shows the tree with the leaves that we use to make a prediction.

4.5.2 M5P Tree

The M5P Tree is an ordinary decision tree, with the difference that in each leaf a regression model is used to predict the values, instead of using the expected value in a leaf. A M5P Tree is built by branching on the attribute that reduces the variation down each branch the most (Witten, 2005). When calling the fraction of training data that reaches a certain node, T , we obtain the Standard Deviation Reduction (SDR) with:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} * sd(T_i) \quad (4.9)$$

Where : T_i = set i , which results from splitting the node with the chosen attribute

We build the tree until there are <5 instances in a leaf, or when $\frac{\sum_i sd(T_i)}{SD(T)} \leq 0.05$.

After building the tree, we prune the tree, just as in regular decision trees. We do this iteratively and continue if and only if the Mean Squared Error decreases. We construct a linear model in each node, using the training data that reach this node. Then, we prune a subtree if and only if:

$$\frac{1}{n_k} \sum_{i \in k} (y_i - \hat{y}_i)^2 \leq \frac{1}{\sum_{\substack{j \in S_k \\ j \neq k}} n_j} \sum_{\substack{j \in S_k \\ j \neq k}} \sum_{i \in j} (y_{ji} - \hat{y}_{ji})^2 \quad (4.10)$$

Where : k = node with number k

j = node with number j

S_k = Subtree of node k

Appendix 9 shows the algorithm of the M5P Tree. The tree has 3 leaves, with 3 corresponding linear models, Appendix 10 shows the tree. Table 4.5 shows the performance of this model. We see that the model performs better than the linear regression model and the REP Tree, as the S value is lower. Again, we test the residuals on normality using the Chi-squared test. The p-value for this test is 0.88, which again leads to the conclusion that there is no reason to doubt that the errors are distributed normally. Figure 4.10 shows the graph with the predictions and observed values.

Table 4.4: Performance indicators M5P Tree

Mean absolute error	0.3796
Root mean squared error (S)	0.4992
Total Number of Instances	236

4.5.3 Random forest

The random forest is an algorithm that uses a bootstrap algorithm (bagging). This means there are more training sets derived from the data, which all generate an independent prediction. Then, the predicted value is the average of the independent predictions.

The random forest generates multiple random trees. For each node in each tree there are k attributes chosen randomly to build the tree with (Liaw, 2002). Section 3.2 explains this process in more detail.

Then, we calculate the best splitting attribute again using the SDR (See Equation 4.9) (Witten, 2005). Appendix 11 shows the algorithm to construct a random forest.

Since there are several trees generated, and the trees are random, we do not be display the trees. The Chi-squared test for the residuals has a p-value of 0.74 in this case. Hence, also for the random forest, there is no reason to assume the residuals are not normally distributed (Appendix 6).

Table 4.6 shows the performance indicators for the random forest. Figure 4.11 shows the plotted values together with the prediction. We observe that S is 0.5126, which is better than the regression model and the REP Tree, but worse than the M5P Tree.

Table 4.5: Performance indicators random forest

Mean absolute error	0.385
Root mean squared error (S)	0.5126
Total Number of Instances	236

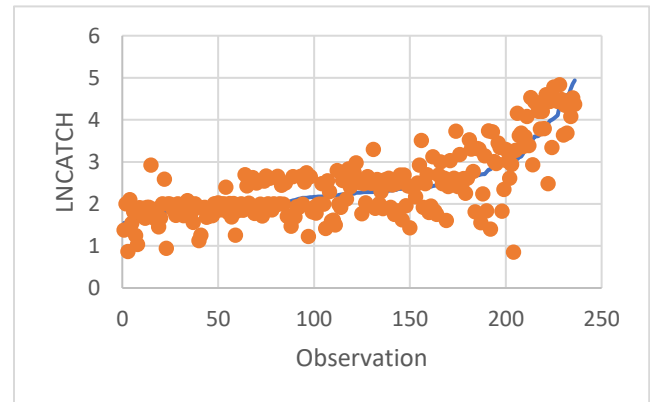


Figure 4.10: Forecasted values together with observed values of the M5P Tree

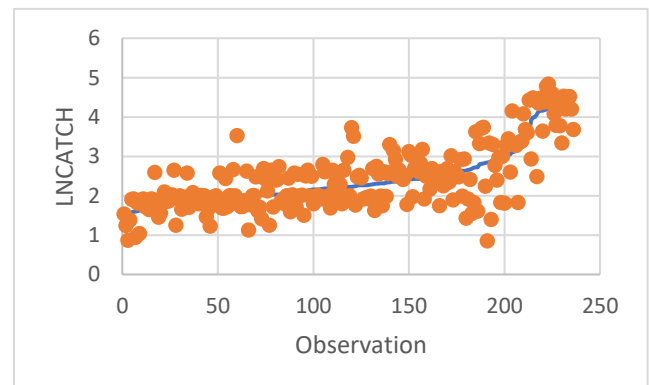


Figure 4.11: Forecasted values together with observed values from the random forest

4.6 Weather forecast accuracy and impact on prediction model

Since it is impossible to perfectly predict the weather, the three variables that are based on weather prediction also need to be treated as random variables. In Section 4.4 and 4.5 we assumed the weather attributes to be deterministic. In this section we analyze the influence of the stochastic weather variables on the forecasting models created in these sections. To do so, we need the forecasting accuracy of the weather forecasts. Therefore, we need to gather information about the historical weather forecasts with respect to the observed weather. We can use these forecasts to see the accuracy of the weather forecasts and use this information in the forecast models for the fish catch. In Section 4.6.1 we start with analyzing the accuracy of the weather forecast. We use this information in section 4.6.2 to see what the influence of this uncertainty is on the models of Section 4.4 and 4.5. In Section 4.6.3 we summarize the results and choose the best forecasting method to predict the fish catch.

4.6.1 Accuracy of the weather forecast

There is enough information available about historical *observed* weather, but there is almost no data about the historical *forecasts* (what was the weather forecast on a given day). The only information available consists of historical forecast of the average temperature. Although we do not use the temperature in the models, it gives an idea on how accurate the weather forecasts are for the other variables. Figure 4.12 shows an example of the observed and the two-day forecast.

It is not exactly clear how many days in advance the forecasting for the fish is needed by LNS, so we make the calculations for 1 until 6 days in advance. The forecasts are made for Tromsø and consists of one-year historical data (01-01-2016 until 01-01-2017). To determine the accuracy of a forecast, we use the variance of the forecast:

$$E(A_j) = \frac{1}{365} \sum_{i=1}^{365} (y_{0,i} - \hat{y}_{j,i})^2 \quad \forall j \in \{1, 2, \dots, 6\},$$

Where : $E(A_j)$ =expected accuracy of the temperature forecast for j days in advance (4.11)

$y_{0,i}$ = observed weather on day i

$\hat{y}_{j,i}$ =forecasted weather j days in advance day i

Figure 4.13 shows the variance for each of the six forecast days. As we can expect, the variance increases until 5 days in advance, and then shows a small decrease. We assume that the other weather variables can be as accurately predicted as the temperature. This assumption has the result that we also can assume that the Coefficient of Variation (CoV) is equal for all the weather attributes and we can calculate the accuracy of the visibility, the wind speed, and the wind direction with CoV of the temperature predictions.

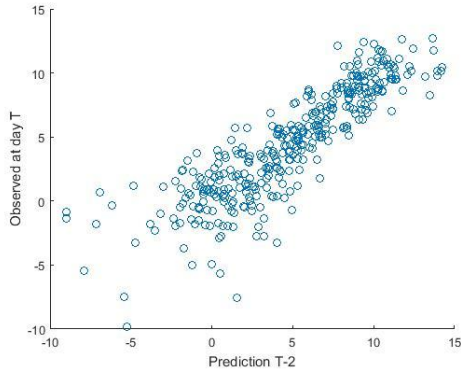


Figure 4.12: Scatterplot observed values and predicted values

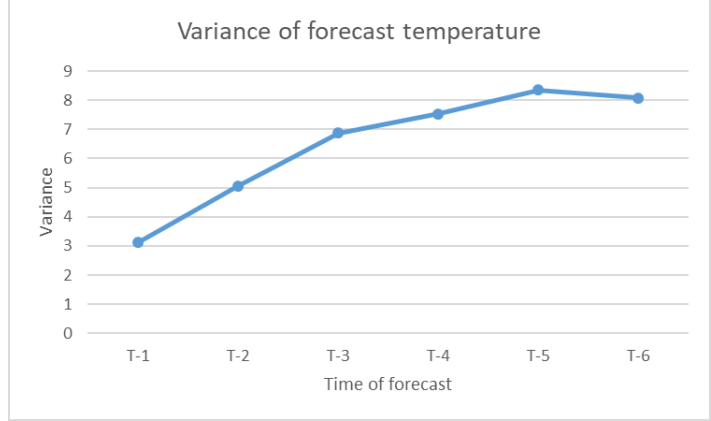


Figure 4.13: Variance of forecast temperature

The CoV is defined as $\frac{\sigma}{\mu}$. The average temperature measured in Tromsø in 2016 is 4.82°C. Then the CoV for the prediction 2 days in advance is $\frac{2.25}{4.82} = 0.467$. We assume that the CoV for the two-day forecast for the visibility is also 0.467. The average visibility used in the model is 8.07 km. Hence, the estimate for the standard deviation is then $0.467 \cdot 8.07 = 3.77$.

In Section 4.4.2 we transformed the variables wind speed and wind direction to create WINDOPT. The random variables wind speed and wind direction are multiplied. To estimate the variance for WINDOPT we need a formula to determine the variance when two random variables are multiplied.

We use (Walpole, 2012):

$$\begin{aligned} Var(XY) &= E[X^2Y^2] - [E(XY)]^2 = \\ &Cov(X^2, Y^2) + E(X^2)E(Y^2) - [E(XY)]^2 = \\ &Cov(X^2, Y^2) + (Var(X) + [E(X)]^2)(Var(Y) + [E(Y)]^2) - [Cov(X, Y) + E(X) + E(Y)]^2 \end{aligned} \quad (4.12)$$

We consider the forecast accurate. The definition of an accurate forecast is that the mean of the forecast is equal to the mean of the observed variable. Therefore, we can assume that the covariance of the observed data reflects the covariance of the forecasts. There is no significant correlation between the windspeed and the wind direction. Therefore, the covariance is equal to 0, dropping the first term of Equation 4.12. Then the variance for the windspeed is equal to $11.56 \cdot 0.467 = 5.40^2 = 29.16$ and the variance of the transformation for the wind direction (see Equation 4.12) is equal to $0.505 \cdot 0.467 = 0.24^2 = 0.0576$.

Then the variance for WINDOPT:

$$Var(WINDOPT) = (29.16 + 11.56)^2 \cdot (0.0576 + 0.505)^2 - (11.56 + 0.505)^2 = 379.26 \quad (4.13)$$

Finally, Table 4.6 shows the means and standard deviations of the weather forecasts we use in the forecasting model for the catch size.

Table 4.6: Expected value and expected standard deviation for VISIBIL and WINDOPT two days in advance

Forecast	E(X)	S _x
VISIBIL	8.07	3.77
WINDOPT	5.98	19.47

4.6.2 Impact of weather uncertainty on the fish forecasting models

In Section 4.4 we created a linear regression model and machine learning models to predict the fish catch per vessel. This regression model has the following equation:

$$\hat{Y} = 0.856 + 0.104606 * L + 0.0376 * V - 0.01046 * A + 0.02113 * W \quad (4.14)$$

In Section 4.4 we considered the Visibility (V) and the WINDOPT (W) deterministic in the models. The visibility and the WINDOPT are, in contradiction to the length (L) and age of the vessel (A), stochastic variables itself. We must consider this in the determination of the variance of the errors of the forecasting models.

Since two of the regressors are stochastic, we calculate the errors with:

$$e_{i,j} = (y_i - E(\hat{y}_{i,j} | v_j, w_j))$$

Where: y_i =observed catch observation i
 $\hat{y}_{i,j}$ =predicted catch observation i, j days in advance
 v_j, w_j =weather forecast for VISIBIL and WINDOPT j days in advance
 $e_{i,j}$ =forecasting error for observation i, j days in advance

(4.15)

Since there are no simple closed formulae to determine the variance of the model with the stochastic influence of the weather attributes, we use Monte Carlo simulation. Before we set up the simulation, we must take into consideration that there is a correlation between VISIBIL and WINDOPT (Appendix 4). This means the variables are not independent and that we estimate WINDOPT when we know VISIBIL. We construct a simple linear regression model where we use WINDOPT as dependent variable and VISIBIL as regressor. The model is significant, although R^2 is only 0.0796. The equation is as follows:

$$WINDOPT = 11.21 - 0.630 * VISIBIL \quad (4.16)$$

The estimated standard deviation of the error is equal to 3.97. Hence, to estimate the WINDOPT:

$$WINDOPT = 11.21 - 0.630 * VISIBIL + F^{-1}(p | 0, 3.97^2) \text{ and}$$

$$w_f = 11.21 - 0.630 * v_f + F^{-1}(p | 0, 3.97^2) \quad (4.17)$$

Since we assumed that the forecasts are accurate, the second equation in Equation 4.17 is true. This means that the variance of the observed weather is the same as the variance of the forecasts. Therefore, the regression equation for the forecasts of WINDOPT is the same as for the observed weather data at time $T=0$.

The Monte Carlo simulation to estimate the accuracy for the forecasting of the fish caught j days in advance consists of 8 steps:

1. Draw a random observation (i) out of the list of all observed catch statistics together with the attributes;
2. Read the attributes LENGTH and AGE of the vessel of observation i ;
3. Read the attribute VISIBIL, that is the observed visibility of observation i ;
4. Draw a random number (p) in the interval $[0,1]$. Then determine from VISIBIL a predicted visibility j day in advance. For $j=2$: $v_f = VISIBIL + F^{-1}(p | 3.77^2)$. Where $F^{-1}(p | S^2)$ is the inverse standard normal distribution with probability p and variance S^2 ;
5. Draw a random number q in the interval $[0,1]$. From the predicted forecast visibility, we can calculate a forecast WINDOPT since these variables are correlated. We use Equation 4.17 with random number q in the interval $[0,1]$ for the probability p ;
6. Use all forecasts and attributes to calculate a prediction with one of the forecasting techniques (Linear regression, REP Tree, M5P Tree);
7. Determine the prediction error and square this error;
8. Sum all the squared prediction errors and divide this by the number of simulation runs, we now have the variance of the forecast error.

Figure 4.14 schematically explains the simulation process.

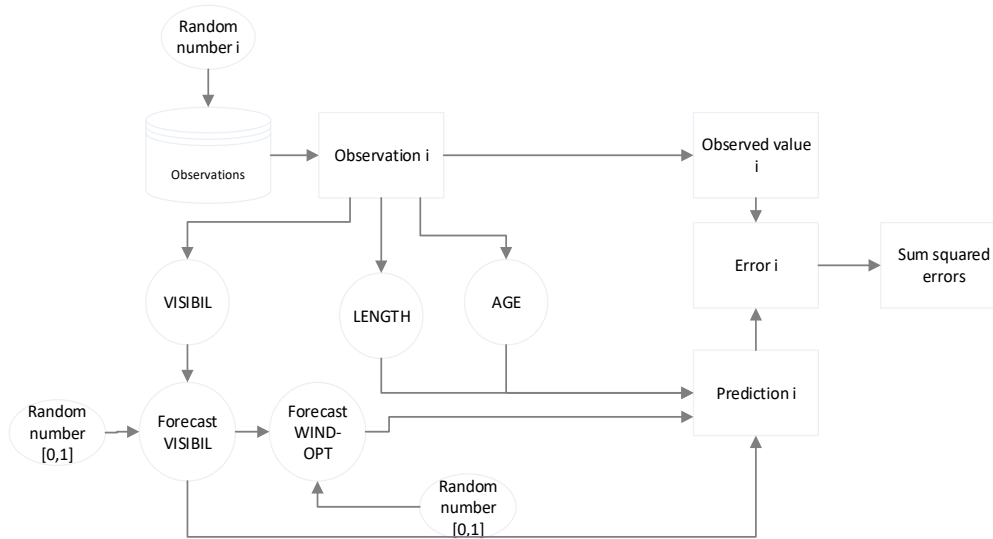


Figure 4.14: Schematic model for Monte Carlo simulation

First, we test the model to ensure its appropriateness. To determine the number of required runs we use the linear regression model that has a real standard deviation of 0.51 when the weather predictors are deterministic. When running the Monte Carlo simulation with all predictors deterministic, the calculated standard deviation of the simulation model converges to the real standard deviation after around 100 simulation runs (Figure 4.15). Hence, we consider 1000 simulation runs an appropriate number. Appendix 12 shows the codes used for the Monte Carlo simulation.

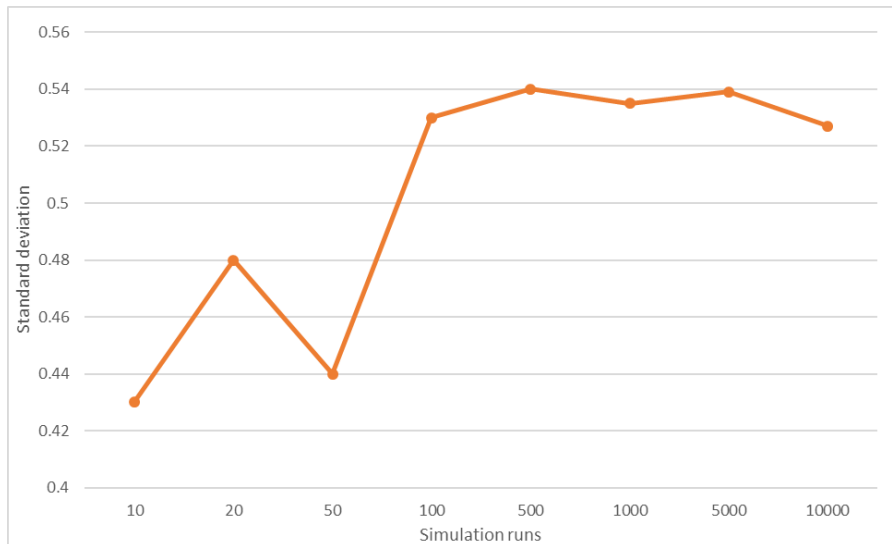


Figure 4.15: Estimated variance for various simulation runs

After running all the simulations, we obtain the results as in Figure 4.16, where the vertical axis shows the estimated standard deviation of the error (S) of the forecasting models. We observe here that the REP Tree incurs the highest increase in standard deviation as the increase from day 0 to day 1 is the largest. This is due to the discrete nature of this tree, where a small deviation in the prediction variables can have the result that the prediction falls in another leaf, changing the prediction value drastically. The linear regression model and the M5P Tree show a similar increase, considering the parallel lines. Although the M5P Tree is also a classification tree, the tree only classifies based on the vessel length. It is not possible to estimate the increase in standard deviation of the random forest since this algorithm does not have a fixed tree. Instead, we assume that the increase of the random forest is the same as the increase of the M5P Tree.

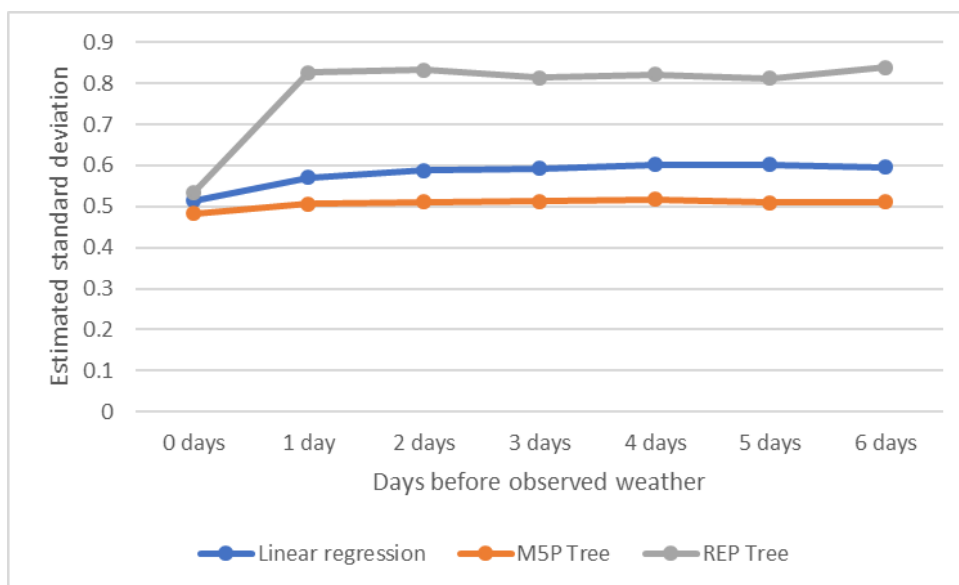


Figure 4.16: Estimated standard deviation based on Monte Carlo simulation

4.7 Selection of best model

We take the linear regression model as base model and compare the estimation of the standard deviation with the other models. The running time for making a forecast is for all algorithms smaller than $O(n^2)$. Given the small size of daily fishing vessels visiting a plant (approximately 4 per day), we consider the running time of all algorithms sufficiently small and we do not compare the methods on the running time.

Figure 4.17 shows the model comparison. We observe that the REP Tree performs on average 40% worse in all scenarios than the linear regression and that the M5P Tree performs best. Therefore, we select the M5P Tree to forecast the supply for LNS. The expected value for the M5P Tree is 2.42. Hence, we expect a random vessel to deliver 11.28 tons of Cod. We need this information in Section 5.4 to determine how many fishing vessels deliver fish to LNS daily.

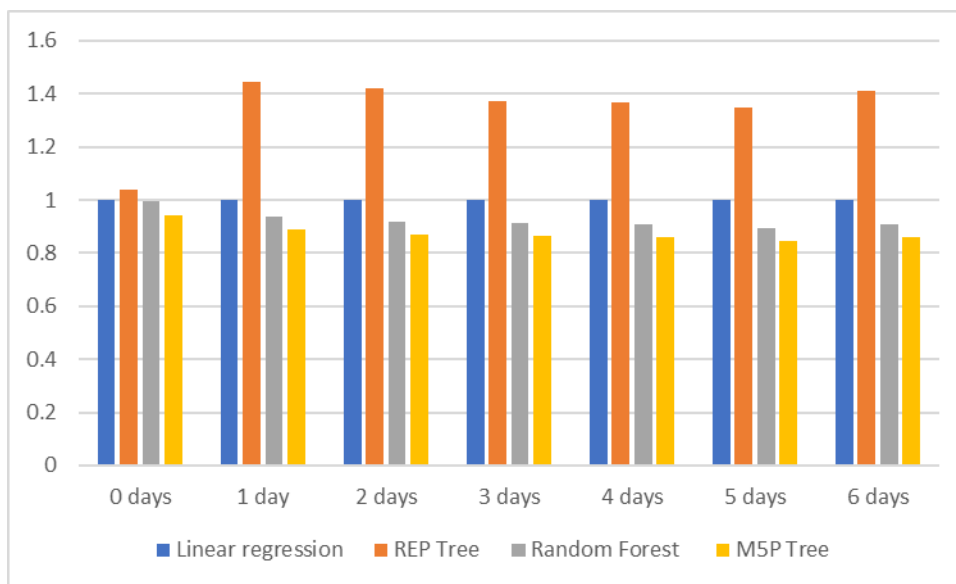


Figure 4.17: Comparison of all forecast methods (linear regression=1.0)

4.8 Conclusions of the reduction of supply uncertainty

In this chapter we started with proposing three possible methods, derived from the literature, which are interesting for LNS to reduce the supply uncertainty of raw material. Since LNS competes on product quality (see Section 1.1) the company needs good quality raw material to produce the best quality end products. Therefore, purchasing raw material from fish farms is not an option for LNS, since the quality of the raw material is not as good as fresh caught sea fish. Also, the fishermen have the most power in the supply chain. As a result, they are not willing to share catch information with LNS. We consider the option to forecast the catch size as the best solution to reduce the supply uncertainty.

We conclude that the region Troms is the best region to base our forecast on, due to the fact the tones of caught fish per kilometer coastline was the highest in this region. We showed that a combination of classification and regression gives the best results, as the forecasting error is the smallest in the M5P Tree. We must note that, although the forecasting models are significant on this dataset, the model can give different results when applying on data of another region or another year. We consider the forecasting error representative for the other regions and years.

5 Using the supply information in the capacity planning

In this chapter we use the acquired distribution of the catch forecast of Chapter 4 to optimize the capacity planning of LNS. To do so, we first look for a method to model the processing plant of LNS in Section 5.1. Section 5.2 explains the way of modelling and why we choose for this method. The model consists of three modules; one to optimize the capacity planning in the processing plant, one to evaluate the planning method, and one to collect data. Section 5.3 explains the first module, which optimizes the capacity planning. Section 5.4 focusses on the evaluation of the model, generating raw material supply, and collecting the output of the model. Finally, Section 5.5 finishes this chapter with the conclusions. This chapter answers research question 4.

5.1 Assumptions about the processes

We must make several assumptions about the production processes of LNS. These assumptions are necessary for two reasons: to scope the model because of our limited time and because LNS cannot provide us with all the necessary data.

5.1.1 Assumptions to scope the model

Although strictly spoken we have variables that are integer within the model, we assume all variables to be continuous. We plan the capacity in hours and not in employees. LNS must pay employees in shifts of 6.6 hours. We always round up since the production plan must be feasible. The maximum rounding difference from hours to employees is thus 6.6 hours per workstation. We checked the impact of this with LNS and the company indicates there is enough extra work to spend these hours productively. Also, we measure all produced products, used raw material and bought raw material in kilograms. Although very strictly spoken we should measure this in the number of fishes, LNS also uses kilograms for these variables.

Next, we omit all constraints with respect to the market demand. Hence, we assume all end products are sold immediately. Although the demand per individual product can differ, the aggregated global demand for fish products is increasing, and predicted to increase in the coming years with 1.1% per year. Beside this, the production is predicted to stay stable (The World Bank, 2013). Therefore, the assumption that all fish can be sold is justified.

We take only the three most purchased species into account, which are Cod, Haddock, and Pollock. These three species account for 96% of total production in 2017. We assume that the production of the other 4%, that consists of 7 remaining species, can be planned easily when the planning is made for the three main species.

Also, we assume the processing times within the facility to be deterministic. The filleting and freezing of the fish is completely deterministic, since this is automated. We justify the assumption that the processing time of the cutting process is also deterministic by providing the fact that the cutting department is occupied by more than 1 cutter simultaneously. This means that, although the variance of the processing time of each cutter on itself might have a high Coefficient of Variation (CoV), the CoV of the whole process (summation of all the individual cutters) decreases, when the number of cutters increases. Besides this, the layout in the processing plants is production oriented, resulting in an already low variation of processing times (Slack, 2008).

5.1.2 Assumptions because of the lack of data

Since LNS cannot provide us with all the required data, we make reasonable assumptions to make up for the lack of data. First, LNS could not provide us with historical data about which ships are visiting the processing plant. Therefore, we assume that the distribution of the ship characteristics in Båtsfjord is equal to the distribution in Troms, where the forecast model is based upon. Second, LNS could not provide us with accurate data about the costs and planned capacity in the processing plant. Therefore, it is not possible to determine the current costs levels in the processing plants and make comparisons. In Chapter 6, we create our own scenario with respect to the current costs situation at the processing plant.

5.2 Choice of the modelling approach

To decide which modelling approach we consider best, we must consider the complexity of the model and the quality of the solution. The first determines the running time, the second determines if the solution is optimal or how close the solution is to optimality. Section 5.2.1 starts with explaining the variables that are present in the model, after which we continue in Section 5.2.2 with explaining the modeling approach.

5.2.1 Variables in the model

Chapter 2 states that the processing times of the filleting and freezing department are deterministic since they are automatized. We justified in Section 5.1 why we consider the processing times in the cutting department to be deterministic. We consider all the variables in the model to be continuous, since we plan the capacity at each department in time, and not in the number of workers.

We also saw in Chapter 2 that the age and quality of the raw material is important. The quality does not change when the fish gets older, but the age of the raw material impacts the production matrix. Fish older than 6 days is always sold. Hence, we have for each specie 3 different qualities and 6 days of age.

Table 5.1 shows the decision variables of the model together with the unit in which they are measured and the number of elements each variable consists of. We need this information in the next section.

Table 5.1: Variables with their units and elements

Variable	Produced products	Used raw material to production	used raw material to sell	Planned capacity
Unit	kg	kg	kg	hours
Elements consist of	9 products	3 species, 3 qualities, 6 days	3 species, 3 qualities, 6 days	5 work stations
Total number of elements	9	54	54	5

5.2.2 Modelling approach: Rolling planning horizon

Exact algorithms give the optimal solution, in contrast to heuristics (Trevisan, 2011). We also must consider the complexity of our model. Therefore, we prefer exact algorithms if the running time is polynomial or less. Given the fact that we only have continuous variables and that LNS does not share workers between working departments (Section 2), we conclude that we can model the deterministic capacity problem in the processing plant as a Linear Programming Problem (LPP).

An LPP can be solved in polynomial running time (Trevisan, 2011). With the deterministic capacity problem, we mean the situation where all the supply of raw material is known. When observing the number of elements of all the variables of Table 5.1 we see why it is important that we only have continuous variables. Integer variables mean we need a Mixed Integer Linear Programming Problem (MILPP). These algorithms do not have a polynomial running time, meaning the running time explodes when the number of variables increase.

We also must take into consideration that the solution of day T depends on the solution of day $T-1$, since the age of the remaining inventory on day $T-1$ is one day older at day T ; we call this a recursion. Figure 5.1 shows this process schematically for the first two days. The symbols between brackets denotes the notation we use in this research. We explain the notation further in Section 5.3 Note that from day $T=1$ onwards, LNS can only change the capacity for day $T+3$. The capacity of days T , $T+1$, and $T+2$ comes from the planning made in day $T-1$. Because of the recursion, it is not possible to solve the problem with closed formulae for only one day, but we can formulate the problem as a rolling planning horizon (Kimms, 1997).

In Section 2.1 we explained that LNS must plan the capacity three days in advance and that LNS knows three days in advance how many, and which ships are coming to deliver fish. Therefore, we choose 3 days as our planning horizon. This means that on day T , given that T is bigger than 0, we can only change the capacity for day $T+3$, since the other days have been planned on day $T-1$ and cannot be changed anymore. To see the result of the planned capacity at day $T-1$, we optimize the production plan by producing the products that result in the highest profit, given the amount of delivered raw material and planned capacity. The planned supply comes from the forecasting model of Chapter 4.

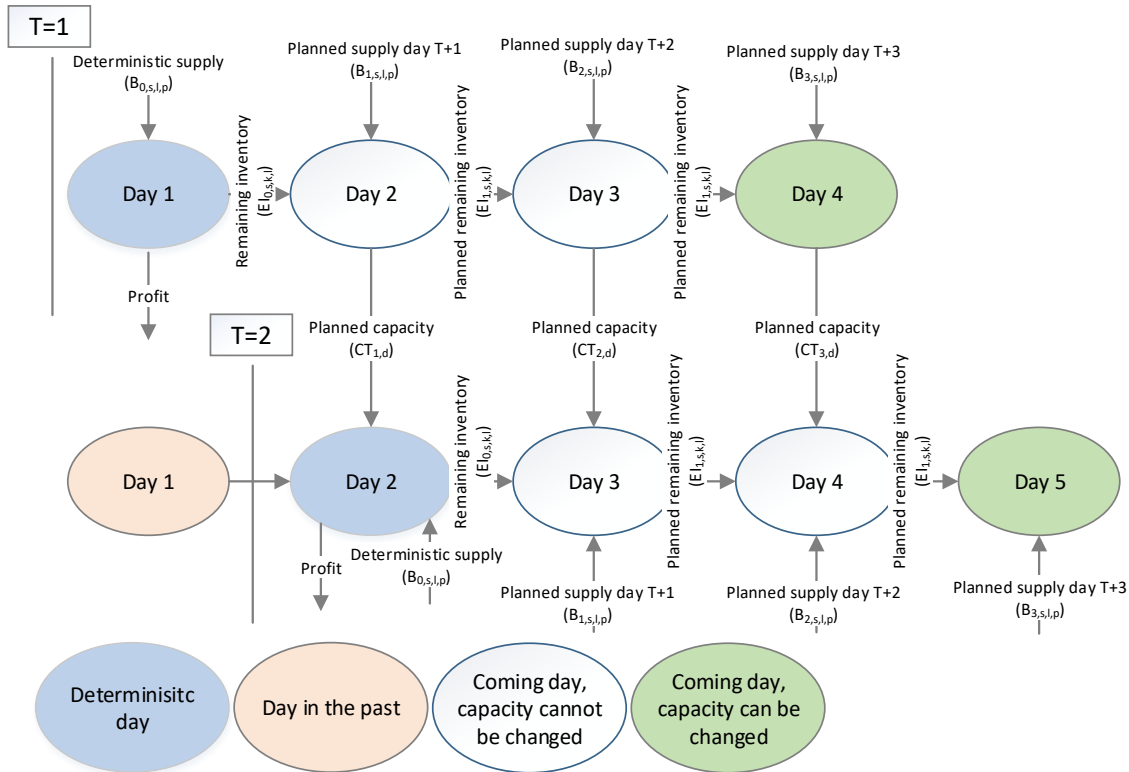


Figure 5.1: Recursion in the calculations

Taking into consideration that we use a stochastic variable into the model (planned supply of raw material) we must adjust the LPP model to cope with this uncertainty. Since we have a prediction for the catch we can use stochastic programming to solve the planning problem each day. Section 3.1 explains this method and literature uses this method to solve capacity planning problems with uncertain supply. An extra benefit of using SPP is that it is formulated as a Linear Programming Problem and that we therefore obtain a dual solution. The dual values tell us how much the profit function increases or decreases when relaxing the right-hand side of a constraint marginally. We use this dual solution to analyze the quality of the planning. Section 5.4.3 explains this in more detail.

5.3 Modelling the processing plant in Båtsfjord

We present the Stochastic Programming Problem (SPP) in this section. This problem is the base of our rolling horizon planning, since this creates the capacity planning each day. We base the model on the information that we gave in Chapter 2, which explains the processes in the processing plant in Båtsfjord. To model the processing plant, we need to know which parameters and variables are of interest and which ones we can omit.

The capacity planning is made in Båtsfjord three days in advance. Our forecast horizon is thus three days. We name the deterministic day in our model day T. Each day of the planning horizon we solve the model, which we present in this section. The model optimizes the production planning, given the planned capacity of the previous day, and plans the capacity for day T+3.

We need different scenarios for the prediction of the fish catch. This is because the actual catch differs from the forecast. When capacity is cheap in comparison with the sales prices of the products, it is optimal to plan more capacity than the expectation of the forecast of the catch and vice versa (Begen, 2003). We call this the opportunity costs. When we plan the capacity only based on the expected forecast, we do not take this phenomenon into account. In Chapter 4 we saw that we have a normal distribution for the forecast of the catch size. The SPP needs discrete inputs for the catch forecast. Since the normal distribution is a continuous distribution, we need to discretize the probabilities to generate the different forecast scenarios.

Table 5.2: Scenarios with their probabilities

Scenario	Z-value	Probability
High	1.029	0.3035
Normal	0	0.393
Low	-1.029	0.3035

Research shows that three scenarios, as shown in Table 5.2, yields an optimal result (Pflug, 2001). The Z-value of Table 5.2 corresponds with the Z-value of the used normal distribution of the catch forecast. For example, a catch forecast of 70,000 kg Cod with a standard deviation of 4,000 yields three forecast scenarios:

1. High: $70,000 + 4,000 * 1.029 = 74,116$ kg with probability 0.3035
2. Normal: $70,000 + 4,000 * 0 = 70,000$ kg with probability 0.393
3. Low: $70,000 + 4,000 * -1.029 = 65,884$ kg with probability 0.3035

First, we present all the variables, indices and parameters in Table 5.3, Table 5.4, and Table 5.5 respectively. The product numbers correspond with the products of Table 2.3. The working stations with the work station of Table 2.6.

Table 5.3: Indices of the SPP

Index	Explanation	Values
i	Day	4
s	Specie	3
k	Age of raw material	6
l	Quality of raw material	3
p	Scenario	3
o	Product	7
d	Work station	5

Table 5.5: Parameters of the SPP

Parameters	Indices	Explanation
SP	s,o	Selling price product per kilogram
K	s,l	Buying price raw material per kilogram
H	d	Wages per hour per employee per department
Ucut	k,l,o	Percentage product that can be cut out of each age/quality combination of the raw material
PT	o,d	Product time per kilogram raw material in department d
q	p	Probability of scenario p
CT	i,d	Planned capacity in the solution of day T-1
IT	s,k,l	Inventory of the deterministic day in the solution of day T-1

Table 5.4: Variables of the SPP

Variable	indices of variable	Explanation
P	i,s,o,p	Kilogram produced of product o, in scenario p
F	i,s,k,l,p	Raw material used for production
S	i,s,k,l,p	Raw material used to sell
I	i,s,k,l,p	Inventory raw material in scenario p end the end of day i
EI	i,s,k,l	Expected inventory at the end of day i
C	i,d	Planned capacity on day T
B	i,s,l,p	Planned delivery of raw material in scenario p if i={1,2,3}, bought raw material if i=0

Next, we formulate the SPP as we solve it for each day of the planning horizon. We omit the explanation of the variables and indices in the model, as we gave them already.

$$\begin{aligned}
&\text{Objective: } \max \left(\sum_{p=1}^3 \left(\sum_{i=0}^3 \sum_s \sum_o P_{i,s,o,p} SP_{s,o} - \sum_{i=0}^3 \sum_s \sum_l B_{i,s,l,p} K_{s,l} - \sum_{i=1}^3 \sum_d C_{i,d} H_d \right) q_p \right) \\
&\text{subject to} \\
&1: EI_{i,s,k,l} = IT_{s,k,l} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall s, l, i = 0, k \in \{2, 3, 4, 5, 6\} \\
&2: EI_{i,s,k,l} = B_{i,s,l,p} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall s, l, i = 0, k = 1 \\
&3: EI_{i,s,k,l} = \sum_{p=1}^3 I_{i,s,k,l,p} q_p \quad \forall i, s, k, l \\
&4: I_{i,s,k,l,p} = EI_{i-1,s,k-1,l} + B_{i,s,l,p} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall i \in \{1, 2, 3\}, s, k = 1, l, p \\
&5: I_{i,s,k,l,p} = EI_{i-1,s,k-1,l} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall i \in \{1, 2, 3\}, s, k \in \{2, 3, 4, 5, 6\}, l, p \\
&6: P_{i,s,o,p} \leq F_{i,s,k,l,p} Ucut_{k,l,o} \quad \forall i, s, k, l, o, p \\
&7: P_{i,s,o,p} \leq S_{i,s,k,l,p} Ucut_{k,l,o} \quad \forall i, s, k, l, o, p \\
&8: \sum_s \sum_m P_{i,s,m} PT_{o,d} \leq C_{i,d} \quad \forall i, d \in \{2, 3, 4, 5\} \\
&9: \sum_s \sum_k \sum_l F_{i,s,k,l} PT_{o,d} \leq C_{i,d} \quad \forall i, d = 1 \\
&10: C_{i,d} \leq \text{workinghours} * \text{max employees} \quad \forall i, d \in \{1, 2\} \\
&11: C_{i,d} = CT_{i-1,d} \quad \forall i \in \{0, 1, 2\}, d
\end{aligned} \tag{5.1}$$

The goal function takes the weighted average of the profits in each scenario of 4 days, where day T is the deterministic day and thus has only one scenario. The profit in each scenario is determined by all the produced product multiplied with the product prices, minus the total price of all the bought fish, minus the planned capacity multiplied with the wages per hour. We note that $B_{i,s,l,p}$ does not depend on the age of the raw material, as the processing plant in Båtsfjord only buys fish from coastal vessels. They deliver fish at the same day. Therefore, all the purchased fish is one day old when delivered to LNS.

Constraints 1 and 2 ensure that in each scenario of the deterministic day, there cannot be more raw material used than available. For fish of 1 day old, this available raw material is equal to the delivered fish on day T . For fish older than 1 day, the available raw material is equal to the rest inventory of the deterministic day of day $T-1$. Since the inventory constraints in day 1 until 3 depend on the actual catch scenario, which is not certain yet, we take the expected inventory to compensate for this lack of knowledge. Literature applies this solution as well. Constraint 3 calculates the expected inventory. Note that in constraints 1 and 2, we take the expected inventory directly, as the inventories are not uncertain at day T . Constraint 4 and 5 make sure that the inventories each day in each scenario equal the expected inventory of the previous day minus all the used fish for production and selling. When the fish is one day old we add the purchased raw material to the right-hand side.

Constraint 6 and 7 ensure the usage of raw material is sufficient to produce all the products. Since we use a lower or equal sign here, it is not mandatory to produce all the products that are possible. This is because it might be optimal to only cut out fresh products when there is no more freezing capacity available. Constraint 8 and 9 make sure that there is enough capacity planned in all the work stations to produce all the products. Constraint 10 sets the maximum production capacity for the filleting and cutting department. We do not know the maximum capacity of the other station, but we showed in Section 2.3.3 that the capacity of the other stations is bigger than the filleting station. Finally, constraint 11 ensures that the capacity in the coming 3 days is equal to the planned capacity of day $T-1$ in the rolling planning horizon.

The model does not use information of the solution of day $T-1$ to forecast the catch in day T , something we expect LNS to do. Due to our limited time and scope, we choose not to research this further.

5.4 Evaluating the planning model

This section focusses on testing the quality of the solution of the model that we constructed in Section 5.3. We stated in Section 5.2.2 that there is a recursion between the days of the rolling planning horizon. Since it is not possible to evaluate the model using simple closed formulae, we use simulation. Section 5.4.1 explains how the supply information is generated. Section 5.4.2 focusses on the simulation process and how we collect the output. Finally, Section 5.4.3 finishes with the explanation of the warmup period, the number of runs, and how we use the output of the model.

5.4.1 Generating planned and realized supply

In Section 5.3 we created a model for the processing plant. Now we implement the supply information of Chapter 4 into this model to generate a supply and a forecast for each day of the rolling planning horizon. We need this information for the variable $B_{i,s,p,l}$ (see Equation 5.1) where we must make a forecast for day $T+1$, $T+2$, $T+3$. We made a forecast model for the number of tones of Cod with head caught by an individual fishing vessel, given a vessel size and age as vessel characteristics and the

visibility and western wind as weather characteristics. The catch per individual vessel is not dependent on the time of the year, as we showed the day of the year was not a significant variable in the model.

We observed in Section 2.3 that the total delivered fish to LNS varies per month. It is important to take these fluctuations into account since the capacity planning depends on the amount of raw material.

To simplify the analysis, we apply cluster analysis to create clusters with a comparable amount of bought fish per month. In this way we do not have to analyze each month individually. We want as few clusters as possible and we consider a cluster acceptable when each observation within a cluster does not significantly differ from the average within that cluster. We test this difference by applying a 1-sample t-test within each cluster, where we test on the biggest outlier in each group. When the biggest outlier is not statistically different from the mean, the other values are also not statistically different. We start with 2 clusters, when one of the clusters is not acceptable, we increase the number of clusters with 1, until all the clusters are acceptable. We obtain 4 clusters, which are shown in Table 5.6. We also apply an ANOVA test to see if the average values between the groups are significant different. The P-value of this test is 0.000. Appendix 13 shows the summaries of all the tests.

Table 5.6: Clusters for analysis

Cluster	Months	Average kilogram per month raw fish	P-value 1-sample T-test on biggest outlier
1	July, August, September	773,258	0.190
2	February, June, October, December	1,291,547	0.058
3	January, March, April, May	1,530,826	0.117
4	November	2,177,315	-

Since our forecasting model provides a forecast per fishing vessel we need to determine how many fishing vessels are delivering fish on an average day in each cluster. The amount of fishing vessels determines how much raw material LNS receives, which influences the amount of required capacity. To determine the average number of fishing vessels per day, we look at the average amount of delivered raw material in each cluster and divide this number by the expected number of kilograms per fishing vessel.

As we only made the prediction model for Cod, we need a method to forecast the other two species as well. We assume that the prediction for Haddock and Pollock is as accurate as our forecasting model for Cod. Because of this assumption we forecast the amount of delivered Haddock and Pollock as a fixed fraction of the amount of forecasted Cod in each cluster. We forecast the number of kilograms Cod per fishing vessel and assume that each fishing vessel consist of these fractions of delivered fish.

For example, in cluster 1 the total bought fish consisted on average of 80% Cod, 9% Haddock, and 11% Pollock. If we forecast a vessel to have 50 tons of Cod, we assume this vessel brings in 50 tons of Cod, 5.8 tons of Haddock, and 7 tons of Pollock. We use the distribution of the weather data and ship information to draw the predictors for the forecasting model, using a random number.

Table 5.7: Predictors and their distribution

Predictor	Expected value	Standard deviation
Vessel length	13.58	5.97
Vessel age	26.11	14.26
Visibility	8.3	1.85
Wind speed from the west	5.98	4.13

Table 5.7 shows the expected values and standard deviations of the prediction variables.

Since LNS could not provide historical data about the number of ships visiting the plant, we must make an estimate of this number. In Section 4.4 we read that the expected catch of Cod for a random vessel is 11,028 kg. We calculate the expected amount of daily fish delivered per specie.

With this we calculate the fractions of each specie and the expected daily number of ships coming to Båtsfjord. This number is the expected kg Cod delivered divided by the 11,028 kg, which is the expected delivery Cod for a random vessel.

Next, we want to know how many fishing vessels we have to generate each day in our evaluation model. Although we do not have historical data about the number of vessels visiting LNS, we expect that there are fluctuations in this number. Since we cannot find a probability distribution for the number of vessels, we have to make an assumption about this. A Poisson distribution is used to count the number of events within a time period. This describes the number of boats visiting the plant per day. Therefore, we choose to draw random numbers out of a Poisson distribution where the expected value is equal to the expected daily shipments of Table 5.8 within each cluster.

Finally, we need the probability of delivering good, regular, and bad quality fish by a random fishing vessel. The quality is always graded for the whole shipment of a vessel. In Section 2.3.1 we showed for 2,597 shipments the quality ratings. We calculate the probability of a good quality shipment to be 0.50, regular quality 0.46 and the probability for a bad quality shipment 0.04.

Then we program the module so that it does the following steps, until the simulation stops:

1. Generate 4 random values out of the corresponding distributions for the visibility, vessel age, wind speed from the west and vessel length.
2. If $T < 4$: Generate a three-day forecast for day $T+1$, $T+2$, and $T+3$ for the fish catch on day T , generate 3 forecast scenarios for each forecasted day. We now have 9 forecasts.
3. If $T < 4$: Generate a realized catch for each day, based on the different standard deviation for each forecasting day of Section 4.7.
4. If $T > 3$: Move up all the realized catches one day, from day T to day $T-1$. Since the standard deviation for the catch is lower at day $T-1$ than at day T (see Section 4.6), we make a new forecast for each day, and generate a new realized catch for day $T+3$ together with a forecast.

5.4.2 The simulation process explained and collecting the output

We explained the first two modules of the model in Section 5.3.1 and 5.3.2. We explain in this section how we connect all the modules and how we run the simulation. We explain the model, starting from $T=4$. We continue from step 4 of the prediction module of Section 5.3.2, where we generated a catch forecast and a realized catch for day $T=0$ until day $T=3$.

Table 5.8: Expected daily shipments in each cluster

Cluster	Expected kg Cod	Expected kg Haddock	Expected kg Pollock	Expected daily shipments
1	43,169.36293	4,994.63545	6,046.469	3.91
2	70,697.12319	12,942.20064	7,451.711	6.41
3	67,539.7343	13,514.89127	2,142.456	6.12
4	69,233.97498	11,826.47109	30,923.07	6.27

5. Retrieve the planned capacities for day T , day $T+1$ and day $T+2$ of the solution of the Stochastic Programming Problem for day $T-1$.
6. Retrieve the remaining inventories of the deterministic day of the solution of the SPP for day $T-1$.
7. Construct the SPP for day T , with the realized and forecasted catch sizes, the planned capacities of the previous day, and the remaining inventory of day $T-1$.
8. Solve the SPP, write the end inventory for day T , the planned capacity for day $T+1$, $T+2$ and $T+3$ to the central database of the optimization module. This information is needed to solve the SPP for the next day.
9. Write the profit of day T (the deterministic day), the capacity planning, and the dual values of the solution to the module Results for analysis later.
10. Increase the day counter with 1
11. Start at step 1 (Section 5.4.2).

Figure 5.2 shows the information flows between the modules and the simulation process schematically. Appendix 14 shows a detailed explanation of the model, where we show all the matrices that we use in the model. We omit the exact code in the appendix, as this takes over 12 pages to show.

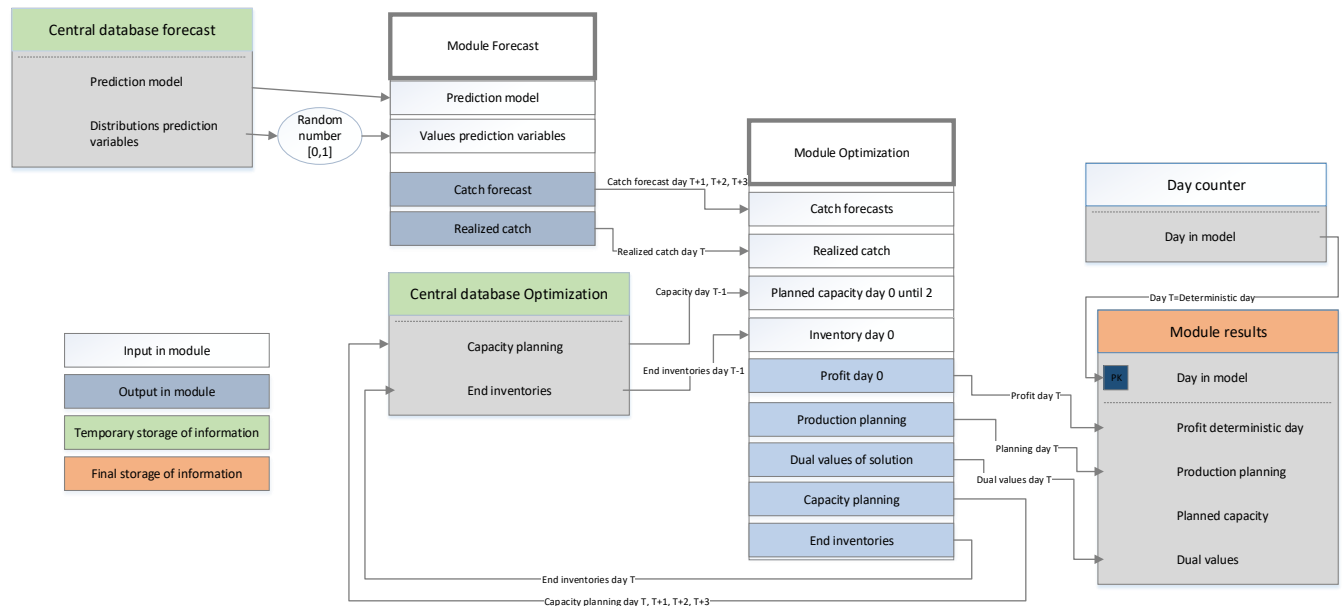


Figure 5.2: Simulation process schematically explained

5.4.3 Sensitivity analysis, warmup period, and number of replications

Finally, we want to know how the model performs with respect to the planned capacity. For this we use the dual values of the SPP. The dual values tell us how much we can vary each decision variable until the current solution is not optimal anymore.

We obtain the dual values for restriction 11 of the SPP. This restriction is always binding for the capacity, which means the dual value is never 0. When this dual value is lower than 0 it means the model planned more capacity than needed. This happens because the real delivery of raw material is lower than

forecasted. Higher than 0 means the model planned too little capacity than needed, due to the forecasting model forecasts less fish than actual delivered. We give two examples in Table 5.9, taking the filleting and cutting station as example. In the first row, the dual values of restriction 9 are positive for both stations, meaning the profit increases when more capacity is planned. However, the filleting station is already planned at its maximum capacity. In the second row, both duals are negative, meaning the profit increases when capacity is reduced (and thus the planned capacity is too high). We look at the minimum value of the corresponding dual to see how much we can decrease the capacity. For the filleting station this is 1131.63 minutes. Hence, we planned $1164.6 - 1584 = -32.97$ minutes too much capacity. For the cutting station this number is 45.23. To calculate how much reduction in profit this planning error incurred, we multiply the planning error with the corresponding dual value of the station.

Table 5.9: Two examples of output simulation model

Profit	Capacity Filleting	Capacity Cutting	Dual Filleting	From	To	Dual cutting	From	To
259007.16	1584.00	9088.97	90.93	1584.00	1584.00	0.61	9043.61	9088.97
164591.68	1164.60	6416.88	-8.33	1131.63	1584.00	-4.17	6371.65	10296.00

Next, we determine the number of required simulation runs based upon a corrected relative error (y') of output of the model. We set this number at 0.048. We define the output as the result of the goal function of the SPP. We find the minimum number of runs with:

$$n_r^* = \min \left(i \geq n : \frac{t_{i-1, 1-\alpha/2}^* \sqrt{S^2(n)/i}}{|\bar{X}|} \leq y' \right) \quad (5.2)$$

We proof in Appendix 15 that we need a minimum of 182 simulation runs for each cluster. As we included data of the warmup period, we expect the actual number of required runs to be even lower, but we maintain 182 as a minimum.

Finally, we need the warmup period for the simulation. The warmup period is necessary since we start our simulation model empty. It takes a certain number of runs so that the simulation is at its normal state. To determine this number of runs that cannot be used, we use Welch's graphical method (Law, 2015). For this method we construct a graph of the moving averages of the outputs of the simulation model for 20 replications of each 500 runs. We then must find from which run length the grey dotted graph is stable. Figure 5.3 shows the result and we can see that from approximately a run length of 50, the graph is stable. We thus set the warmup period at 50 runs. Hence, for each simulation, we discard the first 50 results. The full method is explained in Appendix 16.

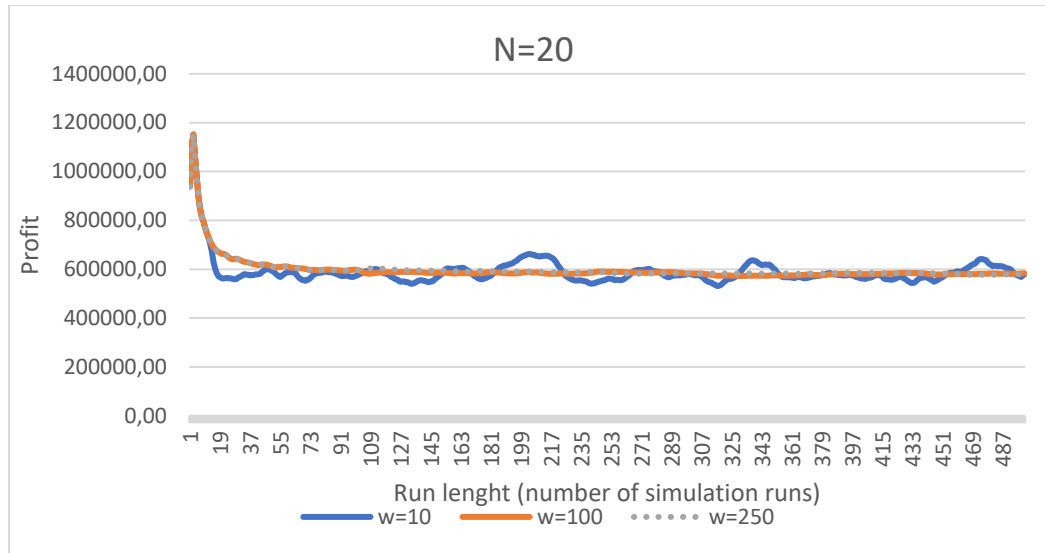


Figure 5.3: Welch's graphical method for 20 simulation runs of length 500

5.5 Conclusions of the plant model and simulation process

In this chapter we provided a method to plan the capacity in the processing plant in Båtsfjord given a forecasted catch size of the three species we investigate. We consider the processing times deterministic, and all the other variables continuous. Therefore, we propose to use Stochastic Programming to plan the capacity. As LNS knows three days in advance how many, and which ships come to deliver fish, and LNS must plan the capacity three days in advance, we plan the capacity for three days in advance.

As we saw in Section 2, the products that LNS can produce out of the raw material depends on the age and quality of the raw material. Therefore, there is a recursion in the days, making it difficult to use closed formulae. We solve this by using a rolling horizon planning, where we plan the capacity for day $T+3$ and optimize the production by producing the products that yield the highest profit.

To analyze the quality of the solution of the Stochastic Programming Problem, we integrate the forecasting model of Chapter 4 in the SPP. We use simulation to generate for each day (T) of the rolling planning horizon a realized and forecasted catch. Literature tells us we need to generate three different scenarios for the forecasted catch, as we do not know the opportunity costs of the capacity.

We conclude that we need a minimum of 182 simulation runs to make the corrected relative error smaller than 0.048. Also, we need 50 simulation runs to warm the model up. Therefore, we need a minimum of $182+50=232$ simulation runs in each cluster.

In the next chapter we present and analyze the results of the simulation outputs of the model we presented in this chapter.

6 Estimating the gains with the new planning method

In this chapter, we answer research question 5, which asked what savings LNS can expect when using the supply information in the capacity planning. To do so, we first give the variables in Section 6.1 that measure the gains and let us measure the performance of the model. In Section 6.2 we analyze the current situation measured in these variables. Section 6.3 analyzes the situation when the forecasting model is implemented in the capacity planning. Section 6.4 continues with the scenario where LNS uses variable capacity, against extra costs. Section 6.5 creates a scenario where LNS purchases an extra filleting machine, to increase the capacity of the bottleneck. Section 6.6 compares all the scenarios and tests whether the differences are significant. Section 6.7 chooses the best scenario and gives the calculations of the expected improvement. Finally, Section 6.8 gives the conclusions of this chapter.

6.1 Analyzing the output of the model: choosing the variables

We consider six output variables important to analyze the performance of the planning method:

1. The total error of the forecast hours, over all departments:

We calculate the forecast error of a department using the values of the dual solution of constraint 11, which set the capacity of day T until T+2 equal to the planned capacity in day T-1. We calculate the forecast hours using:

$$\text{Total forecast error} = \sum_{d=1}^5 |e_d| \quad (6.1)$$

Where : e_d = forecast error in hours at department d

We set the forecasting error at 0 when the dual value of a department is positive and when the maximum available capacity of this department is fully used. This means constraint 10 is the binding constraint, which set the planned capacity lower or equal than the maximum capacity.

2. Total lost profit in NOK:

We calculate the lost profit because of the forecasting error. When the maximum capacity of a workstation is insufficient to produce all available fish and this workstation is used at full capacity, we set the lost profit at 0. Otherwise we calculate the lost profit using:

$$\text{Lost profit} = \sum_{d=1}^5 |e_d| * |dual_d| \quad (6.2)$$

Where : $dual_d$ = the dual value of station d

3. Percentage lost due to overcapacity:

We want to know how much of the lost profit is due to overcapacity, and how much is due to under capacity. This information is useful when we compare this percentage with the percentage of days the maximum capacity is planned in the processing plant. We calculate this number using:

$$\text{Percentage overcapacity} = \frac{\text{Total lost profit}}{\text{Lost profit due to overcapacity}} \quad (6.3)$$

4. Kilogram fillet per cutting hour:

Since this KPI is the only reliable data we have about the current situation in the processing facility in Båtsfjord, we use this KPI to compare the output of the models with the current situation.

Although this variable does not say anything about the profit in the facility, it is useful to see how the forecasting models are performing compared with the current situation.

5. Profit per kilogram fish:

Since the number of kilogram raw material per simulation run is slightly different, we cannot simply compare the models using the generated profit. Therefore, we divide the profit by the amount of raw material delivered to see how much profit is generated out of the raw material. We use the deterministic profit (the profit of day T of the SPP). We consider this the most important variable, since the profit is the goal function of the SPP. Hence, the profit tells us how well the model is planning the capacity for LNS as we concluded in Section 2.4.

We use:

$$\text{Profit per kilogram day } T = \frac{NOK_T}{\sum_{s=1}^3 B_{T,s}}$$

(6.4)

$B_{T,s}$ = bought raw material day T of specie s

NOK_T = NOK profit on day T

6. Probability of running at maximum capacity

Finally, we determine the percentage of days the processing plant is operating at its maximum capacity. With maximum capacity we mean that restriction 10 is binding for the solution of the SPP. This can be for the filleting or cutting department, as we set constraint 10 for those two departments.

6.2 Current cost level and costs level with deterministic catch

LNS could not provide full data about the current capacity planning. We only have the planned capacity in the cutting department, which we presented in Section 2.3.4. We use this information to make a base scenario for variable 4 of Section 6.1, which is the kilogram fillet per cutting hour. Next to this, we simulate the current situation at LNS where the model uses only the information that LNS has available now, which is the average catch per fishing vessel. With this scenario, we compare what the impact of the forecasting model is on the profit levels. We compare different scenarios of the model using the profit per delivered kilogram of raw material.

Finally, we run the model where we omit constraint 11 from the model, so the model can plan the capacity based on the actual delivered raw material, eliminating the uncertainty in supply. We do this to verify that the model creates logical output, as we expect the profit to fall within the interval of this solution and the solution when taking the average catch per vessel as forecast.

We only give the mean values of the output of the model in this section. Appendix 17 and Appendix 18 show the confidence interval per cluster for the kilogram per cutting hour and the profit per kilogram raw material respectively.

6.2.1 Current productivity levels in each cluster

We can describe the current productivity levels best with the amount of kilogram produced fillet per cutting hour. Section 2.4 explains why. Table 6.1 shows these numbers together with the confidence intervals. We base the information on the full production year of 2015. Although the profits are also given in this dataset, it is not clear which costs are extracted to obtain these profits. We have no information about the overhead costs, such as electricity, management costs, rent of the processing plant etc. Therefore, we use the number of kilograms produces fillet per cutting hour as KPI, and compare each scenario based on this number.

Table 6.1: Current productivity according to data LNS

Cluster	Mean kilogram fillet per cutting hour	95% Confidence interval	
		Lower bound	Upper bound
Cluster 1	88.31	80.97	95.65
Cluster 2	101.80	92.40	111.19
Cluster 3	127.77	119.02	136.53
Cluster 4	88.12	78.44	97.81

6.2.2 Costs levels with deterministic catch

We create a situation where we remove the uncertainty in supply for LNS. With the deterministic catch we mean that LNS knows exactly how much raw material of each quality it gets supplied each day, removing the stochasticity of the supply. We achieve this situation by simply omitting constraint 11 out of the SPP. This means that the model plans exactly the amount of required capacity, when all the supply is known at day T. This means that the fluctuations in the delivery of raw material remain, but the model knows exactly how big these fluctuations are. Table 6.2 shows the results. Since there is no forecasting error or lost profit due to forecasting errors, we only give the kilograms per cutting hour and the profit per kilogram raw material. We compare the setups in the coming sections with this situation.

Table 6.2: Summary output deterministic catch

	Cluster			
	1	2	3	4
	Mean per day	Mean per day	Mean per day	Mean per day
Kilogram fillet per cutting hour	145.87	142.38	151.98	152.50
Profit per kilogram raw material	6.49	8.45	9.58	8.05

6.2.3 Costs levels when using average shipment per vessel as forecast

Since we do not have accurate costs information of the processing plant, we create a base scenario using the simulation model. We run the model using the average shipment per vessel as the forecasts for each day of the planning horizon. This means that if we expect 3 ships at day T+1 that the forecast for Cod for day T+1 is $3 \cdot 11028$ (the average Cod size for a vessel). We calculate the forecast for the remaining species and the different forecast scenarios as in Section 5.3.1 and 5.3.2.

Table 6.3 shows the result of the output of the model. We see that the forecast error in cluster 1 is relatively big compared to the other clusters. This is because in cluster 2 until 4, LNS gets delivered more fish, resulting in a higher occupancy rate of the processing plant. In these clusters at 45% of the days the plant was working at its maximum capacity. Hence, when there is much more fish available than can be produced in a single day, the model simply sets the capacity at its maximum. It is not important how much more fish is available, resulting in a smaller forecasting error. This results also in a lower percentage of lost profit due to overcapacity.

We consider this method as a lower bound for the profit, as we expect the profits to be higher when the model can plan the capacity with more accurate supply information.

6.3 Solution with forecasting model

We run the model with the same parameters and setup of the processing plant as in Section 6.2.2. The only difference is that we add the forecasting model for the fish of Section 2.4, instead of forecasting using the average catch for a fishing vessel. Table 6.4 shows the result of the outputs. We observe that the cutting department yields a higher productivity in all clusters. Also, the raw material is used better, given the increased value in the profit per kilogram fish. It seems that in cluster 3 the losses are increased, given the higher lost profit per day. However, this number is only relative to the current solution of the SPP. The solutions can be better than the solution in Section 6.2.2, while yielding a higher lost profit at the same time. The lost profit is for 100% caused by overcapacity in cluster 1, and for 83% to 98% in the other clusters. This tells us that the opportunity costs of producing are high. As a result, the model plans capacity for the scenario with the highest catch (see Section 5.3). This also explains the higher forecasting error compared with the solution of Section 6.2.3. The model plans, on average, more capacity than needed. This results in higher profits.

Table 6.3: Summary output when using average catch as forecast

	Cluster			
	1	2	3	4
	Mean per day	Mean per day	Mean per day	Mean per day
Total forecast error [hours]	70.67	3.13	3.12	3.12
Total lost profit [NOK]	20455.82	1552.31	1558.87	1558.87
Fraction lost profit due to overcapacity	.95	.73	.73	.73
Kilogram fillet per cutting hour	100.33	148.39	136.60	140.22
Profit per kilogram raw material	4.12	5.59	6.70	5.73
Probability of using maximum capacity	.17	.47	.46	.46

Table 6.4: Summary output with forecasting model

	Cluster			
	1	2	3	4
	Mean per day	Mean per day	Mean per day	Mean per day
Total forecast error (hours)	21.21	3.93	15.33	5.05
Total lost profit (NOK)	5668.34	1001.88	3905.51	1273.23
Fraction lost profit due to overcapacity	1.00	.83	.98	.92
Kilogram fillet per cutting hour	137.88	148.89	147.81	147.38
Profit per kilogram raw material	5.57	8.07	8.37	7.61
Probability of using maximum capacity	.02	.91	.83	.97

We further observe that in clusters 2, 3, and 4 the production was restricted due to exceeding of the maximum capacity in 83% to 97% of the days. In 77% and 97% of the cases, this is due to the filleting station. This supports the conclusions of Section 2.4, where we concluded that the filleting machine is the bottleneck.

We observe that the productivity and profit per kilogram raw material of all clusters lay within the deterministic solution and the solution of Section 6.2.3. This is what we anticipated and tells us the model is performing as expected.

6.4 Solution with possibility of overtime

Given the fact that the maximum capacity of one of the departments restricts the production in cluster 2, 3, and 4 for more than 80%, we suggest producing in overtime hours. When LNS uses overtime, it calls in an extra team of workers, which is not necessary to plan three days in advance. These workers can be called in on the beginning of the production day.

The costs of the employee wages are higher when producing in overtime, as the capacity is more flexible. The costs are 1.5 times higher than producing in regular time (Arbeidstilsynet, 2017). We assume that the time that can be produced in overtime is the same as in regular time, hence the maximum total production capacity per day is two times as big than in Section 6.3.

There are two more KPIs we measure with this setup. First, we measure the percentage of days that overtime production is used in at least one department. Next, we measure, with the help of the sensitivity analysis, how much the overtime costs could increase, given the current optimal solution. This would mean that if the overtime costs increase more than this number, overtime production is not profitable anymore. The probability of using maximum capacity gives now the percentage of days that the total capacity, including overtime, is restricting the production.

Table 6.5: Summary output with overtime model

	Cluster			
	1	2	3	4
	Mean per day	Mean per day	Mean per day	Mean per day
Total forecast error (hours)	8.80	5.09	4.83	15.70
Total lost profit (NOK)	2490.42	1355.58	1294.46	4224.43
Fraction lost profit due to overcapacity	0.99	0.98	0.97	0.97
Kilogram fillet per cutting hour	137.17	165.41	167.55	178.65
Profit per kilogram raw material	6.09	8.45	9.80	8.04
Probability of using maximum capacity	.00	.00	.01	.05
Probability of using overtime	.06	.84	.90	.93
Maximum increase overtime costs (NOK per hour)	-6.25	-6.31	-6.26	-6.31

Table 6.5 gives the results. We observe that in clusters 1, 3, and 4 the profit per kilogram fish has increased. Also, the productivity increases in clusters 2, 3, and 4.

6.5 Solution with capacity increase of filleting machine

In Section 6.3 we observe that in most of days where production is restricted by a work station, the filleting line is the bottleneck. In this section we analyze if, and how much benefits LNS will have when an extra filleting line is installed. We keep the possibility of overtime, as this increased the average profit per kilogram raw material in three of the four clusters.

Table 6.6 shows the result of this setup. We observe that the profit increases in cluster 2 and 4. In cluster 1 the profit decreases, which is understandable, as the extra capacity is unnecessary in this cluster. The forecast error increases in all clusters.

Table 6.6: Summary output with capacity increase model

	Cluster			
	1	2	3	4
	Mean	Mean	Mean	Mean
Total forecast error (hours)	9.48	7.80	4.97	16.87
Total lost profit (NOK)	2802.48	2242.09	1428.86	4708.67
Fraction lost profit due to overcapacity	.99	.96	.94	.98
Kilogram fillet per cutting hour	137.36	166.48	153.57	176.87
Profit per kilogram raw material	5.64	9.22	9.79	8.29
Probability of using maximum capacity	.00	.00	.00	.00
Probability of using overtime	.06	.84	.90	.93
Maximum increase overtime costs (NOK per hour)	-1.87	-1.91	-1.88	-0.71

6.6 Comparison of all the setups and testing the significance

In this section we compare the results of all setups. We compare both against the scenario where we assume LNS uses the average forecast of a vessel as prediction for the raw material, and the productivity in the cutting department, derived from actual data of LNS. Section 6.6.1 compares all the average values of the cost levels and the productivity. Section 6.6.2 continues with testing on the significance of the differences. This section starts answering research question 5.

6.6.1 Comparison of the costs levels

We compare the average values of the productivity (kilogram fillet produced per cutting hour) and the profit (profit per kilogram raw material) of the different processing plant setups. Only for the first value we compare the models with the current situation of the processing plant in Båtsfjord. Figure 6.1 shows the productivity of all models in each cluster. To clarify: the current situation corresponds with Section 6.2.1, the deterministic situation with Section 6.2.2, the average catch with Section 6.2.3, the regular forecast with Section 6.3, overtime with Section 6.4, and extra capacity filleting with Section 6.5.

We observe that the productivity for the overtime setup and the increase capacity setup is higher than the deterministic solution in cluster 2, 3, and 4. The explanation for this is that, due to the higher capacity, more raw material can be produced.

When there is not enough capacity, the model gives priority to Haddock and Pollock, as the whole fish prices are lower than for Cod. The cutting time for Haddock and Pollock is higher, resulting in a lower productivity.

We further observe that in only one of the 4 clusters the extra capacity setup yields the highest productivity. It is logical that this setup does not yield the highest productivity in the first cluster, since the extra capacity is not necessary in this scenario, given the low percentage of days where the maximum capacity is used. We also observe that in clusters 3 and 4 the overtime setup gives better results on average than the extra capacity setup.

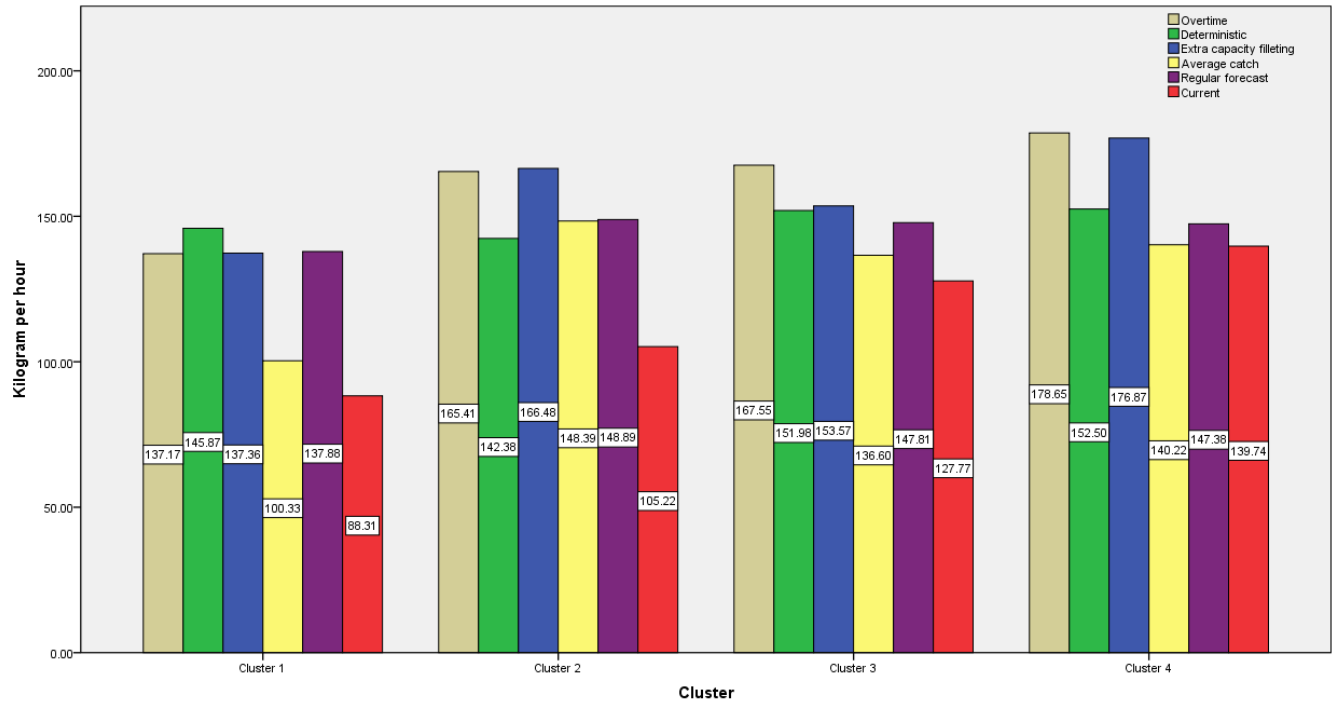


Figure 6.1: Comparison of average productivity in cutting department

Figure 6.2 shows the average profits in each cluster. We observe again that in the first cluster the extra capacity in the form of overtime, or the extra filleting machines yields the smallest differences compared with the regular forecast. In the clusters 3 and 4 these differences are larger. In all clusters all methods yield higher profits than using the average catch per vessel as forecast in the model.

We observe that in cluster 2, 3, and 4 the extra capacity does give the best results. This is logical, since more fish can be produced in regular time, due to the higher capacity of the bottleneck. Hence, the model uses less fish to sell without processing. As a result, the profits increase since taking the fish into production yields higher profit than selling it without processing (see Section 2.3).

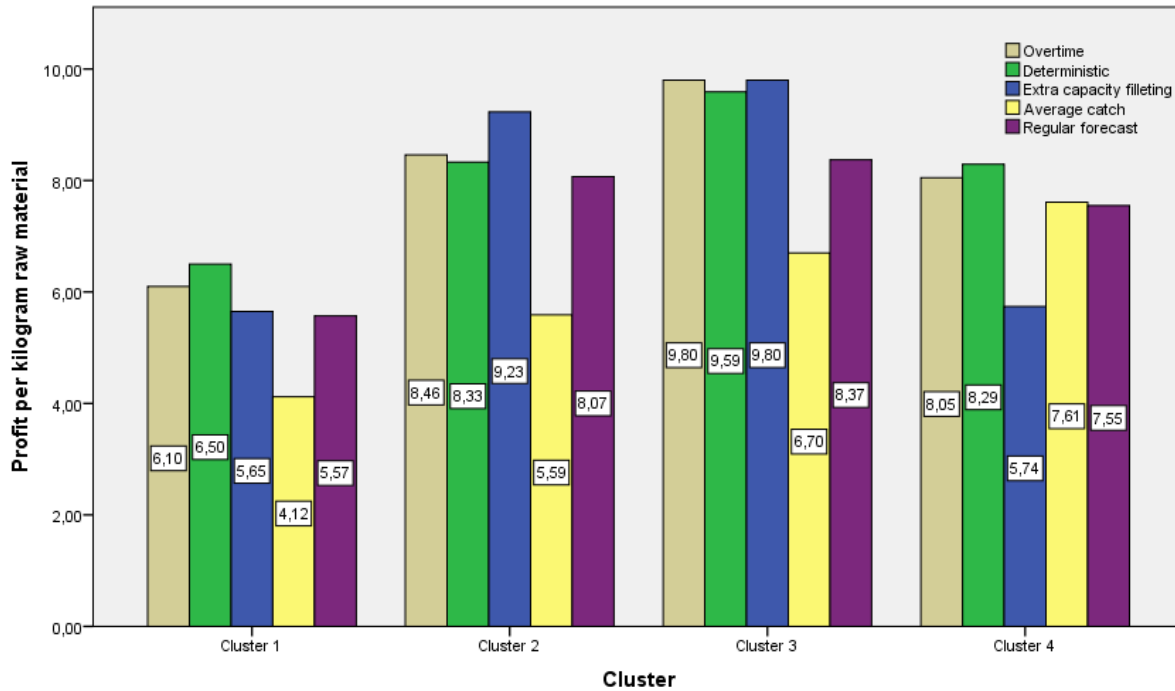


Figure 6.2: Comparison of average profit per kilogram raw material

6.6.2 Testing the significance of the differences

Although the mean values are different in the setups, we must test if these differences are statistically significant. We use ANOVA to test the differences between the methods within each cluster with a maximum error of 0.05. Appendix 19 shows the full results, including the pairwise comparison for all possible combinations in each group.

We provide the results in Table 6.7, which shows if a difference is significant. We give the results for the three methods that are potentially interesting to use: regular forecast, using overtime, and the increased filleting capacity in combination with overtime. Red means the difference is not significant, blue means the difference is significantly lower, and green means the difference is significantly higher.

Table 6.7: Significance table differences in productivity

	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
Regular forecast	Over-time	Extra capacity	Current situation	Over-time	Extra capacity	Current situation	Over-time	Extra capacity	Current situation	Over-time	Extra capacity	Current situation
	Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast	
Over-time	Regular forecast	Extra capacity	Current situation	Regular forecast	Extra capacity	Current situation	Regular forecast	Extra capacity	Current situation	Regular forecast	Extra capacity	Current situation
	Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast	
Extra capacity	Over-time	Regular forecast	Current situation	Over-time	Regular forecast	Current situation	Over-time	Regular forecast	Current situation	Over-time	Regular forecast	Current situation
	Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast		Deter-ministic	Average forecast	

We observe from Table 6.7 that the overtime method and extra capacity method improve the productivity significantly in all clusters, compared with using the average shipment as forecast. When compared with the current situation, all methods are significantly better in all clusters, except the regular forecast in cluster 4. There are no significant differences in the productivity between overtime and extra capacity in cluster 1, 2, and 4. In cluster 3 the extra capacity method is significantly worse than using only overtime. As stated in Section 6.6.1 we expected that the extra capacity setup would yield significant higher results than the overtime method.

Next, we show the significance table of the costs levels in Table 6.8. We observe that the overtime and extra capacity setups are significantly higher than the average forecast in all clusters. Between the first two methods there is no significant difference in any of the clusters, although the mean value of the extra capacity is higher than the overtime method (see Section 6.6.1). Appendix 20 shows the ANOVA with the pairwise comparisons of all methods.

Table 6.8: Significance table differences in profit

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Regular forecast	Overtime	Extra capacity	Overtime	Extra capacity	Overtime	Extra capacity	Overtime	Extra capacity
	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast
Overtime	Regular forecast	Extra capacity	Regular forecast	Extra capacity	Regular forecast	Extra capacity	Regular forecast	Extra capacity
	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast
Extra capacity	Overtime	Regular forecast	Overtime	Regular forecast	Overtime	Regular forecast	Overtime	Regular forecast
	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast	Deterministic	Average forecast

6.7 Choosing the best setup and determination of the expected improvement

In this section, we choose the best method based on the average number of kilograms fillets processed per cutting hour, and the average profit per kilogram delivered raw material. We compare each method with the created base scenario; the method where LNS uses the average shipment as forecast. The extra capacity method needs an investment in the form of an extra cutting machine. We take this investment into account and analyze if the investment is profitable, using the net present value (NPV) of the investment. In Section 6.7.1 we decide which methods are most promising. Section 6.7.2 finalizes with analyzing if the extra filleting machine yields a positive NPV. This section finalizes the answer to research question 5.

6.7.1 Choosing between the three potential methods

We have three potential methods that potentially improve our base scenario. We compare the scenarios based on Table 6.8, which shows if the differences in profit per kilogram fish are significant. We observe that both the overtime, and the extra capacity method significantly improve the profit, and productivity of the cutting station, in all clusters, compared with the base scenario and the current situation. The difference for the regular forecast method is not significant in cluster 4. The overtime method does not require any extra investments. Also, although not significant, the average profit is higher than for the regular forecast method and the productivity is significantly higher in two of the four clusters. Due to

these facts, we choose this method over the regular forecast method, given the overtime costs are within the range of Table 6.5.

With one method eliminated, we focus on the remaining two methods. In the next section we analyze the NPV of the investment of an extra filleting machine, to see if the investment is profitable.

6.7.2 Net present values of the remaining methods

The NPV gives the value of an investment, corrected for the time preference of the cash flows. Money is worth more if received now than it is after 1 year. We calculate the NPV using (Wijst, 2012):

$$NPV(i, N) = \sum_{t=0}^N \frac{R_t}{(1+i)^t}$$

Where: R_t = cash flow in year t
 i = discount rate (the return that could be made when not investing in this project)
 N = Period of the project (in years)

(6.5)

We set the discount rate at 0.1, as this value is generally used and reflects the return that could be earned when investing the money in the open market (Wijst, 2012). The initial investment in year 0 consist of the purchase and installation of the extra filleting machine. The cost of a similar used filleting machine from 2008 is around 2.1 million NOK (Fishmachinery.com, 2017). We add 900.000 NOK for the installation, which makes the total investment 3 million NOK. We assume that the machine has 5 years of production time left and that the rest value of the machine is 0.

The cashflows consist of the cost savings of using the extra filleting line, compared with only using overtime. The differences between these methods are not significant in any cluster, which means 0 is in the confidence intervals of the differences. To analyze the risk of the investment, we use the distribution of the differences of the profits between the situation with the extra filleting line, and the situation where only overtime is used. We set the probability such that the NPV of the 5-year project is exactly 0. Hence, we look for the probability that the investment will break even. We calculate this using:

$$p_{break\ even} = \left\{ p \in [0,1] : \sum_{t=1}^5 \frac{R_t}{(1.1)^t} = 0 \right\}$$

$$R_t = \sum_{c=1}^4 F^{-1}(p, \bar{X}_c, s_c) * B_c$$

Where: $p_{break\ even}$ = Probability of break even
 \bar{X}_c = Mean difference of costs saving cluster c
 s_c = Estimated standard deviation costs saving cluster c
 B_c = Total bought kilograms of fish in cluster c
 $F^{-1}(p, \mu_c, s_c)$ = Inverse of the normal distribution

(6.6)

We solve Equation 6.6 and determine the probability of breakeven to be 0.433. Hence LNS has approximately 43% change of making less money with buying an extra filleting machine, compared with using only overtime. Figure 6.3 shows the probabilities together with the expected NPV.

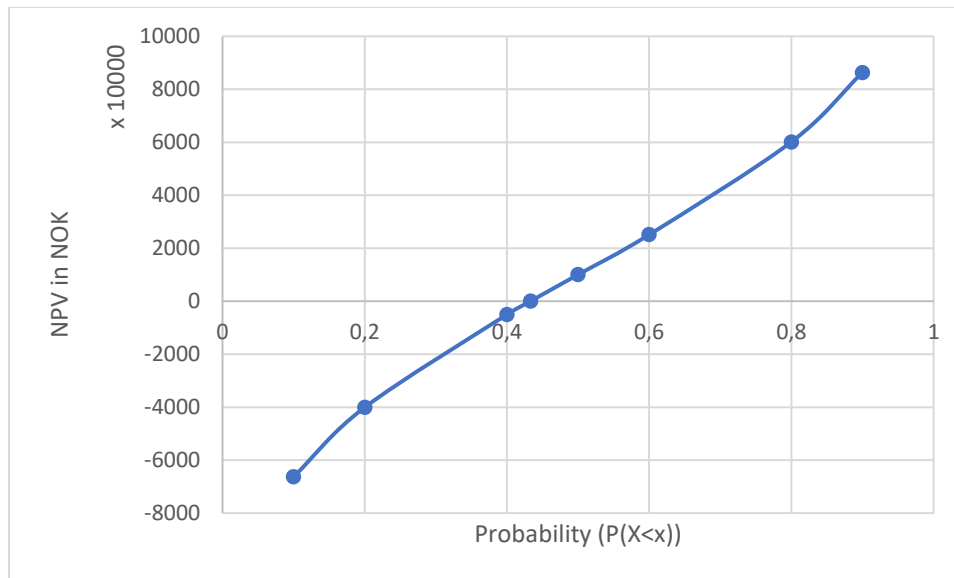


Figure 6.3: Total NPV compared with probabilities

Next, we calculate the cost saving when using overtime, compared with the base scenario of using the average catch as forecast. The expected extra profit, when LNS uses the average shipment as forecast, is 43.4 million NOK per year. The real cost savings are probably lower in practice, as we expect LNS to use a more dynamic, and better planning method than using the average catch per vessel.

Also, the productivity increases when using overtime. When compared with the current situation at LNS, the productivity increases on average with 42%. Compared with the average forecast method this increase is 20%. Table 6.9 shows the confidence intervals of the differences of the productivity and the profit when comparing all the scenarios we created with the current situation and the average forecast method. We observe that the overtime scenario and the extra capacity scenario are indeed better than only using the forecasting method. All three methods significantly improve the current situation, as the lower bound of the confidence intervals are positive for all comparisons.

Table 6.9: Confidence interval increase in productivity and profit

Method	KPI	Confidence interval 95% Current situation		Confidence interval 95% Average forecast	
		Min	Max	Min	Max
Regular forecast	Productivity	14%	28%	3%	8%
	Profit	No data		2%	76%
Overtime	Productivity	15%	59%	13%	23%
	Profit	No data		10%	73%
Extra capacity	Productivity	20%	41%	12%	22%
	Profit	No data		18%	90%

6.8 Conclusions about the gains when using the forecasting method

In this chapter we showed the potential gains for LNS when using the forecasting method, we created in chapter 4 in the capacity planning. For this, we use the model of the processing plant in Båtsfjord we showed in Chapter 5.

We create three scenarios to improve the current situation:

1. Use the forecasting method in the model of the processing plant;
2. Use scenario 1 in combination with the possibility of using overtime in the form of flexible workers against extra costs;
3. Use scenario 2 in combination with an extra filleting machine.

To make a proper analysis we define two important KPIs that we use to compare the different scenarios: the profit per kilogram raw material and the kilogram processed filet per cutting hour. We have real data of LNS for the last mentioned KPI. To generate the profit KPI we create an own scenario, where LNS uses the average catch per vessel as forecast method.

We compare the differences in the productivity and profit between the three scenarios and the real data and average forecast method. We observe that the lower bound of the confidence interval is positive in each scenario, meaning both three methods significantly improve both the profit as the productivity.

Although upper bound of the confidence interval of the profit KPI shows high values (from 73% to 90% increase) we must be critical about this number. We expect that LNS currently uses a better forecasting method than using the average catch per vessel. Since the central planner makes the forecasts based on experience, it is not possible for us to simulate this planning method.

Finally, we conclude that method 2 and 3 perform better than method 1. Method 3 requires an investment. We compare the probability of yielding less profit when using scenario 3 compared to using scenario 2. This probability is 42%. Further, we conclude that the differences in profits between scenario 2 and 3 are symmetric, meaning the probability of saving 100 NOK is the same as losing 100 NOK. As LNS made losses from 2011 until 2015, we expect the willingness to take risks is low. Therefore, we propose to use method 2, as this method does not require investments.

7 Conclusions, recommendations, and discussion

This last chapter finalizes the research by drawing the conclusions, giving the recommendations and discussing the results. Section 7.1 draws the conclusions and answers the main research question. Section 7.2 continuous with the recommendations to LNS. Section 7.3 finishes with the discussion.

7.1 Conclusions

We started this research with our main research question, which was:

“How can information about the raw material supply be used to improve the capacity planning within the processing facilities of LNS to reduce under- and overcapacity of labor and optimize the use of the raw material”

We split this research question in 5 sub questions, which we answered during the research.

We saw that the production processes at LNS have a production-oriented layout, and that all products require the same production steps, with the exception when a product needs to be frozen. Also, we can consider the processing times to be deterministic. This helped us modeling the processing plant using an exact method. Literature shows us that Stochastic Programming is used before to solve comparable problems with uncertain supply. We conclude further that planning the capacity accurately is important for LNS, as too much capacity results in lost money due to extra salary, and too little capacity results in the delay of production of raw material, deteriorating the raw material and making it less profitable to produce. Hence, the profit generated by the raw material reflects the quality of the capacity planning.

Since the processes are deterministic, we expect the number of kilograms processed fillet per cutting hour to be relatively stable. However, we saw big fluctuations in this number, from which we conclude that the current capacity planning is not optimal and has room for improvement.

Out of the three possibilities we derived from the literature to reduce the uncertainty in supply, we chose, based on three criteria we used to measure the quality of the possibility, to predict the fish catch per vessel based on causal models. We created 1 regression model and 3 machine learning models and conclude that the M5P Tree gives the smallest error for our dataset (S is 0.5). Hence, this method is best qualified to predict the supply of a vessel.

We used information of the forecasting model in the Stochastic Programming Model, since this needs the expected catch as input. We conclude that the model performs as expected, since the output of the different methods we created are logical. To analyze the output, we created three methods; a method that uses the average catch of a vessel as forecast, one method that only uses the forecasting model, and one method where we eliminate the uncertainty in supply. We expect the last method to give the highest profits, as there is no uncertainty in this case. We expect the first method to give the lowest profits, as the supply forecasts are the worst in this method. Hence, we expect the performance of the method with the forecasting model to fall within the other two methods. This is indeed the case, justifying the conclusion that the model performs as expected.

We also conclude that the raw material supply is significantly different in some months, making it not possible to base the analysis on a single month, as the raw material supply influences the production planning. Therefore, we create as few clusters as possible, combining months where the raw material supply is not significantly different. We conclude that the capacity in three out of the four clusters is

insufficient to produce all the fish LNS gets supplied into fillet products. As a result, LNS must sell whole fish, or delay the production, resulting in loss in profits. Further we conclude that producing in overtime is easier, since too little planned capacity can be solved using overtime hours. Overtime production increases the productivity in the cutting department significantly, compared with using only the forecasting model and producing in regular time.

Using overtime production in combination with the forecasting model we created generates an increase in profit of 43.4 million NOK. Also, the productivity (kilogram processed filet per cutting hour) in the cutting department increased significantly compared with the current situation at LNS, which we determined using actual data from the processing plant in Båtsfjord. We conclude that reducing the uncertainty in supply, using the forecasting model, is valuable for LNS. We also conclude that using the forecasting model is most profitable in situations where the processing plant is not used close to full capacity, as the forecasting model yields the best improvements in the first cluster, both for the profit as the productivity. This is because when there is more capacity needed than available, it is not necessary anymore to accurately plan the supply. LNS can simply plan the capacity to its maximum level.

7.2 Recommendations

We saw in the conclusions that the forecasting model improves the current situation when we compare the results with the base scenario we created. Both the productivity in the cutting station as the profit per kilogram raw material increased significantly. Therefore, we recommend using this forecasting model in the capacity planning of the processing stations.

The case when using an extra filleting line compared with using overtime hours yields the highest profits. However, LNS needs to invest in an extra filleting line. We calculated the Net Present Value of this investment, over the expected lifetime of the filleting machine, which we determined on five years. Although the expected value of the NPV is positive, we do not recommend investing in the extra filleting line. The expected NPV is approximately 1 million NOK over five years. However, there is a probability of around 40% of losing money on the investment. Since the differences in profit follow a normal distribution, the probability of losing a given amount of money, is the same as saving the same amount of money. The volatility of the NPV is therefore high, as Figure 6.3 shows. Therefore, we do not recommend investing money in the extra filleting capacity. We recommend using the forecasting method we developed in Chapter 4 in combination with the use of overtime against a cost increase of 1.5 times the cost level in regular production time. We showed that this setup increases the productivity both in our created base model and in the real situation at LNS.

When the overtime costs increase with more than 6.25 NOK per hour, compared with the costs of using regular time, the current solution is no longer optimal (see Table 6.5). If this is the case, we advise LNS to stop using overtime, and produce only in regular time using the forecast model we provided in Chapter 4.

7.3 Discussion

In this research we showed the value of reducing supply uncertainty for LNS. This conclusion matches with the information the literature provides us. We showed also that, with our dataset, it is possible to forecast the supply of fishing vessels using weather information and ship characteristics. Although literature states that it is possible to forecast catch sizes with weather information, not all the predictors mentioned in the literature are significant in our model. The moon phase, which the literature considers

a significant predictor, is insignificant in our model. The exact values of the predictors can differ when applied on a different dataset. However, we have no reasons to doubt the forecast errors of our models, as our dataset covered a whole year and there were enough data points to apply a regression model.

We created the forecasting model based on catch and weather information in Troms. We used the forecasting model in the region Finnmark. We did not create or test our forecasting model in the Finnmark region. Therefore, we cannot conclude that the forecasting model in the Finnmark region is as accurate as in the Troms region. Although the forecasting model could be slightly different in the region Finnmark, the goal of our research was to show that it is possible to forecast the catch size and use this information to improve the capacity planning. Also, we only forecasted the Cod catch, and assumed the forecasts for Haddock and Pollock are as accurate as for Cod. This does not have to be true, as it could be possible that Haddock and Pollock are more difficult to predict. We give the same argument as before, stating that we showed the value of forecasting the catch sizes for LNS was the goal of our research.

We did not obtain accurate data of LNS considering the capacity planning in the past. Therefore, it was not possible for us to compare the outcomes of our model accurately with the current situation at LNS. We therefore created an own base scenario, in which we do not use our forecasting model, to compare the different setups with. This base scenario uses the average catch per fishing vessel as forecast for the raw material supply. Although this creates a scenario where LNS plans the capacity without having access to the forecasting model, we expect LNS to use a more effective and accurate way of planning in the current situation. We observe an increase in the profit of 10% to 73% using overtime compared with the base scenario. Since we do not have accurate information of the profit of LNS we cannot conclude that LNS has this potential, but we showed the value of using the forecasting information compared with using the average catch size per vessel.

8 Bibliography

Allen, M. (2015). *HEALTH SERVICES AND DELIVERY RESEARCH*.

Arbeidstilsynet. (2017). Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. (arbeidsmiljøloven).

Begen, M. A. (2003). Development of a Catch Allocation Tool Design for Production Planning at JS McMillan Fisheries. *INFOR: Information Systems and Operational Research*.

Belouch, M. (2017). A Two-Stage Classifier Approach using RepTree. *International Journal of Advanced Computer Science and Applications*.

Britishseafishing. (2016). Retrieved from <http://britishseafishing.co.uk/techniques-and-info/weather-and-sea-conditions/>

Courtney, A. J. (1996). Lunar periodicity in catch rate and reproductive condition of adult Eastern King prawns, *Penaeus plebejus*, in coastal waters of South-eastern Queensland, Australia.

Fishmachinery.com. (2017). Retrieved from <https://www.fishmachinery.com/details/curio-c2011-filleting-machine/>

Fiskeridir. (2009). Tabell over offisielle norske omregningsfaktorer for fisk.

Fiskeridirektoratet. (2017). Retrieved from <http://www.fiskeridir.no/register/fartoyreg/?m=frtyoppl&f=1>

Hameri, A.-P. (2010). Supply chain management in the fishing industry: the case of Iceland. *International Journal of Logistics Research and Applications* .

Hameri, A.-P. (2010). Supply chain management in the fishing industry: the case of Iceland. *International Journal of Logistics Research and Applications* .

Harrell, F. (2001). *Regression Modeling Strategies*.

Henriksen, E. (2013). *Lønnsom foredling av hvitfisk i Norge - hva skal til?* Tromsø.

Jensson, P. (1988). Daily production planning in fish processing firms . *European Journal of Operational Research* .

Kalita, J. (2009). *Decision and Regression Tree Learning*. Retrieved from <http://www.cs.uccs.edu/~jkalita/work/cs586/2013/DecisionTrees.pdf>

Kazemi Zanjani, M. (2008). A multi-stage stochastic programming approach for production planning with uncertainty in the quality of raw materials and demand. *International Journal of Production Research*.

Kimms, A. (1997). *Multi-Level Lot Sizing and Scheduling*. Physica-Verlag Heidelberg.

Larsen, R. J. (2012). *An Introduction to Mathematical Statistics and Its Applications*.

Law, A. M. (2015). *Simulation Modeling and Analysis* .

Li, X. (2012). Are Linear Regression Techniques Appropriate for Analysis When the Dependent Variable Is Not Normally Distributed?

- Liaw, A. (2002). Classification and Regression by randomForest.
- LNS (1). (2017). Råraver førstehåndkjøp.
- LNS (2). (2017). *Innsikt Hvitfisk*.
- LNS (3). (2017). Kvalitet på råstoff. Båtsfjord.
- LNS (4). (2017). Prisliste Hvitfisk.
- LNS (5). (2016). Dashboard Norway Seafoods; Dashboard Master.
- LNS (6). (2017). Kutteplan.
- LNS (7). (2018). Kapasitet per lokasjon.
- LNS (8). (2016). Dashboard Norway Seafoods; Budget2.
- Maritime and Coastguard Agency. (2012, 10 09). *Fishing vessels: classification, registration and inspection*. Retrieved from <https://www.gov.uk/guidance/fishing-vessel-classification-registration-and-inspection>
- Norges råfisklag. (2017). Retrieved from Samlet omsetning 2016 - kvantum omregnet til rund vekt: http://www.rafisklaget.no/portal/page/portal/RafisklagetDokumenter/Aarsomsetning/omsetningsstat_2016.pdf
- Norway Seafoods. (2017). Retrieved from <http://bedrift.norwayseafoods.com/text.cfm?path=538,571,572&lid=1>
- Norway Seafoods AS. (2015). *Konsernregnskapet*.
- Norwayexports. (2017). Retrieved from <http://www.norwayexports.no/sectors/articles/norwegian-seafood-export-sets-new-records-2016/>
- Norwegian Ministry of Fisheries and Coastal Affairs. (2013). *Facts about Fisheries and Aquaculture 2013*.
- O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb.
- Peidro, D. (2008). Quantitative models for supply chain planning.
- Pettersen, K. (2017, 10 25). Interview Central Planner LNS.
- Pflug, C. (2001). Scenario tree generation for multiperiod Financial optimization by optimal discretization.
- Rafisklaget. (2017). Retrieved from <http://www.rafisklaget.no/portal/page/portal/NR/PrisogStatistikk/Minstepris>
- Sakia, R. M. (1992). The Box-Cox Transformation Technique: A review.
- Slack, N. (2008). *Operations Management*.
- Smunt, T. L. (1996). Rough Cut Capacity Planning in a Learning Environment . *IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT*.

- Statistisk sentralbyrå. (2013). *Samlet areal, arealfordelinger og kystlinjens lengde1, etter fylke. 2013*. Retrieved from <http://www.ssb.no/a/aarbok/tab/tab-019.html>
- Statistisk sentralbyrå. (2016, 06 02). *Aquaculture, 2015, preliminary figures*. Retrieved from <https://www.ssb.no/en/jord-skog-jakt-og-fiskeri/statistikker/fiskeoppdrett/aar-forelopige/2016-06-02#content>
- Statistisk sentralbyrå. (2017, 1 12). Retrieved from <https://www.ssb.no/en/jord-skog-jakt-og-fiskeri/statistikker/fiskeri>
- Statistisk sentralbyrå. (2017, 09 06). Retrieved from <https://www.ssb.no/en/nasjonalregnskap-og-konjunkturer/nokkeltall/key-figures-for-national-accounts>
- Statistisk sentralbyrå. (2017). Retrieved from <https://www.ssb.no/303334/total-area-distribution-of-area-and-length-of-coastline-by-county.2017-sy-19>
- Stergiou, K. (1997). Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods . *Fisheries Research* 29.
- Thakur, M. (2017). *Processing mapping in the whitefish supply chain - analysis of information and material flow*. Sintef.
- The World Bank. (2013). *FISH TO 2030 Prospects for Fisheries and Aquaculture*. Washington DC.
- Trevisan, L. (2011, 03 09). Combinatorial Optimization: Exact and Approximate Algorithms.
- Tveterås, R. (2014). *Til nærings- og fiskeridepartementet*. Tromsø.
- Ullman, J. D. (2007). The Running Time.
- V. D. R. Guide JR., R. S. (1997). An evaluation of capacity planning techniques in a remanufacturing environment. *International Journal of Production Research*.
- Vorst, J. G. (1998). Supply Chain Management in Food Chains: Improving Performance by Reducing. *Int. Trans. Opl Res.*
- Walpole, R. E. (2012). *Probability and Statistics for Engineers and Scientists, 9th Edition*. Pearson.
- Walpole, R. E. (2016). *Probability & Statistics for Engineers & Scientists*.
- Wiersema, M. T. (1992). Customer Intimacy and Other Value Disciplines.
- Wijst, N. v. (2012). *Finance: A Quantitative Introduction* . Cambridge.
- Wikstöm, J. (2015). *Evaluating supervised machine learning algorithms to predict recreational fishing success*. KTH ROYAL INSTITUTE OF TECHNOLOGY.
- Witten, I. H. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. Elsevier Inc.

9 Table of content appendices

Appendix 1	Conversation ratios for different fish types	92
Appendix 2	Kolmogorov-Smirnov test response variable	93
Appendix 3	Box-Cox transformation	94
Appendix 4	Correlations between predictors	95
Appendix 5	Model with all predictors	96
Appendix 6	Residual tests for normality	97
Appendix 7	Algorithm for the REP Tree.....	100
Appendix 8	REP Tree.....	101
Appendix 9	Algorithm of the M5P Tree.....	102
Appendix 10	M5P Tree.....	103
Appendix 11	Algorithm of the random forest.....	104
Appendix 12	Code Monte Carlo simulation	105
Appendix 13	Tests for cluster analysis	108
Appendix 14	Detailed explanation of simulation model.....	110
Appendix 15	Determination of the amount of simulation runs	118
Appendix 16	Determination of warmup period.....	120
Appendix 17	Summary statistics kilogram per cutting hour	123
Appendix 18	Summary statistics profit per kilogram raw material	124
Appendix 19	ANOVA productivity cutting department	125
Appendix 20	ANOVA profit per kilogram raw material.....	128

Appendix 1 Conversation ratios for different fish types

Omregningsfaktorer specificeret på produkttilstande / conversion factors specified on product categories																								
NS	F/AO	Filaset	Species	Sløjde produkter / gutted products				Fyngback				Fillet produkter / fillet products*								Andre produkter/other ..				
Kode	Kode			Med hoved	Uden hoved	Med hoved	Uden hoved	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	
9400				Med hoved	Uden hoved	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	
				Rund snit	Fladt snit	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	Med skind	Uden skind	
				Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	
				Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	Guided	
				Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	Head on	Head off	
				Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	Round cut	
				Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	Manuel	
				210	211	214	212	311	311	361	362	510	511	513	512	515	514	521	520	410	412	513	211	
				NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	NS9400 and listandskote	
0811	ELE	AI	European eel	1,10	1,10																			
0821	COE	Havål	European conger	1,10	1,10																			
0899	COX	Uspes. ål	Congers, eels	1,10	1,10																			
0911	GAR	Horngjel	Garfish	1,00	1,00																			
0921	SAU	Makrellgjedde	Atlantic saury	1,00	1,00																			
1011	MOR	Mora	Moras	1,20	1,40																			
1021	USK	Brosme	Tusk	1,20	1,40																			
1022	COD	Torsk	Atlantic cod	1,18	1,50	1,55																		
102201	COD	Skrei	Atlantic cod / Skrei	1,18	1,50	1,80																		
1023	LIN	Lange	Ling	1,20	1,40																			
1024	BLJ	Bjålange	Blue ling	1,20	1,40																			
1025	GFB	Skjelbrosme	Greater forkbeard	1,20	1,40																			
1027	HAD	Hyse	Headbok	1,14	1,40	1,65																		
1032	POK	Sei	Saithe	1,20	1,35	1,60																		
1034	POL	Lyr	Pollock	1,15	1,30																			
1035	POC	Polarorsk	Polar cod	1,20	1,40																			
1036	NOP	Øyepål	Norway pout	1,00	1,00																			
1038	WHB	Karmule	Blue whiting	1,20	1,40																			
1039	WHG	Hvitling	Whiting	1,20	1,40																			
1051	HKE	Lysing	European hake	1,20	1,40																			
1061	RHG	Isqait	Roughthead grenadier	1,20	1,40	2,46 1/2																		
1062	RNG	Skolest	Roundnose grenadier	1,20	1,40																			
1099	GAD	Diverse torskellisk	Gadiformes	1,20	1,40																			
1321	ORY	Orange Roughy	Orange Roughy	1,20	1,65																			
1411	SPF	St. Petersfisk	John dory	1,13																				
1611	HOM	Tagmakrell (Hessmakrell)	Atlantic horse mack.	1,10	1,55																			

* Maskinelt flænsning og manuelt flænsning af fiskekødet

Appendix 2 Kolmogorov-Smirnov test response variable

To test the catch statistics for normality we use the Kolmogorov-Smirnov test. This test measures the maximum distance between the empirical data and the theoretical distribution, which is the normal distribution in our case. The test statistic is calculated with:

$$D = \max_{1 \leq i \leq N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right) \quad (9.1)$$

Where : $F(Y_i)$ = Theoretical cumulative normal distribution

The two hypotheses are then:

H_0 : The catch data follows a normal distribution

H_1 : The catch data follows another distribution than the normal distribution

We choose alpha to be 0.05 and execute the test on the catch data. We obtain the result as in Figure 9.1a We conclude that the p-value is lower than 0.05 and we thus reject the hypothesis that the data is normal distributed.

When taking the natural logarithm of all the data points and executing the test again, we get the result as in Figure 9.1b. We observe that the test statistic is lower from which we can conclude the data is more normally distributed. However, we still must reject H_0

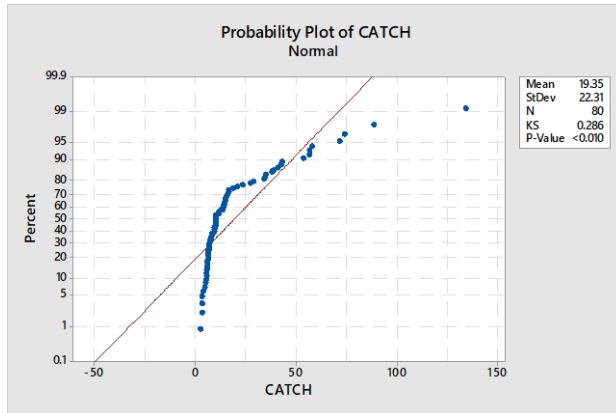


Figure 9.1a: Data before transformation

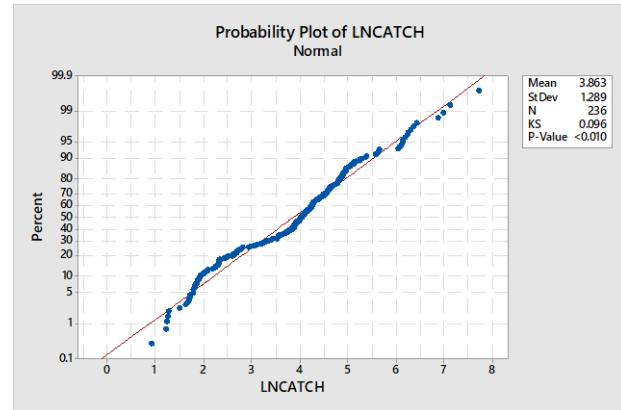


Figure 9.1b: Data after transformation

Appendix 3 Box-Cox transformation

We will try to normalize the data using a transformation. A common used transformation for this is the Box-Cox transformation (Sakia, 1992). The transformation is calculated with:

$$Y'_\lambda = \frac{Y^\lambda - 1}{\lambda} \quad (9.2)$$

$$\lambda \neq 0$$

Since we want to transform the data such that it approximates a normal distribution, we want to choose λ such that the correlation between the catch and the normal theoretical quantities is maximized, hence:

$$\max_{\lambda} \left\{ \rho(\Phi^{-1}(\frac{i-0,5}{n}), \frac{Y_i^\lambda - 1}{\lambda}) \right\} \quad (9.3)$$

When Equation 9.3 the value of λ tends to converge to 0. This means there is a limit of 0/0 in Equation 9.3, and we can apply l'Hôpital's rule to find the value:

$$\lim_{\lambda \rightarrow 0} \frac{Y_i^\lambda - 1}{\lambda} \stackrel{l'Hopital}{=} \frac{\frac{d}{d\lambda}(Y_i^\lambda - 1)}{\frac{d}{d\lambda} \lambda} = \frac{Y_i^\lambda * \ln(Y_i)}{1} = \ln(Y_i) \quad \forall i \quad (9.4)$$

Hence, the best transformation for the data is to take the natural logarithm from the original data.

Appendix 4 Correlations between predictors

	SUN H	MOO N	PRES SURE	HUM IDITY	VISIB IL	WIN DDIR	AVG WIN D	RAIN	AVTE MP	HP	AGE	WIDT H
LENG TH	N.S	N.S.	- 0.138	0.153	N.S.	N.S.	0.13	N.S.	N.S.	N.S.	N.S.	0.896
WIDT H	N.S	N.S.	N.S.	N.S.	N.S.	N.S.	0.167	N.S.	N.S.	N.S.	- 0.225	
AGE	0.152	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.		
HP	N.S	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S	N.S.			
AVGT EMP	0.397	0.244	0.13	- 0.334	0.429	N.S.	N.S.	N.S.				
RAIN	N.S	0.145	- 0.371	0.486	- 0.281	0.244	0.375					
AVG WIN D	N.S	0.143	- 0.536	0.281	N.S.	N.S.						
WIN DDIR	N.S	N.S.	N.S.	0.155	- 0.244							
VISIB L	N.S	0.166	N.S.	- 0.334								
HUM IDITY	-0.42	N.S.	- 0.545									
PRES SURE	0.371	- 0.272										
MOO N	N.S											

Appendix 5 Model with all predictors

Model Summary

S	R ²	R ² (adj)	R ² (pred)
0.498663	67.26%	64.71%	60.26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.535	4.67	0.92	0.360	
LENGTH	0.0456	0.0140	3.27	0.001	6.59
AVGTEMP	-0.00252	0.00599	-0.42	0.674	1.95
RAIN	0.0135	0.0117	1.15	0.251	1.98
AVGWIND	0.0076	0.0100	0.76	0.447	1.67
WINDDIR	0.000558	0.000468	1.19	0.235	1.19
VISIBIL	0.0377	0.0229	1.65	0.100	1.68
HUMID	-0.00524	0.00565	-0.93	0.354	2.46
PRESSURE	-0.00127	0.00442	-0.29	0.774	2.24
SUNH	0.0016	0.0101	0.15	0.878	1.82
MOON	-0.00611	0.00458	-1.33	0.183	1.27
WIDTH	0.2287	0.0630	3.63	0.000	7.51
AGE	-0.00711	0.00300	-2.37	0.019	1.74
HP	-0.000010	0.000079	-0.12	0.903	1.06
MATERIAL					
KOMPOSITT	-0.053	0.212	-0.25	0.802	1.56
PLAST	-0.051	0.136	-0.38	0.706	4.16
STÅL	0.228	0.207	1.10	0.272	4.51
TRE	0.055	0.168	0.33	0.745	3.30

Appendix 6 Residual tests for normality

To test if the residuals are normally distributed we use the Chi squared test. With this test we can test if observed data follows a different distribution than the one we test on. The higher the p-value of this test, the higher the probability that the observed data follows our hypothesized distribution.

The two hypothesizes are:

H_0 : The distribution of the errors follows a normal distribution

H_1 : The distribution of the errors follows another distribution than the normal distribution

The test statistic of this test equals:

$$\chi^2 = \sum_{i=1}^k \frac{o_i - N * p_i}{N * p_i}$$

Where : k = amount of bins used (9.5)
 o_i = observed observations in bin i
 p_i = probability of an observation belonging to bin i

To determine the number of bins, we use $2 \log n$, where n denotes the number of observations in our dataset. N is equal to 236, which means we use 8 bins if we round the number to the nearest integer. That means that we have 8-1 is 7 degrees of freedom. The critical value of a right tailed Chi squared distribution with 7 degrees of freedom is 14.07. Hence if our test statistic is higher than 14.07, we reject H_0 and we conclude that our distribution is not normally distributed.

9.1 Linear regression

Figure 9.2 shows the result of the Chi-squared test for the residuals of the linear regression model. We observe a p-value of 0.98, meaning we do not reject H_0 and there is no reason to assume the residuals are not normally distributed.

bin	from	to	n*p	p(cum)	observed	contribution to chi squared
1	-100000	-0.58584	29.625	0.125	29	0.013185654
2	-0.58584	-0.34331	29.625	0.25	28	0.089135021
3	-0.34331	-0.16195	29.625	0.375	31	0.063818565
4	-0.16195	0.00045	29.625	0.5	29	0.013185654
5	0.00045	0.162847	29.625	0.625	33	0.384493671
6	0.162847	0.34421	29.625	0.75	29	0.013185654
7	0.34421	0.586737	29.625	0.875	25	0.722046414
8	0.586737	10000	29.625	1	32	0.190400844
					Test statistic	1.489451477
					Critical value	14.06714045
					P-Value	0.982672603

Figure 9.2: Chi-squared test linear regression

REP Tree

Figure 9.3 shows the result of the Chi-squared test for the residuals of the REP Tree. The p-value is higher than 0.05. As a result, we can conclude that there is no reason to assume the distribution of the residuals do not follow a normal distribution.

bin	from	to	n*p	p(cum)	observed	contribution to chi squared
1	-100000	-0.59943	29.5	0.125	26	0.415254237
2	-0.59943	-0.3491	29.5	0.25	30	0.008474576
3	-0.3491	-0.1619	29.5	0.375	30	0.008474576
4	-0.1619	0.00572	29.5	0.5	43	6.177966102
5	0.00572	0.173343	29.5	0.625	27	0.211864407
6	0.173343	0.36054	29.5	0.75	20	3.059322034
7	0.36054	0.61087	29.5	0.875	32	0.211864407
8	0.61087	100001	29.5	1	28	0.076271186
					Test staistic	10.16949153
					Critical value	14.06714045
					P-Value	0.179170288

Figure 9.3: Chi-squared test REP Tree

M5P Tree

The p-value of the Chi squared test on the residuals of the M5P Tree is 0.88, as Figure 9.4 shows. Again, we have no reason to doubt that the data does not follow a normal distribution.

bin	from	to	n*p	p(cum)	observed	contribution to chi squared
1	-100000	-0.57637	29.625	0.125	29	0.013185654
2	-0.57637	-0.34585	29.625	0.25	31	0.063818565
3	-0.34585	-0.17346	29.625	0.375	32	0.190400844
4	-0.17346	-0.01911	29.625	0.5	33	0.384493671
5	-0.01911	0.135252	29.625	0.625	32	0.190400844
6	0.135252	0.307636	29.625	0.75	25	0.722046414
7	0.307636	0.538156	29.625	0.875	23	1.481540084
8	0.538156	10000	29.625	1	31	3.045886076
					Test statistic	6.091772152
					Critical value	12.59158724
					P-Value	0.880725179

Figure 9.4: Chi-squared test MSP Tree

Random forest

Finally, the p-value for the test on the residuals of the random forest is equal to 0.74 as Figure 9.5 shows. We again draw the same conclusions as for the other algorithms.

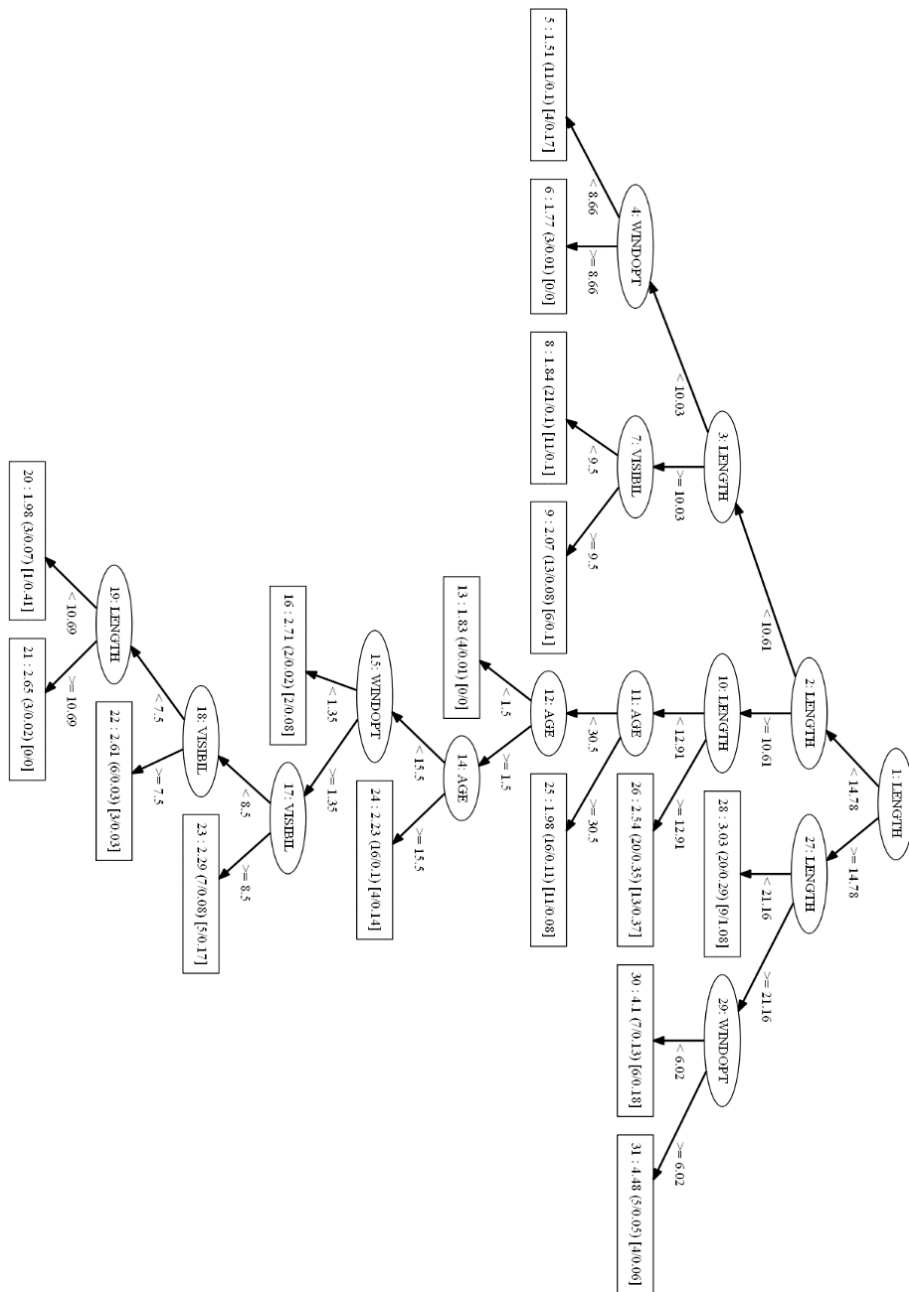
bin	from	to	n*p	p(cum)	observed	contribution to chi squared
1	-100000	-0.58491	29.625	0.125	26	0.443565401
2	-0.58491	-0.34653	29.625	0.25	27	0.232594937
3	-0.34653	-0.16827	29.625	0.375	33	0.384493671
4	-0.16827	-0.00864	29.625	0.5	38	2.367616034
5	-0.00864	0.150979	29.625	0.625	32	0.190400844
6	0.150979	0.329242	29.625	0.75	30	0.004746835
7	0.329242	0.567624	29.625	0.875	25	0.722046414
8	0.567624	10000	29.625	1	25	4.345464135
					Test statistic	8.69092827
					Critical value	12.59158724
					P-Value	0.739237113

Figure 9.5: Chi-squared test random forest

Appendix 7 Algorithm for the REP Tree

1. Split data randomly 10 times in training set and test set and set k to 0
2. Do while $k < 10$
3. Build a fully-grown decision tree, which optimizes Straining based on training data and calculate $Stest$
4. Do while $Stest > Stest'$ //do as long as there is an improvement
5. Set $Stest'$ to ∞ and i to the number of the last node in the tree //start in the last node
6. Do for all nodes in the tree
7. Pick node i and prune, including all branches below it and calculate $Stest''$
8. If $Stest' < Stest''$ update $Stest'$ to the value of $Stest''$ and remember node, update i to $i-1$
9. If all nodes in tree has been examined go to 10, otherwise go to 7
10. If $Stest' < Stest$, update the tree by pruning the node, update $Stest$ to the value of $Stest'$ and go to 4
11. Construct the smallest possible tree with lowest $Stest'$, store value in S_k , update k to the value of $k+1$ and go to 2
12. Calculate the estimated error from the model by taking the average of all S_k

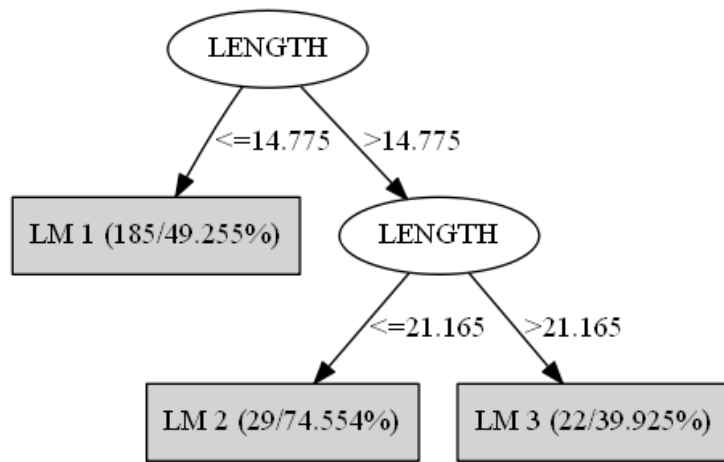
Appendix 8 REP Tree



Appendix 9 Algorithm of the M5P Tree

1. Split data randomly 10 times in training set and test set and set k to 0
2. Do while $k < 10$
3. Set i to 1
4. Create node i, calculate $sd(T)$, set i to 1
5. Do while (number of nodes) $> i$ //do as long there are open nodes
6. Set SDR' to ∞
7. For all available attributes do
8. Calculate SDR' with optimal split for chosen attribute
9. If $SDR' < SDR$, update SDR to the value of SDR' , remember the attribute and calculate $sd(T_i)$
10. If all attributes have been checked, go to 11, otherwise go to 7
11. If (number of attributes node i) < 5 or $sd(T_i) / sd(T) < 0.05$,
12. Then: make node into a leaf, update i to the value of i+1 and go to 4
13. Else: Split node according to attribute from line 8, in optimal way, update i to the value of i+1 and go to 4
- //The tree is now built, and the linear models will be constructed from here on
14. Set i to 1
15. Do while $i < (\text{number of nodes and leaves})$
16. Create linear regression model based on T which minimizes error in training data
17. Store error in value $error_i$
18. update i to the value of i+1 and go to 15
- //From here on, the tree will be pruned
19. Set i to max(nodenumber) //set i to the last node (ignoring the leaves)
20. Do while $i > 0$ //do until the most upper node
21. If $error_i > \text{average}(error_{sk})$, prune subtree S_k
22. Update i to the value of i+1 and go to 20
- //The final tree will be created
23. Create the smallest tree with the lowest total error, store this value in S_k , update k to the value of k+1 and go to 2
24. Calculate the estimated error from the model by taking the average of all S_k

Appendix 10 M5P Tree



Appendix 11 Algorithm of the random forest

1. Split data randomly 10 times in training set and test set and set k to 0
2. Do while $k < 10$
3. Take randomly n samples from the training data and set i to 1
4. While $i < n$ do //do for every sample
5. If no node created in tree i , create node 1 and set m to 0
6. If $m = |\text{amount of nodes in } i|$ update i to the value of $i+1$ and go to 4 //check if there are open nodes, if there are not, start with the next tree
7. do while $|\text{attributes at node } t| > 1$ //continue as long there are attributes at a node
8. Select k random attributes and set j to 1
9. While $j < k$ do
10. Select attribute j , sort values from small to large and call smallest value (i) 1, do for all values
11. Calculate SDR and store attribute and splitting value if $\text{SDR}(i,j) > \text{SDR}(i-1,j)$
12. Split node according to highest $\text{SDR}(i,j)$, calculate S_{training} and keep split if there is a reduction. Name nodes $(t+1)$ and $(t+2)$, update m to the value of $m+2$ and go to 6
- //the random trees are now completed, and the test data will be used from here on
13. Set i to 1
14. Do while $|\text{number of instances test data}| > i$
15. Do for all n samples //do for all the generated random trees
16. Classify instance i according to random tree and store value (C_j)
17. Sum all C_j over j and divide by n //this is the predicted value for instance i
18. Calculated the squared error compared to the observed value and store error in e_i
19. Update i to the value of $i+1$
20. Sum e_i over i and divide by i , store this value in S_k , update k to the value of $k+1$ and go to 2 //this is the standard deviation of the test data
21. Calculate the estimated error from the model by taking the average of all S_k

Appendix 12 Code Monte Carlo simulation

We use Excel VBA for the Monte Carlo simulation to estimate the standard deviation of the error when the weather variation is considered. First, we provide the Excel sheet where all the information is stored in Figure 9.6.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Ship predictors					Linear regression model								
2	LENGTH	VISIBIL	WINDOPT	AGE	LNCATCH	LENGTH	VISIBIL	WINDOPT	AGE	Intercept	REPETITIONS		Variance	0.704991
3	25.33	11	20.53	15	4.830558		0.104606	0.0376	0.02113	-0.01046	0.856	10000	STDEV error	0.839638
4	28.18	9	6.28	29	4.780558	STDEV	4.836698							
5	24.15	10	15.87	40	4.590558									
6	23.1	8	7.09	26	4.520558									
7	38.08	8	11.99	18	4.520558	Prediction model for WINDOPT								
8	27.42	10	2.63	16	4.480558									
9	36.39	7	5.06	15	4.480558	Intercept	*VISIBIL	STDEV						
10	27.49	8	10.25	41	4.440558	WINDOPT	11.21	-0.63	3.972					
11	21.34	10	1.11	29	4.430558	M5P Tree	<	14.775	lm1					
12	27.91	9	4.36	16	4.430558		<	21.165	lm2					
13	22	8	2.02	29	4.380558				lm3					
14	37.9	9	4.12	10	4.380558		M5P Tree							
15	38.05	6	17.33	13	4.370558		LENGTH	VISIBIL	WINDOPT	AGE	Intercept			
16	39.7	10	2.89	29	4.340558	lm1	0.1547	0.0028	0.0142	-0.0104	0.566			
17	27.41	6	7.8	28	4.200558	lm2	0.1011	0.0341	0.0127	-0.0159	1.378			
18	27.43	7	5.06	16	4.200558	lm3	0.044	0.039	0.0298	-0.0062	2.4817			
19	20.93	10	4.73	35	4.150558									

Figure 9.6: Excel sheet for Monte Carlo simulation

First, we provide the code for the linear regression model.

```

Sub sim()

SumE = 0

For i = 1 To Cells(2, 11) 'Read the amount of simulation runs
    rand = Int(Rnd() * 237 + 2) 'Draw a random ship
    predVISIBIL = Cells(rand, 2) + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(3, 7) 'predict the visibility
    If predVISIBIL > 10 Then
        predVISIBIL = 10
    End If
    If predVISIBIL < 0 Then
        predVISIBIL = 0
    End If
    predWINDOPT = 11.21 - 0.63 * predVISIBIL + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(7, 9) 'predict the WINDOPT
    If predWINDOPT < 0 Then
        predWINDOPT = 0
    End If
    Prediction = Cells(2, 10) + Cells(rand, 1) * Cells(2, 6) + predVISIBIL * Cells(2, 7) + Cells(2, 8) * predWINDOPT + Cells(rand, 4) * Cells(2, 9) 'Make a forecast
    E = Cells(rand, 5) - Prediction 'Determine the error
    SumE = SumE + E * E 'Square the error and add this to the sum of errors
Next i

ErrorE = SumE / i 'Calculate the average error
Cells(1, 16) = ErrorE
Cells(2, 16) = Sqr(ErrorE)

End Sub

```

Next, we show the code for the M5P Tree.

```

Sub M5E()

SumE = 0

For i = 1 To Cells(2, 11)

    rand = Int(Rnd() * 237 + 2)
    predVISIBIL = Cells(rand, 2) + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(3, 7)
    If predVISIBIL > 10 Then
        predVISIBIL = 10
    End If
    If predVISIBIL < 0 Then
        predVISIBIL = 0
    End If

    predWINDOPT = 11.21 - 0.63 * predVISIBIL + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(7, 9)
    If predWINDOPT < 0 Then
        predWINDOPT = 0
    End If

    If Cells(rand, 1) < Range("H10") Then 'Check in which leave the prediction falls
        Prediction = Cells(rand, 1) * Range("g14") + Range("h14") * predVISIBIL + Range("i14") * predWINDOPT + Cells(rand, 4) * Range("J14") + Range("K14")
    ElseIf Cells(rand, 1) > Range("H10") And Cells(rand, 1) < Range("H11") Then
        Prediction = Cells(rand, 1) * Range("g15") + Range("h15") * predVISIBIL + Range("i15") * predWINDOPT + Cells(rand, 4) * Range("J15") + Range("K15")
    ElseIf Cells(rand, 1) > Range("H11") Then
        Prediction = Cells(rand, 1) * Range("g16") + Range("h16") * predVISIBIL + Range("i16") * predWINDOPT + Cells(rand, 4) * Range("J16") + Range("K16")
    End If

    E = Cells(rand, 5) - Prediction
    SumE = SumE + E * E
Next i

ErrorE = SumE / i
Cells(1, 16) = ErrorE
Cells(2, 16) = Sqr(ErrorE)
End Sub

```

Finally, the code for the REP Tree.

```

Sub REPTree()
|
SumE = 0

For i = 1 To Cells(2, 11)
    rand = Int(Rnd() * 237 + 2)
    predVISIBIL = Cells(2, 10) + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(3, 7)
    predWINDOPT = 11.21 - 0.63 * predVISIBIL + WorksheetFunction.Norm_S_Inv(Rnd()) * Cells(7, 9)

    'from here the model checks in which leave the prediction falls
    If Cells(rand, 1) < 10.03 Then
        If predWINDOPT < 8.66 Then
            If predVISIBIL < 9.5 Then
                Prediction = 3.46
            Else
                If predVISIBIL > 9.5 Then
                    Prediction = 3.73
                Else
                    If predWINDOPT > 8.66 Then
                        Prediction = 3.85
                    Else
                        If Cells(rand, 1) > 10.03 And Cells(rand, 1) < 10.61 Then
                            If predVISIBIL < 9.5 Then
                                Prediction = 3.92
                            Else
                                Prediction = 4.15
                            End If
                        End If
                    End If
                End If
            End If
        End If
    Else
        If Cells(rand, 1) > 10.61 And Cells(rand, 1) < 14.78 Then
            If Cells(rand, 1) > 12.91 Then
                Prediction = 4.62
            Else
                If Cells(rand, 3) > 30.5 Then
                    Prediction = 4.06
                Else
                    If Cells(rand, 3) < 1.5 Then
                        Prediction = 3.91
                    End If
                End If
            End If
        End If
    End If
End Sub

```

```

Else
If Cells(rand, 3) > 15.5 Then
Prediction = 4.31

Else
If Cells(rand, 1) > 12.23 Then
Prediction = 4.78

Else
If predVISIBIL > 8.5 Then
Prediction = 4.39

Else
If predVISIBIL > 7.5 Then
Prediction = 4.72

Else
If Cells(rand, 1) > 10.69 Then
Prediction = 4.73

Else
Prediction = 4.06
End If
End If
End If
End If
End If
End If
End If
End If

Else
If Cells(rand, 1) > 14.78 Then
If Cells(rand, 1) < 21.16 Then
Prediction = 5.11

Else
If predWINDOPT < 6.02 Then
Prediction = 6.18

Else
Prediction = 6.56

```

```

End If
End If

End If
End If
End If

E = Cells(rand, 5) - Prediction
SumE = SumE + E * E

Next i

ErrorE = SumE / i
Cells(1, 16) = ErrorE
Cells(2, 16) = Sqr(ErrorE)

End Sub

```

Appendix 13 Tests for cluster analysis

To test if the clusters are acceptable, we use the one-sample T test, where we test:

Null hypothesis $H_0: \mu = \mu_x$

Alternative hypothesis $H_1: \mu \neq \mu_x$

Where μ_x is the observation that has the greatest deviation from the mean within the cluster. First we present all the summary statistics in Table 9.1

Table 9.1: Summary statistics clusters

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Cluster 1	3	0	773258	26657	46172	721271	721271	789009	809493	809493
Cluster 2	4	0	1291547	18389	36779	1273158	1273158	1273158	1328325	1346715
Cluster 3	4	0	1530826	84938	169875	1375826	1381328	1515739	1695410	1715997
Cluster 4	1	0	2177315	*	*	2177315	*	2177315	*	2177315

Next, we perform the 3 T-test on the first three clusters. Since cluster 4 only has one observation it is not necessary and not possible to perform a t-test on this cluster.

One-Sample T: Cluster 1 Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
3	773258	46172	26657	(658560, 887956)

μ : mean of Cluster 1

Test

Null hypothesis $H_0: \mu = 721271$

Alternative hypothesis $H_1: \mu \neq 721271$

T-Value	P-Value
1.95	0.190

One-Sample T: Cluster 2 Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
4	1291547	36779	18389	(1233024, 1350070)

μ : mean of Cluster 2

Test

Null hypothesis $H_0: \mu = 1346715$

Alternative hypothesis $H_1: \mu \neq 1346715$

T-Value	P-Value
-3.00	0.058

One-Sample T: Cluster 3 Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
4	1530826	169875	84938	(1260517, 1801135)

μ : mean of Cluster 1

Test

Null hypothesis $H_0: \mu = 1715997$

Alternative hypothesis $H_1: \mu \neq 1715997$

T-Value	P-Value
-2.18	0.117

Finally, we perform an ANOVA on the clusters to confirm that the mean values between the groups are significantly different. The test has p-value of 0.000 and a F-value of 50.94, confirming the mean values are different.

One-way ANOVA: Cluster 1, Cluster 2, Cluster 3, Cluster 4

Method

Null hypothesis All means are equal
 Alternative hypothesis Not all means are equal
 Significance level $\alpha = 0.05$
Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	Cluster 1, Cluster 2, Cluster 3, Cluster 4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	1.81256E+12	6.04187E+11	50.94	0.000
Error	8	94894268643	11861783580		
Total	11	1.90745E+12			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
108912	95.03%	93.16%	*

Means

Factor	N	Mean	StDev	95% CI
Cluster 1	3	773258	46172	(628256, 918260)
Cluster 2	4	1291547	36779	(1165971, 1417122)
Cluster 3	4	1530826	169875	(1405250, 1656401)
Cluster 4	1	2177315	*	(1926164, 2428466)

Pooled StDev = 108912

Appendix 14 Detailed explanation of simulation model

In this appendix, we provide a detailed explanation of the simulation model and the detailed information within the modules. We constructed the model in Delphi. First, we present the whole model again in Figure 9.7.

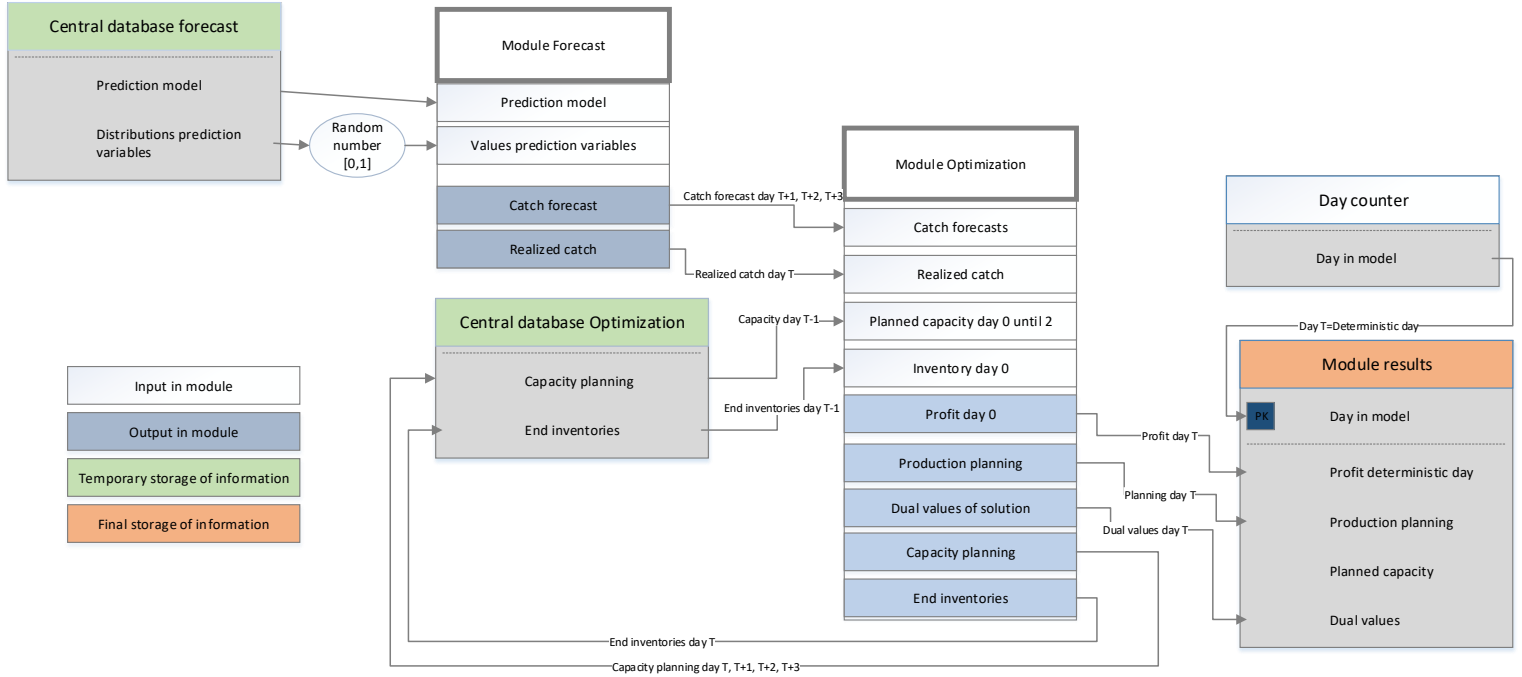


Figure 9.7: Connection of modules within the model

Module forecast

The forecast module generates catch sizes for each specie and quality. For this, it uses the forecasting model of Section 4, together with the values of the distribution of the prediction variables. These are stored in two matrices:

$$\begin{aligned} SHIPINFO &= \begin{bmatrix} \bar{X}_{length} & S_{length} & \bar{X}_{age} & S_{age} \end{bmatrix} \\ SHIPINFO &= \begin{bmatrix} 13.58 & 5.97 & 26.11 & 14.26 \end{bmatrix} \end{aligned} \quad (9.6)$$

$$\begin{aligned} WEATHERINFO &= \begin{bmatrix} \bar{X}_{visibility} & S_{visibility} & \bar{X}_{windopt} & S_{windopt} \end{bmatrix} \\ WEATHERINFO &= \begin{bmatrix} 8.3 & 1.85 & 5.98 & 4.13 \end{bmatrix} \end{aligned} \quad (9.7)$$

To translate the Cod forecast to a realized catch, we need the standard deviation of a ship for day T+1, T+2, and T+3. With the standard deviation of the forecast error we can calculate a forecasting error. The standard deviations are stored in STDEVSHIP:

$$\begin{aligned}
STDEVSHIP &= S_{forecast\ error\ day0} \quad S_{forecast\ error\ day1} \quad S_{forecast\ error\ day2} \quad S_{forecast\ error\ day3} \\
STDEVSHIP &= 0 \quad 0.459 \quad 0.462 \quad 0.471
\end{aligned}
\tag{9.8}$$

Next, we need the information about the probability of a ship delivering a good, regular or bad quality. We store this information in the matrix:

$$\begin{aligned}
QUALPROB &= \begin{bmatrix} P_{good\ quality} & P_{regular\ quality} & P_{bad\ quality} \end{bmatrix} \\
QUALPROB &= \begin{bmatrix} 0.50 & 0.45 & 0.04 \end{bmatrix}
\end{aligned}
\tag{9.9}$$

Also, we need the composition of the fishing vessel with respect to the three species. We use the fractions of the delivered fish to Båtsfjord.

$$\begin{aligned}
FACTOR &= \begin{bmatrix} Cod_{scenario1} & Haddock_{scenario1} & Pollock_{scenario1} \\ Cod_{scenario2} & Haddock_{scenario2} & Pollock_{scenario2} \\ Cod_{scenario3} & Haddock_{scenario3} & Pollock_{scenario3} \\ Cod_{scenario4} & Haddock_{scenario4} & Pollock_{scenario4} \end{bmatrix} \\
FACTOR &= \begin{bmatrix} 0.80 & 0.09 & 0.11 \\ 0.78 & 0.14 & 0.08 \\ 0.62 & 0.10 & 0.14 \\ 0.81 & 0.16 & 0.03 \end{bmatrix}
\end{aligned}
\tag{9.10}$$

The module creates a catch size each day together with a forecast for three days. This information is stored in the matrix TOTFISH.

$$\begin{aligned}
TOTFISH &= [\text{day}, \text{specie}, \text{prediction day}, \text{quality}, \text{scenario}] \\
TOTFISH &= [0..3, 1..3, 1..3, 1..3, 1..3]
\end{aligned}
\tag{9.11}$$

Figure 9.9 shows the information flows within the module forecasting schematically.

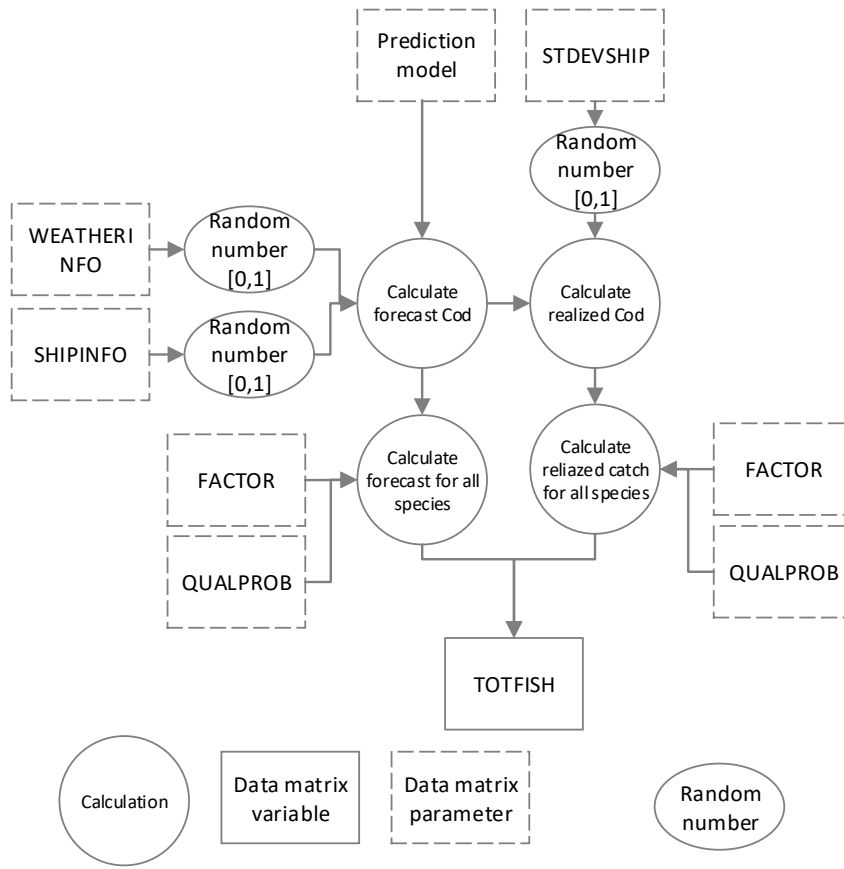


Figure 9.9: Information flows within module forecasting

Module Optimization

After the model creates a forecast, and a realized catch for day T , $T+1$, $T+2$, and $T+3$, we can plan the capacity, the production plan, and the inventory levels using the module Optimization. The Stochastic Programming Problem (SPP) of Section 5.3.1 solves the problem and sets a value for all the decision variables. We present the SPP again in Equation 9.12.

$$\begin{aligned}
& \text{Objective: } \max \left(\sum_{p=1}^3 \left(\sum_{i=0}^3 \sum_s \sum_o P_{i,s,o,p} SP_{s,o} - \sum_{i=0}^3 \sum_s \sum_l B_{i,s,l,p} K_{s,l} - \sum_{i=1}^3 \sum_d C_{i,d} H_d \right) q_p \right) \\
& \text{subject to} \\
& 1: EI_{i,s,k,l} = IT_{s,k,l} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall s, l, i=0, k \in \{2,3,4,5,6\} \\
& 2: EI_{i,s,k,l} = B_{i,s,l,p} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall s, l, i=0, k=1 \\
& 3: EI_{i,s,k,l} = \sum_{p=1}^3 I_{i,s,k,l,p} q_p \quad \forall i, s, k, l \\
& 4: I_{i,s,k,l,p} = EI_{i-1,s,k-1,l} + B_{i,s,l,p} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall i \in \{1,2,3\}, s, k=1, l, p \\
& 5: I_{i,s,k,l,p} = EI_{i-1,s,k-1,l} - (F_{i,s,k,l,p} + S_{i,s,k,l,p}) \quad \forall i \in \{1,2,3\}, s, k \in \{2,3,4,5,6\}, l, p \\
& 6: P_{i,s,o,p} \leq F_{i,s,k,l,p} Utcut_{k,l,o} \quad \forall i, s, k, l, o, p \\
& 7: P_{i,s,o,p} \leq S_{i,s,k,l,p} Utcut_{k,l,o} \quad \forall i, s, k, l, o, p \\
& 8: \sum_s \sum_m P_{i,s,m} PT_{o,d} \leq C_{i,d} \quad \forall i, d \in \{2,3,4,5\} \\
& 9: \sum_s \sum_k \sum_l F_{i,s,k,l} PT_{o,d} \leq C_{i,d} \quad \forall i, d=1 \\
& 10: C_{i,d} \leq \text{workinghours} * \text{max employees} \quad \forall i, d \in \{1,2\} \\
& 11: C_{i,d} = CT_{i-1,d} \quad \forall i \in \{0,1,2\}, d
\end{aligned} \tag{9.12}$$

First, we present all the matrices that we use in the model as decision variable:

1. Used raw material to production at day T:

$$\begin{aligned}
F &= [\text{day} \quad \text{specie} \quad \text{age} \quad \text{quality} \quad \text{scnario}] \\
F &= [0..3 \quad 1..3 \quad 1..6 \quad 1..3 \quad 1..3]
\end{aligned} \tag{9.13}$$

2. Used raw material to sell at day T:

$$\begin{aligned}
S &= [\text{day} \quad \text{specie} \quad \text{age} \quad \text{quality} \quad \text{scnario}] \\
S &= [0..3 \quad 1..3 \quad 1..6 \quad 1..3 \quad 1..3]
\end{aligned} \tag{9.14}$$

3. Produced products at day T:

$$\begin{aligned}
P &= [\text{day} \quad \text{specie} \quad \text{product} \quad \text{scenario}] \\
P &= [0..3 \quad 1..3 \quad 1..7 \quad 1..3]
\end{aligned} \tag{9.15}$$

4. Inventory at the end of day T:

$$\begin{aligned}
I &= [\text{day} \quad \text{specie} \quad \text{age} \quad \text{quality} \quad \text{scenario}] \\
I &= [0..3 \quad 1..3 \quad 1..6 \quad 1..3 \quad 1..3]
\end{aligned} \tag{9.16}$$

5. Expected inventory at the end of day T:

$$\begin{aligned}
EI &= [\text{day} \quad \text{specie} \quad \text{age} \quad \text{quality}] \\
EI &= [0..3 \quad 1..3 \quad 1..6 \quad 1..3]
\end{aligned} \tag{9.17}$$

6. Planned capacity at day T:

$$C = \begin{bmatrix} \text{day} & \text{department} \end{bmatrix}$$

$$C = \begin{bmatrix} 0..3 & 1..7 \end{bmatrix} \quad (9.18)$$

7. Bought raw material at day T:

$$EI = \begin{bmatrix} \text{day} & \text{specie} & \text{quality} & \text{scenario} \end{bmatrix}$$

$$EI = \begin{bmatrix} 0..3 & 1..3 & 1..6 & 1..3 \end{bmatrix} \quad (9.19)$$

Next, we present all the parameters necessary to solve the model.

1. We need the production times per department. Since the fish species differ in weight, the production times are different per specie. The times are stored in the following matrix:
2. There are departments that require more than one employee per production line. For example, the filleting line needs to employees per production line. This information is stored in the following matrix:
3. The maximum capacity per department is needed to make sure there is no more capacity planned than the maximum production capacity. These values are stored in:
4. We need of each age/quality combination of the raw material the possible products that can be cut out and the fractions of the total weight of the fillet. The final matrix that contains these values is called UTCUT. This matrix is calculated by multiplying the matrices that give the possible end products for each age/quality combination, the production matrix and the matrix with the weight fraction for each part of the fish (see Section 2.3.2)

- Next, we need a matrix to store the deterministic end inventory after day T, that is the deterministic inventory that is needed as input for the next day in the rolling horizon:

$$\begin{aligned} ENDINV &= [specie \quad age \quad quality] \\ ENDINV &= [1..3 \quad 1..6 \quad 1..3] \end{aligned} \quad (9.23)$$

Figure 9.10 shows the information flow between all the matrices. The matrix TOTFISH is also a parameter for the SPP and comes from the module Forecasting. After solving the model for day T, where the model thus plans the capacity for day 0, 1, and 3, the values of the decision variables of the inventory levels of day 0 and the capacities of day 1, 2, and 3 are stored. Then we move up one day to day T+1. The values of the matrixes in the Temporary Storage are used as input for day T+1.

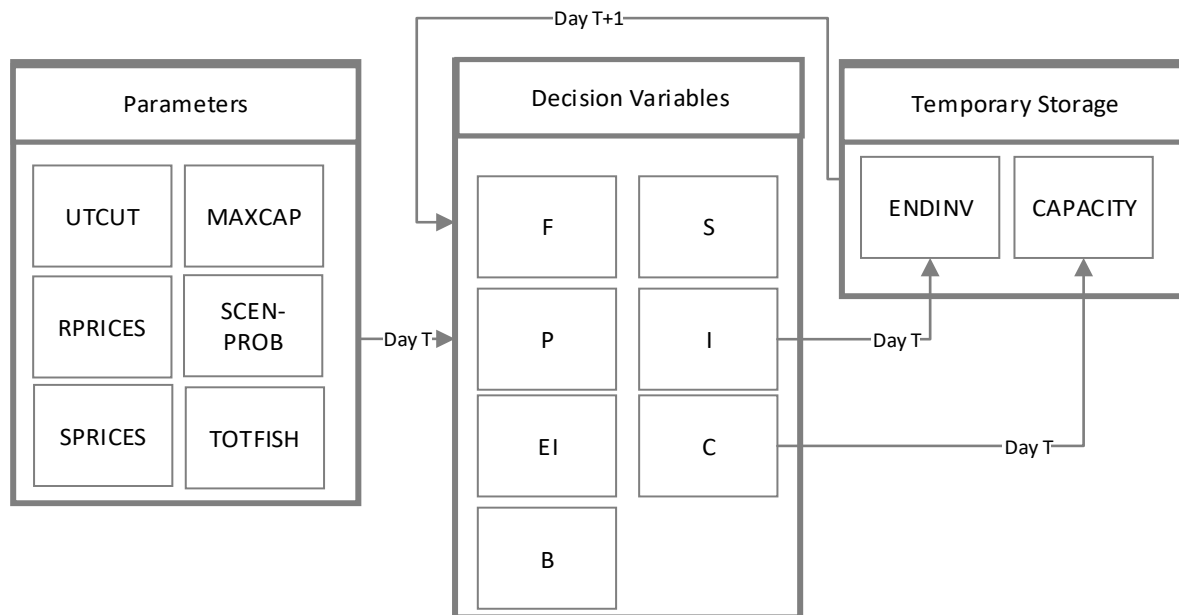


Figure 9.10: Information flow between matrices

Module results

Finally, we need to create a module to store all the information necessary to perform analysis after the simulation runs. We create a module that can write all the necessary information to Excel. We extract the information each day from after solving the SPP. We write the necessary values of the decision variables, together with their dual values to Excel. We do this only for the deterministic day, as we are not interested in the planned capacity. We also want to see when the model plants too much, or too few capacity. We can see this from the dual solution of the SPP. Figure 9.11 shows the information flow between the SPP and Excel.

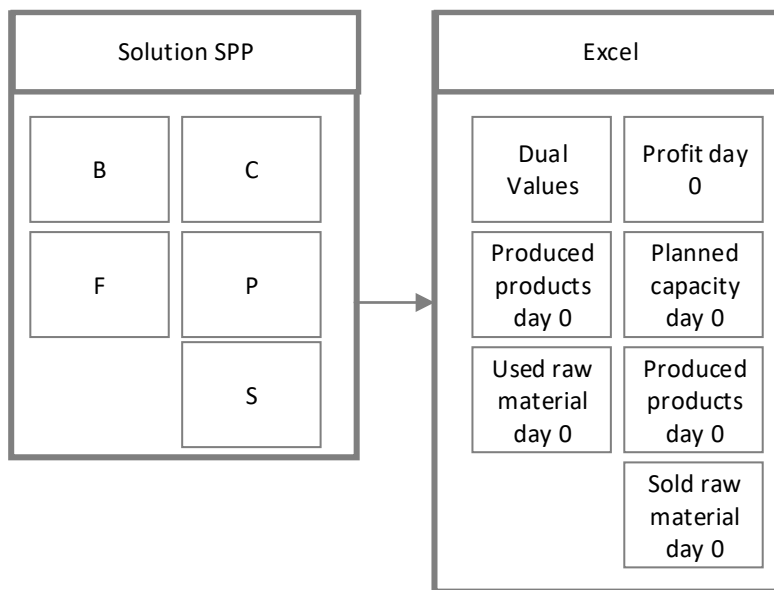


Figure 9.11: Information flow between the SPP and Excel

Appendix 15 Determination of the amount of simulation runs

To determine how many simulation runs we need we use the corrected relative error, which is defined as:

$$y' = \frac{y}{1 - y} \quad (9.24)$$

where: y =relative absolute error

We set y to 0.05, and we search for the amount of simulation runs (n) so that:

$$n_r^* = \min \left(i \geq n : \frac{t_{i-1, 1-\alpha/2} * \sqrt{S^2(n) / i}}{|\bar{X}|} \leq y' \right) \quad (9.25)$$

We calculate y' and set this value at 0.048. We start with 2 simulation runs and determine the Test Statistic. If the Test Statistic is bigger than 0.048, we increase the number of simulation runs with 1 and repeat the previous step. We do this until the Test Statistic is smaller than 0.048. This is the case from 182 simulation runs. Hence, the minimum runs we need is 182. Figure 9.12 shows the results, where the results from 11 simulation runs until 168 are omitted.

Observation	Profit	XBAR	STDEV	T Value	TEST STATISTIC	OK/NOT OK		
1	825989	825989	n.a	n.a	n.a	n.a	Gama	0.05
2	674294	750141.5	75847.5	6.313752	0.638389248	NOT OK	Gamma'	0.047619
3	1213184	904489	160439.2881	2.919986	0.517950365	NOT OK	Alpha	0.05
4	919391	908214.5	113508.8628	2.353363	0.294123918	NOT OK		
5	369272	800426	139100.3878	2.131847	0.370478614	NOT OK		
6	337928	723343	137262.768	2.015048	0.382378923	NOT OK		
7	384402	674922.9	125707.6592	1.94318	0.361926762	NOT OK		
8	363388	635981	115621.278	1.894579	0.344434188	NOT OK		
9	302436	598920.4	108494.4046	1.859548	0.336857022	NOT OK		
10	357099	574738.3	100008.0235	1.833113	0.318973003	NOT OK		
169	589751	585231.1	17651.99282	1.653974	0.049887882	NOT OK		
170	581340	585208.2	17547.86529	1.65392	0.049593911	NOT OK		
171	437895	584346.7	17466.2026	1.653866	0.049434286	NOT OK		
172	444069	583531.2	17383.50022	1.653813	0.049267401	NOT OK		
173	521122	583170.4	17286.49002	1.653761	0.049021214	NOT OK		
174	466965	582502.6	17199.82596	1.653709	0.048829844	NOT OK		
175	596472	582582.4	17101.44512	1.653658	0.048542389	NOT OK		
176	709148	583301.5	17019.19975	1.653607	0.048247903	NOT OK		
177	765273	584329.6	16953.97318	1.653557	0.047976977	NOT OK		
178	803339	585560	16903.29655	1.653508	0.047731634	NOT OK		
179	1057468	588196.4	17014.09406	1.653459	0.047827751	NOT OK		
180	477358	587580.6	16930.50886	1.653411	0.047641271	NOT OK		
181	352996	586284.5	16886.51986	1.653363	0.047621155	NOT OK		
182	471353	585653.1	16805.3494	1.653316	0.047441994	OK		
183	543153	585420.8	16714.87809	1.653269	0.047203976	OK		
184	693957	586010.7	16634.25021	1.653223	0.04692768	OK		
185	563976	585891.6	16544.51974	1.653177	0.046682734	OK		
186	481924	585332.6	16464.82122	1.653132	0.046500947	OK		
187	592364	585370.2	16376.58053	1.653087	0.04624751	OK		
188	485765	584840.4	16297.8521	1.653043	0.046065642	OK		
189	478334	584276.9	16221.18218	1.652999	0.045891941	OK		
190	582105	584265.4	16135.58572	1.652956	0.045649474	OK		
191	518681	583922.1	16054.55635	1.652913	0.045445763	OK		
192	576274	583882.2	15970.76965	1.652871	0.045210512	OK		
193	585331	583889.7	15887.80583	1.652829	0.044973936	OK		

Figure 9.12: Determination of minimum simulation runs

Appendix 16 Determination of warmup period

Since our simulation model starts without any fish in inventory, we need to let the model run for a certain number of runs before the output can be considered representative for the real processing plant. We determine the number of warmup runs with the help of Welch's graphical method.

The method consists of 4 steps:

1. We make n replications of the simulation, where n should be bigger than 5. Each replication has m runs, where m should be bigger than the required number of runs. We choose for m 500 runs.
2. We calculate for each replication n , the average over all the 500 runs. Hence, we have then

$$\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_{499}, \bar{Y}_{500}.$$

3. We calculate the moving average $\bar{Y}_i(w)$ with:

$$\bar{Y}_i(w) = \left\{ \begin{array}{l} \frac{\sum_{s=-w}^w \bar{Y}_{i+s}}{2w+1} \text{ if } i = w+1, \dots, m-w \\ \frac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1} \text{ if } i = 1, \dots, w \end{array} \right\} \quad (9.26)$$

4. We plot the values of $\bar{Y}_i(w)$ for $i=1, 2, \dots, m-w$. We look when the graph has converged to a stable line. We call the value where this stable line starts l . Our warmup period consist of l runs.

We have to plot $\bar{Y}_i(w)$ for several values of w . We choose the values of w to be 10, 100, and 250.

We plot the graphs for $n=5$, $n=10$, and $n=20$. Then we look for where the line is most steady, and we choose l . Figures 9.13a until Figure 9.13c show the results of the plots.

We look at the graph $N=20$ and look for the value l . We observe that for approximately the 50th simulation run, the graph is stable. We thus set the warmup period at 50 simulation runs.

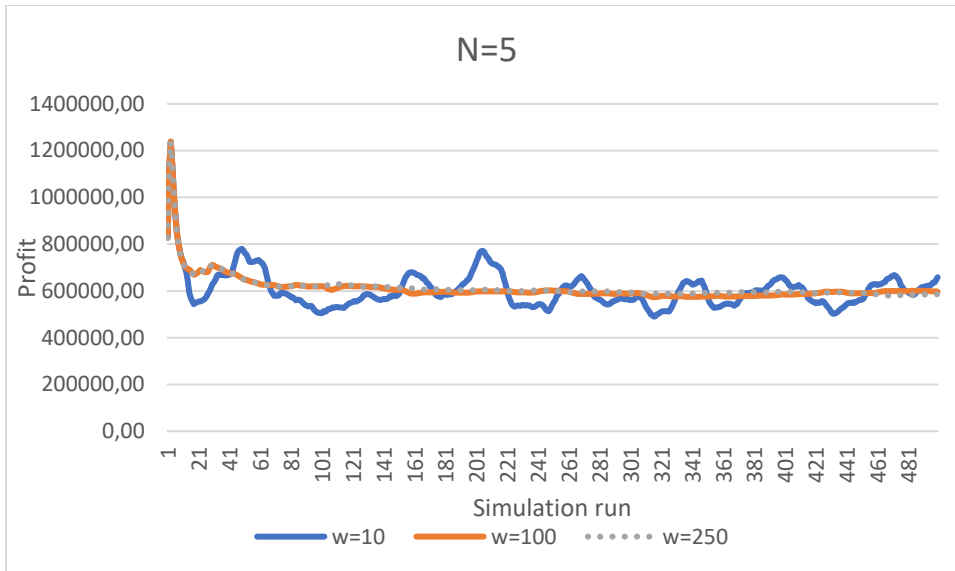


Figure 9.13a: Welch' graphical method for $N=5$

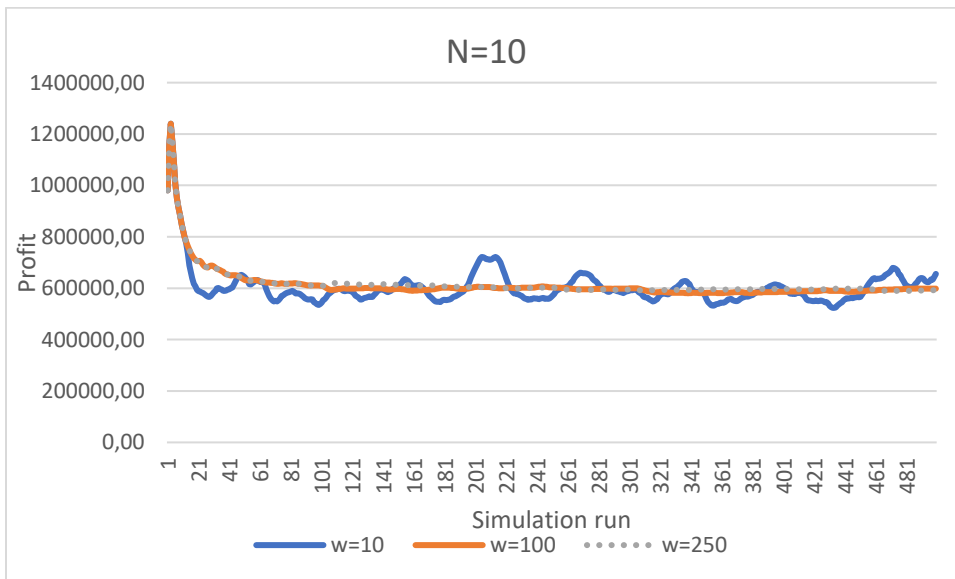


Figure 9.13b: Welch' graphical method for $N=10$

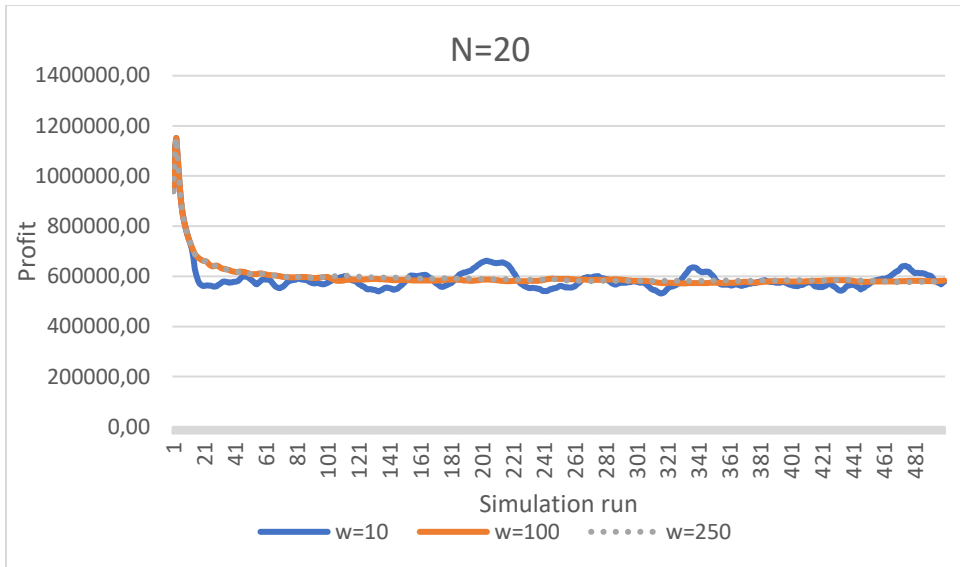


Figure 9.13c: Welch' graphical method for $N=20$

Appendix 17 Summary statistics kilogram per cutting hour

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Kilogram per hour cluster 1	Overtime	238	137.17	11.96	0.78	135.64	138.70	84.68	156.01
	Deterministic	237	145.92	8.94	0.58	144.77	147.06	96.81	158.30
	Extra capacity	238	137.36	12.99	0.84	135.70	139.02	94.53	154.14
	Average forecast ship	238	100.33	38.87	2.52	95.37	105.29	7.14	155.97
	Regular forecast ship	238	137.88	12.97	0.84	136.23	139.54	75.88	154.89
	Current situation	54	88.31	26.90	3.66	80.97	95.65	36.49	166.36
	Total	1243	129.83	27.40	0.78	128.31	131.36	7.14	166.36
Kilogram per hour cluster 2	Overtime	238	165.41	32.51	2.11	161.26	169.56	128.54	261.22
	Deterministic	238	142.38	38.83	2.52	137.42	147.33	0.00	159.43
	Extra capacity	238	166.46	33.47	2.17	162.19	170.74	0.00	280.85
	Average forecast ship	238	148.39	2.86	0.19	148.02	148.75	137.98	153.27
	Regular forecast ship	238	148.89	3.71	0.24	148.41	149.36	134.63	156.77
	Current situation	79	105.22	29.22	3.29	98.68	111.77	57.30	206.87
	Total	1269	151.25	31.24	0.88	149.53	152.97	0.00	280.85
Kilogram per hour cluster 3	Overtime	238	167.50	35.51	2.30	162.97	172.03	0.00	267.12
	Deterministic	238	151.96	6.53	0.42	151.12	152.79	130.31	159.43
	Extra capacity	238	153.57	55.76	3.61	146.45	160.69	0.00	299.49
	Average forecast ship	238	144.06	23.67	1.53	141.04	147.08	0.00	157.81
	Regular forecast ship	233	147.83	4.60	0.30	147.24	148.43	126.58	159.43
	Current situation	80	127.77	39.35	4.40	119.02	136.53	50.59	233.25
	Total	1265	151.41	33.63	0.95	149.55	153.26	0.00	299.49
Kilogram per hour cluster 4	Overtime	238	178.65	39.20	2.54	173.65	183.66	125.19	266.14
	Deterministic	238	152.50	7.44	0.48	151.55	153.45	124.01	159.43
	Extra capacity	238	176.63	39.31	2.55	171.61	181.65	100.06	293.79
	Average forecast ship	238	145.65	17.44	1.13	143.42	147.87	0.00	159.43
	Regular forecast ship	233	147.38	5.59	0.37	146.66	148.10	126.13	158.61
	Current situation	13	140.84	32.46	9.00	121.22	160.45	102.61	199.21
	Total	1198	160.00	30.15	0.87	158.30	161.71	0.00	293.79

Appendix 18 Summary statistics profit per kilogram raw material

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
						Lower Bound	Upper Bound
Profit Cluster 1	Overtime	238	6.10	2.83	0.18	5.74	6.46
	Deterministic	238	6.50	4.47	0.29	5.92	7.07
	Extra capacity	238	5.65	1.88	0.12	5.41	5.89
	Average forecast ship	238	4.12	2.11	0.14	3.85	4.39
	Regular forecast ship	238	5.57	2.94	0.19	5.20	5.95
	Total	1190	5.59	3.09	0.09	5.41	5.76
Profit Cluster 2	Overtime	238	8.46	6.24	0.40	7.66	9.25
	Deterministic	238	8.33	8.49	0.55	7.24	9.41
	Extra capacity	238	9.23	7.05	0.46	8.33	10.13
	Average forecast ship	238	5.59	5.95	0.39	4.83	6.35
	Regular forecast ship	238	8.07	8.52	0.55	6.98	9.16
	Total	1190	7.94	7.42	0.22	7.51	8.36
Profit Cluster 3	Overtime	238	9.80	7.29	0.47	8.87	10.73
	Deterministic	238	9.59	11.11	0.72	8.17	11.01
	Extra capacity	238	9.80	8.32	0.54	8.73	10.86
	Average forecast ship	238	6.70	8.21	0.53	5.66	7.75
	Regular forecast ship	238	8.37	9.66	0.63	7.14	9.61
	Total	1190	8.85	9.08	0.26	8.34	9.37
Profit Cluster 4	Overtime	238	8.05	7.79	0.50	7.05	9.04
	Deterministic	237	8.05	12.92	0.84	6.40	9.71
	Extra capacity	237	8.29	7.02	0.46	7.39	9.19
	Average forecast ship	237	5.74	6.95	0.45	4.85	6.63
	Regular forecast ship	237	7.61	8.42	0.55	6.53	8.68
	Total	1186	7.55	8.93	0.26	7.04	8.06

Appendix 19 ANOVA productivity cutting department

		ANOVA				
		Sum of Squares	df	Mean Square	F	Sig.
Kilogram per hour cluster 1	Between Groups	403317,885	5	80663,577	188,623	,000
	Within Groups	528995,374	1237	427,644		
	Total	932313,260	1242			
Kilogram per hour cluster 2	Between Groups	292208,115	5	58441,623	78,090	,000
	Within Groups	945217,952	1263	748,391		
	Total	1237426,066	1268			
Kilogram per hour cluster 3	Between Groups	123323,144	5	24664,629	23,779	,000
	Within Groups	1305873,937	1259	1037,231		
	Total	1429197,082	1264			
Kilogram per hour cluster 4	Between Groups	252899,039	5	50579,808	72,168	,000
	Within Groups	835426,154	1192	700,861		
	Total	1088325,193	1197			

Multiple Comparisons

Bonferroni

			Mean Difference (I-J)			95% Confidence Interval		
Dependent Variable	(I) ClusterNo	(J) ClusterNo		Std. Error	Sig.	Lower Bound	Upper Bound	
Kilogram per hour cluster 1	Overtime	Deterministic	-8,75 [*]	1,90	,00	-14,33	-3,17	
		Extra capacity	-,19	1,90	1,00	-5,77	5,38	
		Average forecast ship	36,84 [*]	1,90	,00	31,27	42,42	
		Regular forecast ship	-,71	1,90	1,00	-6,29	4,86	
		Current situation	48,86 [*]	3,12	,00	39,69	58,03	
	Deterministic	Overtime	8,75 [*]	1,90	,00	3,17	14,33	
		Extra capacity	8,55 [*]	1,90	,00	2,97	14,13	
		Average forecast ship	45,59 [*]	1,90	,00	40,01	51,17	
		Regular forecast ship	8,03 [*]	1,90	,00	2,45	13,61	
		Current situation	57,61 [*]	3,12	,00	48,43	66,78	
	Extra capacity	Overtime	,19	1,90	1,00	-5,38	5,77	
		Deterministic	-8,55 [*]	1,90	,00	-14,13	-2,97	
		Average forecast ship	37,03 [*]	1,90	,00	31,46	42,61	
		Regular forecast ship	-,52	1,90	1,00	-6,10	5,05	
		Current situation	49,05 [*]	3,12	,00	39,88	58,22	
	Average forecast ship	Overtime	-36,84 [*]	1,90	,00	-42,42	-31,27	
		Deterministic	-45,59 [*]	1,90	,00	-51,17	-40,01	
		Extra capacity	-37,03 [*]	1,90	,00	-42,61	-31,46	
		Regular forecast ship	-37,56 [*]	1,90	,00	-43,13	-31,98	
		Current situation	12,02 [*]	3,12	,00	2,85	21,18	
	Regular forecast ship	Overtime	,71	1,90	1,00	-4,86	6,29	
		Deterministic	-8,03 [*]	1,90	,00	-13,61	-2,45	
		Extra capacity	,52	1,90	1,00	-5,05	6,10	
		Average forecast ship	37,56 [*]	1,90	,00	31,98	43,13	
		Current situation	49,57 [*]	3,12	,00	40,41	58,74	
	Current situation	Overtime	-48,86 [*]	3,12	,00	-58,03	-39,69	
		Deterministic	-57,61 [*]	3,12	,00	-66,78	-48,43	
		Extra capacity	-49,05 [*]	3,12	,00	-58,22	-39,88	
		Average forecast ship	-12,02 [*]	3,12	,00	-21,18	-2,85	
		Regular forecast ship	-49,57 [*]	3,12	,00	-58,74	-40,41	
	Kilogram per hour cluster 2	Overtime	Deterministic	23,04 [*]	2,51	,00	15,66	30,41
			Extra capacity	-1,05	2,51	1,00	-8,43	6,32
			Average forecast ship	17,02 [*]	2,51	,00	9,65	24,40
			Regular forecast ship	16,53 [*]	2,51	,00	9,15	23,90
			Current situation	60,19 [*]	3,55	,00	49,74	70,64
		Deterministic	Overtime	-23,04 [*]	2,51	,00	-30,41	-15,66
			Extra capacity	-24,09 [*]	2,51	,00	-31,46	-16,71
			Average forecast ship	-6,01	2,51	,25	-13,39	1,36
			Regular forecast ship	-6,51	2,51	,14	-13,89	,86
			Current situation	37,15 [*]	3,55	,00	26,71	47,60
Extra capacity		Overtime	1,05	2,51	1,00	-6,32	8,43	
		Deterministic	24,09 [*]	2,51	,00	16,71	31,46	
		Average forecast ship	18,08 [*]	2,51	,00	10,70	25,45	
		Regular forecast ship	17,58 [*]	2,51	,00	10,20	24,95	
		Current situation	61,24 [*]	3,55	,00	50,80	71,69	
Average forecast ship		Overtime	-17,02 [*]	2,51	,00	-24,40	-9,65	
		Deterministic	6,01	2,51	,25	-1,36	13,39	
		Extra capacity	-18,08 [*]	2,51	,00	-25,45	-10,70	
		Regular forecast ship	-,50	2,51	1,00	-7,87	6,88	
		Current situation	43,16 [*]	3,55	,00	32,72	53,61	
Regular forecast ship		Overtime	-16,53 [*]	2,51	,00	-23,90	-9,15	
		Deterministic	6,51	2,51	,14	-,86	13,89	
		Extra capacity	-17,58 [*]	2,51	,00	-24,95	-10,20	
		Average forecast ship	,50	2,51	1,00	-6,88	7,87	
		Current situation	43,66 [*]	3,55	,00	33,22	54,11	
Current situation		Overtime	-60,19 [*]	3,55	,00	-70,64	-49,74	
		Deterministic	-37,15 [*]	3,55	,00	-47,60	-26,71	
		Extra capacity	-61,24 [*]	3,55	,00	-71,69	-50,80	
		Average forecast ship	-43,16 [*]	3,55	,00	-53,61	-32,72	
		Regular forecast ship	-43,66 [*]	3,55	,00	-54,11	-33,22	

Kilogram per hour cluster 3	Overtime	Deterministic	15,54*	2,95	,00	6,86	24,22
		Extra capacity	13,93*	2,95	,00	5,25	22,61
		Average forecast ship	23,44*	2,95	,00	14,76	32,12
		Regular forecast ship	19,67*	2,97	,00	10,94	28,39
		Current situation	39,73*	4,16	,00	27,49	51,97
	Deterministic	Overtime	-15,54*	2,95	,00	-24,22	-6,86
		Extra capacity	-1,61	2,95	1,00	-10,29	7,07
		Average forecast ship	7,90	2,95	,11	-,78	16,58
		Regular forecast ship	4,12	2,97	1,00	-4,60	12,85
		Current situation	24,18*	4,16	,00	11,94	36,42
	Extra capacity	Overtime	-13,93*	2,95	,00	-22,61	-5,25
		Deterministic	1,61	2,95	1,00	-7,07	10,29
		Average forecast ship	9,51*	2,95	,02	,82	18,19
		Regular forecast ship	5,73	2,97	,80	-3,00	14,46
		Current situation	25,79*	4,16	,00	13,55	38,03
	Average forecast ship	Overtime	-23,44*	2,95	,00	-32,12	-14,76
		Deterministic	-7,90	2,95	,11	-16,58	,78
		Extra capacity	-9,51*	2,95	,02	-18,19	-,82
		Regular forecast ship	-3,77	2,97	1,00	-12,50	4,95
		Current situation	16,29*	4,16	,00	4,05	28,53
	Regular forecast ship	Overtime	-19,67*	2,97	,00	-28,39	-10,94
		Deterministic	-4,12	2,97	1,00	-12,85	4,60
		Extra capacity	-5,73	2,97	,80	-14,46	3,00
		Average forecast ship	3,77	2,97	1,00	-4,95	12,50
		Current situation	20,06*	4,17	,00	7,79	32,33
	Current situation	Overtime	-39,73*	4,16	,00	-51,97	-27,49
		Deterministic	-24,18*	4,16	,00	-36,42	-11,94
		Extra capacity	-25,79*	4,16	,00	-38,03	-13,55
		Average forecast ship	-16,29*	4,16	,00	-28,53	-4,05
		Regular forecast ship	-20,06*	4,17	,00	-32,33	-7,79
Kilogram per hour cluster 4	Overtime	Deterministic	26,15*	2,43	,00	19,02	33,29
		Extra capacity	2,02	2,43	1,00	-5,11	9,16
		Average forecast ship	33,01*	2,43	,00	25,87	40,14
		Regular forecast ship	31,27*	2,44	,00	24,09	38,45
		Current situation	37,81*	7,54	,00	15,64	59,99
	Deterministic	Overtime	-26,15*	2,43	,00	-33,29	-19,02
		Extra capacity	-24,13*	2,43	,00	-31,27	-16,99
		Average forecast ship	6,85	2,43	,07	-,29	13,99
		Regular forecast ship	5,12	2,44	,54	-2,06	12,29
		Current situation	11,66	7,54	1,00	-10,52	33,84
	Extra capacity	Overtime	-2,02	2,43	1,00	-9,16	5,11
		Deterministic	24,13*	2,43	,00	16,99	31,27
		Average forecast ship	30,98*	2,43	,00	23,84	38,12
		Regular forecast ship	29,25*	2,44	,00	22,07	36,42
		Current situation	35,79*	7,54	,00	13,61	57,97
	Average forecast ship	Overtime	-33,01*	2,43	,00	-40,14	-25,87
		Deterministic	-6,85	2,43	,07	-13,99	,29
		Extra capacity	-30,98*	2,43	,00	-38,12	-23,84
		Regular forecast ship	-1,73	2,44	1,00	-8,91	5,44
		Current situation	4,81	7,54	1,00	-17,37	26,99
	Regular forecast ship	Overtime	-31,27*	2,44	,00	-38,45	-24,09
		Deterministic	-5,12	2,44	,54	-12,29	2,06
		Extra capacity	-29,25*	2,44	,00	-36,42	-22,07
		Average forecast ship	1,73	2,44	1,00	-5,44	8,91
		Current situation	6,54	7,54	1,00	-15,64	28,73
	Current situation	Overtime	-37,81*	7,54	,00	-59,99	-15,64
		Deterministic	-11,66	7,54	1,00	-33,84	10,52
		Extra capacity	-35,79*	7,54	,00	-57,97	-13,61
		Average forecast ship	-4,81	7,54	1,00	-26,99	17,37
		Regular forecast ship	-6,54	7,54	1,00	-28,73	15,64

*. The mean difference is significant at the 0.05 level.

Appendix 20 ANOVA profit per kilogram raw material

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Cluster 1	Between Groups	769,42	4,00	192,36	21,54	,00
	Within Groups	10580,05	1185,00	8,93		
	Total	11349,47	1189,00			
Cluster 2	Between Groups	1812,56	4,00	453,14	8,43	,00
	Within Groups	63680,37	1185,00	53,74		
	Total	65492,93	1189,00			
Cluster 3	Between Groups	1708,19	4,00	427,05	5,25	,00
	Within Groups	96345,97	1185,00	81,30		
	Total	98054,16	1189,00			
Cluster 4	Between Groups	1028,22	4,00	257,05	3,24	,01
	Within Groups	93571,11	1181,00	79,23		
	Total	94599,33	1185,00			

Multiple Comparisons

Bonferroni

Dependent Variable	(I) Clustemo	(J) Clustemo	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Cluster 1	Overtime	Deterministic	-,40	,27	1,00	-1,17	,37
		Extra capacity	,45	,27	1,00	-,32	1,22
		Average forecast ship	1,97 *	,27	,00	1,20	2,74
		Regular forecast ship	,52	,27	,56	-,25	1,29
	Deterministic	Overtime	,40	,27	1,00	-,37	1,17
		Extra capacity	,85 *	,27	,02	,08	1,62
		Average forecast ship	2,37 *	,27	,00	1,60	3,14
		Regular forecast ship	,92 *	,27	,01	,15	1,69
	Extra capacity	Overtime	-,45	,27	1,00	-1,22	,32
		Deterministic	-,85 *	,27	,02	-1,62	-,08
		Average forecast ship	1,52 *	,27	,00	,75	2,29
		Regular forecast ship	,07	,27	1,00	-,70	,85
	Average forecast ship	Overtime	-1,97 *	,27	,00	-2,74	-1,20
		Deterministic	-2,37 *	,27	,00	-3,14	-1,60
		Extra capacity	-1,52 *	,27	,00	-2,29	-,75
		Regular forecast ship	-1,45 *	,27	,00	-2,22	-,68
	Regular forecast ship	Overtime	-,52	,27	,56	-1,29	,25
		Deterministic	-,92 *	,27	,01	-1,69	-,15
		Extra capacity	-,07	,27	1,00	-,85	,70
		Average forecast ship	1,45 *	,27	,00	,68	2,22
Cluster 2	Overtime	Deterministic	,13	,67	1,00	-1,76	2,02
		Extra capacity	-,77	,67	1,00	-2,66	1,12
		Average forecast ship	2,87 *	,67	,00	,98	4,76
		Regular forecast ship	,39	,67	1,00	-1,50	2,28
	Deterministic	Overtime	-,13	,67	1,00	-2,02	1,76
		Extra capacity	-,90	,67	1,00	-2,79	,99
		Average forecast ship	2,74 *	,67	,00	,85	4,63
		Regular forecast ship	,26	,67	1,00	-1,63	2,15
	Extra capacity	Overtime	,77	,67	1,00	-1,12	2,66
		Deterministic	,90	,67	1,00	-,99	2,79
		Average forecast ship	3,64 *	,67	,00	1,75	5,53
		Regular forecast ship	1,16	,67	,85	-,73	3,05
	Average forecast ship	Overtime	-2,87 *	,67	,00	-4,76	-,98
		Deterministic	-2,74 *	,67	,00	-4,63	-,85
		Extra capacity	-3,64 *	,67	,00	-5,53	-1,75
		Regular forecast ship	-2,48 *	,67	,00	-4,37	-,59
	Regular forecast ship	Overtime	-,39	,67	1,00	-2,28	1,50
		Deterministic	-,26	,67	1,00	-2,15	1,63
		Extra capacity	-1,16	,67	,85	-3,05	,73
		Average forecast ship	2,48 *	,67	,00	,59	4,37

Cluster 3	Overtime	Deterministic	,21	,83	1,00	-2,11	2,54
		Extra capacity	,01	,83	1,00	-2,32	2,33
		Average forecast ship	3,10 *	,83	,00	,77	5,42
		Regular forecast ship	1,43	,83	,84	-,90	3,75
	Deterministic	Overtime	-,21	,83	1,00	-2,54	2,11
		Extra capacity	-,21	,83	1,00	-2,53	2,12
		Average forecast ship	2,88 *	,83	,01	,56	5,21
		Regular forecast ship	1,21	,83	1,00	-1,11	3,54
	Extra capacity	Overtime	-,01	,83	1,00	-2,33	2,32
		Deterministic	,21	,83	1,00	-2,12	2,53
		Average forecast ship	3,09 *	,83	,00	,77	5,42
		Regular forecast ship	1,42	,83	,85	-,90	3,75
	Average forecast ship	Overtime	-3,10 *	,83	,00	-5,42	-,77
		Deterministic	-2,88 *	,83	,01	-5,21	-,56
		Extra capacity	-3,09 *	,83	,00	-5,42	-,77
		Regular forecast ship	-1,67	,83	,44	-3,99	,66
	Regular forecast ship	Overtime	-1,43	,83	,84	-3,75	,90
		Deterministic	-1,21	,83	1,00	-3,54	1,11
		Extra capacity	-1,42	,83	,85	-3,75	,90
		Average forecast ship	1,67	,83	,44	-,66	3,99
Cluster 4	Overtime	Deterministic	,00	,82	1,00	-2,30	2,29
		Extra capacity	-,24	,82	1,00	-2,54	2,05
		Average forecast ship	2,31 *	,82	,05	,01	4,61
		Regular forecast ship	,44	,82	1,00	-1,85	2,74
	Deterministic	Overtime	,00	,82	1,00	-2,29	2,30
		Extra capacity	-,24	,82	1,00	-2,54	2,06
		Average forecast ship	2,31 *	,82	,05	,01	4,61
		Regular forecast ship	,45	,82	1,00	-1,85	2,75
	Extra capacity	Overtime	,24	,82	1,00	-2,05	2,54
		Deterministic	,24	,82	1,00	-2,06	2,54
		Average forecast ship	2,55 *	,82	,02	,25	4,85
		Regular forecast ship	,69	,82	1,00	-1,61	2,99
	Average forecast ship	Overtime	-2,31 *	,82	,05	-4,61	-,01
		Deterministic	-2,31 *	,82	,05	-4,61	-,01
		Extra capacity	-2,55 *	,82	,02	-4,85	-,25
		Regular forecast ship	-1,87	,82	,23	-4,17	,43
	Regular forecast ship	Overtime	-,44	,82	1,00	-2,74	1,85
		Deterministic	-,45	,82	1,00	-2,75	1,85
		Extra capacity	-,69	,82	1,00	-2,99	1,61
		Average forecast ship	1,87	,82	,23	-,43	4,17

*. The mean difference is significant at the 0.05 level.