



Universiteit
Leiden
The Netherlands

Estimation of the impact of uncertainty in ORTEC's Supply Chain Design software

Fugers, I.

Citation

Fugers, I. (2016). *Estimation of the impact of uncertainty in ORTEC's Supply Chain Design software*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

Downloaded from: <https://hdl.handle.net/1887/3597097>

Note: To cite this publication please use the final published version (if applicable).

MATHEMATICAL INSTITUTE

MASTER THESIS

STATISTICAL SCIENCE FOR THE LIFE AND BEHAVIOURAL SCIENCES

Estimation of the impact of uncertainty in ORTEC's Supply Chain Design software

Author:
Ivo Fugers

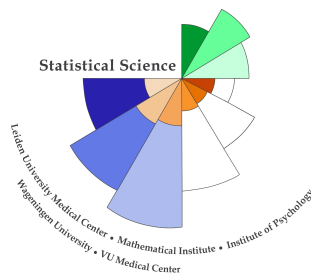
First Supervisor:
Dr. Floske Spieksma
Mathematical Institute, Leiden University

Second Supervisor:
Jawad Elomari, PhD
ORTEC

October 2016



Universiteit
Leiden
The Netherlands



ORTEC
OPTIMIZE YOUR WORLD

ABSTRACT

The supply chain network of companies that produce goods and deliver them to customers, can be captured in a mathematical model. ORTEC Supply Chain Design (OSCD) uses a mixed integer mathematical program to design a supply chain network while optimizing a certain output (e.g. minimize cost). The input parameters of such a program, like transport cost and customer demand, not only have to be estimated, but they also have ranges of uncertainty. Consequently, the output (the network design) is uncertain as well. In this work the impact of that uncertainty on the output, in the mathematical model of OSCD, is investigated. Generic effects that hold for several case studies of OSCD are presented. In addition a tool for precise estimation in unique cases is proposed. This work could be used to further extend the relationship between input and output of more complex case studies of OSCD.

ACKNOWLEDGMENTS

The nine months I spent at ORTEC have been a valuable experience for me. I am happy to say that after my graduation I will be back to start working as a Data Science consultant. During my time at ORTEC, there were several people who helped me make this project to a success. First of all there were Pascha Iljin and Frans van Helden, who gave me this opportunity and who made sure that I got a head start with my activities. During my project I was guided by Jawad Elomari. With his in-depth understanding of the problem, broad knowledge base, helicopter perspective and patience, he helped me through the tougher parts of the project. Floske Spijksma, my first supervisor, has been the perfect critical look at my work. Not speaking the 'ORTEC language' she has always looked at my work with an open mind, asking the right questions and encouraging me to formulate my thoughts and results in an understandable way. Lastly, I should mention my parents, who have consistently been there for me with the necessary moral and financial support.

CONTENTS

I	INTRODUCTION & LITERATURE	6
1	INTRODUCTION	7
2	LITERATURE OVERVIEW	9
2.1	ORTEC Supply Chain Design	9
2.2	Mixed Integer Programming	10
2.3	Sensitivity based on mathematical model	12
2.4	Sensitivity Analysis	12
2.5	Design of Experiments	14
2.6	Applications of Global Sensitivity Analysis Methods	15
II	METHODS	16
3	SOLUTION APPROACH	17
3.1	Solution part 1	17
3.2	Solution part 2	18
3.2.1	Analysis Methods	19
III	RESULTS	21
4	CASE STUDY 1: LACTALIS	22
4.1	Description	22
4.2	Results	23
4.2.1	Optimal Cost	24
4.2.2	DC Allocation	26
5	CASE STUDY 2: MEDUX	30
5.1	Description	30
5.2	Results	31
5.2.1	Optimal Cost	32
5.2.2	DC Allocation	35
6	SOLUTION PART 2: GENERIC EFFECTS IN OSCD	39
6.1	Generic regression model on optimal cost	39
6.2	Generic regression model on DC allocation	41
IV	CONCLUSION	45
7	CONCLUSION	46
7.1	Conclusions solution part 1	46
7.2	Conclusions solution part 2	47
7.3	Future Work	47
	Appendices	51
A	MATHEMATICAL MODEL	52
A.1	Sets	52
A.2	Parameters	53
A.2.1	Capacity Parameters	53

- A.2.2 Revenue and Cost Parameters 53
- A.2.3 Other Parameters 54
- A.3 Variables 54
 - A.3.1 Decision Variables 54
 - A.3.2 Derived Variables 54
- A.4 Constraints 54
 - A.4.1 General constraints 54
 - A.4.2 Demand constraints 55
 - A.4.3 Supply constraints 55
 - A.4.4 Flow constraints 55
 - A.4.5 Production constraints 55
 - A.4.6 Location constraints 55
 - A.4.7 Stock constraints 55
 - A.4.8 Cost Equations 56
- A.5 Objective 56

Part I

INTRODUCTION & LITERATURE

INTRODUCTION

ORTEC Supply Chain Design (OSCD) is ORTEC's own supply chain optimization software. In OSCD, a supply chain can be modeled completely. A set of characteristics or requirements is defined by input parameters. These include typical features such as capacities, cost and market demand. The values of these parameters are used by OSCD to optimize the supply chain network, through a set of deterministic linear functions. The software optimizes the solution by minimizing cost that consist of several parameters (e.g. transport, supply, handling, fixed, inventory, import or storage cost). OSCD is able to offer insight in the optimal performance of all processes in a supply chain. Results of the solution can give answers into questions such as:

- What is the optimal number and size of distribution centers in the network?
- What is the best location for a production facility?
- How large should the truck fleet size be?

Answers to these questions can be used to make strategic decisions about the design of the supply chain network.

In OSCD, and in any general supply chain optimization, input parameters have a degree of uncertainty. Consequently, the output is uncertain as well. Users of OSCD wish to understand the relationship between the uncertainty of the input and the level of uncertainty in the output. The optimization algorithm solves a model that represents a complicated set of interacting relations. The goal of this project is to find a method to estimate the effect of uncertainty in input parameters on the optimal solution and to identify any generic effects that occur in several case studies of OSCD.

ORTECs customers make decisions based on these optimized networks. This makes it important to understand the uncertainty relationship. For example, they might decide to open another distribution center, a decision that could cost €120.000, such as in case study 1 of this report. Kim et al. (2011) present an optimization model that enables the selection of biomass conversion technologies in a biofuel supply chain. The case is known for extreme demand and supply fluctuations that can affect the profitability of the network. In that example, wrong assignment can be costly by not meeting demand or not optimally assigning all supply to different conversion technologies.

Currently, ORTEC consultants run lots of scenarios and observe the changes in order to gain insight in the effect of input uncertainty. This is done in an unstructured way and based on trial and error. This process needs to be repeated every time something changes.

In the literature, experiments are setup to obtain sensitivity measures. They use design of experiments or have some probabilistic aspect (i.e. Monte Carlo simulation). The latter is used since it is capable of capturing non-linear relationships (Wang et al.). These methods are effective but case specific, since the probabilistic nature of sampling is done based on historic data of that case. In addition, it is computationally expensive and they must be repeated every time something changes.

This work is aimed at designing a method that efficiently estimates the uncertainty in the output of any case in OSCD. The efficiency of a method refers to the fact that it should obtain results with a limited computational budget. Based on previous experience from ORTEC consultants with uncertainty in OSCD, it is known that customer demand and cost parameters (e.g. transport or location cost) are most sensitive to fluctuations of the value. For this project, these parameters will be of interest. In addition, this work is aimed at getting insight in generic effects that are present in the mathematical model of OSCD. The optimized output is studied in terms of optimal cost and change in the network of allocated distribution centers (DC's). The latter is chosen since it represents important information for strategic decisions. Summarizing, the objectives of this project are:

1. Design and implementation of a method that quantifies the importance and effect of uncertain parameters. "Importance" refers to the impact on the optimal solution.
2. Investigate to what extent the effects in the OSCD mathematical model are generalizable to other instances of the model. In other words, how are the input parameters generally related to the output parameters?

These two objectives should be pursued in a way that they can be applied to any generic supply chain within the OSCD software. Objective 1 will be tested on two recent case studies where OSCD was used to design a supply chain network. With the results of objective 1, expectations about a third case will be formulated. This will be the basis for objective 2. An additional analysis on the third case will confirm if these expectations are correct.

Within the OSCD context, procedures for running these experiments will be built, as well as the means to analyze it (e.g. a linear regression procedure). The procedures will be built in such a way that can be used by future customers.

This report is organized as follows: chapter 2 (Literature Overview) gives an overview of the relevant literature and it will elaborate on the functionality of OSCD. Based on this, chapter 3 (Solution Approach) describes the methods used in the experiments. In chapter 4 (Case Study 1), 5 (Case Study 2) and 6 (Generic effects in OSCD) this solution will be applied and analyzed in three case studies. Chapter 7 (Conclusion), the last chapter, contains the conclusions of this work.

LITERATURE OVERVIEW

2.1 ORTEC SUPPLY CHAIN DESIGN

OSCD is used for supply chain optimization studies and offers decision support at a strategic and tactical level. The program is capable of handling a generic supply chain. All details about the network have to be specified by the user. A simplified network is shown in Figure 2. This network has one supply facility that transports products by boat to a Distribution Center (DC). This DC transports the products by truck to customers.

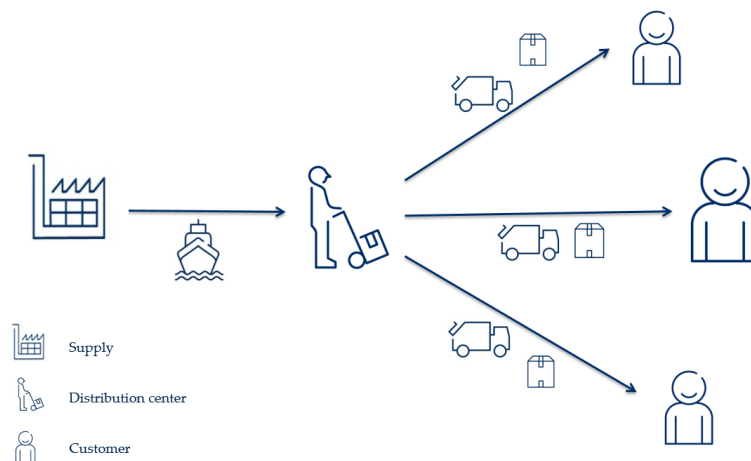


Figure 2: Simplified example of a Supply Chain Network

In reality, the software is capable of handling a diverse set of structures. OSCD optimizes a flow of products through a supply chain network. All transport, storage, production and supply variables are allowed to change in the model, as long as they meet the specified constraints (e.g. storage capacity of a truck). There can be numerous supply locations with different products. The software can handle different transport modalities for every route or even multiple possibilities for one route. The mathematical solver seeks to minimize or maximize an objective taking all cost parameters into account. That objective is usually to minimize cost, but can also be to maximize revenue, minimize CO₂ emission or something else specified by the user. The result is a set of decision variables that chooses which route the flow of products will follow to reach a customer. A visualization of such network is shown in Figure 3. The model has two supply locations (on the right side of the figure) and several DC's and customers in France, Belgium and the Netherlands.

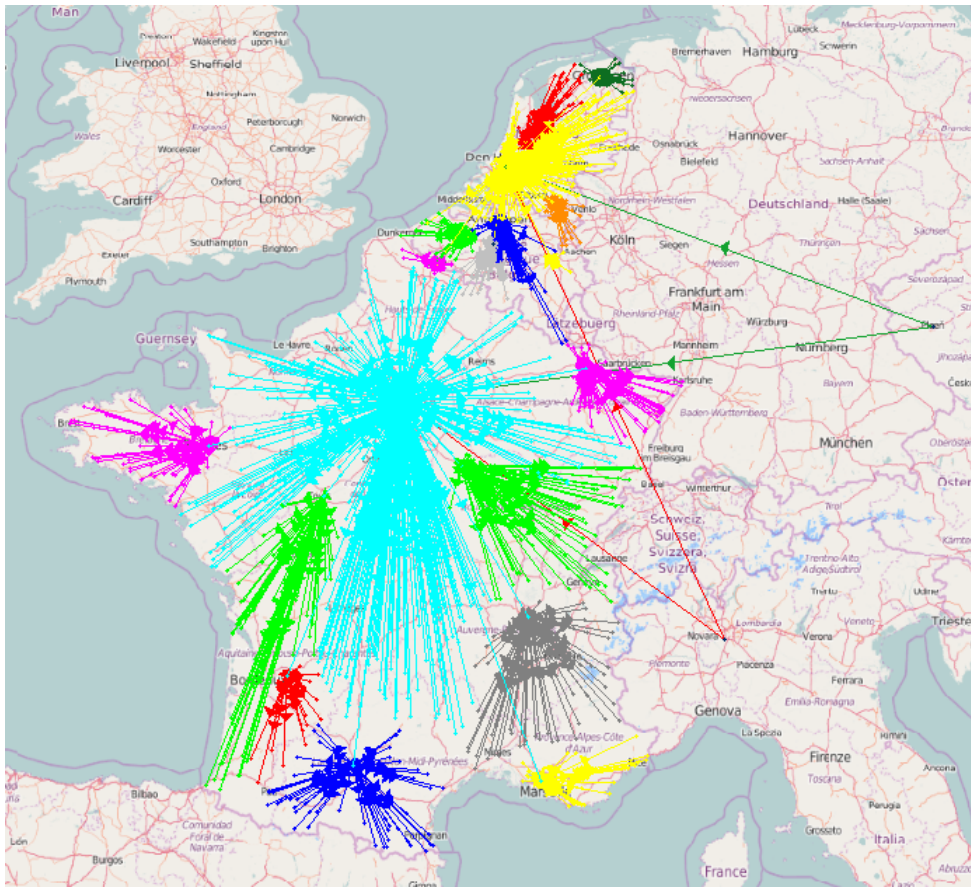


Figure 3: Example of visualized output of supply chain network

OSCD is a Mixed Integer Linear Program (see 2.2 for theory). The mathematical model that is used by the program is specified in appendix A.

Uncertainty in a mathematical program can be assessed through various methods. A simple scenario run can test different values for an input parameter. Robust optimization methods find a solution that is best amongst worst-case realizations of the data. Stochastic programming methods determine a solution that is best on average over the range of uncertainty (Busschers, 2012). An obvious disadvantage of the last two methods is computation time. As the number of uncertain parameters increases, the complexity of the problem increases exponentially (Zhang et al., 2015). In this work the interest lies in finding a quick method that can identify the ranges of sensitivity with respect to a specific outcome. This chapter will first present the specifics of a Mixed Integer Linear Program. Secondly, various sensitivity analysis methods are discussed. Thirdly, some relevant experimental designs will be mentioned. Finally, several application of these methods will be discussed.

2.2 MIXED INTEGER PROGRAMMING

Any mathematical program consists of the following elements:

- Parameters: these contain all input data of the model, for example all DC's and their fixed and variable cost;
- Decision variables: these contain all decisions potentially made by the program. For example, the allocation of customers to DC's;
- Constraints: these are at the heart of the mathematical program and describe the restrictions on the model. For example: all customers should at least be provided with their minimum demand every period;
- Objective function: this is the function describing the objective of OSCD. For minimizing the total cost, this function gives the total cost of a solution.

In this project the uncertainty of the input parameters and the constraints are of interest. A linear programming (LP) problem is an optimization problem that optimizes a linear function that is subject to certain linear constraints. The general formulation is as follows:

$$\max \sum_{j=1}^n c_j x_j$$

Subject to:

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = 1, \dots, m)$$

$$x_j \geq 0 \quad (j = 1, \dots, n).$$

In this work, the interest is in estimating the effect of uncertainty in the right-hand side of the constraint (demand, denoted by b_i) and in the coefficients of the objective function (cost, denoted by c_j). OSCD makes use of mixed integer linear programming (MIP). A MIP has the additional constraint that some coefficients in the objective function must be of integer value. This adds the further restriction on the set of feasible solutions:

$$x_j \in \mathbb{Z} \quad (j = 1, \dots, n).$$

In the specific context of OSCD this last constraint is not only integer, but also binary which replaces the latter restriction with:

$$x_j \in \{0, 1\} \quad (j = 1, \dots, n).$$

Integer and binary constraints make it more difficult to optimize the linear program. The simplex method, an optimization algorithm, finds an optimum by iteratively computing a solution better than the current one. This is not possible for a MIP. An alternative would be to solve the problem by trying all binary values for the constraints. This is not feasible as it will lead to 2^p possible assignments (with p constraints). The Branch and Bound algorithm is able to solve a MIP by lowering the complexity of the problem by relaxing all constraints, while identifying which ones are already at integer (or binary) values. The branching algorithm, for a binary constraint, is as follows (Hendriks, 1991):

1. Solve the relaxed approach (with the Simplex algorithm);
2. If there are variables z_i with non-binary values, pick one of them and recursively solve each relaxation with $z = 0$ and $z = 1$;
3. Stop when the solution only contains binary values.

The mathematical model of OSCD is formulated in appendix A.

2.3 SENSITIVITY BASED ON MATHEMATICAL MODEL

The sensitivity of decision variables and constraints can be derived from the mathematical model. Sensitivity in the constraints is defined by the dual prices. A dual price is the amount of benefit in the objective function, if a constraint is weakened by one unit (Taha, 2007). For example, the transport capacity of a truck can increase making the transport cost per product relatively cheaper. Sensitivity in a decision variable is defined by the reduced cost. In this example the objective improves by relaxing one of the constraints in the model. The reduced cost is the reduction in revenue if a zero-coefficient is forced to be non-zero. This cost arises, because the resource that is used, could have been spent in a better way. In addition, one can look at the ranges at which a variable can be changed, without its affecting the optimal solution (Taha, 2007).

These measures can be informative but for this project the main interest is not in decision variables but in cost parameters. The mathematical model does not reveal sensitivity information about these parameters (or their interactions). With the help of sensitivity analysis methods (i.e. designed experiments) sensitivity estimates of those parameters can be found.

2.4 SENSITIVITY ANALYSIS

This project aims to quantify the direct relationship between a set of input parameters and an optimized outcome. To really understand cause-and-effect relationships, deliberate changes to the input parameters should be made, while observing the effect on the output. For a model with n parameters, the difference (δ) in output parameter y can be explained by differences in input parameters. An important distinction is the effect of individual parameters (main effect) and the effect of input parameters when they are changing simultaneously (interaction effect). A main effect is the change in output if one input parameter is changed by one unit. Interaction occurs when the level of change of one input parameter is affected by the level of another input parameter (Fox, 2008). To illustrate this consider the following example:

A supply chain optimizer finds the optimal number of DC's for a soft-drink company, based on numerous input parameters such as location cost, transport cost and demand forecasts. The effect of location cost is illustrated in the plot below. When the location cost is higher, the optimal number of DC's decreases (for this example, the amount of DC's is assumed to be continuous). This effect is different for different levels of transport cost, as is illustrated in Figure 4.

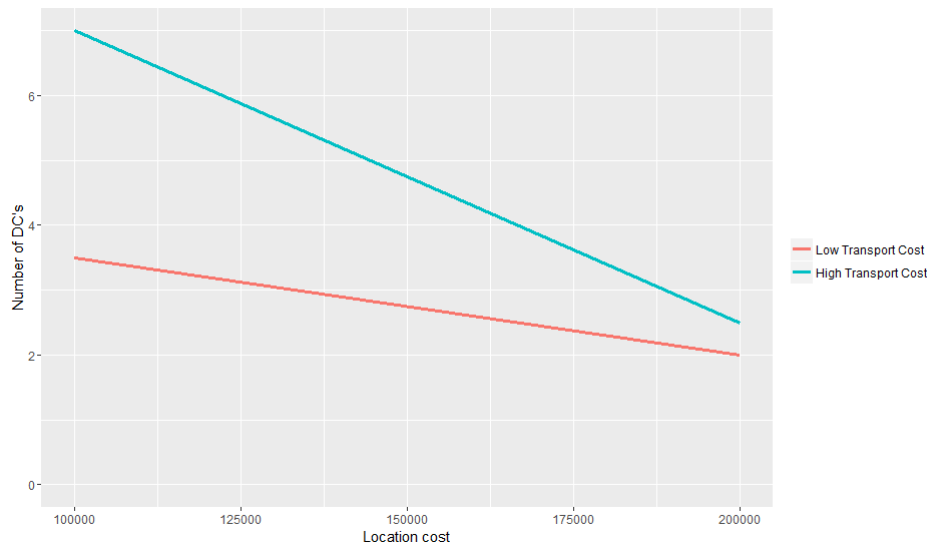


Figure 4: Example of interaction effect in supply chain optimization

In this case, the input parameters location cost and transport cost are interacting. If no interaction were present, the two lines in Figure 4 would be parallel.

Sensitivity Analysis (SA) is a method that enables insight into complex mathematical models such as an optimization problem. SA reveals to what extent the outcome changes as a result of a change in the input parameters (Bogenovo, Plischke 2015). In a standard setting, SA studies the model response to a selected range of input parameters. Local SA methods (such as nominal range SA, automated differentiation) test individual parameter effects while, global SA methods test parameter effects with respect to the entire set of parameter distributions (Sobol's method, regression and ANOVA) (Cadini et al., 2012). In other words, local SA methods only test for main effects while global methods also take interaction effects into account. Clearly, global SA methods are the ones that are of interest for this project.

The difference (δ) in output parameter y (e.g. number of DC's) can be explained by the following equation:

$$\delta_y = \sum_{i=1}^n \phi_i + \sum_{i<j} \phi_{i,j} + \sum_{i<j<k} \phi_{i,j,k} + \dots + \phi_{1,2,\dots,n} \quad (1)$$

The ϕ_i represents all n main effects associated with the n input parameters. $\phi_{i,j}$ are the two-way interaction terms, $\phi_{i,j,k}$ are the three-way interaction terms, etc. (all the way up to the n -way interaction term). Estimates of three-way interaction terms (and higher) are generally not of interest as they have a small impact on the response and are hard to interpret (Montgomery, Runger, 2011). In the results of this project, it will be shown that these high-way interactions have indeed little to no effect on the output.

The extent to which higher way interactions are identifiable, is dependent on the design of the executed computer experiments (see next section).

2.5 DESIGN OF EXPERIMENTS

Experimental design can allow for finding the structure of main and interaction effects. Some different approaches of designed experiments:

- One-at-a-Time design: Change one parameter at a time and observe the output. This will enable one to estimate the main effects, but they might be confounded with interaction effects.
- Factorial design: Varies over all combinations of uncertain parameters. The ranges over which the parameters are varied, have to be specified in advance. This enables one to estimate both main and interaction effects. The downside is the computational price. Varying every parameter at 2 different levels leads to 2^n combinations, where n is the number of parameters. This can quickly become infeasible.
- Fractional factorial design: This design does not cover all possible combinations but a part of it (fraction). As a result the researcher has to choose which interaction effects have to be covered. This can be done based on expert knowledge and previous tests. This design will identify all main effects but only part of the interactions. It can still uncover important relations and is less computationally expensive (Montgomery, 2011).

Fractional design will allow for efficient estimation of most important effects without having to deal with too much computation time. The smaller the experimental design (i.e. the smaller the fraction) the less effects can be estimated. With less effects to be estimated, there will be less insight in the relation between input parameters and changes in output (i.e. less explained variance).

Experimenting with a mathematical model results in model responses corresponding to changes in the input parameters. Analysis of Variance (ANOVA) is a statistical technique that is capable of quantifying the importance of parameters by computing the induced variance in the model. In the context of computer experiments, a similar method, called Sobol's method (Sobol, 2001) is known. There is an important difference between ANOVA and Sobol's method. ANOVA tests the importance of parameters by comparing the parameter explained variance to the residuals (the unexplained part in the model). Sobol's method, in contrast, compares parameter variance with the explained part of the model. Sobol's method does not test for significance but rather quantifies each parameter's importance relative to the whole model. In computer experiments the unexplained part is usually known as the part that is deliberately left out of the model (e.g. to reduce computation effort). Another important difference is that Sobol's method uses a quasi-random simulation approach for testing different values of input parameters. In case of ANOVA, it is more common to use designed experiments (with fixed levels of parameters). As mentioned, the simulation approach has the advantage of the possibility to identify non-linear relationship (while being more computationally expensive, Zhang et al. (2015)).

2.6 APPLICATIONS OF GLOBAL SENSITIVITY ANALYSIS METHODS

Kim et al. (2011) used SA to find the dominant input parameters in a biomass supply chain network. From 14 parameters, 5 were selected to be dominant, using the global SA technique Sobol's method. These 5 parameters were used to create different scenarios based on different values over the range of uncertainty of the dominant parameters. As a comparison, they conducted a Monte Carlo simulation including all 14 parameters. This simulation showed that the 5 dominant parameters explained 96% of the variance in the output. This result indicates that a global SA method successfully narrows down the complexity of the problem (in this case from 14 parameters to 5). The full simulation approach (which tests all main parameters at a lot of different values) is extremely computationally expensive compared to a sensitivity experiment (which only tests the variables at a few values).

Another work of Kim et al (2011) uses a simulation with various demand levels. Afterwards, SA was used to test robustness of the output in terms of net profit. They showed that, for supply chain network design, a simulation based approach can provide insight in alternative outcomes as a result of uncertainty in demand.

Wang et al. (2009) compared two types of global SA, one which used experiments and another which used simulation. They created various input scenario's for a supply chain network optimization, according to either a factorial design (experiment) or Importance Measure sampling (simulation). A factorial design with two levels tests the input levels at the two extreme values of the predefined uncertain ranges. This can be extended by taking more levels within these ranges. Importance Measure sampling is similar, but instead of testing at several levels within the range, it takes a random sample of values within the range. This random sample is executed according to a suitable probability distribution. The authors applied regression analysis on the results of the simulation. Wang et al. showed that importance measure sampling is capable of capturing non-linear effects, but is computationally more expensive to perform (i.e. the probabilistic nature requires more runs). In their example, both methods effectively captured the relevant effects. In OSCD, it is expected that the relationship between the input parameters and the optimal value is linear (to a certain extent), based on experience by consultants who use and implement the software. This will be confirmed in the result section of this report.

Deleris et al. (2005) used simulation to cover the probabilistic nature of certain scenarios in a supply chain optimization. Using SA, they selected four hazardous scenarios (employee strike, shortage of supply, political instability, hurricane) that influence the performance of the supply chain. The performance output that was selected, is loss of transported volume of the product. The authors ran a simulation creating scenarios by including the hazardous events with a certain probability. This simulation enabled them to analyze the expected loss of volume.

Part II

METHODS

SOLUTION APPROACH

The goal of this project is to develop an efficient method that correctly estimates the effect of uncertainty of the main parameters in a MIP that optimizes a supply chain network. Previous research uses several SA techniques to achieve similar results. In this project the focus is on achievable computation time. To obtain this goal the solution is split into two parts. The first part is an experimental method that estimates the uncertainty of all main parameters and their interactions completely. The second part uses the findings of the first part to construct a generic model to explain the relationship between input parameters and the output.

3.1 SOLUTION PART 1

The experiment is set up such that it yields insights in the following issues:

1. Identify important main effects;
2. Identify important interactions;
3. Test to what extent higher order interactions are important;
4. Test whether the relationship between input and output is linear;
5. Test whether the method produces stable results over different instances of the same model.

To achieve this, all main parameters are tested at 4 different levels (-75% , -50% , $+50\%$ and $+100\%$). A full factorial design is used such that all combinations of all levels for each parameter is tested once. With a regression analysis all main and interaction effects can be quantified and tested. For every run, the optimal cost and the DC allocation are stored. The first one is used directly as outcome variable, the latter is used to compute 'DC change'. This outcome variable represents the difference in DC allocation with respect to a base case run. The base case run has all levels at their original value. For an experiment with 4 main parameters (e.g. transport cost, location cost, supply cost and demand) this leads to an experiment of 257 runs ($4^{\text{levels}} + 1 \text{ base case} = 257 \text{ runs}$). This experiment enables to get insight in issues 1 to 4.

To test the stability of the method (insight 5), the experiment needs to be replicated. This solution approach will be tested on two case studies, based on data of recent clients of ORTEC. Every case in OSCD has a unique structure. Although the overall

mathematical model is the same, applications can differ a lot. E.g. certain cost that are known in one case may not be present in another or proportionally much higher. In order to gain insight in the stability of the method the experiments are repeated with different instances of the case studies. New instances are created by changing the demand. This is a realistic way to generate new instances since changes in demand is a likely thing to happen (e.g. next year has different demand than this year). Demand is specified for each customer in a supply chain. By adding (or subtracting) a specific amount on each customer demand, a new instance is created. The random changes are made according to three different probability distributions:

- Uniform distribution with $min = -30\%$ and $max = +30\%$. With this distribution the total demand change has an expected value of 0% .
- Skewed uniform distribution with $min = -30\%$ and $max = +100\%$. With this distribution the total expected demand change is $+35\%$.
- Truncated Normal distribution with $\mu = 65, \sigma = 30, min = 30$ and $max = \infty$. The expected change of the total demand is -35% in this case. Figure 5 shows the three probability distributions.

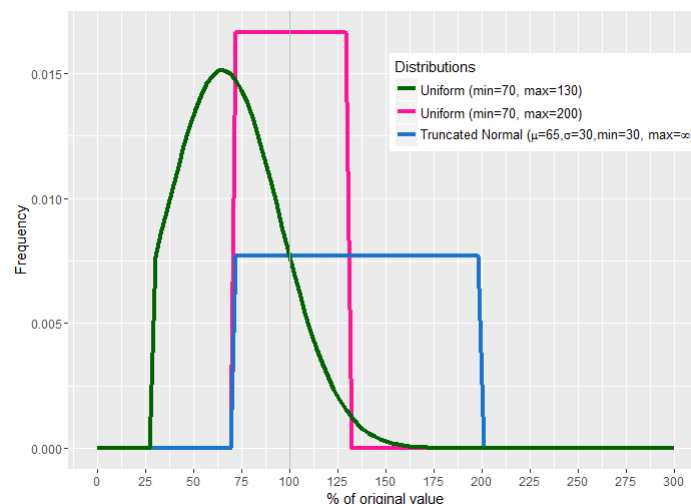


Figure 5: Density of distributions used for generating new instances.

All three instance types are generated 4 times. Including the base case run (with original demand) this gives 13 experiments per case study. Sampling from the uniform distributions is done through standard sampling functions in R. Sampling from the truncated normal distribution is done using an R package called TruncNorm. This random sampler uses a mixed accept-rejection algorithm to mimic the behavior of a truncated normal distribution (according to the work of J. Geweke, 1991).

3.2 SOLUTION PART 2

The results of part 1 will be used to investigate the possibility of finding any generic effects that occur in the mathematical model of OSCD. Generic effects can not be

estimated directly as the models can differ a lot. For example, the second case study has twice to cost of the first. In order to investigate any generic effect the outcome will be transformed such that different case studies are approximately on the same scale. This will be done by transforming optimal cost to a percentage difference with the base-case value. The results of the two case-studies will be used to formulate expectations on the third case. Figure 6 shows the overall structure of this work.

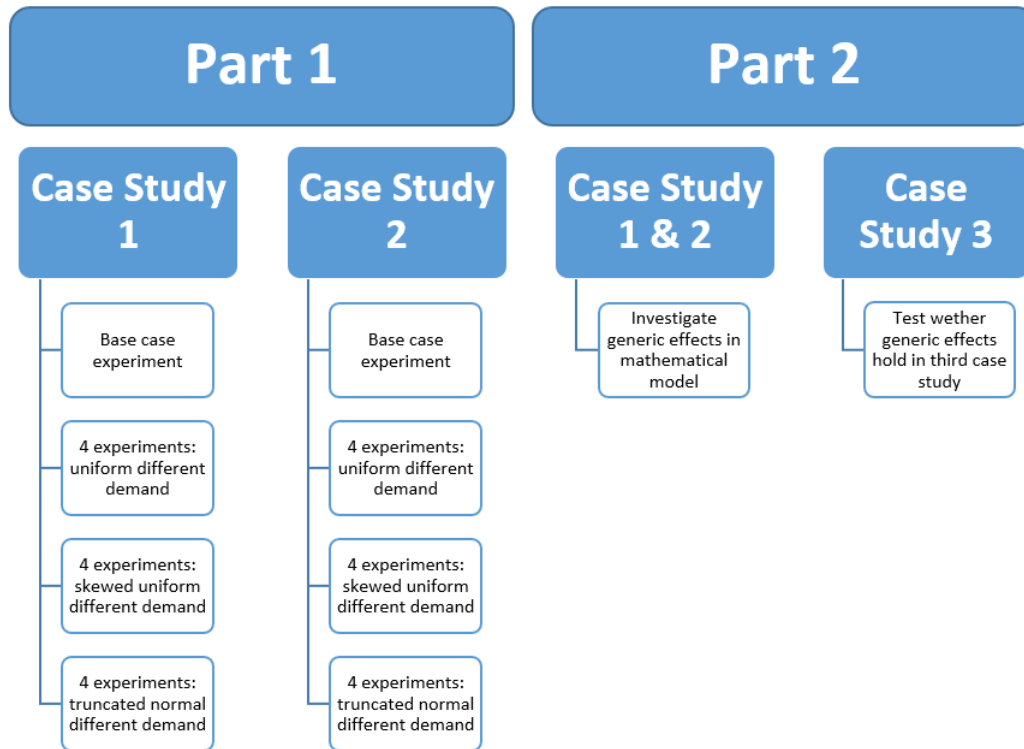


Figure 6: Structure of experiments and case studies.

3.2.1 Analysis Methods

The experiments are setup and executed in AIMMS within the OSCD software package. The end product of this project will be a library in OSCD. Analysis in that library is done by creating the necessary functionalities (such as the generation of an experimental design or the execution of a linear regression capable of computing the explained variance per parameter).

The analysis of all results in this project is done using the statistical software package R. The experiments are analyzed using regression analysis. For optimal cost a multiple linear regression is used. In linear regression, optimal estimates are found by minimizing the sum of squares (Fox, 2008), i.e. the squared distance between observed and estimated outcome. The DC allocation is quantified by the number of DC's that are different from the base case solution. DC allocation has a categorical nature, which is a result of the fact that values are always integers (you can't have half a DC more). The categorical nature of DC allocation makes an ordinary regression analysis not suitable. Instead, an ordinal logistic regression is used. Ordinal logistic regres-

sion can capture the categorical nature of the outcome variable while also accounting for the inherent ordering of the levels (in contrast to multinomial logistic regression) (Kleinbaum & Klein, 2010).

In this work the cumulative link model is used. Christensen (2015) formulates this model as follows:

The model takes a response variable Y_i , that can have ordered $j = 1, \dots, J$ categories. In DC Change the categories are the amount of DC's different from the basecase solution. These categories are naturally ordered, for example ranging from 4 DC's less than the basecase solution to 10 DC's more than the basecase solution. The cumulative distribution of an observation being in category j is:

$$\gamma_{ij} = P(Y_i \leq j) = \pi_{i1} + \dots + \pi_{ij}, \quad (2)$$

Where π_{ij} is the probability that the i th observation falls in category j . The logit link function is defined by $\text{logit}(\pi) = \log[\pi/(1 - \pi)]$. Using this link function changes to formulation to:

$$\text{logit}(\gamma_{ij}) = \text{logit}(P(Y_i \leq j)) = \log \frac{P(Y_i \leq j)}{1 - P(Y_i \leq j)}. \quad (3)$$

The cumulative logit model can be written as a regression equation:

$$\text{logit}(\gamma_{ij}) = \alpha_j - x_i^T \beta. \quad (4)$$

where x_i is a vector of explanatory variables of the i th observation and β is a vector with regression estimates. The intercept of each class j is denoted by α_j . With this model, for every observation a most likely class can be computed (Christensen, 2015).

The importance of parameters in a linear regression is quantified by computing the explained variance. The explained variance per parameter gives an interpretable value (a percentage). For ordinal logistic regression such a quantification is harder to make. Instead, the parameter estimate values can be compared, since all input parameters are on the same scale (going from -75% to $+100\%$). In addition, the prediction accuracy of the model when used for new data can be used to test the model performance.

Part III

RESULTS

CASE STUDY 1: LACTALIS

4.1 DESCRIPTION

Lactalis produces and supplies dairy products world wide. They make products such as cheese, milk, yoghurt, butter and cream. ORTEC has helped them with optimizing the supply chain of the Canadian subsidiary, called Parmalat. In this case, the client requested insight in the impact of different future scenario's on DC allocation and cost. Different scenario's had a different demand of specific products or different possible DC locations. For the experiment, this case was used at base case level with all DC's available. The network consist of 14 DC's and 870 customers. In the original case, production was divided over several locations. For the purpose of this study, it was decided to combine all production in one location so that the network would be able to close a DC (which is not possible if it is also a production plant). The locations are shown in Figure 7.

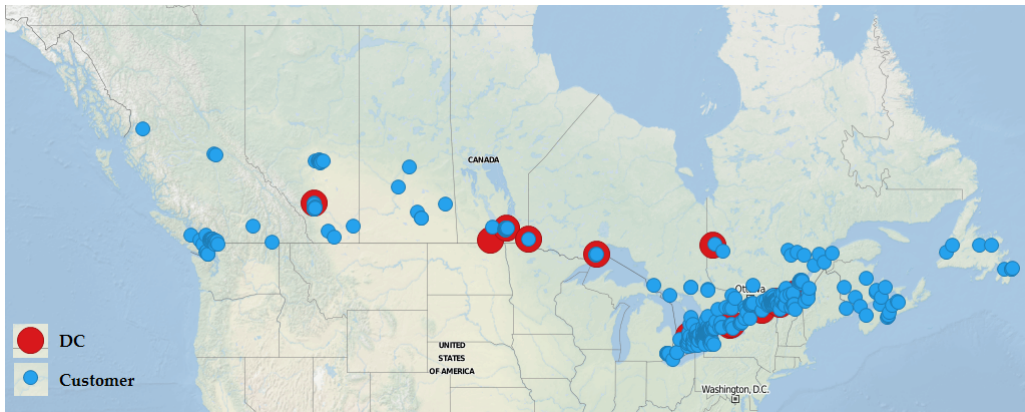


Figure 7: Map with locations of DC's and customers of Lactalis in Canada.

This case has values for the main parameters transport cost, location cost, handling cost and demand. These values will be configured in the experiment. This resulted in a 4 level factorial design with 4 parameters, leading to 256 configurations, excluding the run with base values. As described in chapter 3, the experiment is repeated in different instances, where the customer demand is changed. The aggregated results of all instances are presented in the next section.

4.2 RESULTS

Figures 8 and 9 show the results of the exploratory experiment where the input parameters are changed one at the time. Optimal cost is relatively most impacted by changes in transport cost or demand. DC allocation is only affected by demand in this experiment. In later experiments, it is shown that other parameters have an effect when changed simultaneously.

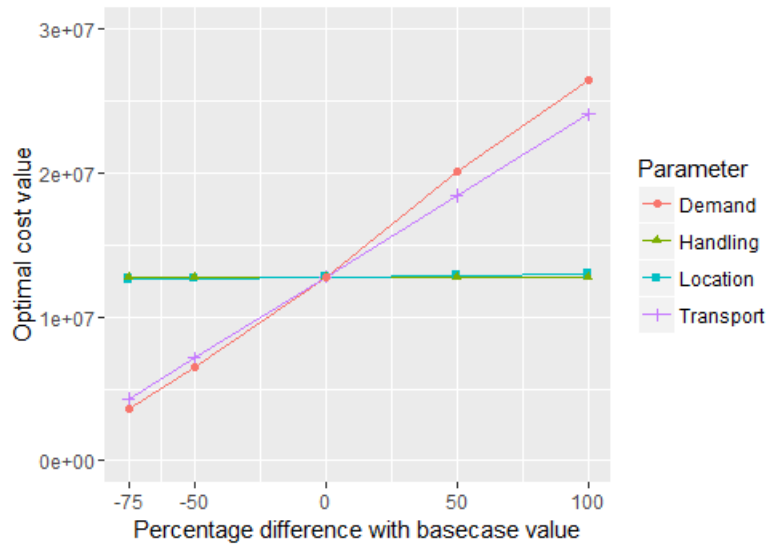


Figure 8: Change in the optimal cost as a result of one at the time changes in input parameters.

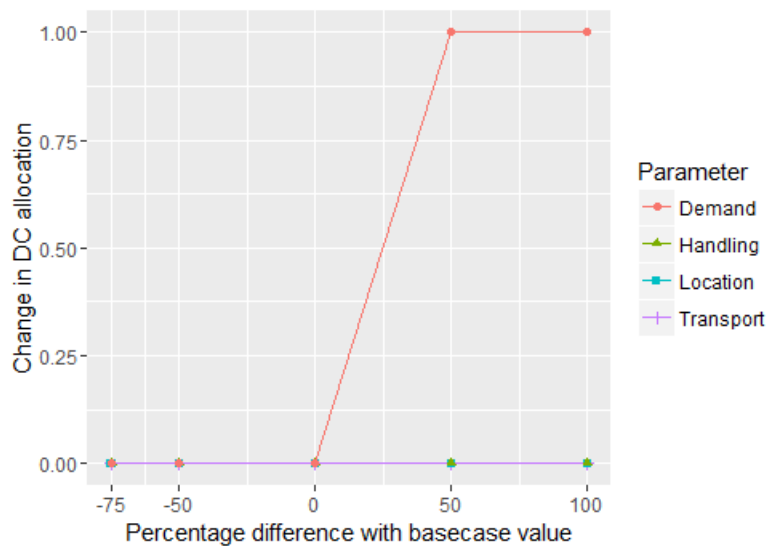


Figure 9: Change in DC allocation as a result of one at the time changes in input parameters.

4.2.1 *Optimal Cost*

Table 1 shows the outcome of a regression analysis with the optimal cost as a dependent variable. It shows that the main effects transport cost, location cost and demand are significant ($p < 0.05$), as well as the interaction effect between transport cost and demand. In addition, the 3 and 4-way interaction estimates are either very small, or non-existent. The parameter estimates represent the amount of change in optimal cost as a result of a one-unit change in the input parameters. This model explains the variance almost perfectly (99.995% of variance explained).

Table 1: Regression table optimal cost

Parameter	Estimate	Std. Error	t value	p value	% of variance
(Intercept)	1.29e+07	6.36e+03	2,023.93	0.00	
Transport	1.14e+05	8.87e+01	1,285.84	0.00	38.23
Location	1.85e+03	8.87e+01	20.80	0.00	0.01
Handling	4.45e-01	8.87e+01	0.01	1.00	0.00
Demand	1.25e+05	8.87e+01	1,403.30	0.00	45.09
Transport:Location	2.06e+00	1.24e+00	1.67	0.10	0.00
Transport:Handling	-5.40e-04	1.24e+00	-0.00	1.00	0.00
Transport:Demand	1.12e+03	1.24e+00	904.90	0.00	16.67
Location:Handling	-5.40e-04	1.24e+00	-0.00	1.00	0.00
Location:Demand	2.06e+00	1.24e+00	1.67	0.10	0.00
Handling:Demand	-5.40e-04	1.24e+00	-0.00	1.00	0.00
Transport:Location:Handling	6.55e-07	1.72e-02	0.00	1.00	0.00
Transport:Location:Demand	-2.73e-02	1.72e-02	-1.59	0.11	0.00
Transport:Handling:Demand	6.55e-07	1.72e-02	0.00	1.00	0.00
Location:Handling:Demand	6.55e-07	1.72e-02	0.00	1.00	0.00
Transport:Location:- Handling:Demand	-7.93e-10	2.40e-04	-0.00	1.00	0.00

The same analysis was done on the output of other instances. Figures 10 and 11 show how the regression coefficients and explained variance per parameter differ over the instances. For the regression coefficients, only the significant effects are shown. For the explained variance, only the parameters with more than 1% are shown. The figures show that the coefficient estimates vary over different instances but are relatively stable within the same type of instance. The skewed distributions (truncated normal and skewed uniform) show more change. This is to be expected, as the total demand at base level is different for these instances. The explained variance per parameter shows to be quite stable over different instances.

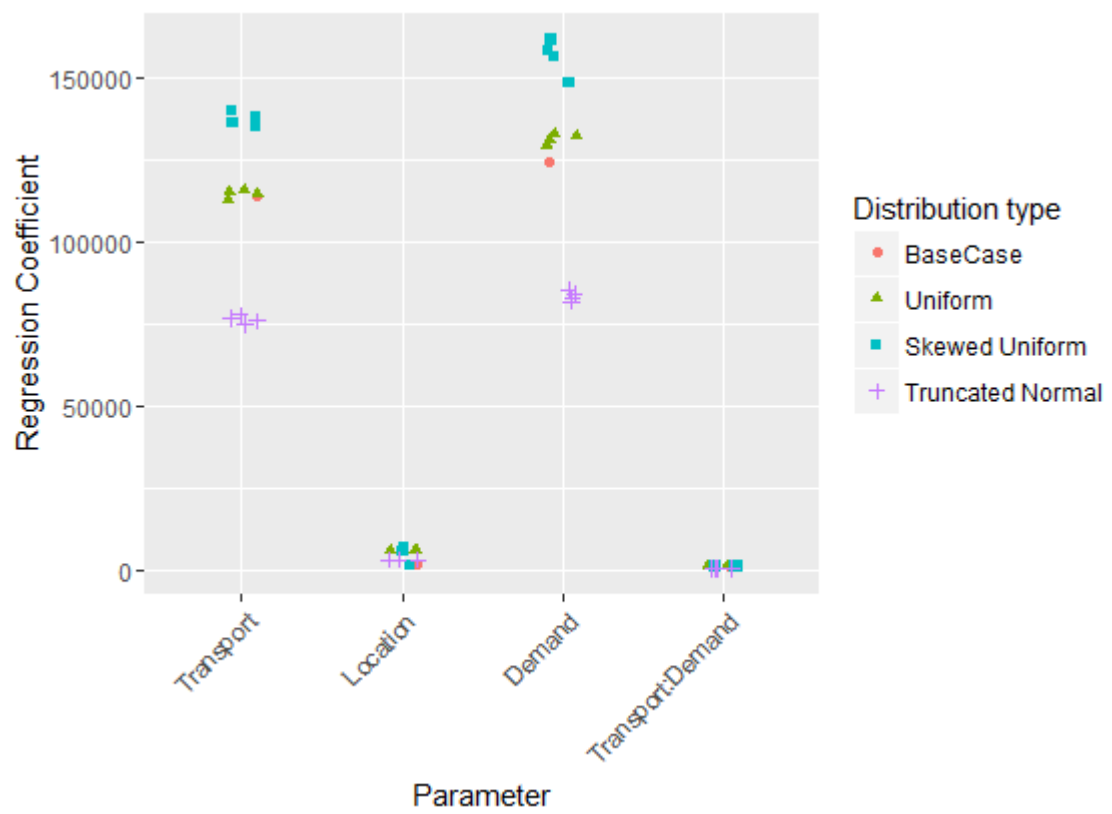


Figure 10: Regression estimates per instance for the optimal cost.

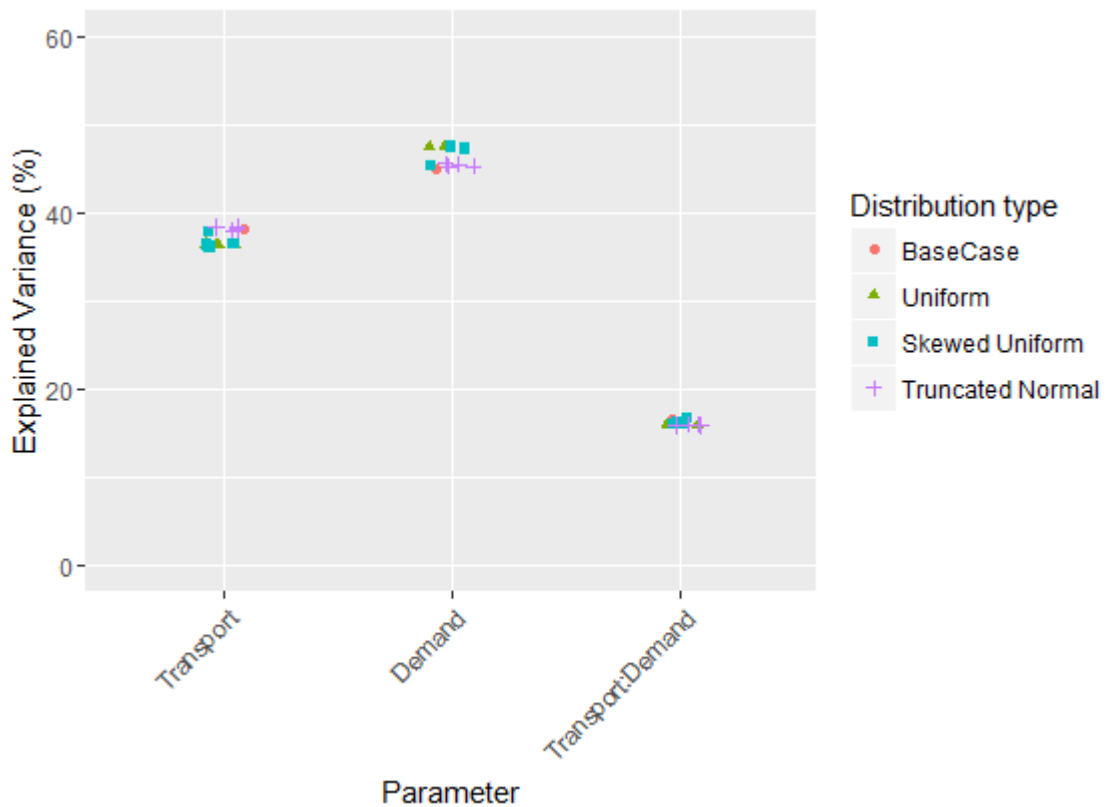


Figure 11: Explained variance per significant parameter per instance, for the optimal cost.

4.2.2 DC Allocation

The same type of analysis has been applied to the allocation of DC's. Lactalis showed to have a stable set of DC's. Figure 9 with the changes of the OAT analysis revealed that even with a lot of change (+100%) in the input parameters, the amount of DC's only changes by 1 center. In other words, in most of the experiments the amount of DC's is equal. This makes it hard to fit an ordinal logistic regression model and the analysis turned out to be problematic. Ordinal Logistic Regression is fitted using maximum likelihood. During the analysis it occurred that the optimization algorithm for fitting an ordinal logistic regression did not converge. This is caused by the fact that the model can't distinguish the outcome variable by its input parameters because there is too little variation in the outcome. The result is that the model can't find the standard errors of the parameters estimates and consequently no corresponding p value. The main objective of this work is to identify and quantify sensitivity of the input parameters. Although a linear regression does not capture the non-linear behavior of DC allocation, it can identify parameters that influence it and quantify them by their impact. For simplicity and interpretation purposes a linear regression is applied. Although this is not the proper analysis for such an outcome, it does give insight in which input parameters influence the amount of DC's.

Table 2 shows the results of a regression analysis with change in DC allocation as output. The same main effects are significant but now all their interaction effects as well, including the three-way interaction. The dependent variable in this analysis, DC allocation change, has a different scale than optimal costs. In all configurations of an experiment, DC change varies from -1 to 1 , while the optimal cost changes from approximately 120 thousand to 5 million. This leads to the fact that the estimates seem small. The estimates of the interaction coefficients are even smaller. Transport:Location for example is equal to 0.00002 . Both values taken at their largest value, means that the value of the input parameter should be multiplied by 10.000 ($100 * 100$). This means that the interaction in this model causes a maximum change of $10.000 * 0.00002 = 0.2$ DC's. This model explains 72.80% of the variance.

Table 2: Regression table DC allocation.

Parameter	Estimate	Std. Error	t value	p value	% of variance
(Intercept)	-1.01e-01	1.01e-02	-10.04	0.00	
Transport	1.33e-03	1.41e-04	9.44	0.00	10.74
Location	-1.49e-03	1.41e-04	-10.60	0.00	9.56
Handling	-4.95e-07	1.41e-04	-0.00	1.00	0.00
Demand	1.33e-03	1.41e-04	9.44	0.00	10.74
Transport:Location	1.95e-05	1.96e-06	9.94	0.00	9.92
Transport:Handling	6.00e-10	1.96e-06	0.00	1.00	0.00
Transport:Demand	-1.72e-05	1.96e-06	-8.76	0.00	10.87
Location:Handling	6.00e-10	1.96e-06	0.00	1.00	0.00
Location:Demand	1.95e-05	1.96e-06	9.94	0.00	9.92
Handling:Demand	6.00e-10	1.96e-06	0.00	1.00	0.00
Transport:Location:Handling	-7.27e-13	2.73e-08	-0.00	1.00	0.00
Transport:Location:Demand	-2.52e-07	2.73e-08	-9.23	0.00	10.04
Transport:Handling:Demand	-7.27e-13	2.73e-08	-0.00	1.00	0.00
Location:Handling:Demand	-7.27e-13	2.73e-08	-0.00	1.00	0.00
Transport:Location:- Handling:Demand	8.82e-16	3.81e-10	0.00	1.00	0.00

Figures 12 and 13 show the difference of estimates and explained variance over all instances. Again, the results within a type of instance is quite stable. There are some large differences in the estimates of demand.

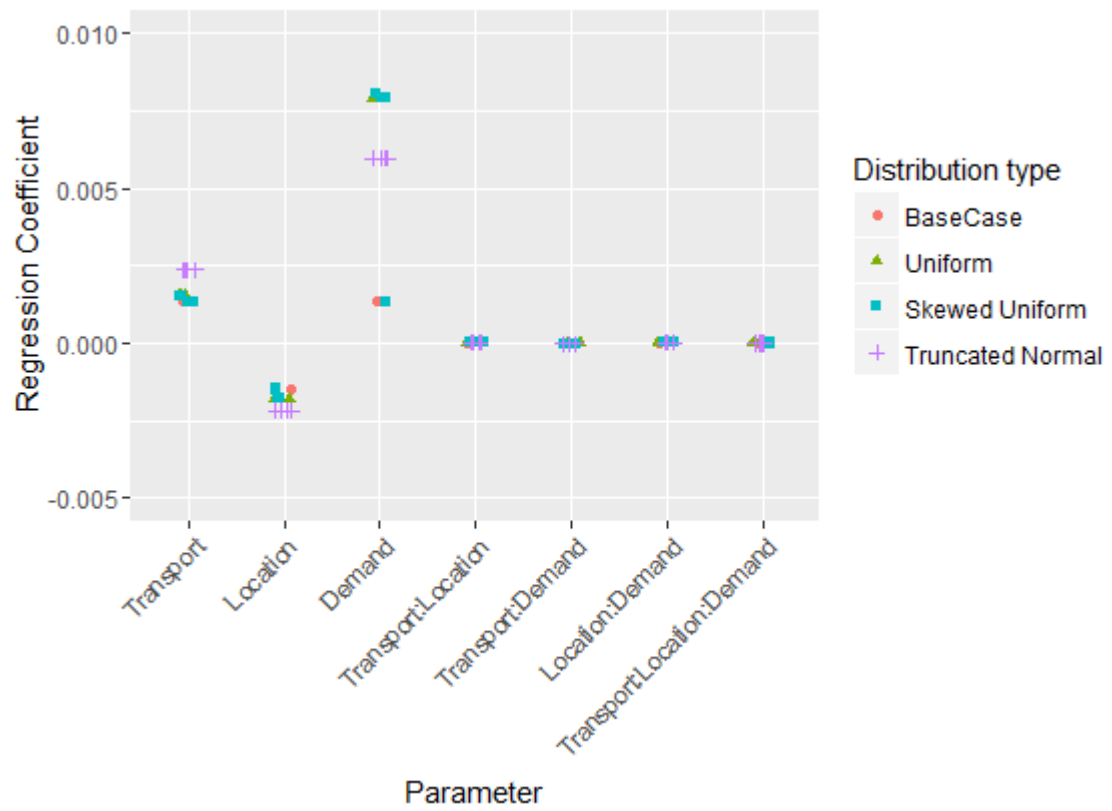


Figure 12: Significant regression coefficients per instance for DC allocation.

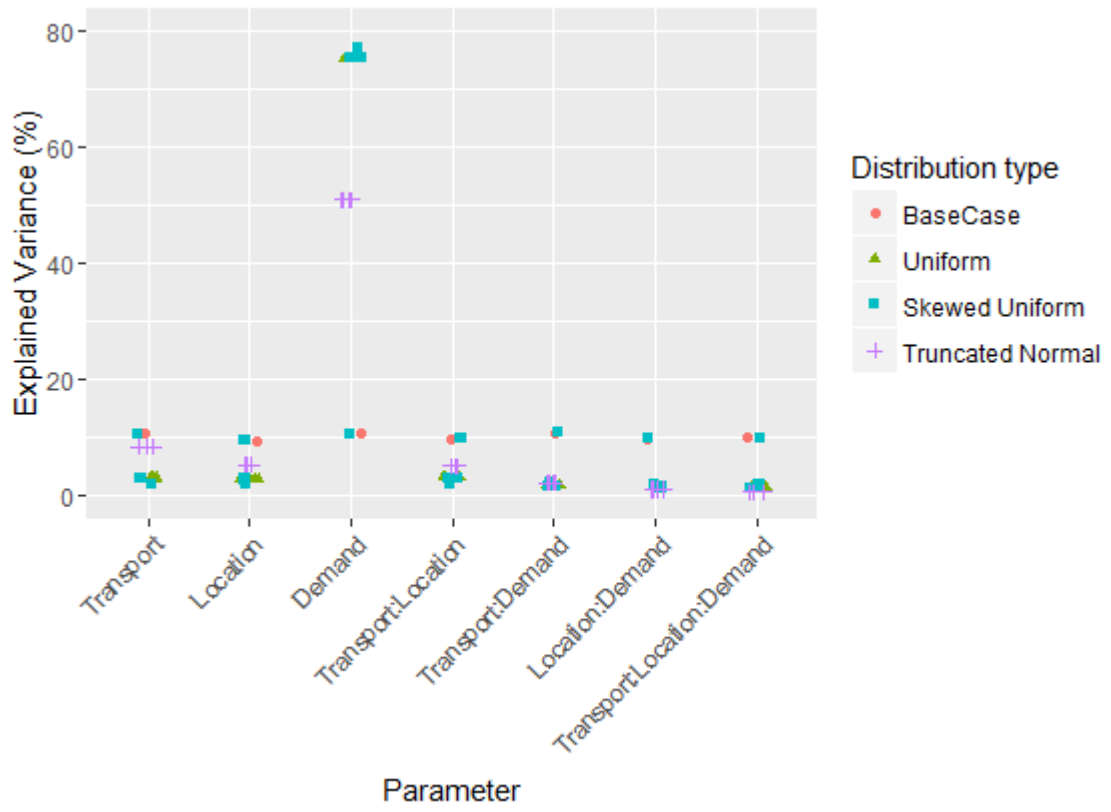


Figure 13: Explained variance per significant parameter per instance for DC allocation.

CASE STUDY 2: MEDUX

5.1 DESCRIPTION

Medux is an organization active in the Dutch health-care industry. They mainly focus on offering health and wellness products and delivering them to peoples homes. Recently, ORTEC helped Medux with a strategic decision about their supply chain network, using OSCD. Medux has several DC's in the Netherlands and wanted to know how they could spend their resources optimally. Most importantly, they wanted to know whether some of the DC's could be closed, while still being able to deliver efficiently throughout the Netherlands. This strategic decision makes this case suitable for a sensitivity experiment. E.g. Medux would be interested in knowing to what extent optimal allocation of DC's is stable over different ranges of the input parameters. Since Medux uses one of ORTEC's products (ORTEC Service Planning) a lot of data are available on their demand and transport over the last few years.

The network consists of 17 DC's and 797 customers. A map with the locations is shown in Figure 14. The DC's consist of current centers that are used by Medux and possible new locations for a DC. The network, optimized by OSCD, is allowed to use any center in the set, the only restriction being the fixed cost required for using a center.

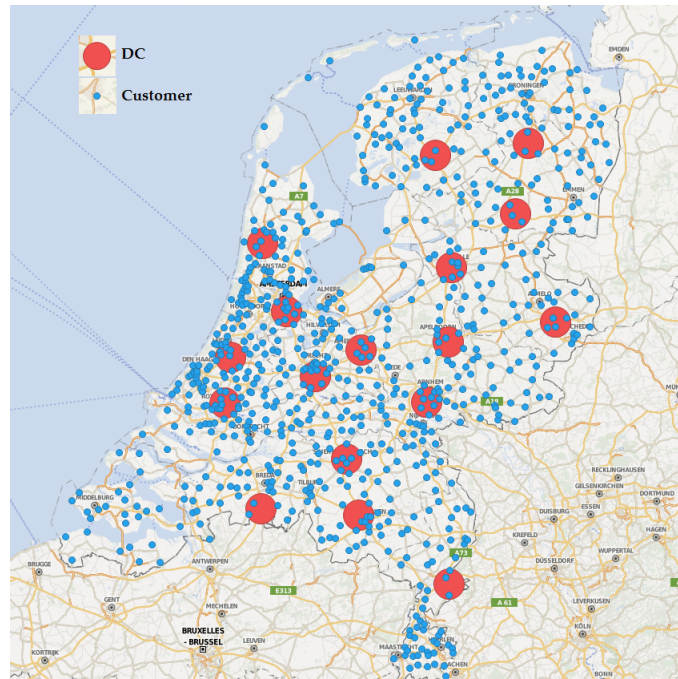


Figure 14: Map with locations of DC's and customers of Medux

This case has values for the main parameters transport cost, supply cost, location cost and demand. Again, this results in a factorial design with 256 configurations, not counting the run with base values. The results for all instances are shown in the next section.

5.2 RESULTS

Figures 15 and 16 shows the results of an exploratory experiment where the input parameters are changed one at the time. The optimal cost is relatively most impacted by changes in transport cost or demand. DC allocation is affected by demand, transport cost and location cost.

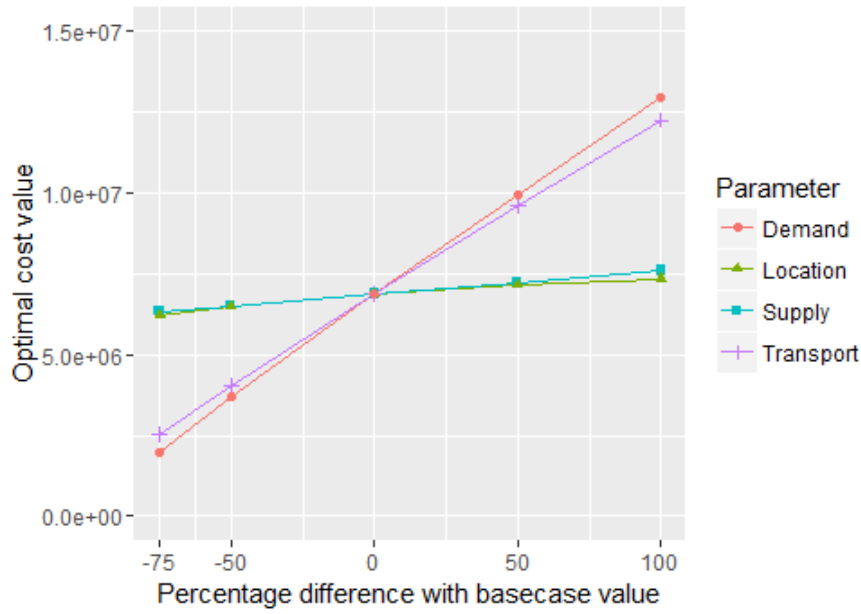


Figure 15: Change in optimal cost as result of one at the time changes in input parameters.

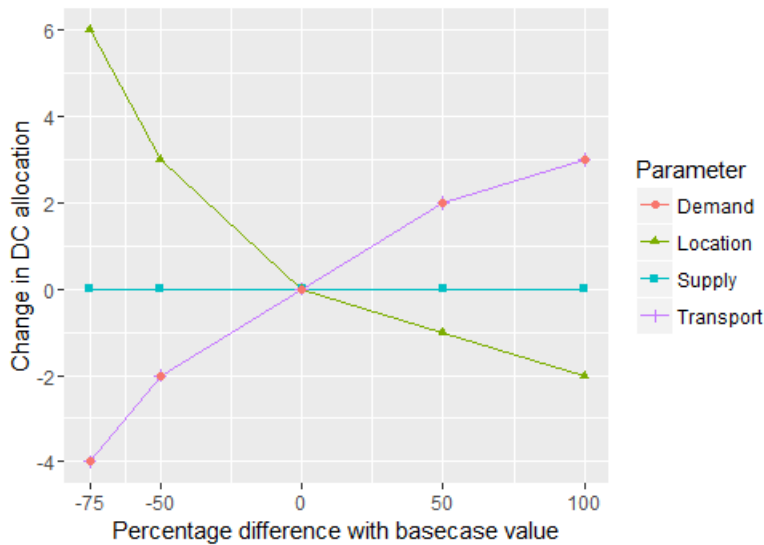


Figure 16: Change in DC allocation as result of one at the time changes in input parameters.

5.2.1 Optimal Cost

The regression table with the optimal cost as a dependent variable is shown in table 3. All main effects except for the supply cost are significant, although the explained variance of the location cost is relatively small. The same applies to the significant interaction terms that include location cost. This regression model explains 99.97%, an almost perfect fit.

Table 3: Regression table optimal cost.

Parameter	Estimate	Std. Error	t value	p value	% of variance
(Intercept)	5.87e+06	7.33e+03	800.44	0.00	
Transport	5.48e+04	1.02e+02	535.49	0.00	41.33
Supply	2.05e+01	1.02e+02	0.20	0.84	0.00
Location	5.02e+03	1.02e+02	49.14	0.00	0.36
Demand	5.47e+04	1.02e+02	534.95	0.00	41.25
Transport:Supply	-1.26e-01	1.42e+00	-0.09	0.93	0.00
Transport:Location	3.32e+01	1.42e+00	23.34	0.00	0.07
Transport:Demand	5.19e+02	1.42e+00	364.46	0.00	16.87
Supply:Location	-3.81e-01	1.42e+00	-0.27	0.79	0.00
Supply:Demand	-3.61e-01	1.42e+00	-0.25	0.80	0.00
Location:Demand	3.40e+01	1.42e+00	23.87	0.00	0.08
Transport:Supply:Location	1.64e-03	1.98e-02	0.08	0.93	0.00
Transport:Supply:Demand	1.90e-03	1.98e-02	0.10	0.92	0.00
Transport:Location:Demand	1.52e-01	1.98e-02	7.64	0.00	0.01
Supply:Location:Demand	5.59e-03	1.98e-02	0.28	0.78	0.00
Transport:Supply:- Location:Demand	-2.38e-05	2.76e-04	-0.09	0.93	0.00

Figures 17 and 18 show the stability of those estimates over all instances. Again, the regression coefficients change between different types of instances, while they remain stable within one type instance. The explained variance estimates show to be stable over all instances.

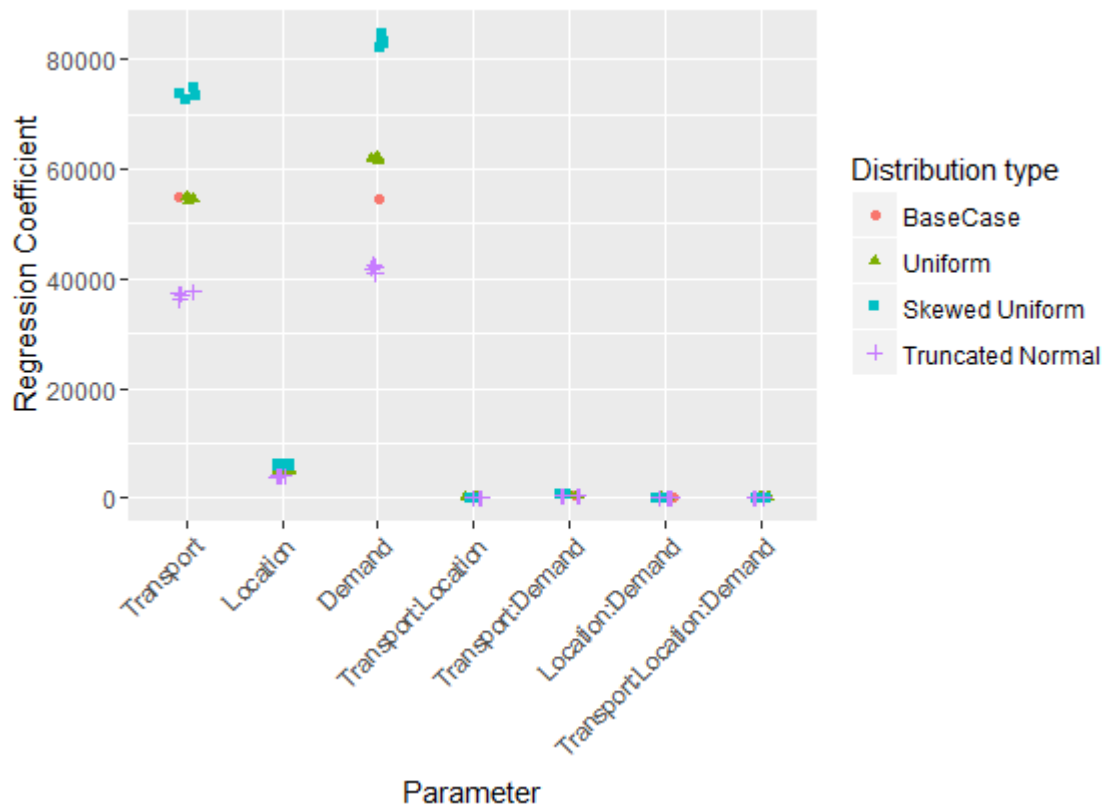


Figure 17: Significant regression estimates per instance for optimal cost.

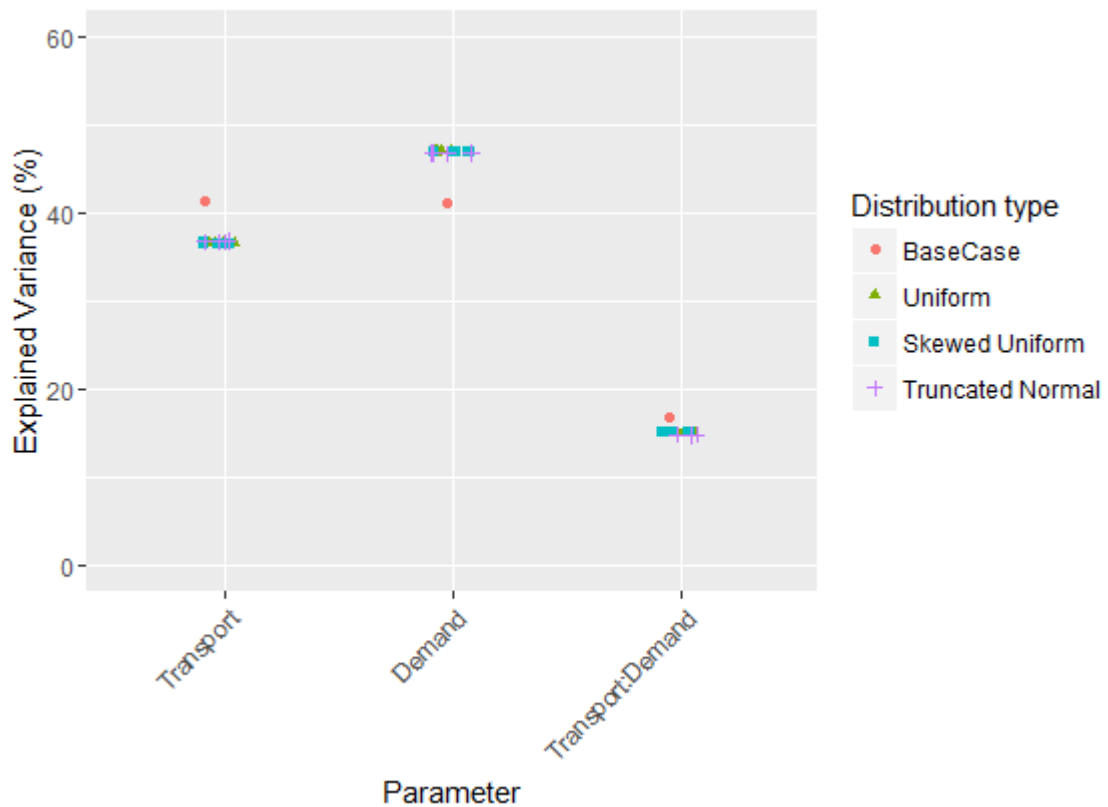


Figure 18: Explained variance per significant parameter per instance for optimal cost.

5.2.2 DC Allocation

The ordinal logistic regression table with DC change as output, can be found in table 4 and 5. The change in DC allocation for this case ranges from -5 to 11 . The logodds between two classes is calculated by taking the corresponding intercept estimate and add that to the product of the input parameter values and their estimates (from table 5). The exponent of this value leads to the odds between the categories. By doing this, the most likely class for every observation can be found.

Transport cost, location cost and demand all have significant main and interaction effects. In addition, transport cost and demand have positive estimates (with equal magnitude), indicating that the higher the value, the more likely it is to end up in a higher category (i.e. have more DC's). For supply cost, this value is negative leading to an opposite effect. An estimate of a parameter represents the increase in logodds (or decrease for negative values) if a parameter is increased by 1%.

Table 4: Intercepts of ordinal logistic regression model on DC Change

Parameter	Estimate	Std. Error	z value	p value
-5 -4	-8.23	0.62	-13.37	0.00
-4 -3	-5.59	0.50	-11.25	0.00
-3 -2	-5.08	0.48	-10.47	0.00
-2 -1	-2.10	0.46	-4.58	0.00
-1 0	0.05	0.42	0.12	0.90
0 1	1.33	0.43	3.07	0.00
1 2	1.70	0.43	3.92	0.00
2 3	4.60	0.52	8.76	0.00
3 4	7.60	0.61	12.45	0.00
4 5	7.97	0.63	12.72	0.00
5 6	8.91	0.67	13.23	0.00
6 7	11.49	1.01	11.41	0.00
7 8	14.53	1.41	10.28	0.00
8 10	17.58	1.48	11.85	0.00
10 11	24.51	2.23	10.99	0.00

Table 5: Ordinal logistic regression model on DC change

Parameter	Estimate	exp(Estimate)	Std. Error	z value	p value
Transport	7.64e-02	1.08e+00	0.01	13.91	5.27e-44
Location	-7.47e-02	9.28e-01	0.01	-13.78	3.30e-43
Supply	-1.36e-05	1.00e+00	0.00	-0.01	9.96e-01
Demand	7.64e-02	1.08e+00	0.01	13.91	5.27e-44
Transport:Location	-1.48e-04	1.00e+00	0.00	-3.03	2.45e-03
Transport:Supply	9.22e-08	1.00e+00	0.00	0.00	9.98e-01
Transport:Demand	2.01e-04	1.00e+00	0.00	3.83	1.27e-04
Location:Supply	-1.15e-07	1.00e+00	0.00	-0.00	9.98e-01
Location:Demand	-1.48e-04	1.00e+00	0.00	-3.03	2.45e-03
Supply:Demand	9.22e-08	1.00e+00	0.00	0.00	9.98e-01
Transport:Location:Supply	1.52e-09	1.00e+00	0.00	0.00	9.98e-01
Transport:Location:Demand	-9.79e-07	1.00e+00	0.00	-1.46	1.45e-01
Transport:Supply:Demand	-1.18e-09	1.00e+00	0.00	-0.00	9.98e-01
Location:Supply:Demand	1.52e-09	1.00e+00	0.00	0.00	9.98e-01
Transport:Location:- Supply:Demand	-1.66e-11	1.00e+00	0.00	-0.00	9.98e-01

For Lactalis a linear regression was applied to DC change. As a comparison, the same type of regression is applied to this case as well (see table 6). The same parameters are significant and the direction of the estimates is the same (i.e. Location has a negative effect). The magnitude of the impact is different. This is caused by the fact that the range of DC change for Medux is much wider. This model captures the

variance of the outcome relatively better than the Lactalis case, as the total explained variance is 94.6%.

In Figure 12 the stability of those estimates is plotted for all instances. The points are all relatively close to each other, indicating that the experiments reveal similar results both within and between different types of instances.

Table 6: Linear regression model on DC Change

Parameter	Estimate	Std. Error	z value	p value
Transport	7.64e-02	5.49e-03	13.91	0.00
Supply	-1.36e-05	2.67e-03	-0.01	1.00
Location	-7.47e-02	5.42e-03	-13.78	0.00
Demand	7.64e-02	5.49e-03	13.91	0.00
Transport:Supply	9.22e-08	3.46e-05	0.00	1.00
Transport:Location	-1.48e-04	4.90e-05	-3.03	0.00
Transport:Demand	2.01e-04	5.25e-05	3.83	0.00
Supply:Location	-1.15e-07	4.19e-05	-0.00	1.00
Supply:Demand	9.22e-08	3.46e-05	0.00	1.00
Location:Demand	-1.48e-04	4.90e-05	-3.03	0.00
Transport:Supply:Location	1.52e-09	5.17e-07	0.00	1.00
Transport:Supply:Demand	-1.18e-09	4.65e-07	-0.00	1.00
Transport:Location:Demand	-9.79e-07	6.72e-07	-1.46	0.15
Supply:Location:Demand	1.52e-09	5.17e-07	0.00	1.00
Transport:Supply:Location:Demand	-1.66e-11	6.61e-09	-0.00	1.00

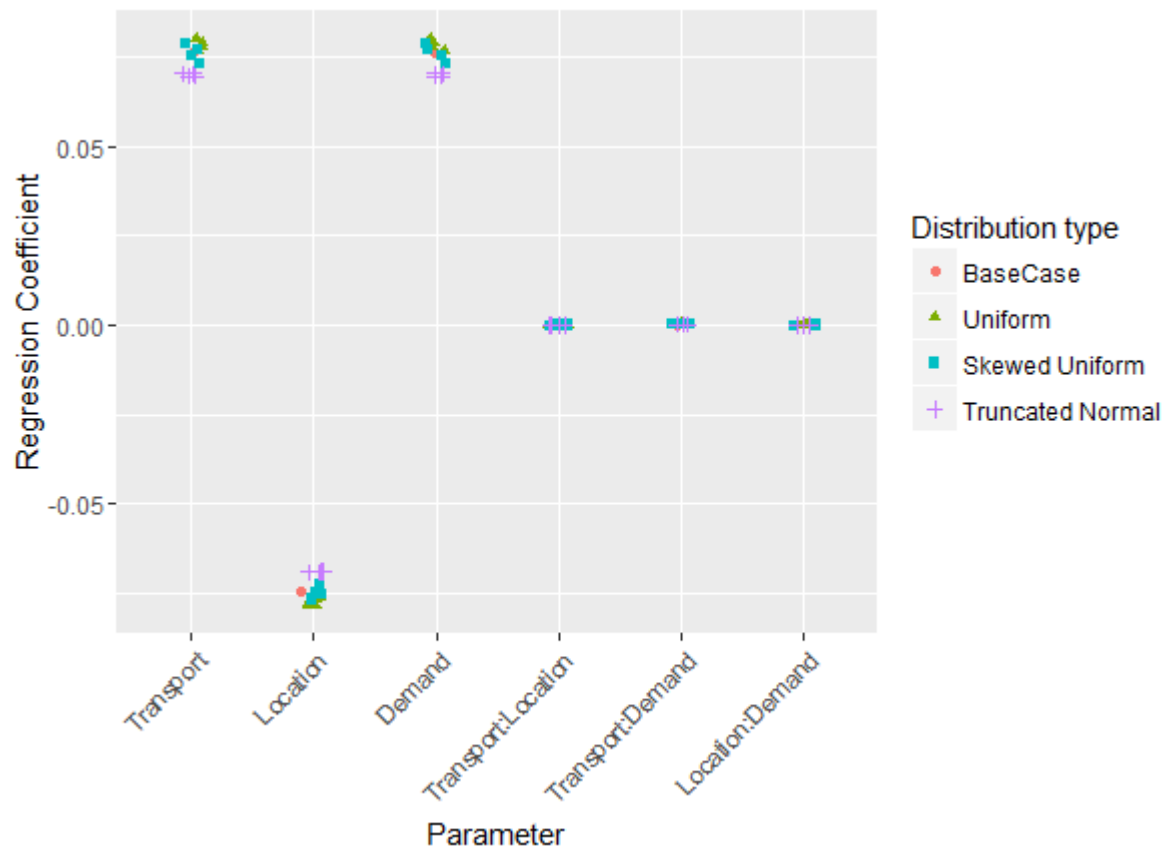


Figure 19: Significant regression coefficients per instance for DC allocation.

SOLUTION PART 2: GENERIC EFFECTS IN OSCD

In this chapter, generic models for the input-output relationship in OSCD will be investigated. These models will be built, based on the results in the two case studies (Lactalis and Medux). The models will be compared to the results of a third case study. This third case study contains data from Isreal Chemicals Ltd (ICL). ICL is an international company that develops, produces and markets chemical products used for agriculture, food and engineering. ORTEC helped them with their supply chain network in the United Kingdom and Ireland. The ICL network consists of 2177 customers and 7 DC's. The locations in the ICL network are shown in Figure 20.

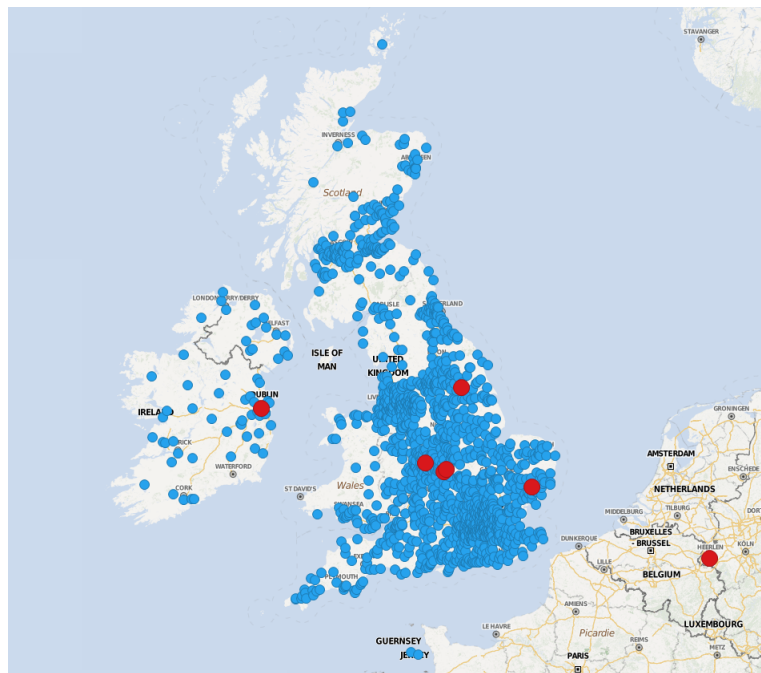


Figure 20: Map with locations of DC's and customers of ICL

6.1 GENERIC REGRESSION MODEL ON OPTIMAL COST

It is expected that the parameters in the mathematical model of OSCD have an impact on the optimal cost that is relatively similar in different cases. To confirm this, a regression model is created that includes both Lactalis and Medux data. For this analysis a new outcome variable is created. The optimal cost is not suitable for analysis as they

differ in size between Lactalis en Medux (i.e. the Lactalis network has approximately twice the cost of the Medux network). Instead, a new variable is created that represents the proportional change of the optimal cost (in percentages), as compared to its base case value. Table 7 show the regression table on the relative change in optimal cost in both Medux and Lactalis. The adjusted R^2 of this model is 0.99. The handling and supply cost are left out of the model as they only explained a small part of the variance in the regression models and they are not present in both cases. The estimates represent the percentage change as a result of a 1% change in the input parameter. For example, with an increase of 1% change in transport cost, the optimal cost increases by 0.85%.

Table 7: Regression Model on relative change in optimal cost for Lactalis and Medux combined

Parameter	Estimate	Std. Error	t value	Pr(> t)	% of variance
(Intercept)	3.42e-01	4.88e-01	0.70	0.48	
Transport	8.46e-01	6.80e-03	124.40	0.00	36.33
Location	6.09e-02	6.80e-03	8.95	0.00	0.20
Demand	9.68e-01	6.80e-03	142.33	0.00	46.99
Transport:Location	2.87e-04	9.47e-05	3.03	0.00	0.02
Transport:Demand	8.17e-03	9.47e-05	86.19	0.00	15.36
Location:Demand	5.11e-04	9.47e-05	5.39	0.00	0.06
Transport:Location:Demand	1.31e-06	1.32e-06	0.99	0.32	0.00

The regression model based on the two case studies is used to predict values in the ICL case. For the ICL case a similar experiment (including main cost parameters with 4 different values) is conducted. Figure 21 shows a plot of the predicted values versus the observed values. This plot suggests that the prediction is quite accurate. The correlation between the observed and predicted values is 0.999. The blue line represents perfect predictions. As the observed cost get higher, the prediction model overestimates the actual value (suggested by the fact that the black dots are above the blue line).

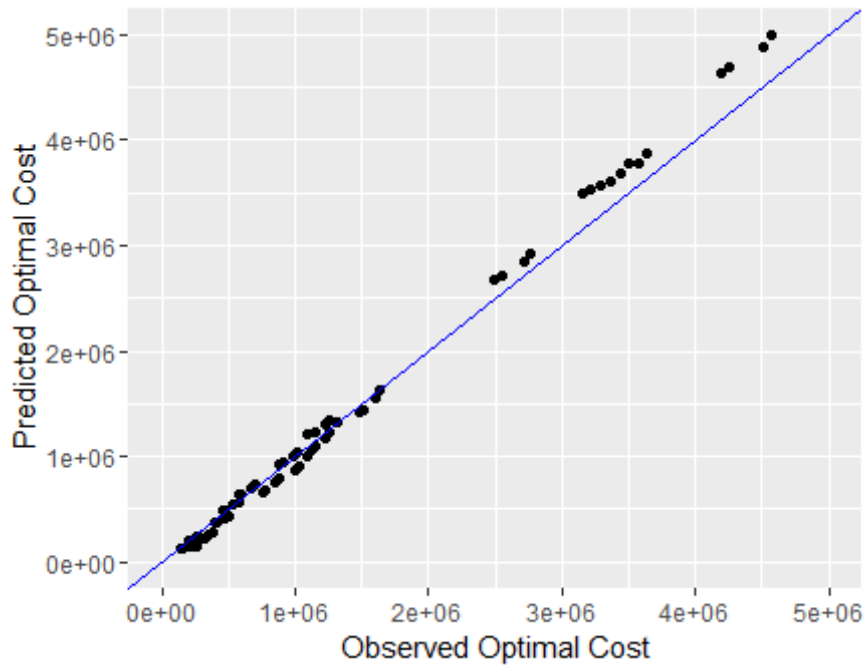


Figure 21: Observed and predicted values for optimal cost versus for the ICL case.

6.2 GENERIC REGRESSION MODEL ON DC ALLOCATION

A similar analysis has been done for DC change. An ordinal regression is applied to the collaborate results of DC allocation (DC Change) of both cases. The regression model is shown in table 8 and 9. The logodds between two classes is calculated by taking the corresponding intercept estimate and add that to the product of the input parameter values and their estimates (from table 8). The odds are computed by taking to exponent of the logodds. By doing this, the most likely class for every observation can be found. This is done in order to formulate predictions in the ICL case. Transport cost and demand have positive estimates, indicating that the higher the value, the more likely it is to end up in a higher category (i.e. have more DC's). For supply cost, this value is negative leading to an opposite effect.

Table 8: Ordinal logistic regression model on DC change cost for Lactalis and Medux combined

Parameter	Estimate	exp(Estimate)	Std. Error	z value	p value
Transport	1.56e-02	1.02e+00	0.00	11.77	5.90e-32
Location	-1.50e-02	9.85e-01	0.00	-11.45	2.46e-30
Demand	2.29e-02	1.02e+00	0.00	14.77	2.27e-49
Transport:Location	1.18e-06	1.00e+00	0.00	0.07	9.44e-01
Transport:Demand	2.83e-05	1.00e+00	0.00	1.63	1.02e-01
Location:Demand	-1.89e-05	1.00e+00	0.00	-1.11	2.65e-01
Transport:Location:Demand	8.89e-07	1.00e+00	0.00	3.63	2.83e-04

Table 9: Intercepts of ordinal logistic regression model on DC Change

Class	Estimate	Std. Error	z value	p value
-5 -4	-3.64	0.20	-18.52	0.00
-4 -3	-2.90	0.17	-17.48	0.00
-3 -2	-2.74	0.16	-17.02	0.00
-2 -1	-2.15	0.15	-14.58	0.00
-1 0	-1.56	0.14	-11.26	0.00
0 1	0.42	0.13	3.20	0.00
1 2	2.59	0.17	15.18	0.00
2 3	3.08	0.18	16.90	0.00
3 4	3.75	0.21	18.22	0.00
4 5	3.86	0.21	18.35	0.00
5 6	4.12	0.22	18.54	0.00
6 7	4.43	0.24	18.51	0.00
7 8	4.63	0.25	18.37	0.00
8 10	5.09	0.29	17.74	0.00
10 11	6.98	0.55	12.77	0.00

Figure 22 shows a plot of the observed values versus the predicted ones of DC allocation. The observed values are spread (jittered) around their value to prevent the dots from overlapping. For this plot, DC change has been transformed to represent the amount of DC's in the solution. There is clearly more variation between observed and predicted than with optimal cost. The percentage of correctly predicted observations is 63%. The correlation between the observed and predicted values is 0.58. Table 10 shows a cross-table of the observed and predicted values. The plot (and the table) suggests that in some extreme cases the model both over- and underestimates the actual amount of DC's.

mize a different loss function. The difference in performance on observed-predicted correlation and classification accuracy is related to the differences in these loss functions.

Table 11: Regression Model on DC change in optimal cost for Lactalis and Medux combined

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.53e-01	9.86e-02	2.57	0.01
Transport	1.84e-02	1.38e-03	13.34	0.00
Location	-1.78e-02	1.38e-03	-12.91	0.00
Demand	2.15e-02	1.38e-03	15.65	0.00
Transport:Location	-4.79e-05	1.92e-05	-2.50	0.01
Transport:Demand	5.06e-05	1.92e-05	2.64	0.01
Location:Demand	-5.14e-05	1.92e-05	-2.68	0.01
Transport:Location:Demand	2.68e-07	2.67e-07	1.01	0.32

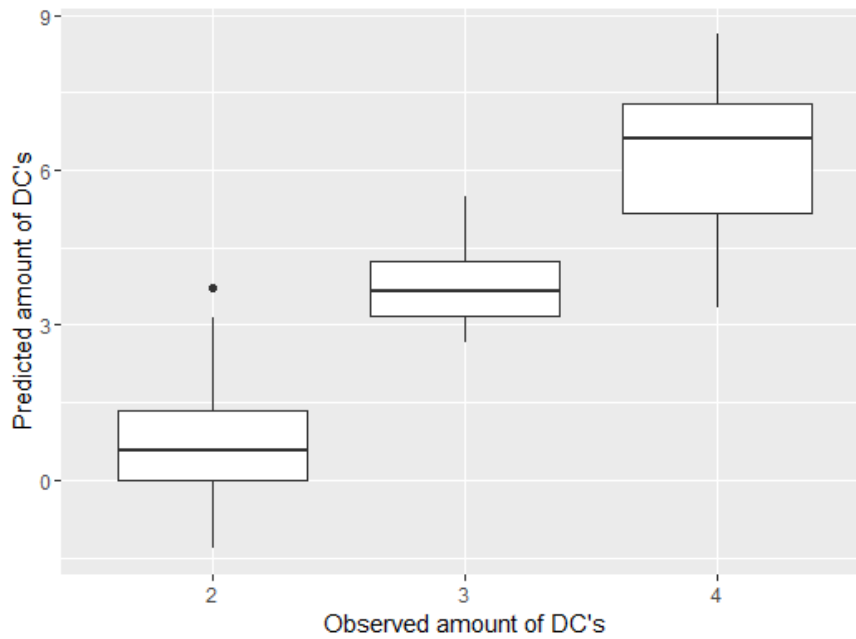


Figure 23: Boxplot of predicted values per observed amount of DC's, based on linear regression model.

Part IV

CONCLUSION

CONCLUSION

Users of OSCD wish to design a supply chain network that is optimal for a given set of parameter values. However, those input parameters have uncertainty and, consequently, the output is uncertain as well. In this work the relationship between the uncertainty of the input and the level of uncertainty in the output, is investigated.

This work came up with a thorough analysis of two case studies. With the results of these case studies some conclusions can be made about generic effects in the mathematical model of OSCD. In addition, this project provided a tool that can be used by consultants and clients who work with OSCD. This tool can run sensitivity experiments, analyze the results, visualize the effects and communicate the detailed results.

In case studies, it was shown that different configurations of a case study could lead to different network designs (e.g. for Medux the number of DC's could increase by 10 locations). Such differences in optimal designs have a large impact on the actual cost.

In the literature various methods to compute the impact of uncertainty, i.e. sensitivity methods, are used. In this work a factorial experimental design with a matching regression analysis has been used. This was chosen, since it can capture the complexity of the relationship (i.e. main effects and high-way interactions) without putting too much pressure on the computational budget. This method has been applied to two case studies and is tested on a third. The means for applying these methods automatically are built in the optimization software AIMMS and can be used in future projects on new OSCD cases. The tool is capable of performing and analyzing the experiments on the main cost parameters or on parameters manually chosen by the user. In addition to the tool, this work aimed at finding any generic effects that are present in OSCD. This can help consultants and users better understand the mathematical model, without having to perform experiments.

7.1 CONCLUSIONS SOLUTION PART 1

Part 1 of the solution consisted of two case studies. In these two case studies all main cost parameters and the demand parameter were tested for their sensitivity on the output. This was done with a factorial experiment with 4 different levels. The results showed that the high-way interaction (3-way and up) were mostly insignificant and had always little explained variance.

The results also showed that all effects of input parameters on optimal costs behave linearly. A 2-level experiment would therefore be sufficient. This does not apply to the

change in DC's as the number of DC's is bounded below and above (e.g. a minimum number of DC's is required for any network). In order to get a sufficient idea of the effects, a 2-level experiment with wide ranges (e.g. -75% and + 100%) can obtain interpretable results. A small range might change too little in the design to capture any differences in DC allocation. This design will not capture the correct estimates of input parameter on DC change (since it behaves non-linearly) but it will give insight in which parameters influence DC change most and by what amount, at the used levels.

Finally, part 1 revealed estimates are relatively stable over different instances. Estimates of coefficients on optimal cost/ DC allocation show some variation but that can be explained by the different amount of flow (of products) in the network.

7.2 CONCLUSIONS SOLUTION PART 2

In part 2 of the solution the same case studies were analyzed. The goal of this analysis was to find generic effects that occurred over both cases. To achieve this, regression models were built on data where both cases were combined. For the optimal cost, the outcome variable is transformed to represent the relative change to the base case solution. This is done, since models in OSCD can have a different size (and therefore varying magnitudes of cost). The generic models are tested by using them to predict values in a third case. An experiment with this third case is executed to verify the predictions.

The generic effects on the optimal cost seem relatively constant over different OSCD cases. Prediction in the third case is shown to be accurate, since the predicted values are close to the observed values. For DC allocation the results are less constant. The signs of the effects are similar, but the magnitude is different. This can be explained by the fact that the possible changes in DC allocation in both models are varying. A supply chain network design in OSCD will find the optimal selection of DC's, but the amount of DC's is influenced by several characteristics of the case study (e.g. DC size constraints, amount of DC's available, transport constraints, etc.). Using the model in the third case shows some predictive power but it overestimates the amount of DC's that the model will change.

7.3 FUTURE WORK

For feasibility purposes, the case studies that are used, are relatively small (a single run in an experiment solves within one minute). In addition, the main cost parameters were aggregated per topic. For example, transport cost has several underlying parameters. In the exploratory phase of this work, it was found that those underlying parameters have varying magnitudes of impact on the optimal solution. For future work, it will be interesting to further develop insight in the impact of individual input parameters. In addition, larger models (with more input parameters and a larger network) are worth investigating. For larger models it can be beneficial to experiment according to a iterative procedure that tests all parameters at one level before running the full experiment.

OSCD has more parameters that can have uncertainty. However, they are not investigated in this study. The model assumes that products in the supply network are

delivered using a 'milk run'. This means that every customer is visited and delivered a specific dropsize (amount of product delivered at one visit) while also delivering to other customers which have a distance (inter customer distance). The inter customer distance and the dropsize are examples of input parameters that have to be estimated. In future work these parameters could be included in the study.

To increase the predictive ability of the generic models, more case studies can be used to train the model. This is especially relevant for DC allocation and possibly as well for new output variables that have not been investigated in this work (e.g. net profit or CO₂ emission). It can also be relevant to add additional cases to investigate the behavior of optimal cost, when changes in input parameters are outside of the range that is used in this work (-75% to +100%).

BIBLIOGRAPHY

- [1] Bogenovo, E., & Plischke, E. (2015). Sensitivity analysis :A review of recent advances, *European Journal of Operational Research*, 248, p. 869-887.
- [2] Busschers, R. (2012). Generating alternative solutions for a supply chain optimization model. Master thesis.
- [3] Cadini, F., De Sanctis, J., Cherubini, A., Zio E., Riva, M., & Guadagnini, A. (2012). Nominal Range Sensitivity Analysis of peak radionuclide concentrations in randomly heterogeneous aquifers. *Annals of Nuclear Energy*, 47, p. 166-172.
- [4] Christensen, R.H.B. (2015). Analysis of ordinal data with cumulative link models: estimation with R-package ordinal. Retrieved September 30, 2016, from https://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf
- [5] Deleris, L.A., & Erhun, F. (2005). Risk management in supply networks using Monte-Carlo simulation. *Proceedings of the Winter Simulation Conference*, 1643-1649.
- [6] Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, Ed. E. Keramidas and S. Kaufman, pp. 571-8. Fairfax Station, VA: Interface Foundation of North America.
- [7] Hendriks, H.B., & Van Beek, P. (1991). *Optimaliseringstechnieken*, Third edition, chapters 4 & 6.
- [8] Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Second edition, chapter 5 & 6.
- [9] Kim, J, Realff, M.J., Lee, J.H., Whittaker, C. & Furtner, L. (2011). Design of biomass processing network for biofuel production using an MILP model. *Biomass and Bioenergy*, 35, p. 853-871.
- [10] Kim, J., Realff, M.J., & Lee, J.H. (2011). Optimal design and global sensitivity analysis of biomass supply chain networks for biofuels under uncertainty. *Computers and Chemical Engineering*, 35, p. 11738-1751.
- [11] Kleinbaum, D.G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. Third edition, chapter 13.
- [12] Montgomery, D., & Runger, G.C. (2011). *Applied Statistics and Probability for Engineers*, Fifth Edition, chapters 13 & 14.
- [13] Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55, 271-280.

- [14] Taha, H.A. (2007). *Operations Research: An Introduction*. 8th edition, chapter 3.
- [15] Van de Heijden, I. (2013). *Parameter Uncertainty in the Downstream Supply Chain of Oil and Gas*. Master thesis.
- [16] Wang, X., Yoo, M., Glardon, R., & Furbringer, J. (2009). Performance and sensitivity analysis of supply chain alternative configurations - A case study. *Computers and Industrial Engineering*. International Conference on, Troyes, 2009, 708-713.
- [17] Zhang, X.Y., Trame, M.N., Lesko, L.J., & Schmidt, S. (2015). Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *Center for Pharmacometrics and Systems Pharmacology*, 4, 69-79.
- [18] Zhang, Z., Yang, X., Oseledets, I.V., Karniadakis, G.E., & Daniel, L. (2015). Enabling High-Dimensional Hierarchical Uncertainty Quantification by ANOVA and Tensor-Train Decomposition. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 34, 63-76.

Appendices



MATHEMATICAL MODEL

A description of the generic supply chain that can be handled by OSCD is presented below. This representation does not cover the full functionality of OSCD. This version is a simplified version to enable the reader to understand the structure of the model.

A.1 SETS

The model is capable of handling different products, that are produced in facilities. Warehouses, distribution centers and customers are specified by locations. Transportation modes (e.g. Truck or Boat) can have different capacities and costs.

$l \in L$	Locations
$f \in F$	Facilities
$p \in P$	Products
$tm \in TM$	Transportation Mode
$pm \in PM$	Production Modes
$t \in T$	Time Period
$g \in G = P \cup C$	Ingredients

A.2 PARAMETERS

A.2.1 Capacity Parameters

$Demand_{p,l,t}^{max}$	Maximum Demand for product p at location l in time period t
$Demand_{p,l,t}^{min}$	Minimum Demand for product p at location l in time period t
$Supply_{p,l,t}^{max}$	Maximum Supply for product p at location l in time period t
$Supply_{p,l,t}^{min}$	Minimum Supply for product p at location l in time period t
$Stock_{p,l,t}^{max}$	Maximum inventory level of product p at location l at the end of time period t
$Stock_{p,l,t}^{min}$	Minimum inventory level of product p at location l at the end of time period t
$Flow_{p,tm,l_1,l_2,t}^{max}$	Maximum flow of product p with transport mode tm from l_1 to l_2 during time period t
$Flow_{p,tm,l_1,l_2,t}^{min}$	Minimum flow of product p with transport mode tm from l_1 to l_2 during time period t
$Prod_{pm,f,l,t}^{max}$	Maximum production in virtual mode pm on facility f at location l during time period t
$Prod_{pm,f,l,t}^{min}$	Minimum production in virtual mode pm on facility f at location l during time period t

A.2.2 Revenue and Cost Parameters

$RDemand_{p,l,t}$	Sales price of product p at location l during time period t
$CSup_{p,l,t}$	Cost of supply for product p at location l at time t
$CStock_{p,l,t}$	Variable stock cost of product p at location l at time t
$CTrans_{p,tm,l_1,l_2,t}^{var}$	Variable transport cost to move product p with transport mode tm from l_1 to l_2 at time t
$CTrans_{tm,l,t}^{fix}$	Fixed transport cost for using transport mode tm at location l during time period t
$CLoc_{p,tm,l,t}^{Handin}$	Handling cost for product p arriving with transport mode tm at location l at time t
$CLoc_{p,tm,l,t}^{Handout}$	Handling cost for product p leaving with transport mode tm at location l at time t
$CFac_{f,l,t}^{fix}$	Fixed cost for operating facility f at location l during time period t
$CLoc_{l,t}^{fix}$	Fixed cost for using location l during time period t

A.2.3 Other Parameters

$Loss_{p,tm,l_1,l_2,t}$	Loss factor occurring when transporting p with tm from l_1 to l_2 in time period t
$BOM_{g,pm,f,l,t}$	Bill of materials parameter indicating for each in-going and out-going ingredient g how much is needed to create a virtual unit of pm on facility f on location l in time period t
$Stock_{p,l}^{start}$	Initial stock of product p at location l at the beginning of the first time period ($t=0$).

A.3 VARIABLES

A.3.1 Decision Variables

$Flow_{p,tm,l,l_1,t} \in \mathbb{R}^+$	Flow of product p going out of location l headed to l_1 with transport mode tm during time period t
$Prod_{pm,f,l,t} \in \mathbb{R}^+$	Production in production mode pm in facility f on location l in time period t
$Supply_{p,l,t} \in \mathbb{R}^+$	Total supply of product p at location l during time period t
$Demand_{t,p,l} \in \mathbb{R}^+$	Total demand of product p at location l during time period t
$InUse_{l,t}^{Loc} \in \{0,1\}$	Variable that indicates whether a location l is active (1) in time period t or not (0). Activity refers to the location being open or closed.
$InUse_{f,l,t}^{Fac} \in \{0,1\}$	Variable that indicates whether a facility f at location l is active (1) in time period t or not (0)

A.3.2 Derived Variables

$Flow_{p,tm,l_1,l,t}^{in} \in \mathbb{R}^+$	Flow of product p coming into location l from l_1 with transport mode tm during time period t
$Stock_{p,l,t}^{in} \in \mathbb{R}^+$	Stock build up of product p on location l during time period t
$Stock_{p,l,t}^{out} \in \mathbb{R}^+$	Stock build down of product p on location l during time period t
$ProdOut_{p,f,l,t} \in \mathbb{R}^+$	Amount of product p that is produced on facility f in location l during time period t

A.4 CONSTRAINTS

A.4.1 General constraints

$Variables \geq 0$
$Binary\ variables \in \{0,1\}$

A.4.2 Demand constraints

The demand has to fulfill the demand limitations.

$$Demand_{p,l,t}^{min} \leq DemAll_{p,l,t} \leq Demand_{p,l,t}^{max} \quad \forall p, l, t$$

A.4.3 Supply constraints

The supply has to fulfill the demand limitations.

$$Supply_{p,l,t}^{min} + \sum_i Supply_{p,l,t,i}^{Delta_{min}} \leq Supply_{p,l,t} \quad \forall p, l, t$$

$$Supply_{p,l,t} \leq Supply_{p,l,t}^{max} + \sum_i Supply_{p,l,t,i}^{Delta_{max}} \quad \forall p, l, t$$

A.4.4 Flow constraints

Flow into a location is related to the Flow out of the originating location, taken into account the (optional) loss during the transport.

$$Flow_{p,tm,l_1,l,t}^{in} = Flow_{p,tm,l_1,l_2,t} * Loss_{p,tm,l_1,l_2,t} \quad \forall p, tm, l_1, l_2, t$$

A.4.5 Production constraints

$$ProdOut_{p,f,l,t} = \sum_{pm} (Prod_{pm,f,l,t} * BOM_{p,pm,f,l,t}) \quad \forall p, pm, f, l, t | BOM_{p,pm,f,l,t} > 0$$

$$ProdUse_{g,f,l,t} = \sum_{pm} (Prod_{pm,f,l,t} * BOM_{g,pm,f,l,t}) \quad \forall g, pm, f, l, t | BOM_{g,pm,f,l,t} < 0$$

A.4.6 Location constraints

Locations with flow are active by definition. The flow of products in a location has to be equal to the flow out:

$$\sum_{tm,l_1} Flow_{p,l_1,l,tm}^{in} + \sum_f ProdOut_{t,l,p,f} + Supply_{t,l,p} + SurplusBalance(t, l, p) =$$

$$\sum_{tm,l_2} Flow_{t,l,l_2,p,tm}^o + \sum_f ProdIn_{t,l,p,f} + Sales_{t,l,p} + ShortageBalance(t, l, p)$$

A.4.7 Stock constraints

Inventory allows enterprises to maintain supply with fluctuations in demand. By adding and taking stock, the stock level per period changes.

$$Stock_{p,l,t}^{in} - Stock_{p,l,t}^{out} = Stock_{p,l,t}^{level} - Stock_{p,l,t-1}^{level} \quad \forall p, l, t$$

The stock has to be within the accepted limits.

$$Stock_{p,l,t}^{min} + \sum_i Stock_{p,l,t,i} \leq Stock_{p,l,t}^{level} \quad \forall p, l, t$$

A.4.8 Cost Equations

All costs associated with moving flow through the supply chain:

- Total Variable Transportation costs:

$$TTC_t^{var} = \sum_{p,tm,l_1,l_2} (Flow_{p,tm,l_1,l_2,t} * CTrans_{p,tm,l_1,l_2,t}^{var}) \quad \forall t$$

- Total Fixed Transportation costs:

$$TTC_t^{fix} = \sum_{p,tm,l_1,l_2} (InUse_{p,tm,l_1,l_2,t}^{Flow} * CTrans_{p,tm,l_1,l_2,t}^{fix}) \quad \forall t$$

- Total Production costs:

$$TPC_t = \sum_{pm,f,l} (Prod_{pm,f,l,t} * \sum_g BOM_{g,pm,f,l,t}) \quad \forall t$$

- Total Supply costs:

$$TSC_t = \sum_{p,l} (Supply_{p,l,t} * CSu_{p,l,t}) \quad \forall t$$

- Total Inventory costs:

$$TIC_t = \sum_{p,l} \left(\frac{Stock_{p,l,t} + Stock_{p,l,t-1}}{2} * CStock_{p,l,t} \right) \quad \forall t$$

- Total Location Variable costs:

$$TLC_t^{var} = \sum_{tm,p,l_1,l_2} (Flow_{p,tm,l_1,l_2,t,i} * CLoc_{p,tm,l_1,t}^{Handout} + Flow_{p,tm,l_1,l_2,t,i}^{in} * CLoc_{p,tm,l_2,t}^{Handin}) \quad \forall t$$

- Total Location Fixed costs:

$$TLC_t^{fix} = \sum_l (InUse_{l,t}^{Loc} * CLoc_{l,t}^{fix}) + \sum_{f,l} (InUse_{f,l,t}^{fac} * CFac_{f,l,t}^{fix}) \quad \forall t$$

- Total Revenue:

$$TR_t = \sum_{p,l} (DemAll_{p,l,t} * RDemand_{p,l,t}) \quad \forall t$$

A.5 OBJECTIVE

The main objective function minimizes costs:

$$MIN \sum_t (TLC_t^{fix} + TLC_t^{var} + TSC_t + TPC_t + TTC_t^{fix} + TTC_t^{var}).$$

Alternatively, the following objective function maximizes profit:

$$\text{MAX} \sum_t (TR_t - TLC_t^{fix} - TLC_t^{var} - TSC_t - TPC_t - TTC_t^{fix} - TTC_t^{var}).$$