J. Bulters

The distribution process of multi-dose vaccines

A robust optimization approach to supply uncertainty

TU Delft

The distribution process of multi-dose vaccines

A robust optimization approach to supply uncertainty

By

J. Bulters

Master Thesis

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Department Maritime and Transport Technology of Faculty Mechanical, Maritime and Materials Engineering of Delft University of Technology
to be defended publicly on Tuesday March 7, 2023 at 13:00 AM

Student number:	4189981
MSc track:	Multi-Machine Engineering
Report number:	2023.MME.8767
Supervisor:	Dr.ir N. Pourmohammad-Zia Dr.ir F. Schulte
Thesis committee:	Dr.ir. F. Schulte, TU Delft committee Chair, 3ME Dr.ir. N. Pourmohammad-Zia, TU Delft committee member, CiTG X. Tang, TU Delft committee member, 3mE
Date:	March 7, 2023

TU Delft

Preface

This thesis treats the distribution process of vaccines, starting in a time where the virus really hits the world and actively dominates life by causing serious diseases, death and fear. At the same time it has a huge influence on the social life in the form of serious deregulating lockdowns. Starting this project in the middle of the 'c-word' pandemic combined with my interest for optimizing processes, quickly ignited the interest of the treated problem.

I personally experienced that this lockdown really hampered the real life interaction with the supervisors and members of the group for reflection and mirroring , having on the spot feedback, and learning from others. All these I would have loved to have more of, but regretfully limited by the lockdowns. A pity, but it is what it is.

At the same time I just got a fantastic daughter, who created a lot of mental energy. And yes she asked of course for a lot of attention and care but also supported me in the creation of practical knowledge on logistics: The supply chain of diapers, having the right food at the right time, organising the baby care and last but not least how to get enough sleep! (The latter has failed me several times).

It is has been a weird, but fruitful, coincidence: working on a distribution scheme for a vaccine at the very moment the virus was spreading and in using an algorithm that is inspired by the genetic evolution process that shows similarities to how the virus finds its way.

In this thesis you find the result of this work , something I am proud of, certainly given the circumstances in which it was created. I hope you will enjoy the reading!

Summary

The distribution chain of two-dose vaccines by an air carrier (KLM cargo) with its practical features is modelled to study the influence of supply uncertainty. First the goal is to find an efficient solution approach for the model which gives good quality solutions in reasonable computation time. Secondly with the developed solution method the supply uncertainty is taken into account using robust optimization.

Solving the model using an exact solution method leads to a too large increase of computation time with increase of the model size for the purpose of robust analysis. Nevertheless, given the high reliability of the model the results of this commercial model are used as a reference.

To circumvent the high computation time, two alternative solution methods are developed and implemented. Respectively the genetic algorithm and the rolling horizon method.

A basic implementation of the genetic algorithm does not provide adequate results and gets trapped in a local optimum. An analysis shows that measures are needed to increase the flexibility and stimulate the algorithm to find so called “transaction-less” events. To achieve this, a toolbox is developed. The toolbox consists of analysis tools (measurement of convergence rate, sparsity and a diversity measurement) and tools intended to improve the convergence and accuracy of the solution. These latter are an adapted mutation operator that stimulates the number of transaction-less events (sparsity) and an approach for directed mutation. Each measure by itself has a positive effect on the initial convergence rate, but the algorithm still gets trapped at a somewhat improved local optimum at a level of approximately 10% above the optimum solution. A strong improvement is found by combining these measures of the toolbox resulting in solutions that approach the optimal solution within less than 5% for a single destination at a very high convergence rate. The best found combination of measures are implemented to solve a multi-destination problem. The results prove to be not as good as the result for the single destination: the gap to the optimal solution is roughly 10% and the convergence rate is somewhat slower. Probably this is due to the fact that the boundary conditions impose a reduction of flexibility for the multi-destination setting. On the other hand this finding might provide a base for worthwhile future work. Since this is solver is not (yet) suitable to be used in the robustness analysis. Three different implementations of the rolling horizon method were made: (i) the straight forward (myopic) approach, (ii) extending the time window with relaxed periods and (iii) a shifting rolling horizon. The straight forward approach is significantly improved by using the relaxed and shifting methods. Considering the accuracy and the computation time the rolling horizon with shifting is the best solution method. This method is used for the analysis of robust optimization

To study the effect of the supply uncertainty, where the exact probability distribution is unknown, robust optimization is applied. The uncertainty resides in a polyhedral uncertainty set. The influence of supply uncertainty is studied over a specified practical interval. Over this interval the total cost proves to be sensitive to variation in the ordering amount. The shifting rolling horizon method is used to compute a flight and order plan given an uncertainty in the supply amount. To protect the solution against supply uncertainty a high level of conservatism proves to be required. This is achieved by using gamma values in the range of 0.7 to 0.9. For a Γ of 0.9 the price of the robustness is 20% higher than the optimal costs in which no uncertainty is taken into account. If lower costs are desired , at the price of somewhat less protection against uncertainty, but still at a significant robustness level, the flight and order scenario that fits a Γ of 0.8 and even 0.7 can be used lowering the estimated cost with respectively 4% and 10%.

List of tables

List of figures

Content

1. Introduction

Immunization is an effective way to prevent or decrease the spreading of contagious and transferable diseases such as the current SARS-Cov-2 virus. A controlled method for achieving immunity is by means of vaccines. Vaccines consist of several components, where the most important component is a weakened version of the germ that causes the disease or a small characteristic piece of its protein. By injecting this substance the immune system of a person produces antibodies against this germ to develop immunity against the disease, without the negative effects of suffering from the disease [2].

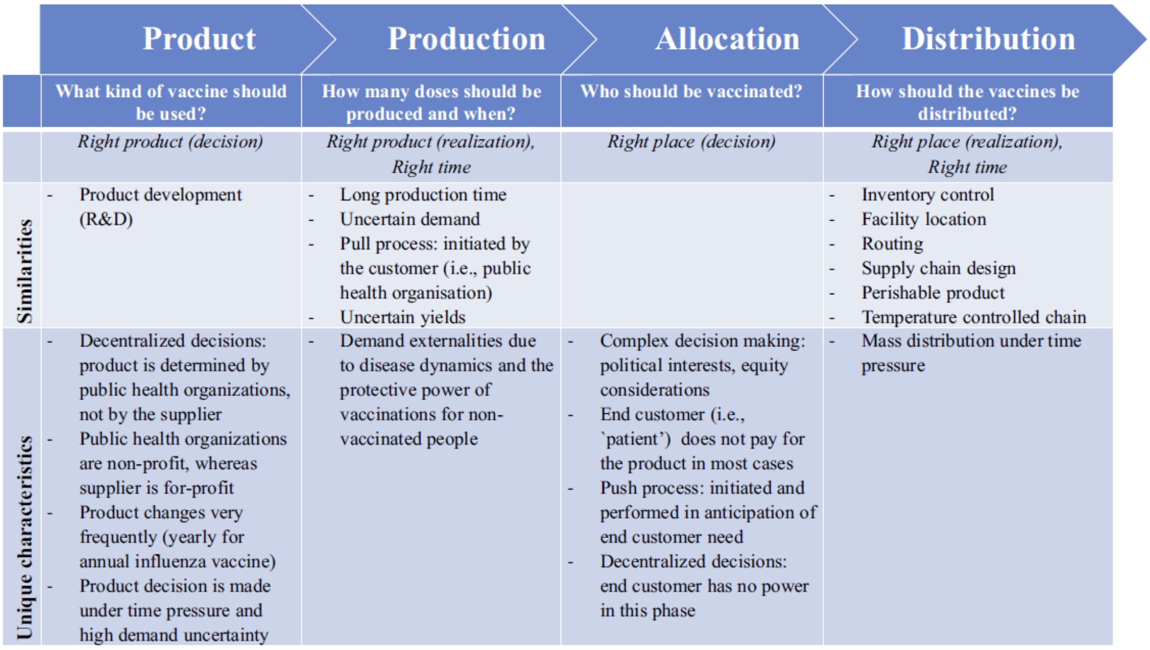
Before a person can be vaccinated, a supply chain is necessary. This supply chain is the network of a vaccine producing company and its suppliers as well as other parties involved to produce, store, distribute and administer the product to the final user [3]. In the distribution network the product is transported from the production facility via the storage facilities eventually to the end consumer. What makes the supply chain of vaccines (and other pharmaceuticals) unique is the set of stringent requirements which must be fulfilled to enable a safe administration at the point of use. E.g. the vaccines must be transported and stored under strict physical conditions.

1.1. The distribution process of vaccines

Once the vaccines have been allocated to a population in a region, the distribution of the vaccines can start. The population receives the vaccine at a point of dispensing (POD). To ensure safety and efficiency during the vaccination procedure the POD's should be replenished timely. In general the replenishment occurs from the national stockpile through various lower level distribution centres to the POD. So the distribution of vaccines consist of storage on different levels and transportation. The result is a multilevel distribution strategy of which the details depend on the size of the vaccinating region, the amount of vaccines to be distributed and available transportation modes. Basically the question which should be answered in the distribution phase is:

How can the allocated vaccine reach the population in the most efficient way under the stringent conditions?

An important aspect of using multiple distribution centres is the uncertainty from the supply side. The effect of this uncertainty can be damped since the distribution centres can be used as a buffer. Another vulnerable aspect of the distribution process is the conservation of the vaccines (by using refrigerators and other cooling instruments). Not complying to the requirements will result in wasted vaccines. Besides dealing with the packaging of the vaccines, the distribution centres and transport vehicles should fulfil the necessary requirements. The following Figure 1.1 summarizes the differences with general supply chains and the unique characteristics of the vaccine supply chain:



	Product	Production	Allocation	Distribution
	What kind of vaccine should be used?	How many doses should be produced and when?	Who should be vaccinated?	How should the vaccines be distributed?
	Right product (decision)	*Right product (realization), Right time*	*Right place (decision)*	*Right place (realization), Right time*
Similarities	- Product development (R&D)	- Long production time - Uncertain demand - Pull process: initiated by the customer (i.e., public health organisation) - Uncertain yields		- Inventory control - Facility location - Routing - Supply chain design - Perishable product - Temperature controlled chain
Unique characteristics	- Decentralized decisions: product is determined by public health organizations, not by the supplier - Public health organizations are non-profit, whereas supplier is for-profit - Product changes very frequently (yearly for annual influenza vaccine) - Product decision is made under time pressure and high demand uncertainty	- Demand externalities due to disease dynamics and the protective power of vaccinations for non-vaccinated people	- Complex decision making: political interests, equity considerations - End customer (i.e., `patient') does not pay for the product in most cases - Push process: initiated and performed in anticipation of end customer need - Decentralized decisions: end customer has no power in this phase	- Mass distribution under time pressure

Figure 1-1: Similarities and unique characteristics of the vaccine supply chain [4]

Also in the vaccine supply chain choices for each component resonate through to the other components. For example, the size of the vial should be chosen carefully as it has influence on the amount of vaccine to be allocated, the inventory size and the transport. This requires an overall coordination to take these interdependencies into account. In this process various decision makers are involved. These decisionmakers can have different interests which may affect the efficiency of the supply chain. This makes the vaccine supply chain a unique process where uncertainty, time pressure and stringent requirements interplay.

1.2. Scope and research question

The scope of this project is the distribution of a multi-dose vaccine similar to the BioNtech/Pfizer [5] vaccine by means of air transport. In specific the role of KLM as link between the production facility and the national warehouses is considered. The distribution process consists of the supply from the production facility, the storage facilities at the airport, the air transport and the storage at the national warehouses located at the airports. The process is driven by the demands on national level. Given the strict requirements in combination with the possible variations in supply, for this distribution process a robust approach is required that protects against this supply uncertainty. To solve this, the research question to be answered is:

"What is the influence of supply uncertainty on the distribution process of multi-dose vaccines?"

Detailing of this research question:

- What is the structure of the distribution process?
- What model underlies the distribution process?
- What is an efficient solution method to solve the model?

- What is a robust scenario that protects against the supply uncertainty?

For this the challenge is to adapt or implement a solution method which is able to approximate good quality solutions in little computation time as compared to a commercial solver. For this development it will be necessary to construct measuring tools for analysis of details of the applied methods, to create and adapt the method.

1.3. The structure of this report

After introducing the vaccine supply chain and its challenges, in Chapter 2 a concise literature survey will be presented. In Chapter 3 the distribution process and the details will be described and subsequently modelled. In Chapter 4 solution methods will be presented. It will start with a brief introduction on the commercial solver after which two alternative solution methods are presented. A significant part of the thesis pays attention to the analysis and development of the genetic algorithm. Followed by introducing an alternative, the rolling horizon method. This method proves to be computational better fit for the robust analysis. The chapter ends with the robust optimization approach. The results of the application of the solution methods and the robust approach are presented in Chapter 5 and will be discussed in Chapter 6 together with directions for future work. The main findings and conclusions on the distribution process that is protected against the supply uncertainty will be summarized in Chapter 7.

2. Literature review

In this chapter an overview of literature concerning the vaccine supply chain, multi-period inventory management, robust optimization and solution methods is presented.

A complete overview of the vaccine supply chains is presented in the literature review of Duijzer et al [4]. A comparison is made between supply chains in general and the vaccine supply chain in specific. Unique to the vaccine supply chain is the aspect of mass distribution under time pressure and accounting for perishability.

Optimization of the distribution of vaccines often focuses on the location of the distribution centres in different countries. The location of the distribution centres and points of use are based on routing decisions. An example of optimizing this part of the supply chain for vaccines is given by Georgiadis et al. [6]. Using a decomposition method based on distances of the distribution centres and points of use a mixed integer linear programming model is solved for a multi-echelon vaccine supply chain with a short planning horizon. A similar approach is followed by Tavana et al. [7] for equitable distribution of COVID-19 vaccines in developing countries, in specific in India, considering different refrigeration methods. Both use short planning horizons of respectively 14 and 8 time periods, treat single dose vaccines and do not take uncertainty explicitly into account.

For proper optimization of the distribution process multi-period inventory decisions should be taken into account to prevent myopic outcomes [8]. This field gets plenty of attention in the literature [9,10,11,12,13,14].

Soysal et al. [9] and Chu et al. [10] minimize the inventory cost for the supply chain of arbitrary products taking into account the inventory cost of all considered nodes applying a vendor managed inventory characterized by the knowledge of the supplier on the inventory levels at the demand nodes. For the distribution process the model should be able to determine whether products are waste and should be disposed. Soysal does take perishability into account but for perishable inventories different, more insightful formulations have been proposed. Alipour et al. [11], Gunpinar et al. [12] and Rohmer et al. [13] track the remaining shelf life or the product age as an extra index to the inventory variable.

Many described inventory routing problems [9,12,13,14] deal with multiple delivery points in a single route. The distribution network we have to look at uses the hub-and-spoke network of directed flights. This is a type of network which has a single hub as centre point and in principle does not consider transport between the endpoints described by e.g. Hsu et al [15]. Different measures for the transport cost can be found. Woo [16] groups the transport cost per unit, Rohmer et al. [13] computes a fixed charge for a routing decision and Rinaldi et al. [17] uses a combination of the aforementioned.

Decisions based on deterministic optimization models can be very sensitive to the input parameters and thus result in wrong outcomes, see Ben-tal and Nemirovski [18]. To provide a better ground for decision-making and to reduce the probability of a severe outcome, uncertainty should be taken into account. The two main approaches used for optimization of uncertain models are robust optimization and stochastic optimization. They differ in the way the probability distribution of the uncertain data is treated. Stochastic optimization has two major drawbacks [18]. The first drawback is the need for knowledge of the exact probability distribution. The second drawback is that by taking the details from the probability distribution into account the solution can become computational intractable [18]. Due to the nature of the problem considered in this thesis, where the probability distributions of the considered parameters are assumed to be unknown, from the two main approaches, the robust optimization method will be applied.

The schemes of distribution should be protected against uncertainties. Robust optimization is a method developed to serve this purpose. Soyster [19] started exploring the field of robust optimization based on the worst-case scenarios. Ben-tal and Nemirovski [18], Ben-tal et al. [20] and Soyster [19] himself point out that this method produces ultra-conservative solutions. Continuing the work of Soyster two less conservative main approaches are developed by the groups of respectively Ben-tal [18] and Bertsimas [21]. Ben-tal et al. use an ellipsoidal uncertainty set. This results in a non-linear set of equations. While for the method developed and Bertsimas et al., using a polyhedral uncertainty set, the set of equations remain linear. Bertsimas and Thiele [20] extend this robust optimization to a broad variety of networks to protect the distribution process against demand uncertainty. From the work of [22] Chu et al. [10] extend the formulation to include the supply uncertainty.

Multi-period optimization problems take the entire planning horizon into account. This can lead to significant computation time. Producing reliable results in a reasonable computation time is the focus of the development of solution methods. For small to moderate size models often commercial solvers are used to provide the optimal solutions. The issue of large computation times in specific holds for problems which are formulated as mixed integer program (MIP) models. Optimization problems formulated as a MIP belong to the class of NP-hard problems [13]. There does not exists an algorithm which can be used to solve MIP models in polynomial time. To approximate the optimal solution, heuristic methods are applied. A widely applied heuristic method is the genetic algorithm.

The work on genetic algorithms was initiated by Holland [23]. He formulated the (intermediate) solution to a problem in terms of genes that are assembled in a so called chromosome. The main requirement is that the chromosome describes a unique solution [24]. Based on the evolution principle of Darwin, 'survival of the fittest', the population of chromosomes evolves with the aim to find improved solutions with advancing generations. The chromosomes are changed by recombination, mutation and selection. In early stages of the development of the genetic algorithm these chromosomes consisted of binary valued genes. Later also the use of numerical valued genes was implemented(real coded genetic algorithm) to reduce the computation time by eliminating the translation of binary values to the real valued solution [25]. In this thesis the real-coded genetic algorithm will be used.

Recombination strategies which are applied in real coded genetic algorithms often use quantitative relations between two genes to determine the values for the genes of the offspring. An example of this is the linear interpolation between to genes ("whole arithmetic cross-over") [24].

In the cases where the gene values cannot be quantitative combined, (e.g. in sequential decision making, found in the inventory routing where the sequence of numbers determines the route [24]), parts of the chromosomes are exchanged by applying single or multiple cut points thus creating offspring. Additional to the recombination strategy, changes in the chromosome can be introduced by mutation: Changing the value of the gene by a random process.

The resulting offspring will be evaluated to assign a 'fitness' value. The fitness value quantifies the quality of the solution presented by one chromosome compared to the rest of the population. After this a selection strategy is applied based on the fitness values of the chromosomes [26].

An often mentioned serious disadvantage of the genetic algorithm is the possibility of 'premature convergence' [27]. In this case the algorithm is trapped in a local optimum. Often this is caused by chromosome which are too much alike [27]. This is an aspect of the evolution process which should be adapted.

The genetic algorithm is applied to a wide variety of problems. For the subset of inventory and distributions problems the details of the application of the method may differ but the generic essence remains the same.

In [26] a genetic algorithm is applied on a multi-plant capacitated lot sizing problem. The production amount of multiple production plants for multiple products and time periods are represented as a chromosome and compared with different (meta)-heuristics. It is concluded that the GA performs better, in particular on convergence.

Tamer et al. [29] discuss an inventory routing problem which is solved with a hybridized genetic algorithm. The chromosome representation consists of the delivery amounts to the destinations in multiple periods. The routing part is solved by a different heuristic method. Feasibility of the model is ensured by using repairing mechanisms for non-feasible chromosomes. For this strategy the gap to the optimal solution is small for smaller problems and increases with model size to 20%.

Dolgui et al. [30] present a non-revisiting genetic algorithm where the solutions are placed in a search tree. If a chromosome is already placed in the search tree, the chromosome is mutated to a place in the search space which has not yet been discovered. No specific chromosome representation is presented on the considered variables. Instead of considering the real values of the genes, the values a gene can take lie in a discretized interval between zero and one, which coincide with the lower and upper bound values for the real solution of the problem. The level of discretization can be adjusted to the computing power available, where high level discretization leads to more refined solutions.

Another versatile method as alternative to the genetic algorithm is the rolling horizon heuristic. The rolling horizon heuristic decomposes the time period into smaller sub-problems that are solved subsequently. The solution of the solved sub-problem is the input for the following sub-problem.

Glomb et al [31] use variations of the rolling horizon time decomposition to solve a lot sizing problem. The basic algorithm consists of a fixed number of time periods which are solved subsequently to optimality. This approach does not provide good quality solutions given that the optimality for a sub-problem does not necessarily belong to the solution of the global optimum. To overcome this issue, an algorithm is used in which the sub-problems are allowed to overlap. The overlap allows to re-optimize the overlapping part of the previous sub-problem proving to result in a better quality solution. The influence of the amount of overlap on the quality of the solution is studied by Hartleb and Schmidt [32] for a vehicle scheduling problem. To approximate the global optimum for a large scale model a large overlap compared to the length of the solved sub-problem is required.

In [6] the MILP is solved using a decomposition method where hubs and points of use are grouped based on distance. With a rolling horizon method disturbances in the supply chain are taken into account by reacting accordingly to the changes without using the formal principles of robust optimization.

2.1.1. Literature table

Table 2-1: An overview of studies from literature related to the topic at hand. Indicated are the elements covered by the respective papers as well as the elements that are covered in this work.

Literature	Ref.	Location allocation	Distribution	Vaccines	Perishability	Genetic Algorithm	Roling Horizon	Double dose	Refrigeration	Uncertainty
This work			x	x	x	x	x	x	x	x
Georgiadis	[6]	x	x	x			x			
Tavana etal	[7]	x	x	x					x	
Soysal	[9]		x							x
Chu	[10]		x							x
Gunpinar	[12]		x		x					x
Bertsimas	[21]		X							x
Mohammadi	[26]		x			x				x
Tamer	[29]		x			x				
Dolgui	[30]		x		x	x				

2.1.2. Scientific gap and contribution

In this work the vaccine supply chain is modelled, taking into account the perishability of the double dose vaccines including strict requirements on the delivery of the second dose. In addition to different works on the distribution of vaccines that consider multiple storage facilities, in this work the operational cost for the different storage facilities are taken into account. Furthermore here the uncertainty in the supply chain is taken into account using a robust optimization strategy. In preparation for the robust optimization strategy, the mutation process of the genetic algorithm has to be improved to obtain higher convergence rates and avoid trapping in local optima. For this a toolbox that enables analysis of the convergence rate and diversity is required. To have an alternative to the genetic algorithm, the rolling horizon heuristic has to be implemented.

3. Modelling the distribution process of vaccines

This chapter provides the description and assumptions of the considered distribution process and the modelling as a mixed integer programming model.

3.1. Description of the distribution process and assumptions

The model describes the role of KLM in the distribution of vaccines. KLM is the link in the vaccine supply chain between the production facility and the national warehouses of several countries. At the production facility the vaccines are filled in vials that contain multiple vaccination doses. Multiple vials are packaged in designated boxes, so called vaccine units (VU). The boxes are designed such that the vaccines are protected against external influences (such as light and heat) and prevent the vials from breaking. Once the vaccines have been produced and packaged accordingly the vaccine units will first be stored in a warehouse at the production facility. KLM can demand these vaccine units when KLM has the available resources to distribute or to store in KLM's own facility. This procedure is referred to as "the ordering of vaccines". In the next step the vaccine units will be distributed by KLM to several locations. KLM has a fleet of several planes to transport the vaccines to the required destination and offers two types of storage facilities. The locations are assumed to be national warehouses located at airports of different countries. The generic flow of vaccines is depicted in the following Figures 3-1 and 3-2:

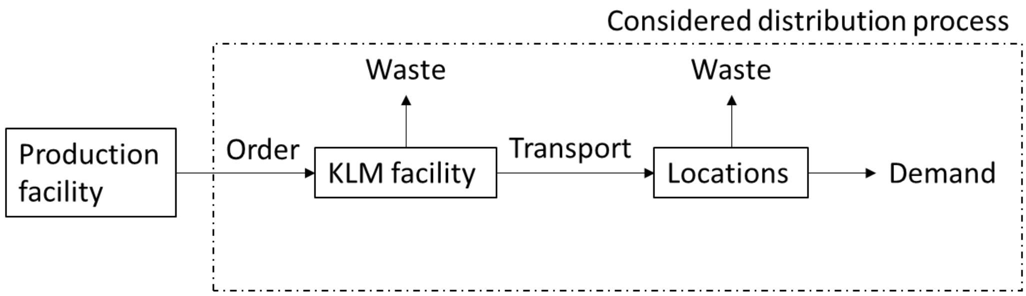


Figure 3-1: Generic flow diagram of the vaccines in the considered distribution process

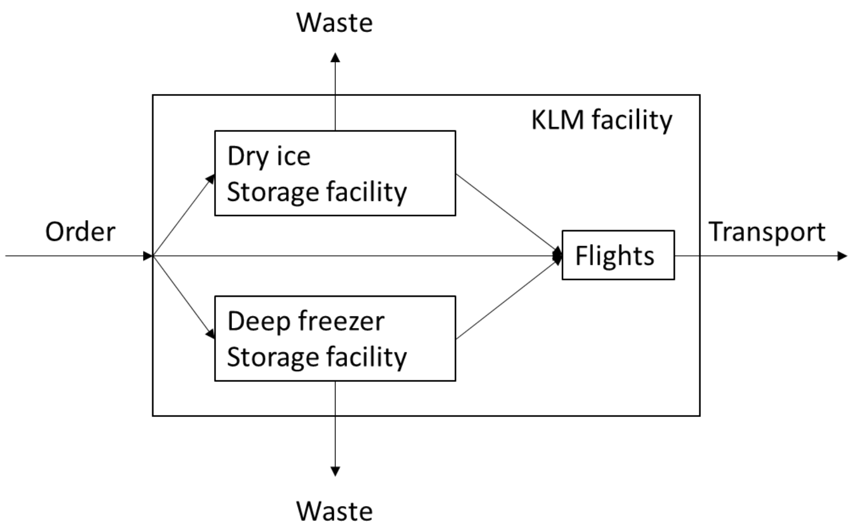


Figure 3-2: Generic flow of vaccines inside the KLM facility

The arrows show the flow of the vaccines. The vaccines enter the KLM facility by placing an order at the production facility. The vaccines leave the KLM facility by transport or as waste. The transport couples the locations to the KLM facility. The vaccines at the locations leave the destination by fulfilling the demand or again as waste.

The flow of vaccines inside the KLM facility is shown in Figure 3.2 The figure shows the incoming order amount that can be stored in the two type of storage facilities or directly leave the KLM facility as transport by means of flights.

The transport can consist of a partition from the dry ice storage facility and a one from the deep freezer storage facility. The transport to the location is conducted by a scheduled passenger plane with a fixed capacity. The amount a location can receive in one period depends on the flight frequency.

3.1.1. Constraints on the vaccine use and its storage

The type of vaccine in this model is based on the characteristics of the Pfizer\BioNtech vaccine. The vaccine is a two dose vaccine where the prescribed time between the first and the second dose is three weeks. This implies that it is required that the quantity of vaccines used to fulfil the demand of the first dose in a certain period at a certain destination also needs to available three weeks later at that same destination to fulfil the demand for the second dose.

To ensure the efficacy of the vaccine, the vaccine should be maintained at a temperature of – 70 degrees Celsius. During transport this is accomplished by using dry ice. The temperature of the vaccines in the storage facilities of KLM can be maintained by either dry ice combined with a passive cool box (PCB)   in one type of storage, or using deep freezers in the other facility where no dry ice is required. The vaccine units in the deep freezer facility are stored per vaccine unit. For the storage facility regarding dry ice, a number of vaccine units are stored in a passive cool box which contains dry ice. The passive cool box has a certain capacity for vaccine units which must be utilized. Both type of methods to maintain the temperature of the vaccine have different limitations. The vaccine can remain in the deep freezer for a period of 6 months. The dry ice of the passive cool box must be replaced every ten days, with a maximum of three dry ice (re)placements, so resulting in a maximum storage time here is 30 days. Once the vaccines have been stored in the dry ice storage facility, the vaccines cannot be stored again in a deep freezer. For modelling convenience and a conservative approach the replacement of dry ice coincides with a period, this implies that the shelf life for vaccines stored with dry ice is three periods. The shelf life for vaccine stored in the deep freezers is 6 months, which translates to 26 periods.

3.1.2. Goal of the model: Cost optimization

The goal of the model is to minimize the cost involved with the distribution of vaccine units for KLM. The model is a multi-period inventory control problem where the planning horizon is a year. The planning horizon is divided in a number of weeks. Where one period is equal to one week.

3.1.3. Cost elements

The ordering cost are the cost for placing an order, preparing for receiving and other general ordering cost. These costs are independent of the order size.
The inventory holding cost take the inventory levels of both storage facilities at KLM into account where the inventory cost are per unit stored. The deep freezer facility require an investment, this investment are included for the respective storage facility.
The shortage cost will be imposed when the demand for the first dose cannot be satisfied, where the costs are based on the amount of vaccine units which cannot be delivered to the specified location.
The wastage cost considered in this model is based on the vaccines which exceed the shelf life.
The transport cost to a destination are fixed per flight , independent of the amount of vaccines transported, and only dependent on the distance.
The handling cost are the cost used for re-icing the PCB which remain in inventory and icing the PCB with vaccine units from the deep freezers used for transport.

Model assumptions:

- The demand for the first dose in each period is known and is given in vaccine units.
- Products are depleted on a FIFO basis from the respective inventories
- The amounts of vaccines are the number vaccine units containing a fixed amount of vaccines.
- Ordered vaccines have no lead time and arrive fresh at the KLM facility.
- The vaccines have a fixed shelf life
- No shortages occur at the KLM facility.
- The dry ice in the passive cool boxes needs to be replaced at the end of each period.
- The dry ice can be replaced three times, thus the maximum stay of a box is four periods.
- Vaccine units from the deep freezer used for transport are equipped with dry ice in a single box.
- Shortages at the destination can only occur for the first dose.
- Shortages must be fulfilled in the following period
- Shortages can only occur if there is no inventory at hand.
- All planes have a fixed capacity in terms of vaccine units.
- The transportation time is negligible compared to the period length.
- The total demand for all destinations must be fulfilled in the specified planning horizon

The model will determine:

- The total cost for the distribution process
- The amount of vaccines which need to be ordered to fulfil the demand in each period
- The amount of transport in each period to each destination

And:

- The inventory level of both storage facilities at KLM and at different locations in each period.
- The amount of vaccine units transported each period
- The amount of shortage for the first dose at each destination location in each period
- The amount of demand for the second dose for each destination in each period
- The amount of dry ice required each period
- The number of flights in each period to each destination

3.2. Nomenclature of the MIP formulation

The following notations, decision variables and input parameters are used to formulate the model as a MIP.

Set notations:

$l \in L$	l is a location belonging to the set of all locations L. $L = \{0,1, \dots, L\}$
$p_d \in \mathcal{P}d$	p_d is the product age regarding dry ice, $\mathcal{P}d$ is the set of product ages.
$p_f \in \mathcal{P}f$	p_f is the product age in the deep freezers, $\mathcal{P}f$ is the set of product ages.
$\mathcal{P}_d$	$\{0,1,..,p_{d,max}\}$
$\mathcal{P}_f$	$\{0,1,..,p_{f,max}\}$
$t \in H$	t is a time period of the planning horizon set H. $H = \{1, \dots, H\}$
$\mathbb{Z}^+$	Set of all positive real integers including 0

Decision variables:

$A_{di,t}$	PCB's requiring dry ice in period t
$B_{di,t}$	Vaccine units requiring dry ice in period t
$d_{t,pd}^{l,1}$	Demand for first dose fulfilled with product age p_d at destination l in period t
$d_{t,pd}^{l,2}$	Demand for second dose fulfilled with product age p_d at destination l in period t
$D_t^{l,2}$	Demand for second dose at destination l in period t
$f_{t,l}$	Number of flights in period t to destination l
I_{t,p_d}^A	Inventory level of storage facility A in period t with product age p_d
I_{t,p_f}^B	Inventory level of storage facility B in period t with product age p_f
$I_{t,p_d}^{dest,l}$	Inventory level at destination l in period t with product age pd
S_t^l	Shortage level at destination l at time period t
T_t^l	Amount of transported Vaccine units in period t to destination l
U_t	Total order amount in period t
u_t^A	Amount of ordered PCB destined for storage A in period t
u_t^B	Amount of ordered vaccine units (VU) destined for storage B in period t
v_t	Binary order value for each period t {1,0} which is 1 if ordered, else 0.
w_{t,l,p_d}^A	Transported PCB from storage A in period t with product age p_d to dest. l
w_{t,l,p_f}^B	Transported Vaccine units from storage B in period t with product age p_f to dest. l l
$w_{t,l,p_d}^{B,pd}$	$w_{t,l,pf}^B$ with product age p_d
$Waste_t^A$	Wastage of storage facility A at time period t
$Waste_t^B$	Wastage of storage facility B at time period t
$Waste_{t,l}^L$	Wastage of for destination l at time period t
$y_{t,l}$	Number of flights in period t to destination l

Input parameters:

$A_{di,t}$	PCB's requiring dry ice in period t
C_A	Capacity storage A
C_{cb}	Capacity passive cool box
C_{df}	Capacity deep freezer
$C_{L,l}$	Capacity storage facility at location
C_{pf}	Capacity passenger flight
cdi	Dry ice cost

d_l	Distance to destination l
$D_t^{l,1}$	Demand for first dose at destination l in period t
$DI_{t,available}$	Amount of available dry ice
$DI_{t,required}$	Amount of dry ice used in period t
DIF_A	Dry ice factor for PCB in storage A
e	Icing cost
$f_{flight,l}$	Flight frequency to location l
h^A	Holding cost storage facility A
h^{df}	Holding cost storage facility B
K	Order cost
M	Very large number for the Big-M method
$p_{f,max}$	Maximum shelf life in deep freezer
$p_{d,max}$	Maximum shelf life regarding dry ice
Q_t^{ext}	Maximum available supply
r_t	Period between two vaccine doses
re	Re-icing cost
tc	Transport cost
ω	Wastage cost
θ	Number of deep freezers
φ	Penalty cost for shortage

3.3. MIP formulation of the distribution problem

The model will be formulated as a mixed integer program (MIP) according to the model description. The total costs consists of the contribution of the different cost elements. The different cost elements and forthcoming boundary conditions and constraints will be described in the following.

3.3.1. Order cost

In one period one order can be placed. The order cost K are fixed for each order and independent of the amount ordered. Whether an order is placed in a certain time period is indicated by the variable v_t : a binary value that adopts the value 1 if an order is placed in period t and 0 if there is no order in the respective period. The order cost summed over all the periods are described as:

$$Order\ Cost(OC) = \sum_t v_t * K \tag{1}$$

The constraint deciding if v_t is 1 or 0 in (mixed) integer programming can be accomplished by the following equation(2) with the big M method [33].

$$U_t - M * v_t \leq 0, \quad M \gg U_t, \forall t \in H \tag{2}$$

In this equation U_t is the total amount of ordered vaccines in period t. M is a number which is much greater than U_t. If $U_t > 0$, v_t is 1 and if U_t is 0, v_t is 0. The total order amount is bounded by the amount the supplier has available: “The available supply” Q_t^{ext} . This is formulated as:

$$U_t \leq Q_t^{ext}, \forall t \in H \tag{3}$$

3.3.2. Inventory, shortage and wastage cost

Before determining the cost and taking the perishability into account the general relations between the different inventories are described.

General inventory

From the flow diagram in Figure 3-2 the inventory levels at KLM depend on the amount of vaccine units ordered from the production facility and the amount of vaccine units which leave the KLM facility by transport or due to wastage. The total amount of ordered vaccines (U_t) consists of vaccine units which are designated to be stored in storage facility 'A' with passive cool boxes (PCB's) containing dry ice and vaccine units which are designated to be stored in the deep freezer storage facility. For vaccine units stored in storage facility 'A', the passive cool box is filled to its fixed capacity (C_{cb}). The total amount of ordered vaccine units is divided over the designated storage facilities and is given by:

$$U_t = C_{cb} * u_t^A + u_t^B \tag{4}$$

Where u_t^A are the incoming PCB filled with C_{cb} vaccine units and for storage facility 'A' and u_t^B are the amount of vaccine units designated to the deep freezer storage facility in period t.

As mentioned in the model description, the transportation of vaccine units has to be performed by using PCB. For facility A, the amount of PCB leaving the facility to destination l with C_{cb} vaccine units is denoted as $w_{t,l}^A$. The vaccine units leaving facility B have to be transferred to a different PCB which fits just one vaccine unit. The number of vaccine units leaving facility B is then determined as $w_{t,l}^B$. The amount of vaccine units coming from storage facility A and B used for transport to each destination is:

$$T_t^l = C_{cb} * w_{t,l}^A + w_{t,l}^B, \forall t \in H, \forall l \in L \tag{5}$$

The total inventory in the respective storage facilities at the KLM facility is the difference between the ordered vaccine units and the transported vaccine units and wasted vaccine units. For storage facility A the inventory level at the end of period t considering PCB filled with vaccine units at the end of each period t is determined as equation (6) with:

$$I_t^{A,KLM} = I_{t-1}^{A,KLM} + u_t^A - \sum_{l \in L} w_{t,l}^A - Waste_{A,t}, \forall t \in H \tag{6}$$

For storage facility B the vaccine units are considered as these vaccine units are directly stored in deep freezers. The inventory level of storage facility B at the end of period t is defined in equation (7) with $\forall t \in T$:

$$I_t^{B,KLM} = I_{t-1}^{B,KLM} + u_t^B - \sum_{l \in L} w_{t,l}^B - Waste_{B,t}, \forall t \in H \tag{7}$$

The demand for each destination consists of the demand for the first and the second dose and possible shortages (S_t^l) on the first dose from the previous period. The total demand (D_t^l) for location l in period t is:

$$D_t^l = D_t^{l,1} + D_t^{l,2} + S_{t-1}^l \tag{8}$$

The demand for the first dose and second dose are coupled by the fulfilment of the first dose and the prescribed number of periods between the first and second dose: r_t. The coupling of the first and second dose is given by the following equation as:

$$D_t^{l,2} = D_{t-r_t}^{l,1} - S_{t-r_t}^{l}, t \leq r_t, \forall t \in H, \forall l \in L \tag{9}$$

The inventory level at the end of period t at location ($I_t^{dest,l}$) is determined by the inventory from the previous period, the transport amount, the demand and possible shortage for the first dose. This results in the following material balance for the locations as:

$$I_t^{dest,l} = I_{t-1}^{dest,l} + T_t^l - D_t^l + S_t^l - Waste_t^l, \forall t \in H, \forall l \in L \tag{10}$$

Perishability

The vaccines in the model are assumed to have a fixed shelf life. The length of the shelf life is influenced by the type of storage in which the vaccine units are kept. Each storage facility has imposes a maximum shelf life for the vaccines. The maximum shelf life is denoted by $\mathcal{P}_{d,max}$ and $\mathcal{P}_{f,max}$ for the vaccines stored in the storage facilities regarding dry ice and the deep freezer storage facility respectively. The amount of periods the vaccines remain in a specific storage facility is traced by using the product age. The product age of stored vaccine units with dry ice is denoted as p_d. The product age of the vaccine units kept in the deep freezer storage is denoted as p_f.

The administration of the product age is based on the method used in [14] where the product age and the period are denoted by different indices. Besides tracing the product age of the inventory at the KLM facility, the product age of the vaccines used for transport should be taken into account for the administration of the product ages at the destinations. The products arrive with product age 0 at the KLM facility. The product age progression of the vaccines stored in storage facility A is determined as:

$$I_{t,0}^{A} = u_t^A - \sum_{l \in L} w_{t,l,pd}^{A}, \qquad for\ p_d = 0 \tag{12}$$

$$I_{t,p_d}^{A} = I_{t-1,p_{d-1}}^{A} - \sum_{l \in L} w_{t,l,p_d}^{A}, for\ 1 < p_d < p_{d,max} \tag{13}$$

$$I_{t,p_d}^{A} = I_{t-1,p_{d-1}}^{A} - \sum_{l \in L} w_{t,l,p_d}^{A} - Waste_{A,t} = 0, for\ p_d = p_{d,max}, \tag{14}$$

$$I_t^{A,KLM} = \sum_{p_d \in \mathcal{P}_d} I_{t,pd}^{A}, \tag{15}$$

The vaccines that reach the expiration age are still included, since these will exit the model as the waste in this same period. Summing the product age specific inventories over the product ages results in the general inventory balance.

The vaccines stored in storage facility B have a considerably longer shelf life due to the constant temperature of the deep freezers. The product age progression of the vaccines stored in the deep freezers is similar to the product age progression of the vaccines in storage facility A. The product age progression of vaccines in storage facility B is defined as equations (17-19) with $\forall t \in H$

$$I_{t,p_f}^{B} = u_t^{B} - \sum_{l\in L}\left(w_{t,l,p_f}^{B}\right) \qquad , for\ p_f = 0, \tag{16}$$
$$I_{t,p_f}^{B} = I_{t-1,p_f-1}^{B} - \sum_{l\in L}(w_{t,l,p_f}^{B}) \qquad for\ 0 < p_f < p_{f,max}, \tag{17}$$

$$I_{t,p_f}^{B} = I_{t-1,p_f-1}^{B} - \sum_{l\in L}(w_{t,l,p_f}^{B}) - Waste_{B,t}, \qquad for\ p_f = p_{f,max} \tag{18}$$

$$I_t^{B,KLM} = \sum_{p_f\in\mathcal{P}_f} I_{t,p_{df}}^{B} \qquad \forall t \in H \tag{19}$$

When vaccines stored in the deep freezer are issued for transport, the vaccines need dry ice for the transport. To relate the product age administration of the values of p_{df} and p_d. The vaccine units which leave the deep freezer younger than ($p_{f,max} - \ p_{d,max}$) are assumed to be fresh in terms of product age p_d. The following relations are proposed where p_f relates to p_d with:

$$p_d = 0, \qquad for\ p_f \le p_{f,max} - p_{d,max} \tag{20}$$
$$p_d = p_{d,max} - \left(p_{f,max} - p_f\right), \qquad for\ p_{f,max} - p_{d,max} < p_f \le \ p_{f,max} \tag{21}$$

The inventory at the locations takes the product ages of the incoming vaccines from transport and the inventory from the previous period into account. The vaccines at the locations are stored in similar conditions to the inventory of storage facility A, defining the product age as p_d. Except for the FIFO assumption, the demand fulfilment is independent from the product age. Considering that the inventory is age specific and the demand is not, variables are introduced for the age specific demand fulfilment for the first and second dose as respectively $d_{t,p_d}^{l,1}$ and $d_{t,p_d}^{l,2}$. The following expressions are used to translate the general inventory to the age specific inventory.

For the demand of the first dose and second dose it holds that the amounts are fulfilled with age specific products as:

$$D_t^{l,1} - S_t^{l} = \sum_{p_{di}=0}^{p_{d,max}} d_{t,p_d}^{l,1}, \tag{22}$$
$$D_t^{l,2} = \sum_{p_{di}=0}^{p_{d,max}} d_{t,p_d}^{l,2}, \tag{23}$$

The formulation relating the age specific demand fulfilment of the first dose and the shortage allows for situations where there is inventory at hand and shortage can occur. To impose that shortages can only occur if there is no inventory, a binary decision variable is introduced as q_t and the following equations are implemented following the big M method [33]:

$$S_t^{l} - q_t * M \le 0, \tag{24}$$
$$\sum_{p_{di}=0}^{p_{di,max}} I_{t,p_d}^{l} - (1 - q_t) * M \le 0 \tag{25}$$

Using the relations for the age specific demand fulfilment in the age specific inventory, the age specific inventories at the end of each period for all locations are written as:

$$I_{t,p_d}^{l} = \ C_{cb} * w_{t,l,pd}^{A} + w_{t,l,pd}^{B} - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2}, \qquad for\ p_{di} = 0 \tag{26}$$
$$I_{t,p_d}^{l} = \ I_{t-1,p_d-1}^{l} + C_{cb} * w_{t,l,pd}^{A} + w_{t,l,pd}^{B} - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2}, \qquad for\ 0 < p_d < p_{d.max} \tag{27}$$

$$I_{t,p_d}^{l} = I_{t-1,p_d-1}^{l} + C_{cb} * w_{t,l,pd}^{A} + w_{t,l,pd}^{B} - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2} - Waste_t^l = 0, \qquad for\ p_d = p_{d.max} \tag{28}$$

$$I_t^{dest,l} = \sum_{p_{di}=0}^{p_{di,max}} I_{t,p_d}^{l}, \tag{29}$$

Capacity constraints

The storage facilities have a certain capacity which cannot be exceeded. For storage facility 'A' and the storage facilities of the destinations the capacity depend on the available space. The capacity of storage facility B depends on the amount of acquired deep freezers. The capacity constraints are given as:

$$I_t^{A,KLM} \leq C_A \tag{30}$$
$$I_t^{B,KLM} \leq \theta * C_{df} \tag{31}$$
$$I_t^{dest,l} \leq C_{L,l} \tag{32}$$

Determining the cost of inventory ,shortage and wastage

The considered cost for inventory are applied to the inventory at the KLM facility. For each period a unit holding cost is incurred depending on the type of storage facility. The unit holding cost for storage facility A are h^A and for the deep freezer storage facility are h^{df}. The total inventory cost for the planning period are than denoted as:

$$\sum_{t \in H} \left(h^A * I_t^{A,KLM} + h^{df} * I_t^{B,KLM} \right) = Inventory\ cost \tag{33}$$

The demand for the first dose which cannot be fulfilled, i.e. the shortage, at each location for each period, has a unit penalty cost φ. The total cost for shortage of all locations over the planning horizon are determined as:

$$\varphi * \sum_{t \in H} \sum_{l \in L} S_t^l = Shortage\ cost \tag{34}$$

The vaccines which have reached the maximum shelf life and are not used to fulfil demand leave the model as waste. For each vaccine unit leaving the system as waste a waste cost ω is incurred. The total waste cost are:

$$\omega * \sum_{t \in H} (C_{cb} * Waste_{A.t} + Waste_{B.t} + \sum_{l \in L} Waste_{t,l}) = Waste\ cost \tag{35}$$

3.3.3. Dry ice and handling cost

Dry ice is required for the transport and storage of vaccines. The required amount of dry ice is taken into account for the inventory in storage facility A and the amount of transported vaccine units from the deep freezer storage facility B ($w_{t,l,pd}^{B}$). Handling operations are considered to replace the dry ice for the passive cool box of storage facility A and to transfer the vaccine units from the deep freezer to a passive cool box equipped with dry ice. The respective handling operations are referred to as ‘re-icing’ and ‘icing’. The passive cool box stored in storage facility A require a different amount of dry ice compared to the passive cool box used for transport from the deep freezer storage facility. A dry ice factor (DIF_A) for the PCB from storage facility A is used, where it assumed that this PCB requires less dry ice than a PCB issued with a single vaccine unit. The number for the dry ice factor is a parameter setting. The required amount of dry ice for the inventory at storage facility A and the transport from the deep freezer storage facility are denoted as $A_{di,t}$ and $B_{di,t}$. The following expressions determine the required amount of dry ice and an imposed boundary which limits the available amount of dry ice

$$A_{di,t} = \sum_{p_d=0}^{p_{d,max}} (DIF_A * I_{t,p_d}^A) \tag{36}$$

$$B_{di,t} = \sum_{p_d=0}^{p_{d,max}} w_{t,l,pd}^B \tag{37}$$

$$DI_{t,required} = A_{di,t} + B_{di,t} \tag{38}$$

$$DI_{t,required} \leq DI_{t,available} \tag{39}$$

A unit cost c_{di} is incurred for the required amount of dry ice. The handling cost incur a unit cost of re and e for the re-icing and icing operations respectively. The cost contribution of the dry ice cost and the handling cost is determined by the following expressions:

$$\sum_{t\in H} DI_{t,required} * c_{di} = Dry\ ice\ cost \tag{40}$$

$$\sum_{t\in H} \left(I_{t,p_d}^A * re + B_{di,t} * e\right) = Handling\ cost \tag{41}$$

3.3.4. Transport cost

The transported amount of vaccines to a location couples the inventories at KLM to the inventory at the respective location. The transport amount consists of vaccines which are available at the KLM facility including the ordered amount. The amount which can be transported to a location in period t depends on the number of assigned flights to a destination in period t $(y_{t,l})$ with a flight capacity C_{pf}. The number of assigned flights is bounded by the frequency of flights to a location in a period, referred to as the flight frequency $(f_{flight,l})$. The model determines the number of assigned flights and with this determines the maximum transport capacity to the respective destination. The following equations show the transport amount and the implied boundaries:

$$T_{t,l} = \sum_{pd\in Pd} C_{cb} * w_{t,l,pd}^A + w_{t,l,pd}^B \tag{42}$$

$$0 \leq y_{t,l} \leq f_{flight,l}, \forall t \in H \text{ and } \forall l \in L \tag{43}$$

$$T_{t,l} \leq y_{t,l} * C_{pf} \tag{44}$$

The contribution of the transport cost to the total cost depends on the number of assigned flights to a location, the distance of the location d_l and the cost of transport per km of the flight flown tc. The total transport costs are defined as:

$$\sum_t \sum_l y_{t,l} * d_l * tc = Transport\ cost \tag{45}$$

3.3.5. Total cost of the distribution process

With all components of the cost function defined, the cost function is determined by adding the cost equations (1),(33), (34) ,(35),(40),(41) and (45) as:

$$\begin{aligned} \text{Total cost} \ = & \overbrace{\sum_{t\in H} v_t * K}^{Order\ cost} + \overbrace{\sum_{t\in H} \left(h^A * I_t^{A,KLM} + h^{df} * I_t^{B,KLM}\right)}^{Inventory\ cost} + \overbrace{\varphi * \sum_{t\in H} \sum_{l\in L} S_t^l}^{Shortage\ cost} \\ & + \overbrace{\omega * \sum_{t\in H} (C_{cb} * Waste_{A.t} + Waste_{B.t} + \sum_{l\in L} Waste_{t,l})}^{Wastage\ cost} + \overbrace{\sum_{t\in H} DI_{t,used} * cdi}^{Dry\ ice\ cost} \\ & + \overbrace{\sum_{t\in H} \left(I_{t,p_d}^A * re + \ B_{di,t} * e\right)}^{Handling\ cost} + \overbrace{\sum_{t\in H} \sum_{l\in L} y_{t,l} * d_l * tc}^{Transport\ cost} \end{aligned} \tag{46}$$

The set of formulated constraints, boundary conditions and cost components leads to a model which can be implemented in a solution method. In the following chapter different suitable solution methods will be discussed.

4. Solution methods for the distribution process

In this section we will discuss three solution methods to solve the distribution of the vaccines. First a commercial MIP solver will be introduced shortly, after this a genetic algorithm will be discussed intensively and subsequently a rolling horizon approach will be presented.

The vaccine distribution problem is formulated as a mixed integer problem (MIP) in the previous chapter. The solution to the model resides in a (multi-dimensional) space which is bounded by specified constraints. The space which satisfies all constraints is referred to as the 'feasible region' of the model. This means that any combination of decision variables allowed by the constraints can be a solution to the model. Compared to problems without the integrality requirement mixed integer problems are in general difficult to solve. An approach for solving mixed integer problems could be by lifting the integrality requirement and rounding the values off to the nearest integer value. Lifting the integrality of the decision variables is referred to as 'relaxation' of the MIP model. Although rounding off the solution might provide reasonable solutions, it is not certain it provides the optimal solution[13].

4.1. Exact solution approach with a commercial solver

Often used solution methods for mixed integer programming models are "branch and bound" and "cutting planes" [34]. With the technological advancement and computational power of the current computers, moderate and larger sized models are solved using software based on these methods. In this thesis a commercially available solver is used to provide solutions for the formulated mixed integer programming model. In this thesis Gurobi is used, this an advanced commercial solver for mathematical programming that is applied for solving a wide variety of operations research problems. The Gurobi solver is generally considered to provide accurate results and is able to calculate and prove the optimal solutions. The drawback of this instrument is that it is computational intensive for large systems. Nevertheless the results with this solver will be used as a reference for comparing results in further work with alternative solution methods.

It is not publicly known how the Gurobi solver exactly works. What ís public is that mixed integer programming problems are solved by using a combination of solution methods. The main solution methods are Branch and Bound, cutting planes and a heuristic method, specifically of the heuristic part no background is available since the details are classified. For details on the branch and bound and cutting planes method see references [34].

4.2. Genetic algorithm approach

Genetic algorithm is an evolutionary based search method [23]. The working principle of the genetic algorithm is inspired by Darwin's principles of the biological evolution process: 'The survival of the fittest'. The algorithm follows Darwin's principles of evolution by evolving a population of chromosomes in a sequence of generations heading for improved solutions. In the Genetic Algorithm each chromosome, describes a unique solution of the problem. The chromosomes consist of a sequence of genes that represent the values of the variables of the problem at hand. The chromosomes are changed and result offspring by so called genetic operators. The genetic operators that can induce changes in chromosomes are recombination and mutation. Finally a selection process creates the population that constitutes the next generation.

The algorithm runs as follows: The algorithm starts with creation of an initial population. Most often the initial population is generated randomly to spread the chromosomes over the search space in which the optimal solution resides [35]. After forming the initial population, parents are selected from the population to generate offspring. The parents are selected based on the 'fitness' of the respective chromosome. The next step is mutation, where individual chromosomes from the available population are selected to change a number of genes. After recombination and mutation, selection occurs. Depending on the chosen selection strategy, the individuals that are part of the next generation are selected.

The evolution process is continued until a stopping criterion is met e.g. until the imposed number of generations is achieved or a quality criterium is met. The schematic overview of the functioning of the genetic algorithm is presented in the following figure.

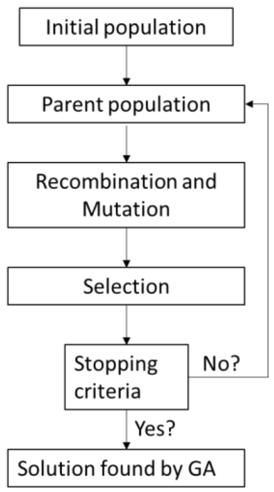


Figure 4-1: Schematic overview of the genetic algorithm

In the following section the application of the genetic algorithm to the distribution problem at hand will be presented. First the chromosome representation is presented, then the genetic operators are introduced.

4.2.1. The chromosome representation of the problem

The genes of the chromosome describe an encoded unique solution to the problem. Besides containing a unique solution, a requirement is that decoding the fitness of the chromosome should be retrieved in reasonable time [30]. Therefor in the chosen chromosome representation explicit variables are chosen as the genes

In this case the genes that constitute the chromosome are chosen as the values for the order amount (u_t) and the transport amount to each destination (T_t^l). Specifically these two type of genes are chosen since the order amount and transport amount are the elements constitute directly the solution of the problem.

The problem is multi-period thus the length of the chromosome is set to the planning horizon. This results in a chromosome with one row describing the order amount and L rows that describe the transport amount to the L destinations. The number of columns equals the dimension of the planning horizon (H). The chromosome representation is depicted in the following figure:

$t =$	0	1	2	3	$\cdots$	H
u_t						
T_t^1		gen				
$\vdots$						
T_t^L						

chromosome

Figure 4-2: Chromosome representation with the periodic order amount (u_t) and transport amount to each destination (T_t^l) over the planning horizon (H). The population consist of N of the chromosomes.

4.2.2. Decreasing the search space: Bounds on the genes

Before discussing the genetic operators, the bounds for the genes will be determined. Besides aiming for feasible solutions, the goal of the bounds is to aid the genetic algorithm in exploring the search space. Defining the bounds should be done carefully, since over-defining the bounds will interfere and might hamper with the search ability of the genetic algorithm [F. Schulte, “Personal communication”, February 2022].

An addition to the upper bounds

While complying with the constraints specified for the MIP solver does result in feasible solutions, the search space can be decreased somewhat. The optimal solution from the MIP solver never contains waste for neither the destination nor the KLM storage facility. This fits with the intuitive desire of ”not ordering more than required”. The search space can be decreased by imposing conditional upper bounds for both the order amount and the transport amount. The addition to the upper bound for the transport amount to a destination is determined as follows: the sum of the demand over a period equal to the shelf life (transporting more would result waste) corrected for the inventory at hand at that respective destination. The addition to the upper bound for the order amount follows the same line of argumentation, where it is the sum over all destinations of the previous introduced transport upper bounds minus the inventory at hand at the KLM storage facility.

It should be noted that it is not to say that there is no occurrence of wastage at the KLM facility in solutions found by the GA. Wastage at KLM can occur due to a mismatch of the order amount and the transport amounts.

In summary , from the model formulation and the previous section, the bounds for the genes are:

- **Bounds on Transport (T_t^l)**
    - Upper bound is the minimum value of:
        - Flight capacity
        - Available products at KLM including the ordered amount
        - Destination specific demand over the time span of the product age minus the age specific inventory at the destination

        - Storage capacity at the respective destination
    - Lower Bound:
        - Shortages from the previous period and the demand for the second dose which cannot be fulfilled from the inventory at the destination
- **Bounds on Order amount ($\boldsymbol{u_t}$)**
    - Upper bound is the minimum value of:
        - Available supply from supplier (Q_t^{ext})
        - Sum of demands over all destinations over the time span of the product age minus the age specific inventory at the destinations minus the inventory at KLM.
    - Lower bound:
        - Shortages from the previous period and the demand for the second dose for all destinations, which cannot be fulfilled from the inventory at KLM

4.2.3. Creating the initial population of chromosomes

The initial population is spread as much as possible over the search space. During the formation of the initial population chromosomes are randomly generated while complying with the gene specific upper and lower bounds. After a chromosome is generated, the cost (and thus the “fitness”) of the respective chromosome is determined.

For the initial population the search can be guided by specifying a minimal order amount in the case an order occurs. The idea behind the minimum order quantity is to decrease number of order occurrences.

The minimal order amount can be chosen as the ratio between the ordering cost and the inventory cost per product formulated as:

$$Min\ order\ amount = \frac{K}{h^A} \tag{47}$$

4.2.4. How chromosomes change: Genetic operators

Recombination

The idea of recombination is to exploit the search space. Recombination occurs by recombining the chromosomes of two parent chromosomes which result in offspring. In this thesis parents are selected by roulette wheel selection. Roulette wheel selection is a selection method where parents with a better fitness value have a higher chance to be selected for recombination. The probability for a chromosome to be selected as parents is:

$$P\ (Chrom_i) = \frac{fitness_{Chrom_i}}{\sum_{k \in N_{pop}} fitness_{Chrom_k}} \tag{48}$$

When the parents are selected, the so called cross-over rate determines if the parents are allowed to recombine. The cross-over rate is a constant parameter typically in the range of [0.6-1]. This parameter is compared with a random number drawn from the [0,1] interval. If this random number is less than the cross-over rate, the parents are allowed to recombine. If the parents are not allowed to recombine, the parents will take themselves the place of the offspring in the population.

The three respective recombination strategies that will be evaluated in this thesis are whole arithmetic cross-over, single point cross-over and multipoint cross-over. Each of these recombination methods is separately evaluated, to come to a best choice in the end.

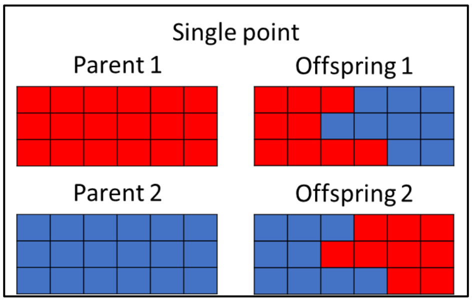


Figure 4-3: single point cross-over, for each row a cut point is chosen

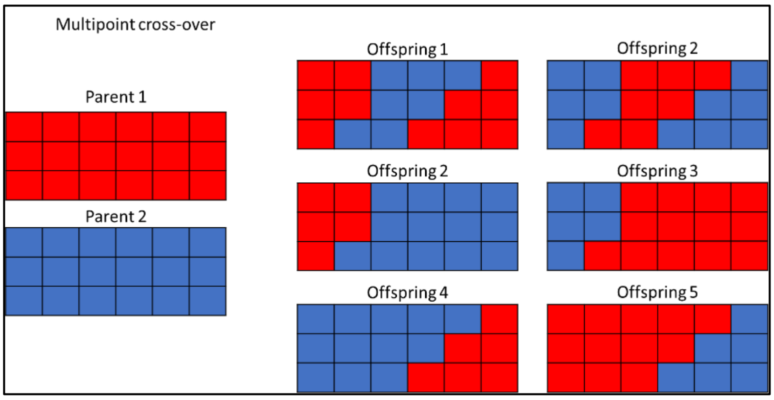


Figure 4-4: Multi point cross-over with 2 cut points per row

Each recombination strategy is applied independent to each row of the chromosome, see Figures 4-3 and 4-4. The reason to apply the recombination row-wise is to allow the model to break inter dependencies of the genes in one period. The difference in recombination strategies is as follows:

Whole arithmetic cross-over

Whole arithmetic cross-over is a recombination strategy often used in genetic algorithms which are coded using the real values of the variables. This recombination works as follows. Two blank offspring chromosomes are created and the values of the genes are interpolated between the corresponding gene values of both parents. For each set of rows the genes are combined using a number $y_{row,i}$ that is randomly drawn from the interval [0,1]. The value of $y_{row,i}$ determines the contribution of both genes parents to each offspring as follows:

$$Offspring^1_{gene,i,t} = y_{row,i} * P^1_{gene,i,t} + (1 - y_{row,i}) * P^2_{gene,i,t} \tag{49}$$

$$Offspring^2_{gene,i,t} = y_{row,i} * P^2_{gene,i,t} + (1 - y_{row,i}) * P^1_{gene,i,t} \tag{50}$$

Where $P^1_{gene,i,t}$ and $P^2_{gene,i,t}$ are the gene values from the chromosome of parent 1 and 2 respectively.

Single point and multi point cross-over

Where the arithmetic cross-over calculates new values of the genes for each chromosome, with single point and multipoint cross-over the values of the genes remain intact. Mostly: If the bounds do not permit the generated value, the value is adapted to the bounds. For single point cross-over one cut point ($c_{row,i}$) is randomly chosen from the length of the planning horizon. Single point cross-over generates two offspring. For multipoint cross-over multiple cut points are chosen. The number of offspring generated with the multipoint cross-over depends on the number of cut points as: $\left(2^{(\text{ \#cutpoints+1})} - 2\right)$ The schematic overview of the two different cross-over strategies are depicted in Figures 4-3 and 4-4.

After the recombination has provided the offspring, each generated offspring is evaluated. The evaluation provides corrects the chromosome if a certain bound is exceeded If a gene exceeds the upper bound, the value of the specific gene will be changed to the upper bound, in the case the value of a gene is less than the lower bound, the value will be set to the lower bound. Subsequently the cost for the chromosomes is evaluated.

Mutation

Where with recombination the chosen values for the genes mostly remain the same, unless bounds are exceeded, the goal for mutation is to create a change in a chromosome. Each gene has a probability P_{mut} to change its value. If a gene is subject to mutation. The new value for the respective gene is randomly drawn from the gene specific uniform interval [LB,UB]

The intermediate population: adding elitist(s)

After applying the genetic operators, the intermediate population consists of the resulting offspring. This intermediate population is now extended with a prescribed number of best performing chromosomes from the parent population. This strategy is called elitism. The elitism variable determines how many of the best performing solutions are moved to the next generation.

Selection of chromosomes for the next population

The next step is a selection process, applied to the intermediate population in order to reduce population to the prescribed dimension. This is the parent population of the next generation. The selection process consists of several selection criteria which will be described in the following.

Uniqueness

The following step in the selection process is to discard duplicate chromosomes, such that only unique chromosomes remain.

Diversity

An often mentioned disadvantage of the GA is its possible premature convergence to a local optimum This can partly be overcome by not only selecting the best found solution thus far, but also allow for chromosomes which perform less to be selected. This way a certain level of diversity is introduced. The total cost relate to the values of the genes in the chromosome. The values of the genes differ from chromosome to chromosome (phenotypic diversity) [35]. Consequently in the first implementation the differences in total cost are used as a measure for a diverse selection. (Neglecting for the moment that the chromosome structure does not uniquely relate to the total cost.). The available number of spots for “less” performing chromosomes is a parameter setting in this implementation of the genetic algorithm.

4.2.5. Measuring Performance

How does the GA perform? Heuristic methods are often applied to problems which are hard to solve for commercial solvers. The idea of applying a heuristic method is to approximate the optimal solutions. Two important quality factors that have to be considered are:

- the gap to the exact solution approach
- How fast does the GA reach the solution.

Gap to exact solution

Gaps from literature show that a reasonable gap to the optimal solution is between 0 and 20 percent [26,30].

Convergence rate

A performance criterium is how fast the optimal solution is approached per generation step.

A first pragmatic impression of the convergence rate is obtained visually from the graph of the cost versus the generation. In specific how rapid do the cost decrease, especially in the first generations, and when does the algorithm reach a plateau.
More quantitative measures can be useful. An intuitive measure to the visually obtained one is the ratio of the step in the cost over the previous cost:

$$Conv.rate_1 = \frac{Cost_{i-1} - Cost_i}{Cost_{i-1}} \tag{51}$$

A measure that results a constant value when the relative improvement of steps is constant would be preferred, in principle. This can be achieved by using (an approximation) of the exact solution and take the ratio of the step in the cost over the distance of the cost of the previous step to the exact solution. This relative convergence considers the optimal solution, the solution of the current generation and the solution from the previous generation. The optimal solution from the commercial solver is chosen as a reference to give a better insight in the progression to a better solution of the genetic algorithm. Whereas taking the zero line as reference for the relative convergence would result in a relative convergence rate which does not truly reflect the convergence. In formula:

$$Conv.rate_2 = \frac{Cost_{i-1} - Cost_i}{Cost_{i-1} - Cost_{optimal}} \tag{52}$$

This method will be used in this thesis. (By the way: Care should be taken that this value can explode when the value of the cost are close to the optimal solution.)

Computation time

Computing time is an important factor, especially in the case of the larger problems we typically deal with. The measure is straightforward: How much time is required to evaluate a certain number of generations? Or to achieve a solution with a specified percentage of accuracy.

4.3. Rolling horizon method

Rolling horizon is a method which solves the model by dividing the considered planning horizon into smaller time windows. The aim is to decrease the computation time by consecutively solving the reduced planning horizons for the model until the total planning horizon is reached. The time windows are consecutively solved, where the solution of the previous time window is the input for the following time window.

In this thesis three approaches of the rolling horizon method will be discussed. All three approaches share the decomposition of the total planning horizon. The differences between the approaches are on how the decomposed time windows are solved. The three approaches are: the straight forward method where the time windows are solved for the given decomposition size, the second method considers relaxed future periods in solving the decomposed problem. In this method in each iteration an additional time period is added, but for this time period the integrality constraint is relieved. The integrality for this time period is imposed in the next iteration of this method. The third method is an extended version of the second method. The difference is the addition of overlapping of the decomposed time windows. Examples of the approaches are depicted in the following figure:

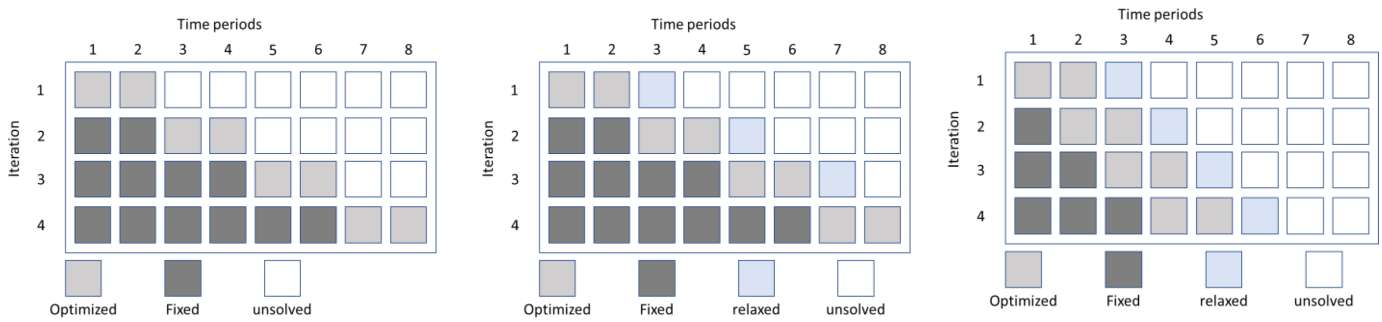


Figure 4-5: Rolling horizon approaches: from left to right: Straight forward RH approach, considering future periods in the RH approach and overlapping the consecutive solved time periods.

The variables involved for solving the rolling horizon approaches are the number of time periods which are optimized with the integrality constraints, the extension of this time window for the number of periods which are relaxed and the shift amount for overlap.

The time window which is optimized with integrality is denoted by η, the extended time window which includes the relaxed periods is denoted by μ and the shift is simply denoted by 'step'.

4.4. Robust optimization

The approach to study the influence of supply uncertainty on the distribution process is robust optimization, a technique where detailed information of the probability distribution of the uncertainty is not needed. The goal of robust optimization is to find an optimal solution which is feasible for all instances of the uncertain parameters.

4.4.1. Formalism of robust optimization:

The generic optimization model consists of a cost function subject to a set of constraints and boundary conditions. The model can be formulated as:

$$\text{Min } \mathbf{c'x}$$

$$subject\ to:$$

$$\boldsymbol{Ax} \leq \boldsymbol{b}$$

For robust optimization the uncertainty resides in the matrix **A** that consists of elements a_{ij}.

The goal of robust optimization is to find a solution which is feasible for all values of the uncertain parameters. Although the distribution of the uncertainty is unknown, with robust optimization uncertainty can be taken into account. This is accomplished by specifying an interval in which the uncertainty resides. The interval of the uncertainty is given as: $[\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$ in which the uncertain parameter a_{ij} resides. $\bar{a}_{ij}$ is used to denote the nominal value and $\hat{a}_{ij}$ is the 'half length' value.

Less conservative approaches [18] and [21] use the uncertainty interval to determine the scaled deviation of the uncertain parameter. The value of the scaled interval lies in the interval of [-1,1] where the scaled deviation is determined as:

$$z_{ij} = \frac{a_{ij} - \bar{a}_{ij}}{\hat{a}_{ij}} \tag{53}$$

By imposing a boundary condition on the scaled deviation conservative solutions can be avoided. With the imposed boundary conditions the number of equations to formulate the model increases. In this work the polyhedral uncertainty set as described by Bertsimas et al. [21] is used. The method of Bertsimas remains linear when imposing the boundary conditions for the scaled deviation as opposed to methods which use ellipsoidal uncertainty sets. The generic application of the polyhedral uncertainty set works as follows:

For each row of the product $\boldsymbol{Ax}$ that contains uncertainty it should hold that:

$$\sum_j \bar{a}_{ij} x_{ij} + \sum_j z_{ij} \hat{a}_{ij} x_{ij} \leq b_i, \forall i \tag{54}$$

The idea of robust optimization is to maximize the uncertainty term $\sum_j z_{ij} \hat{a}_{ij} x_{ij}$ under the conditions formulated by Bertsimas et al [21] in equation (55).

$$\sum_j \bar{a}_{ij} x_{ij} + \max\left(\sum_j z_{ij} \hat{a}_{ij} x_{ij}\right) \leq b_i, \sum_j z_{ij} \leq \Gamma_i, \forall i \tag{55}$$

The total allowed amount of scaled deviations of all uncertain parameters for each row i is limited by the parameter Γ_i, this Γ_i is called the 'budget of uncertainty'.

This constraint cannot be solved directly due to the maximization term. Using the principles of duality theory, proven by Bertsimas et al. [21] the maximization term can be converted into a minimization term. The resulting maximization problem summarizing equation 55 and the interval of z_{ij} leads to:

$$\max \left(\sum_j z_{ij}\hat{a}_{ij}x_{ij}\right) \leq b_i \tag{56}$$

$$subject\ to:$$

$$\sum_j z_{ij} \leq \Gamma_i, \forall i \tag{57}$$

$$|z_{ij}| \leq 1 \tag{58}$$

The dual of this auxiliary problem results in a minimization problem:

$$min \sum_j p_{ij} + q_i * \Gamma_i \tag{59}$$

$$subject\ to:$$

$$q_i + p_{ij} \geq \hat{a}_{ij}|x_i| \tag{60}$$

$$p_{ij} \geq 0 \tag{61}$$

$$q_i \geq 0 \tag{62}$$

The implementation of the dual formulation is implemented in the original uncertain equation (55) and results in the robust counter part of the model:

$$\sum_j \bar{a}_{ij}x_{ij} + q_i * \Gamma_i + \sum_j p_{ij} \ \leq b_i, \forall i \tag{63}$$

$$q_i + p_{ij} \geq \hat{a}_{ij}y_i \tag{64}$$

$$-y \leq x \leq y \tag{65}$$

$$q_i \geq 0, p_{ij} \geq 0, y \geq 0 \tag{66}$$

The $q_i * \Gamma_i + \sum_j p_{ij}$ -term protects the solution against variation of the uncertain parameter a_{ij} in a defined interval.

4.4.2. Application of robust optimization:

The available production capacity at the supplier, and thus the available supply is allowed to fluctuate. This uncertainty will have an effect on the total costs. Limits on the available supply are the following, (i) the production capacity per period can be equal or larger than the transport capacity per period and (ii) the minimum amount required for a feasible model. The implementation of robust optimization starts with the constraints in which the uncertainty takes place. From the model formulation the constraint which requires a robust counterpart is equation (Chapter 3) that with neglecting the deep freezer results in:

$$u_t \leq Q_t^{ext} \tag{67}$$

As Q_t^{ext} is uncertain, Q_t^{ext} can be written as:

$$Q_t^{ext} = \bar{Q}_t^{ext} + z_t * \hat{Q}_t^{ext} \tag{68}$$

Where $\bar{Q}_t^{ext}$ is the nominal value of Q_t^{ext} and $\hat{Q}_t^{ext}$ is the half length of the interval in which Q_t^{ext} resides.

From initial experiments it can be seen that using a production capacity higher than the total transport capacity per period has no influence on the total costs. Hence, the interval in which the uncertainty resides is one sided [10] and defined as: $[\bar{Q}_t^{ext} - \hat{Q}_t^{ext}, \bar{Q}_t^{ext}]$.

From the theory of Bertsimas et al [21] the constraints regarding the order amount (u_t) can be written as:

$$u_t - \bar{Q}_t^{ext} + q_t * \Gamma_t + \sum_j p_{tj} \leq 0 \tag{69}$$

$$q_t + p_{tj} \geq \hat{Q}_t^{ext} \tag{70}$$

$$q_t \geq 0, p_{tj} \geq 0, \tag{71}$$

5. Numerical results

In this chapter the numerical results for the exact solution approach, the two developed solution methods; the genetic algorithm and the rolling horizon method. The two developed solution methods should be qualitative to be able to perform the robust optimization for the supply uncertainty with which this chapter will finish.

5.1. Results with exact solution approach

The results of the exact solution approach will be presented with a sensitivity analysis after which the model will be simplified based on the findings from the sensitivity analysis. The reason to opt for an alternative solution method will be made clear in the Section 5.1.3. which concerns the computation time.

5.1.1. Sensitivity analysis

How does the model behave with the variation of the order costs for a moderate size model? For this purpose the number of destinations is set to 4 and the planning horizon is set to 23 periods. Here the interest is in the behaviour of the model when minimizing the total cost with different parameter settings. Choices of the model are the number of order occurrences, the number of flights to each destination, amount of inventory stored and the amount of shortages.

The influence of the order cost (K) with varying available supply on respectively the total costs, the number of order occurrences and the inventory costs is presented in the graphs shown in Figure 5-1 and 5-2:

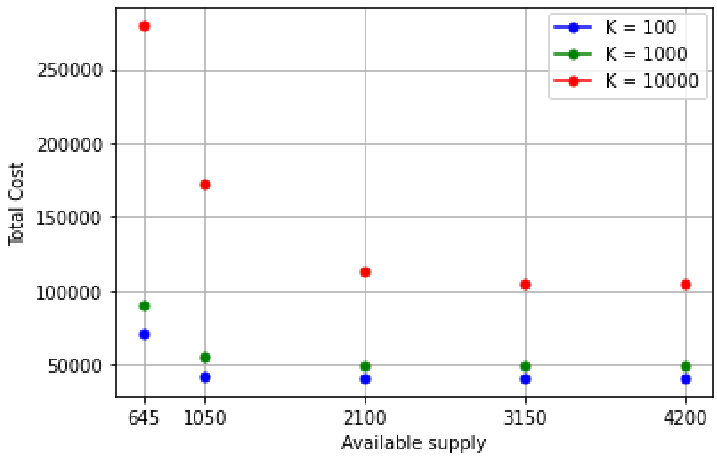


Figure 5-1: Influence of the order cost K on the total costs in dependence of the available supply

In Figure 5-1 the influence of the order cost on the total costs is shown in dependence of the available supply. For varying the order with a decade has only a slight influence on the total costs. The total costs plateau off with increasing available supply. For varying with a decade in this plateau regime   typically the total costs increase with a factor of 2. The increase of total costs at low available supply is more dependent on the order costs. For the order costs of K = 1000, used largely in the remainder of this work, the total costs are insensitive to the available supply. This is different for the number of order occurrences in dependence of the available supply, see Figure 5-2, where the number of required order occurrences is plotted versus the available supply for different order costs. The number of order occurrences shows to be much more dependent on the available supply while the plateau is shifted to higher values of available supply. The increase of order costs strongly reduces the number of orders. The number of order occurrences is always less than the length of the planning horizon. Typically in 25-50% of the time periods an order occurs.

The model has the possibility to store the vaccines using dry ice or deep freezers. In Figure 5-3 the effect on the inventory costs for both storage facilities is shown versus the available supply for varying order costs K. Remarkably only for very high levels of available supply and order costs the deep freezer storage is used. Since this situation is unlikely the deep freezer facility will be neglected for the remainder of this work. The inventory will be kept only using dry ice.

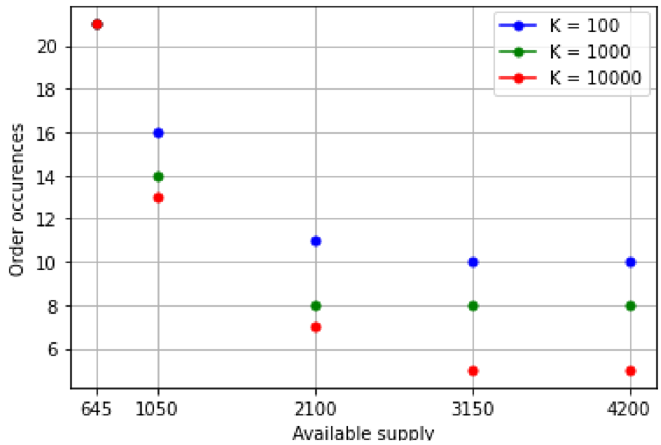


Figure 5-2: Influence of the order cost on the number of order occurrences

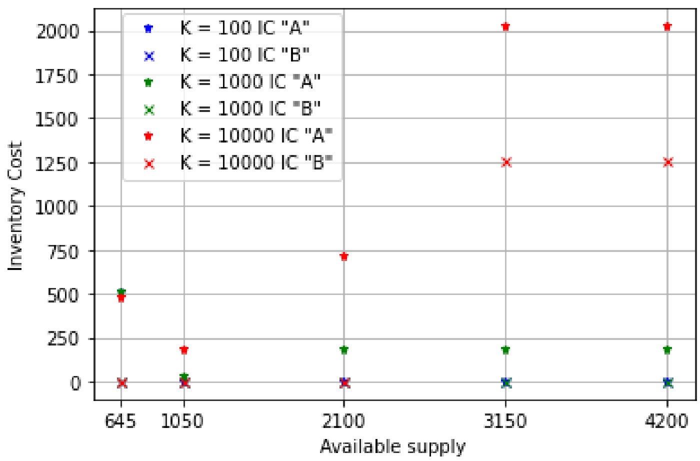


Figure 5-3: Influence of the order cost on the inventory cost for the dry ice storage facility(A) and the deep freezer facility(B)

5.1.2. Solutions used as reference

Using the commercial solver two reference cases are created that are used to develop and adapt alternative solution methods. First the case for a single destination is considered followed by four destinations. For both reference cases a constant demand rate for the first dose of 100 is used for 52 weeks. The planning horizon is extend to 56 weeks. To develop the solution methods no boundary is imposed on the amount that can be ordered from the supply side. Furthermore the maximum number of available flights per period is set to 2 for each destination.

For one destination :

The computation show that the minimum total costs for transporting to one destination are 44250 monetary units. There are 18 order occurrences and 35 flights to the destination. There are no inventory costs, which implies that periods in which an order occurs and transport occurs coincide.

For multi-destinations:

Increasing the number of destinations does not have an influence on the order occurrences and show a similar pattern, where the periods in which is ordered and transport occurs also coincide.

For all situations modelled, the optimal solution does not produces waste.

The opportunity of using the deep freezer storage facility increases the computation time due to the quantity of decision variables involved. Furthermore, the calculations show that there is no utilization of this deep freezer storage facility. Therefor in the remainder of this report the use of the deep freezer facility will be neglected.

In Figure 5-4 the updated schematic flow of the vaccines is shown:

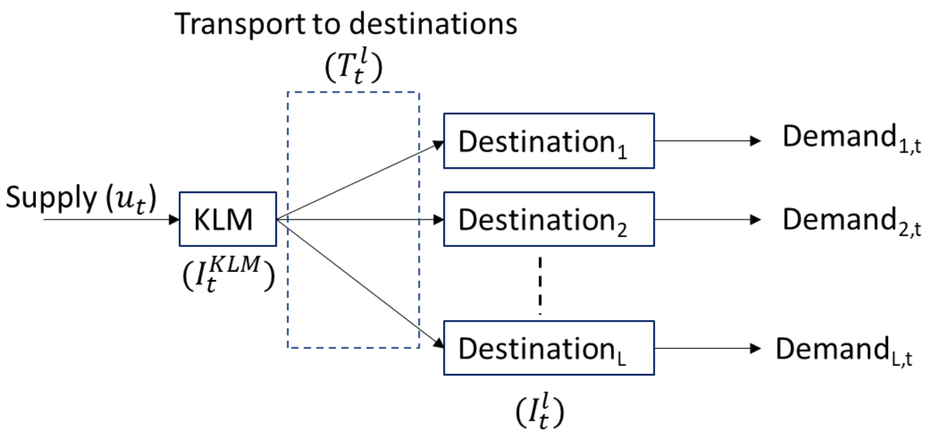


Figure 5-4: Updated schematic overview of the distribution process: Neglecting the deep freezer storage facility

Based on the updated distribution process and the implication that the handling cost and the cost for dry ice depend on the inventory at the dry ice storage facility, the cost for inventory, handling and dry ice can be combined in the holding cost and is updated to:

$$h^{A,new} = DIF * cdi + re + h^{A,old} = h^A \tag{72}$$

From this new definition for the holding cost, the equations and the forthcoming constraints regarding the dry ice cost and the handling cost can be eliminated.

5.1.3. Computation time

Computation time is a critical factor in solving problems. For several settings for the number of destination and the amount of available supply the calculation time is determined. To provide an insight in the relation between the size of the model and the computation time a constant demand rate of 100 will be considered for the first dose in each period. The size of the model is related to the number of destinations and the length of the planning horizon, affecting the number of decision variables and constraints. Two instances of available supply will be considered:

- The available supply is set to the minimum required for a feasible solution based on the assumption that the total demand must be satisfied
- The available supply is equal to the transport amount.

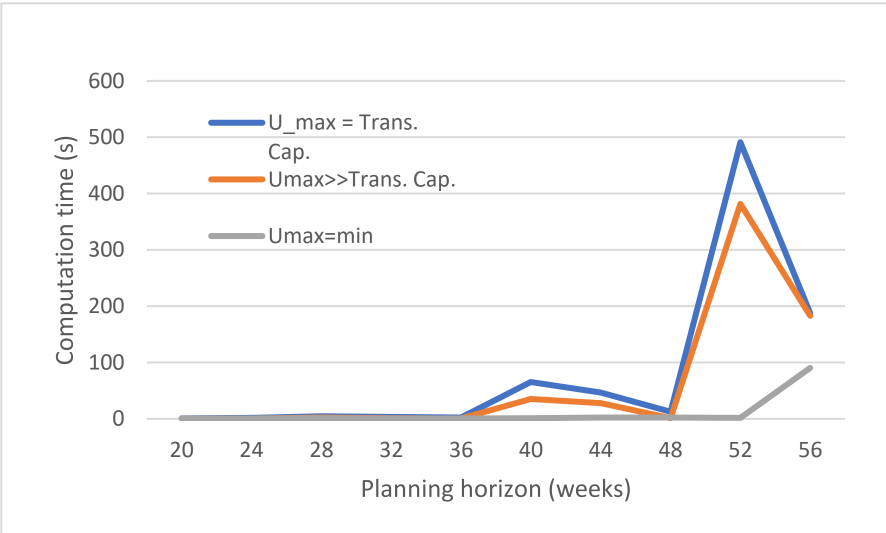


Figure 5-5: Computation time of commercial solver for single destination vs planning horizon for different amounts of supply availability

For a single destination the computation time versus the planning horizon is presented in Figure 5-5.

The figure shows that the computation time for relative small problems is low. The computation time increases slightly with the increase of time periods if the minimal amount of available supply is set to the minimum for a working model. For the instances where the supply is more than sufficient (Umax > transport capacity), the calculation time shows an increasing trend with increasing planning horizon. On the other hand the computation time for some specific planning horizons are low, they show dips, e.g. for planning 44 and 48, after which the computation time increases again. The source of this phenomenon is unclear and will not be sorted out.

The dependence of the computation time on the planning horizon for an extended number of destinations is determined. In Figure 5-6 the results are shown for the available supply equal to the transport capacity. In Figure 5-7 the results are show for the minimum amount of required available supply. A limit on the computation time of 7200 seconds is applied.

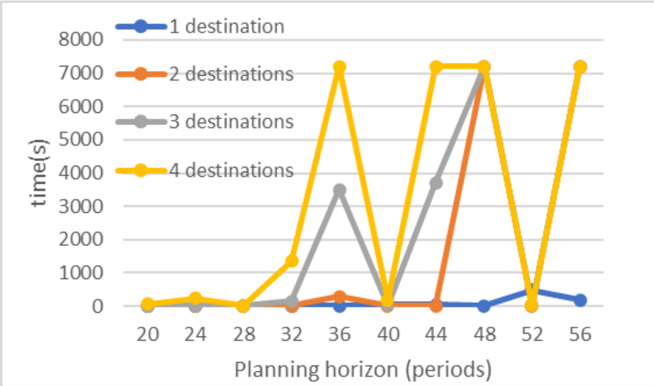


Figure 5-6: Computation time of commercial solver for multiple destinations vs planning horizon comparing with a supply availability for each period which equals the transport capacity

A curious behaviour can be seen for the instances where the available supply is equal to the transport capacity. On average the trend is that the computation increases with as well the number of destinations as with the length of the planning horizon. Just to mention, but not to elaborate on here, is a curious behaviour: dips in computation time seem to coincide irrespective of the number of destinations.

In the case the available supply is equal to the minimum amount required, a more pronounced relation between the model size and the computation time can be seen. This is the result of a more 'tight' available supply. For this case the computation time increases much more rapid with increasing the number of destinations as well as increasing the length of the planning horizon. Actually the imposed limit on the computation time of 7200 seconds is met early and disguises the actual effect of increasing the model size.

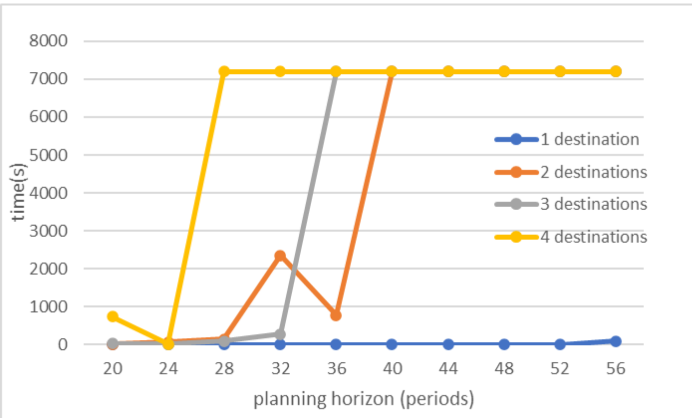


Figure 5-7: Computation time of commercial solver of multiple destinations vs planning horizon with a minimum supply availability for each period for a feasible model

Apart from the size of the model also the parameter setting, in this case the available supply, has a strong effect on the computation time. This can be explained, partly: , due to the limited available supply in almost each period an order occurs. This increases the number of decisions the model must explore. As the destinations only differ in transport cost, this suggests that the model considers all possible options for each destination.

The general trend of large increasing computation times motivates the use of alternative solution methods.

5.2. Results of genetic algorithm parameters

In this section the influence of the parameter selection and settings of the genetic algorithm will be discussed. First a basic implementation of the genetic algorithm will be presented for a small sized model. Based on the results of the basic implementation the mutation operator will be adjusted for an improved functioning of the genetic algorithm.

5.2.1. A first basic implementation and evaluation of the genetic algorithm

To build an understanding of the behaviour of the model and to guide adaptations of the algorithm, it is helpful to simplify the model and thus reduce the computational time. For this purpose the model will be simplified to the situation with a single destination with a constant demand for each period.

5.2.2. The influence of algorithm parameters and genetic operators:

For each recombination strategy experiments are conducted. The parameters which are varied are the cross-over rate, the mutation rate, the number of elitism and the number for phenotypic diversity. Most experiments are conducted varying one parameter and keeping the remaining parameters constant. Each parameter setting is tested for 10 runs. Unless stated otherwise the best performing parameter setting will be presented for each recombination strategy. The less performing results can be found in the appendix. What follows now are the results of the subsequent parameter variations.

Cross-over rate:

The cross-over rate describes the probability of the recombination of two parents . The cross-over rate is varied between 0 and 1 with increments of 0.2. The results show limited sensitivity to the exact value of the cross-over rate in the range from 0.2 – 0.8. In Figure 5-8 the best solution is presented for a cross-over rate of 0.8 for each recombination strategy (results for the other cross-over rates showing similar trends can be found in appendix B. All recombination strategies show rapid initial convergence. After the initial convergence the solution is trapped (or converges only very slowly) at a plateau level approximately 20% above the optimum solution for all strategies. There exists a difference in convergence rate: The multipoint strategy converges the fastest while the arithmetic strategy converges only very slowly.

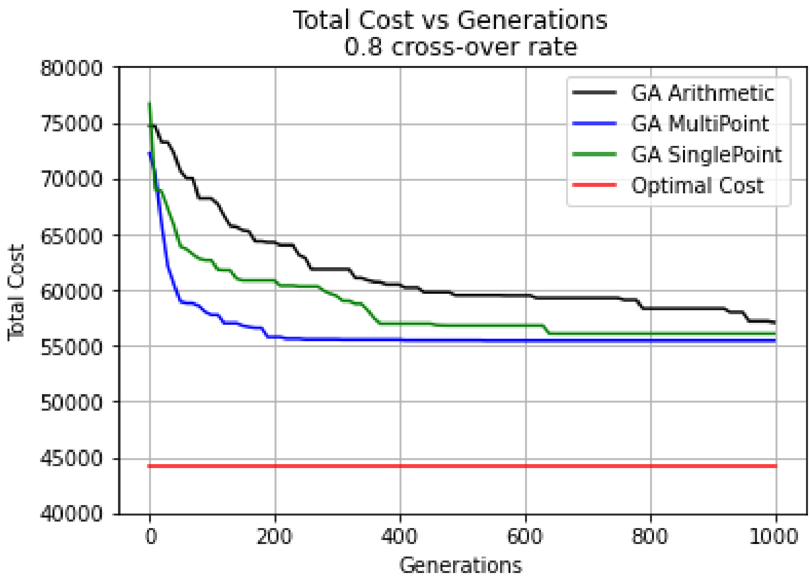


Figure 5-8: Total cost vs generations for three different cross-over strategies for a cross-over rate of 0.8, mutation rate 0.01, elitism 10, phenotypic diversity 30%

In conclusion different recombination strategies do differ in convergence rate, but all get locked at more or less the same deviation from the exact solution. A typical remedy to avoid early trapping of the solution is introducing mutations.

Mutation rate:

The probability that a gene is selected for mutation is called the mutation rate. Following the choices made in literature [24,27] also here the mutation rate is varied in the range between 1/(# genes) and 0.2. The effect is evaluated for the three different recombination strategies. The result of varying the mutation rate for the arithmetic cross-over strategy is presented in Figure 5-9.

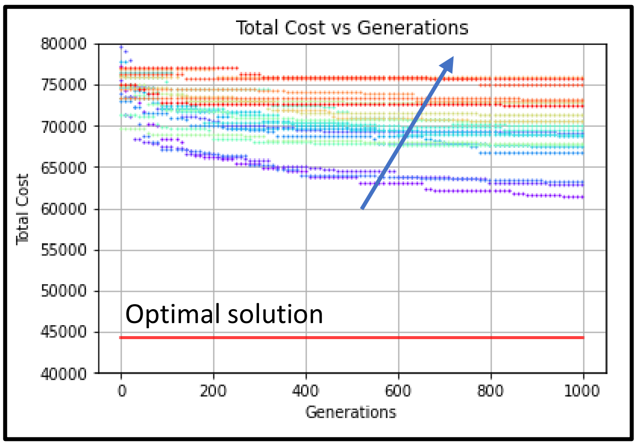


Figure 5-9: Total cost versus generations with different mutation rates for the arithmetic cross-over. The arrow depicts the increase of mutation rate

Intuitively one would expect increasing the mutation rate would increase the flexibility of the model. But unexpectedly the experiment shows the convergence rate drops tremendously with increasing the mutation rate and the distance to the optimal solution is increased! This same influence is , to a somewhat lesser extent, found for the other recombination strategies.

All in all , varying the mutation rate according to this protocol does not help to increase the performance of the model for neither of the recombination strategies. For completeness the effect varying elitism is illustrated in the next section.

Elitism:

The influence of elitism with the three recombination strategies is tested by varying the amount of elitism from 0 to the number of the total population. In Figure 5-10 the total cost versus the level of elitism is presented for each of the three recombination strategies. With the exception of introducing 1 elitist solution to the selection of the population the total cost is hardly influenced by increasing the elitism parameter. With varying the elitism parameter the gap to the optimal solution remains, even for the best results, more than 20%.

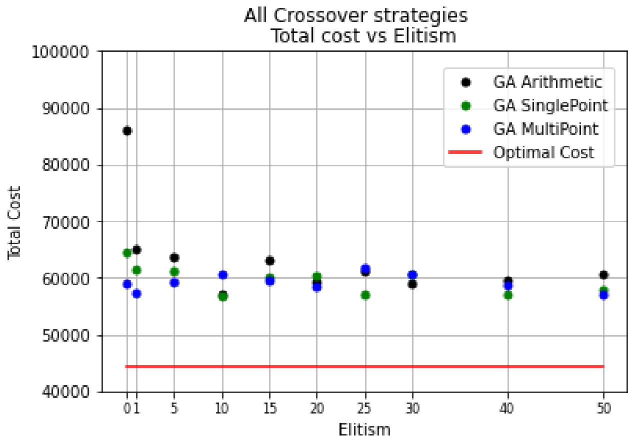


Figure 5-10: Total cost vs level of elitism for all recombination strategies

5.2.3. Discussion: How to improve?

On the one hand the initial convergence rate is significantly influenced by varying the parameter settings of the genetic operators. The multi point and single point strategies perform (by far) the best, so these are the strategies of choice. But on the other hand considering the gap between the total costs of the different recombination strategies and the optimal solution, the current GA does not provide adequate results yet. The magnitude of the gap (>20%) is similar for all three strategies . Different parameter settings of the genetic operators prove not to affect this gap.

Most probably the high initial convergence rate is not so much caused by the influence of the genetic operators but is caused by the random generation of the initial population that offers the space to improve rapidly in the earlier generations.

The question is : why is the solution trapped in a local optimum?
A qualitative analysis as well as literature provides a base for improvement. Mentioned as possible causes for premature convergence are: lack of diversity [26] and “blocking” of the algorithm [23]. Here we will focus for the moment on limitations of the used mutation operator in exploring the search space [21].

Increasing the flexibility

The idea behind the functioning of the genetic operators is to create the ability to explore the entire search space. The current implementation of the genetic algorithm and the respective operators may be too rigid. In specific the mutation operator may not facilitate proper exploration of the search space.

The current mutation operator only allows changes for the genes which are chosen for mutation or exceed the bounds of the respective gene. By following manually imposed mutations that logically can be inferred to increase the quality of the solution, it is observed that the positive change is negated because of corrections that are imposed based on the forthcoming boundary violations. This way the cost of the chromosome increases again.

Furthermore the current boundaries limit the effect of the mutation operator. In specific the transport gen is limited by the upper boundary created by the products available at KLM.

Another impression was that chromosomes show increasing resemblance after several generations.

Based on this first of all an adaptation of the mutation operator could be the way to go. Details and results of an adapted mutation operator will be presented in the next section. Furthermore we will focus on maintaining sufficient diversity of the chromosomes.

5.2.4. Giving the algorithm more flexibility and allowing for diversity

The results from the previous section showed that the basic implementation of the genetic algorithm does not provide adequate results. Probably this is due to the fact that this implementation is too rigid. In this chapter several approaches to increase the flexibility of the algorithm will be proposed and implemented. The results will be discussed.

Introducing methods to increase flexibility

The methods proposed in this section are concerned with the mutation operator and the selection process. First an alternative mutation operator will be proposed. Secondly dealing with the development of sparsity in the chromosomes (the fraction of “transaction-less” genes) proves to need attention and will be treated.

An alternative mutation operator

In the initial mutation operator genes are obliged to change in the case the bound on that specific gene is exceeded. To accommodate for this the mutation operator is adapted by adding three features.

Firstly, the current mutation operator only allows changes for the genes which are selected for mutation or exceed the bound of the respective gene. In the improved mutation operator any arbitrary gene following the mutated one can be changed . The amount these genes are allowed to change is determined by the change of the mutated gene and the respective bounds.

The second added feature allows other genes than the mutated one to change in the same period where the mutation takes place. These genes can adapt their value to the changes the mutated gene induces (In the old situation only the value of the actual mutated gene can change)

The third additional feature considers one type of genes: the transport amount. The upper bounds regarding the products available at KLM is increased to the possible number of products available at KLM, this equals the upper bound for the order amount of the considered period. Furthermore a destination for which the transport amount is mutated is prioritized in transport over the other destinations, so giving it the “first choice” of the available products.

Combining the three features, the change of the genes in a certain period is captured in a mutation vector ($\Delta^{\boldsymbol{mut}}$). The values of the mutation vector describe the magnitude the genes are allowed to change and in which direction. The mutation vector is constructed as:

$$\Delta^{\boldsymbol{mut}} = \begin{bmatrix} \Delta_u^{mut} \\ \Delta_{Trans,l}^{mut} \\ \vdots \\ \Delta_{Trans,L}^{mut} \end{bmatrix} = \begin{bmatrix} u_{t_{mut}}^{original} - u_{t_{mut}}^{mut} \\ T_{t_{mut},l}^{original} - T_{t_{mut},l}^{mut} \\ \vdots \\ T_{t_{mut},L}^{original} - T_{t_{mut},L}^{mut} \end{bmatrix} \tag{73}$$

Where $u_{t_{mut}}^{original}$ is the value before mutation has occurred in the mutation period (t_{mut}) and $u_{t_{mut}}^{mut}$ is the value due to the mutation. The same structure is applied to the transport amounts.

The magnitude of genes are chosen from a specified interval. Depending on whether the value in the mutation vector is positive or negative, the intervals are constructed in the following way.

$if\ \Delta_{gene}^{mut} \leq 0$:

The interval consists of $[\min(LB_{gene}, value_{gene,t} - abs(\Delta_{gene}^{mut})), value_{gene,t}\ ]$

$if\ \Delta_{gene}^{mut} > 0$:

The interval consists of $[value_{gene,t}, \max(UB_{gene}, value_{gene,t} + abs(\Delta_{gene}^{mut}))\ ]$

Additional : multi-gene mutation

First the proposed mutation operator is applied to a single gene. In later parts of the thesis the mutation operator is allowed multiple genes to be mutated again based on the principle of the mutation rate in one chromosome.

5.2.5. Increasing the quantity of “transaction-less” genes or “sparsity”

An intriguing observation is that the details of the chromosome that describes the solution found with the genetic algorithm, shows differences with that of the optimal solution generated with Gurobi. A remarkable difference is the number of “transaction-less” genes (zeros) for both cases.

Analysing the evolution of transaction-less events

In Figure 5-11 the development of the average number of transaction-less genes (“zero’s”) with increasing generation is illustrated for the three recombination strategies and compared with that of the optimal solution.

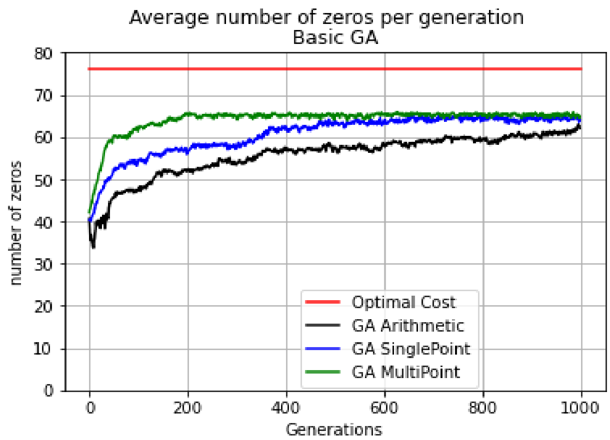


Figure 5-11: Comparing the average number of zeros in a chromosome per generation for the different recombination strategies and the optimal chromosome for the basic GA implementation.

In all strategies the first generations show a steep rise of the number of zero’s. So in this stage the average number of transaction-less events increases rapidly. The steepness of the rise directly correlates with the initial convergence rate of the three subsequent strategies (see Figure 5-8). After the initial increase the level of zero's plateaus off, significantly below that of the optimal chromosome. The optimal chromosome proves to be more dominated by zero valued genes. The level of dominance of transaction-less genes or zeros will be referred to as "sparsity".

So at the start of the evolution process the number of zeros is low, but the algorithm is able in this stage to create new zeros at a significant rate. After this initial stage it plateaus off, so somehow zeros are not formed anymore at a sufficient rate or disappear, thus balancing at a level below the optimum. In the current form the genetic operators are not able to compensate for this.

A qualitative, somewhat handwaving, explanation of this phenomenon is in the following reasoning: After the recombination process of parents intermediate chromosomes are formed. In the next step the values of these genes are compared to the boundary conditions. Subsequently the value of each gene is adapted to comply with the lower, and higher bounds. Those genes for which the value is already non- zero or for which the zero is exchanged after the mutation by a value larger than zero (due to an imposed lower bound) will not, or only slowly change to zero via mutation or cross-over.

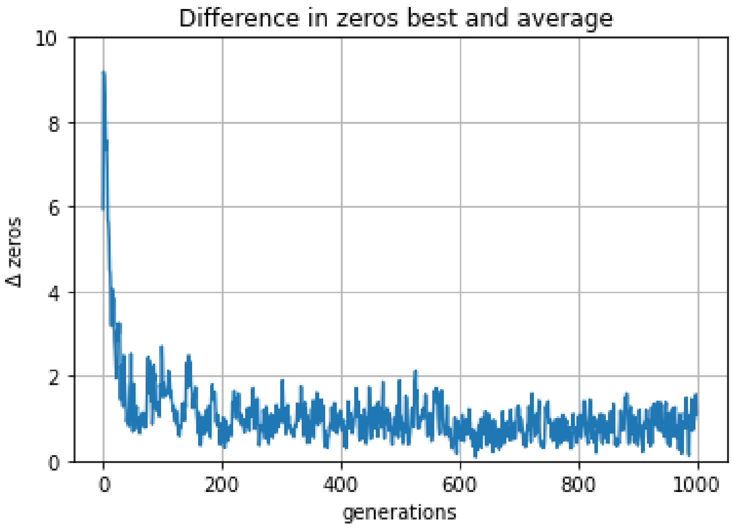


Figure 5-12: The difference between the number of zeros for the best chromosome of the population and the average number of zeros of the respective population in each generation

Probably this is caused by the following:
At the start of the algorithm the structure of the chromosomes is randomly created, so the sequence of zeros is randomly spread and non-correlated over the chromosomes. This facilitates an increase in the number of zeros in specific chromosomes by cross-over, while at the same time the number of zeros in other chromosomes will decrease. The latter will result in higher costs so these chromosomes have a higher chance to be discarded in the subsequent selection process.
After some generations the average number of zero's in the population has increased, but at the same time becomes more structured / correlated over the chromosomes. The cross-over process will not help anymore to create more zeros and the creation of zeros is inferred to become mainly dependent on the mutation process .

In the current mutation process the value of a mutated gene is selected randomly from an interval spanned by the lower and upper bound. So the chance to create a zero by this process is much smaller than that for a non-zero value. Furthermore in the case a lower bound is met (larger than zero) , the chance that the resulting gene will take a value of zero is none. Overall this results only a very weak tendency to (re-)form transaction-less genes.

So in conclusion: Improving on the creation of zeros is required when comparing the results with those of the optimal model. So the measure to repair this is to stimulate the formation of transaction-less events and increase the sparsity, specifically by the adapting the mutation process.

Adapting the mutation process to create transaction-less events: Bias towards a "zero"

As stated above in the current mutation process the chance to create a zero is much smaller than that for a non-zero value. The probability for creating a zero in the mutation process is increased by biasing the probability distribution. The bias is created as follows:

Under the condition that the lower bound of the gene equals zero, the added probability for a zero value (P0) will get a pre-set value chosen on beforehand in the range from 0 to 0.5 (variation will be done in steps of 0.1). While a non-zero value can be picked with the complementary probability 1-P0, the value is with uniform probability drawn from the range between the lower and higher bound .

An additional route to introduce zeros: Discretization of the interval.

Another way to increase the probability for generating a value at the boundary (so also a zero) is achieved by discretising a dummy interval between zero and a value somewhat higher than the upper bound. The values of the lower and upper bound are projected on this dummy interval. This results in a sequence of values starting with the lower bound, the values from the discretized interval and ends with the upper bound. The mutation operator uniformly draws a value from this sequence. Adjusting the number of the discretization steps between the two values influences the chance of using one of the boundaries as the value for the mutated gene. The following figure shows the projection of the lower and upper bound on a discretized interval with different levels of discretization (k).

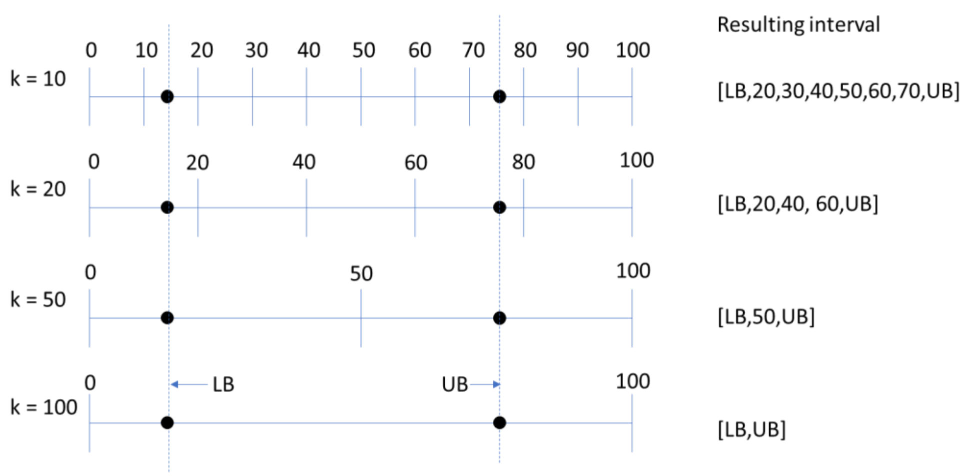


Figure 5-13: Discretized intervals for different levels of discretization (k) with the projection of the lower and upper bound. This results in an interval from which the to be mutated value can be chosen.

This is in line with the reasoning of the function for the mutation operator, which main goal is to explore the search space [23]. The level of discretization is randomly chosen by the model. Where high level of discretization allows for small refinements in the solution, low level of discretization allows for more exploring behaviour.

5.2.6. A method to measure diversity based on the chromosome structure

Diversity is supposed to enable a diverse population what would avoid premature convergence. To more or less quantify this ait it is worthwhile to introduce a measurement: a diversity parameter.

Next to quantifying and thus giving an insight in the evolution process this parameter can/will also be actively used to control the evolution process.

Measuring diversity:

Of the several methods to measure diversity two methods are considered in this thesis. One is based on total cost and the other on quantifying chromosome similarity. The diversity based on cost is straightforward: determined by the value of the total cost as discussed in Chapter 4.

A more relevant measure of diversity is created in the following way: In a population the chromosomes are compared to the best one. The chromosomes are compared on gen level, based on the order occurrences and the number of flights. If a shared value is found a 1 is given, else a 0 will be given. Summing those and dividing by the number of genes provides a number for chromosome similarity. For this measure it are not the amounts that are chosen as a base for the shared value, because whether a flight or order occurs or not presses in general more heavy on the total than variants in the respective amounts

An example of the diversity measurement and the total cost evolution is shown in the next figures. The diversity measurement based on the chromosome similarity is used to calculate the chromosome similarity for a situation where the genetic algorithm was performed while using phenotypic diversity.

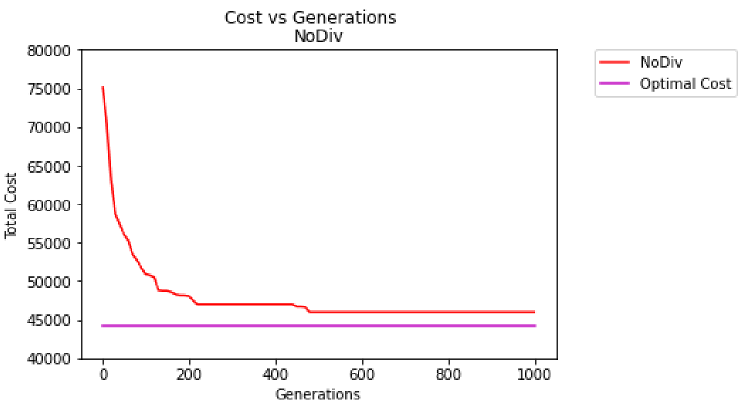


Figure 5-14: Cost progression over periods

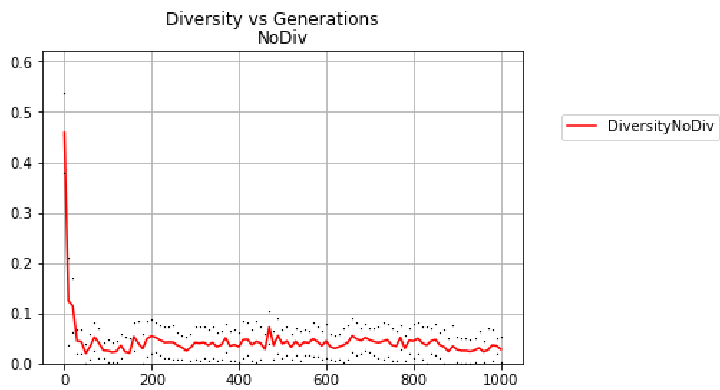


Figure 5-15: Chromosome similarity diversity based on the total cost diversity

Figure 5-14 show the cost progression over the generations. The convergence is in the first generations is high, but rapidly decreased after roughly 120 generations. The cost progression correlates to some extent with the development of the diversity of the population over the generations shown in Figure 5-15. This figure shows that a higher level of diversity can be found in the 20-40 generations, afterwards the diversity rapidly drops. This might be a reason why the genetic algorithm is having difficulties improving the solution.
Another observation from this graph is that it seems that the genetic algorithm improves if in previous generations the diversity increases. This is based on the relative convergence rate and the

diversity depicted in the following figure where the dashed line is the relative convergence and the solid line is the diversity based on chromosome similarity.

Figure 5-16 shows that if the genetic algorithm improvement with the spikes of the relative convergence rate. In the generations just previous to this improvement there is an increased diversity. This could be coincidental but is enough ground to explore this direction.

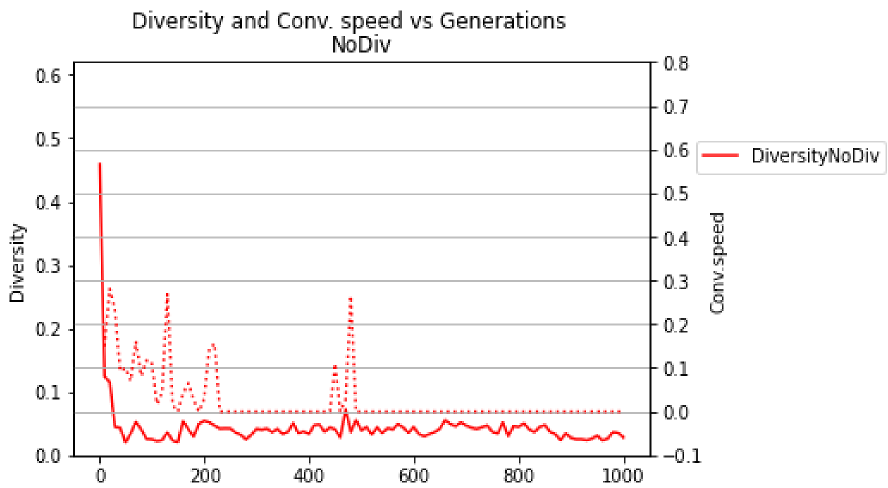


Figure 5-16: Diversity and relative convergence rate versus the generations

The diversity based on chromosome similarity is implemented as follows: Based on the number of best performing chromosomes and the population size, a number of spots is available for less performing chromosomes. The spots are filled with chromosomes which have a chromosome similarity ,compared to the best performing chromosome of the respective generation, which is less than a threshold given by the diversity parameter.

By the way, in the situation where the number of available spots is higher than the number of unique chromosomes which satisfy the threshold, random chromosomes will be mutated until the threshold is met and the available spots are filled . In the situation where the chromosomes eligible for the available spot selection outnumber the available spots, the spots are randomly assigned with a decreasing probability over the increasing total cost.

5.2.7. Directed mutation

Increasing Performance by directed mutation

The Genetic algorithm does not know explicitly where to apply a mutation. To guide the algorithm where to apply mutation, directed mutation could be applied. The summed contribution of all the genes constitutes the fitness function. The idea behind using the directed mutation is that when there is a strong cost increase over a number of time periods, this cost increase is caused in a certain range of periods before. The steps are: to identify these regions with high cost increase and subsequently apply mutation in a timespan before the cost increase is the highest. For this procedure it is convenient normalize the cumulated cost. The cumulative cost are normalized over the vaccines which are or should be available at the respective time period. The algorithm can be guided based on the normalized cost cumulation over the time periods. The following figure shows an example of the normalized cumulative cost evolution over the time periods.

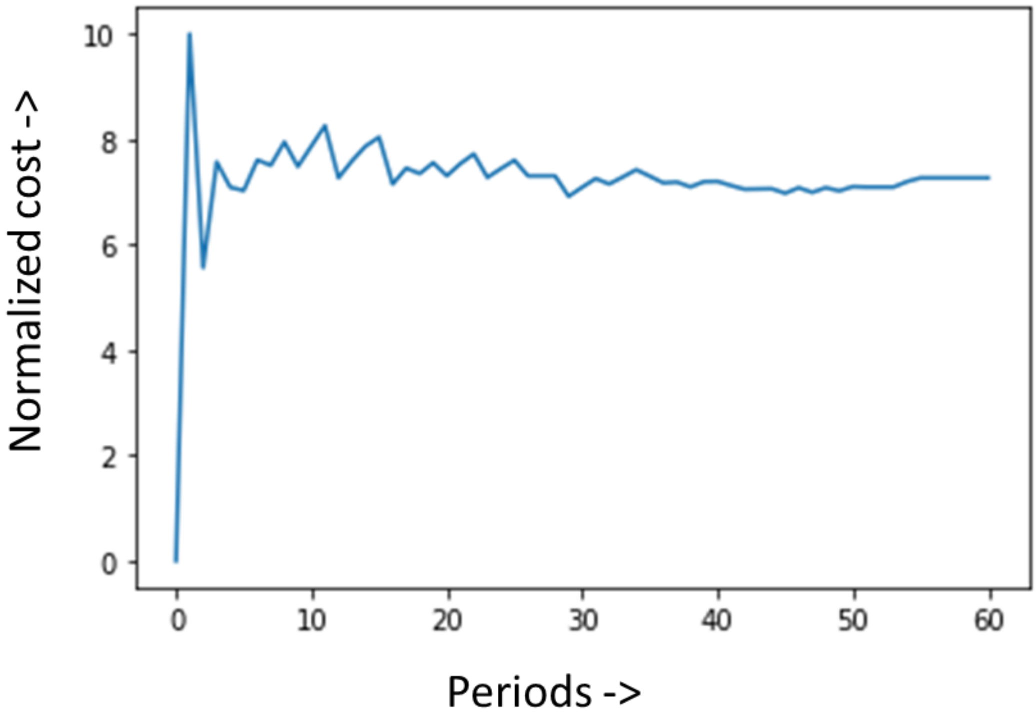


Figure 5-17: Example of normalized cost over the periods for a random chromosome

To find the periods of highest cost increase a sample period with length n is used. 's' is a parameter setting that can be varied. The regions for which the cost increase, over this length s, is the highest are sought for.

To affect the periods in which the cost increase is the highest, the periods before and the respective periods should be targeted for mutation. The periods with the highest cost increase are noted as $t^{highest}$. The periods which are considered for mutation are determined by the width parameter τ. The interval in which mutation should occur is then $[t^{highest} - \tau, t^{highest}]$. From Figure 5-17 of the normalized cost function, the highest cost increase is often in the first period due to the initialization. To allow for more mutation possibilities, the genetic algorithm determines the n highest cost increases $t^{highest,n}$. The model constructs the mutation interval for each $t^{highest,n}$ in which mutation occurs. After which each mutation interval is mutated separately from the original chromosome of which the intervals are determined. Applying the directed mutation to each chromosome of the population would lead to a significant increase in computation time. To retain a reasonable computation time the application of the directed mutation is limited to the best 10 chromosomes in every 10 generations.

5.2.8. Results of applying the flexibility improvements

Most of the measures introduced in the previous section separately only give moderate effects on the evolution process. The strength is found in combing the different measures that increase the flexibility. This is presented at the end of the section. To show the effect of the separate measures as well as for completeness first these results will be addressed shortly.

Results for the improved mutation operator

As only a single gene is selected for mutation, the mutation rate has no influence. The parameters which will be varied are the cross-over rate and the elitism parameter. The cross-over rate is varied between 0.6 and 1, where the elitism is set the range from 10 to 100 % in steps of 10 % .

The best performing recombination strategy is the multipoint with a cross-over rate of 1. The result in the following figure shows the cross-over rate of 1 with several levels of elitism.

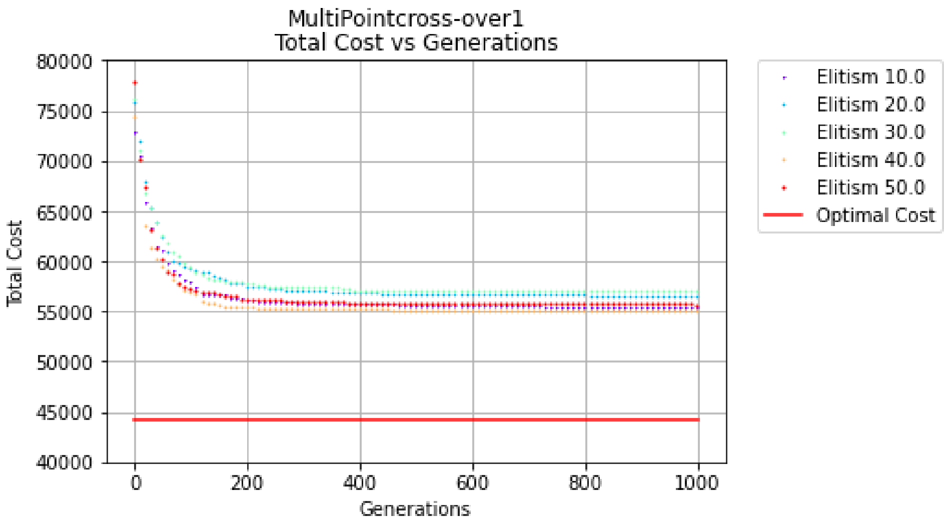


Figure 5-18: Influence of the elitism and the improved mutation operator on the total cost

The first observation is that the algorithm converges very fast initially, nearly independent of the level of elitism (a very slight optimum around a level 0f 30%) and after 200 generations plateaus off to a rigid plateau . Also here only a tiny improvement in the gap to the solution found by the MIP solver is realised.

The same trend is observed for all three cross-over operators. This means the GA is still getting stuck in a local optimum. The improved mutation shows promising results with, although tiny, improvements to the basic implementation of the mutation operator. The use of the improved mutation operator will be further exploited with the use of the bias towards a zero and the use of discretized intervals.

Increasing the sparsity

The idea behind this method is to push the genetic algorithm out of the local optimum by increasing the probability for a gene to obtain the value zero that will reduce the number of transactions. The increased probability for a zero is increased from 0 to 0.5 with increments of 0.1. In the following the results for the multipoint cross-over will be presented as the cost versus the generations and the final total costs for each increased probability for a zero. The difference in average zeros per chromosome in each population will be compared for each recombination strategy.

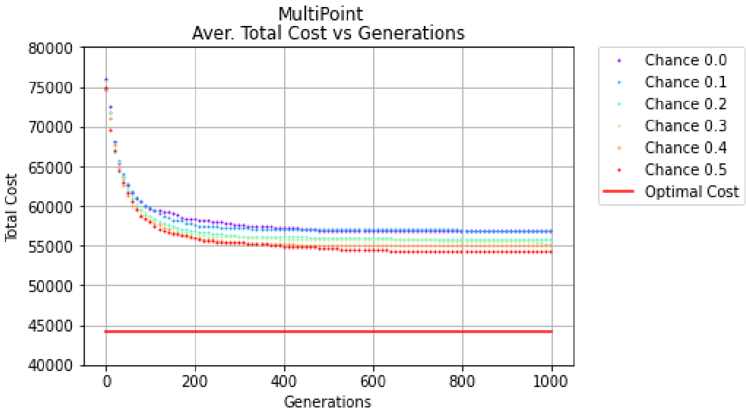


Figure 5-19: Bias towards a 0 for different probabilities applied with the multipoint cross-over showing the total cost vs generations

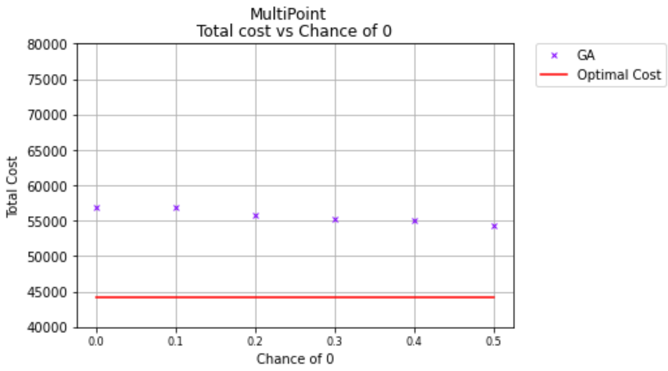


Figure 5-20: Final cost for different probabilities for bias towards a zero

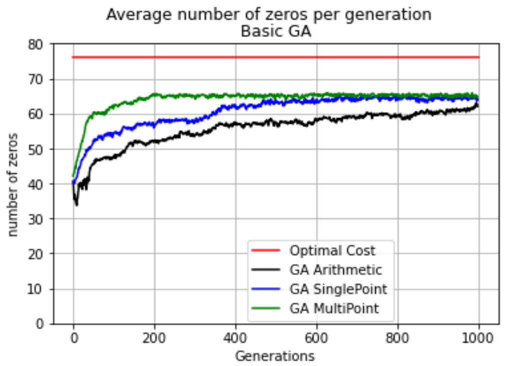


Figure 5-21: Basic GA implementation: Comparing the average number of zeros in a chromosome of the basic GA implementation per generation for the different recombination strategies and the optimal chromosome.

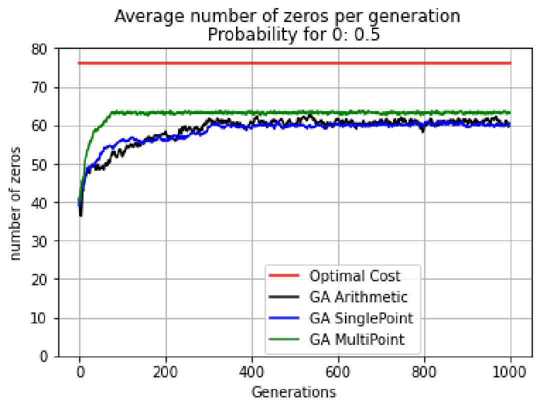


Figure 5-22: Bias towards a zero: Comparing the average number of zeros in a chromosome per generation with a bias towards a zero for the different recombination strategies and the optimal chromosome.

The final solution is still significantly above the optimum solution shown in Figure 5-21. Although the intention was to push the generation of zeros, this proves to be hardly effective , see Figure 5-22 that shows the development of the number of zeros over the generations. So it can be concluded that applying only a bias towards a zero is not strong enough.

Discretized interval

The last proposed method to improve the search for zeros is by implementing the discretized interval where the algorithm randomly chooses the level of discretization for each mutation. Similar to the previous methods the multipoint shows superior performance compared to the rest. The best found solution with multipoint is shown in the following figure:
The application of the discretized interval with varying levels of discretization show a significant improvement in both the gap to the optimality and the increase of zeros in the population for each generation shown in Figure 5-24.

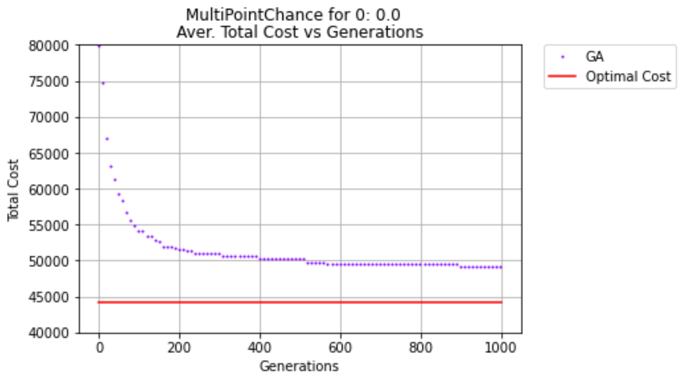


Figure 5-23: Applying the discretized interval method to the multipoint recombination strategy showing the total cost versus the generations

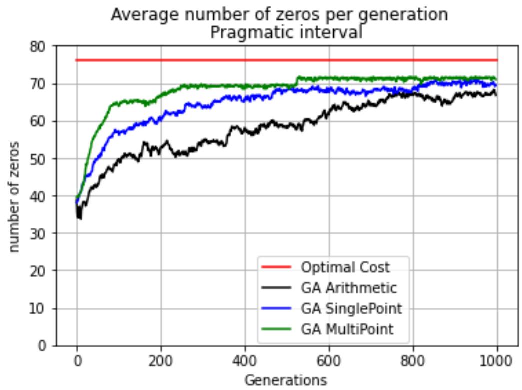


Figure 5-24: Discretized interval The average number of zeros in a chromosome per generation for the different recombination strategies with the discretized intervall method and the optimal chromosome

So all in all the separate measures to increase the flexibility and to create more “transaction-less“ events are effective in creating a high initial convergence rate, but only very limited in avoiding trapping in a local optimum. The final results is improved only slightly towards the optimal solution. It seems that none of these measures generates sufficient flexibility to avoid the local optimum.. At least: when used separately. Since the measures taken are different in nature it seems worthwhile to combine the protocols

Combination of bias towards a zero and the discretized interval:

The combination of combining the bias toward a zero and the discretized interval improves the convergence rate and results in a better solution compared to no increased chance for zero.

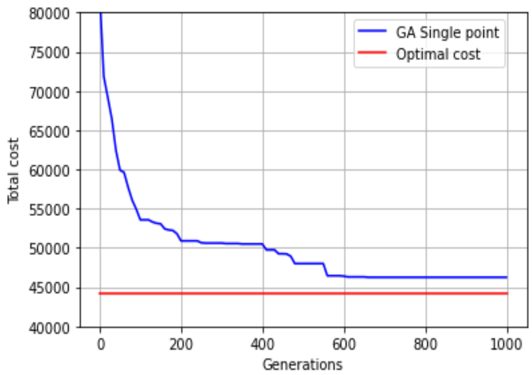


Figure 5-26: Best found solution with single point recombination and the combination of bias towards a zero and the discretized interval method.

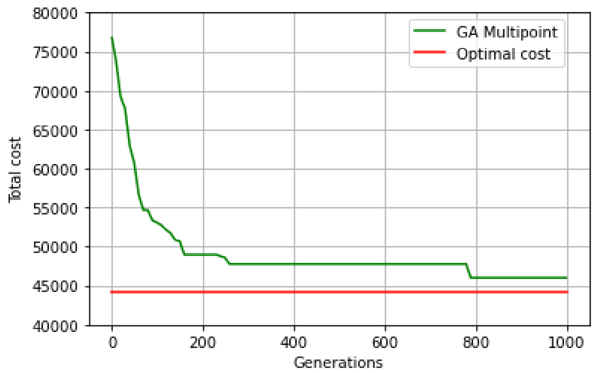


Figure 5-25: Best found solution with multipoint recombination and the combination of bias towards a zero and the discretized interval method.

Directed mutation

The results for directed mutation improve the convergence of the genetic algorithm further. The length of the search interval and the number of periods before the search interval have little influence on the convergence and the gap to the optimal solution of the genetic algorithm.

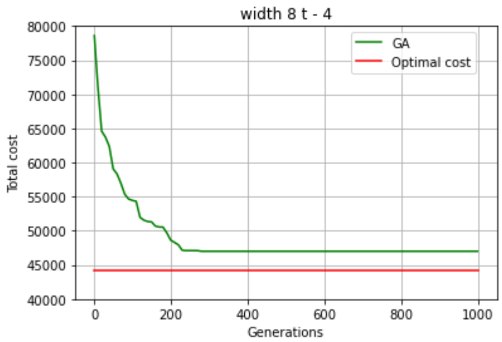


Figure 5-27: Directed mutation: Total cost vs generations for the directed mutation with a search interval of length 8 and an interval of length 12 in which mutation is applied.

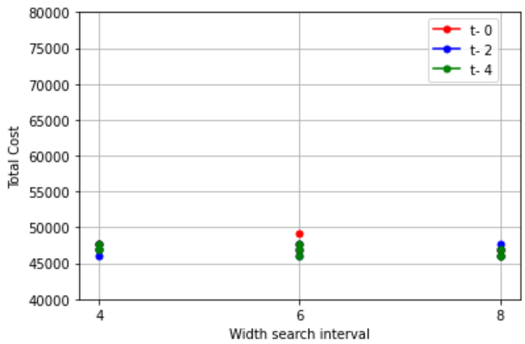


Figure 5-28: Directed mutation: Total cost for varying the width of the search interval and the number of periods added to interval in which mutation is applied

Based on the Figures 5-27 and 5-28 there is a tendency increasing the interval upfront for the mutation leads to better solutions.

5.2.9. Results of the adapted GA on the multi-destination problem

In this section the proposed genetic algorithm will be extended to multiple- destinations. The best performing mutation operator designed in the previous chapter will be implemented. Based on the increased size of the chromosome the improved mutation operator will be combined with the ability to choose multiple genes for mutation. The recombination is the multipoint cross-over based on the overall best performance in the previous chapter.

Varying the diversity parameter and the number of spots available the best solution was found for 0.7 chromosome similarity and 15 spots available for the less performing chromosomes respectively.

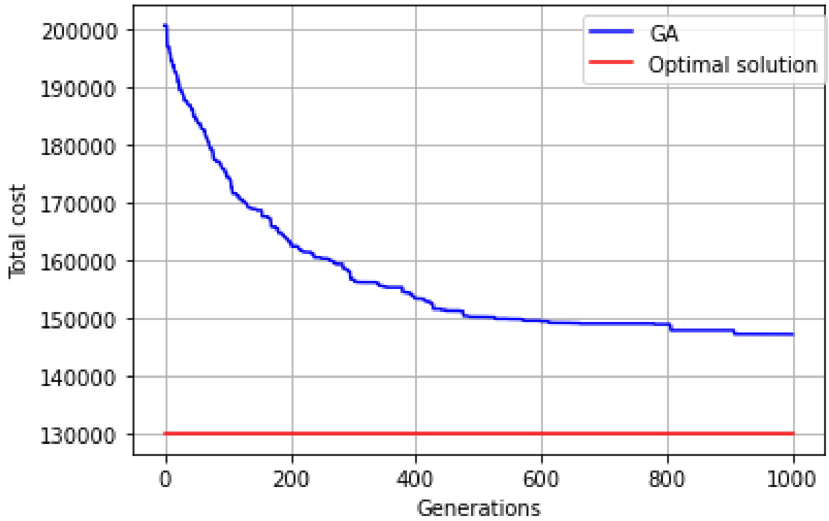


Figure 5-29: Multi-destination: Total cost vs generations for the best found solution with the genetic algorithm applied to the multi destination problem

The convergence of the solution shows a similar pattern to the convergence for a single destination. In the beginning the convergence is higher and gradually decreases. The difference is that there is no clear plateau formation yet as can be found for the single destination. The gap to the optimal solution is larger compared to the gap for a single destination. Here the gap after a 1000 generations is 11%. Observation of the chromosome evolution, shows that the genetic algorithm behaves more rigid compared to the single destination. The genetic algorithm has more difficulties in combining order occurrences it seems.

The accuracy can be increased by choosing for a larger number of the population size or extending the number of generations. The disadvantage of extending the population size and the number of generations is the significant increase in computation time.

5.3. Results for varying rolling horizon parameters

The influence of the parameters related to the rolling horizon approach on the total cost and the computation time will be presented in the following.

The rolling horizon approaches are tested for a model with 4 destinations and a planning horizon of 56 weeks. Two instances of available supply will be considered which are the minimum amount required for a feasible model and the total transport capacity per period. The minimum amount of available supply is estimated with the use of the commercial solver. For the considered situation the minimum available supply is approximated to be 765. The total transport capacity per period is 2400.

The parameters which can be adjusted are the number of time periods which are optimized with the integrality constraints (time window), the total number of time periods including the relaxed time periods (mu) and the step size which determines how much time periods the next to be solved problem is shifted(step size).

Considering the computation time : the time limit for each sub-problem is set to 600 seconds.

5.3.1. Straight forward approach:

The considered time windows are 4,8 and 12 time periods. The results for both instances of available supply are shown in the following table:

Table 5-1: The cost and computation time for different time windows and available supply

Q_t^{ext} = 2400		
Time window	Total Cost	Computation Time (s)
4	150800	1.2
8	144500	58
12	130000	2.1
Commercial solver	130000	7200
Q_t^{ext} = 765		
Time window	Total Cost	Computation Time (s)
4	Infeasible	-
8	Infeasible	-
12	Infeasible	-
Commercial solver	220200	7200

If the available supply is set to the transport capacity shorter time windows lead to higher cost. The results show for increasing the time window the total cost decrease. Optimizing for short time windows does not take future decisions into account which lead to higher costs. This same mechanism is the reason why this approach does not provide feasible solutions for the minimum amount of available supply.

Note that for the available supply equal to the transport capacity the computation time is an encouraging orders of magnitude less than that of the commercial solver.

5.3.2. Rolling horizon with relaxation and no overlap

Based on the myopic result for a time window of 4 periods, the considered time windows are 8 and 12. Varying the number of periods which are taken into account for the relaxation and the results are presented in the following table:

Table 5-2: The dependence of the total cost and computation time on the parameter μ for different values of the time window and for the available supply equal to the transport capacity and the numerical minimum amount.

Time window	μ	Cost	Computation time (s)		Time window	μ	Cost	Computation time (s)
Q_t^{ext}= *2400*					Q_t^{ext}= *765*			
8	12	1300000	3.3		8	12	Infeasible	-
	20	1300000	2.0			20	Infeasible	-
	24	1300000	5.3			24	Infeasible	-
	30	1300000	7.0			30	Infeasible	-
	34	1300000	7.3			34	220600	26
	40	1300000	7.8			40	220828	29
12	20	1300000	9.3		12	20	Infeasible	-
	24	1300000	10			24	Infeasible	-
	30	1300000	13			30	Infeasible	-
	34	1300000	12			34	219250	644
	40	1300000	15			40	218650	74
Commercial solver		1300000	7200		Commercial solver		220200	7200

Setting the available supply to the minimum required, this approach only provides feasible solutions for adding a large number of relaxed periods. In contrast in the case the available supply is equal to the transport capacity the addition of relaxed periods provides the same solution as the commercial solver for all values of mu. This again, within a fraction of the computation time compared with the commercial solver. An increase in the computation time can be noted with the increase of the μ value. For the minimum available supply a larger extension is required to calculate the solution. But when allowing for a larger value of μ does lead to an improved solution and the performance is even better than that of the commercial solver.

5.3.3. Rolling horizon with relaxation and shifting:

The results of the rolling horizon approach including relaxation and shifting are presented in the Figures 5-30 – 5-33. The shift of the time windows is varied in a range from 2 till 6 for the time window of size 8, and in a range from 2 till 10 for the time window of size 12.

For the case where the available supply is equal to the transport capacity the total costs are similar to the costs found with the commercial solver. For both settings of the time window good quality solutions are found. In both cases there is no influence of the µ value on the solution. (With the exception of using a shift of 10 time periods in the time window 12 results in a minor increase of the costs). The results are presented in Figure 5-30

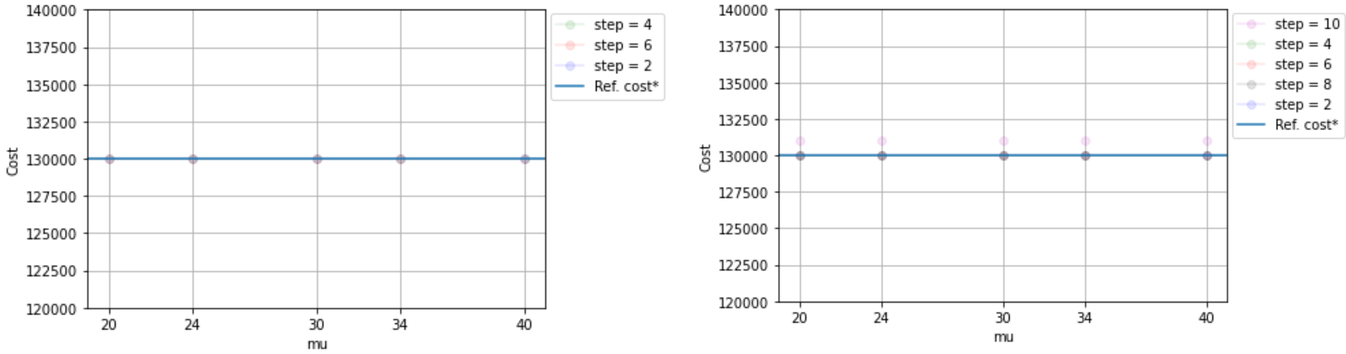


Figure 5-30: Rolling horizon(iii) Transport capacity: Total cost versus µ for time window length 8 (left) and 12(right) for the available supply set to the transport capacity for different shift amounts

Setting the available supply equal to the required minimum amount, similar quality of solutions compared to the commercial solver are found. The solutions are even slightly better than those of the commercial solver. The concrete costs are dependent on the value of µ and the shift amount. To obtain a solution a high value of µ is required; If the shift is too large and the µ value is low, the algorithm does not find feasible solutions. Results are presented in Figure 5-31

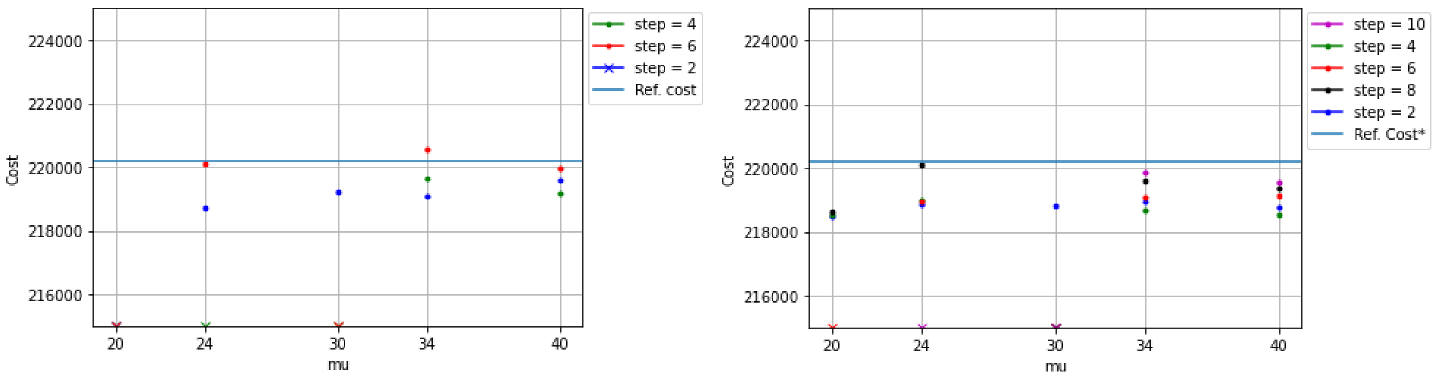


Figure 5-31:Rolling horizon(iii) Minimum amount : Total cost versus µ for time window length 8 (left) and 12(right) for the available supply set to the minimum amount required for different shift amounts. Infeasible solutions in figure are denoted with an 'x'.

The computation remains running smoothly.
For the situation where the available supply is equal to the transport capacity the computation runs two to three decades faster than that of the commercial solver. The value of µ has no strong effect on the computation time. While the shift amount has a stronger effect. For the time window of length 8 the computation time increases with decreasing shift amount. A similar trend can be seen for the time window with length 12, although step size 4 is an extreme outlier for each value of mu. The results for the computation time are shown in Figure 5-32

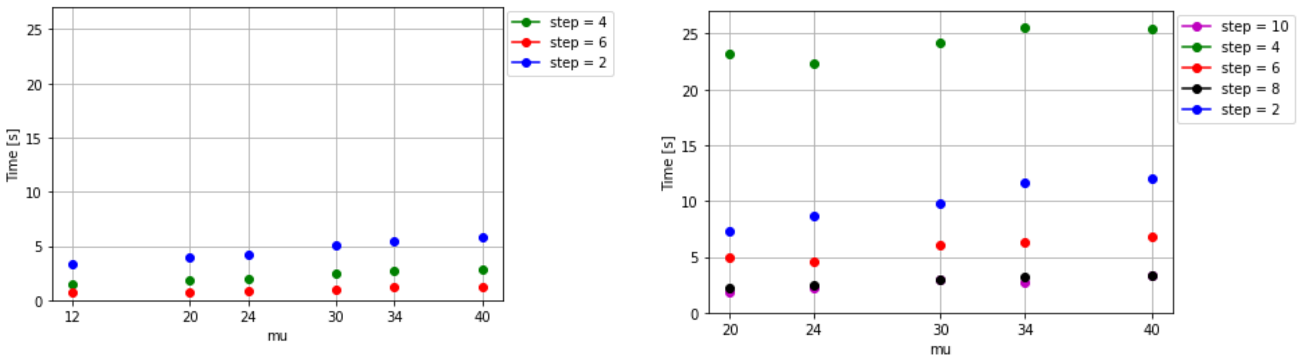


Figure 5-32: Rolling horizon(iii) Transport capacity: Computation time versus µ for time window length 8 (left) and 12(right) for the available supply set to the transport capacity for different step sizes

For the situation where the amount of available supply is equal to the minimum amount required the computation time is generally somewhat larger but still is 1.5 to 3 decades faster. see Figures 5-33. As discussed above for small values of µ not all shift amounts provide feasible solutions. Increasing the value of µ has a positive effect on the computation time for a window length of 12, somewhat depending on the shift amount (see Figure 5-33 (right)). For a window length of 8 the computation time decreases with increasing the shift amount, while there is negligible influence of the µ value on the computation time.

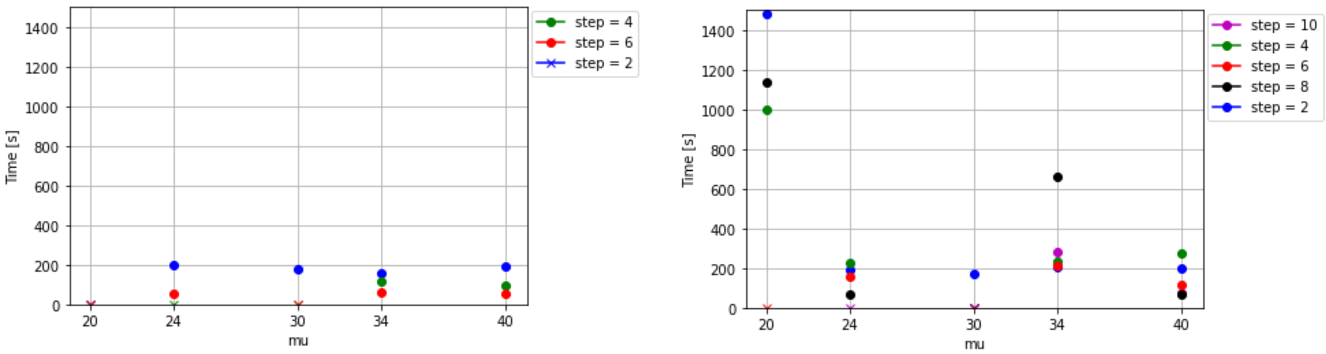


Figure 5-33: Rolling horizon(iii) Minimum amount: Computation time versus µ for time window length 8 (left) and 12(right) for the available supply set to the minimum amount, for different step sizes.

5.4. Results of Robust optimization

In this section the results of robust optimization are presented. Robust optimization is applied on the situation with four destinations, a constant demand rate of 100 for all destinations and a planning horizon of 21 weeks. The resulting size of the model requires a fast algorithm, so the model will be solved using the rolling horizon method including relaxation and shifting. The time window of 12 and a shift amount of 2 are used.

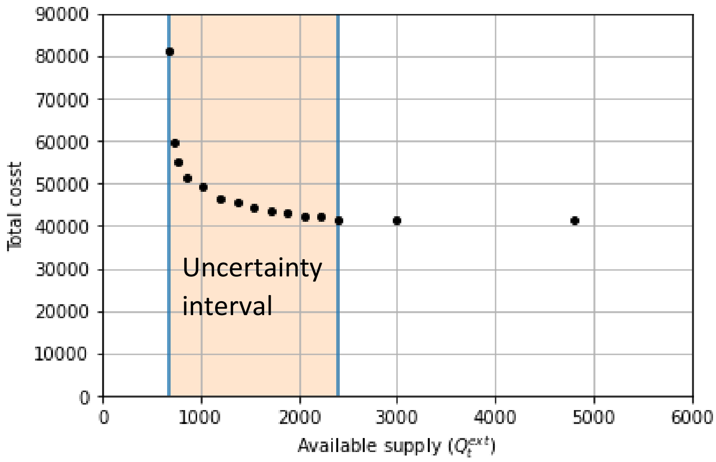


Figure 5-34: Influence of the available supply on the total costs. The interval ranging from the minimum amount required up to the transport capacity is used as the uncertainty interval for the robust optimization.

First the influence of the available supply on the total costs is calculated (see Figure 5-34). The figure shows that for an available supply larger than the transport capacity there is no effect on the total costs. But it also demonstrates that the total costs are largely influenced by the available supply in the range of interest i.e. from the minimum amount required up to the transport capacity. Consequently the supply uncertainty in this range will have a large effect on the total cost. This interval of the available supply, ranging from the minimum amount required up to the transport capacity, will be used as the uncertainty interval for the robust optimization.

Using the method described in Section 4.4 on robust optimization first the optimal robust solutions for different values of Γ are determined. The values of Γ range from 0 to 1.

In figure 5-35 the optimal costs found with the rolling horizon method for different Γ- values are presented. With increasing the Γ- value the costs increase. The increase of costs for increasing Γ-values is mainly caused by the increase of order costs and shortage costs. For a Γ-value of 1.0 the transport cost increase significantly. There are no wastage costs and in four cases ($\Gamma = 0.2, 0.6, 0.9$ and 1) tiny inventory costs are charged.

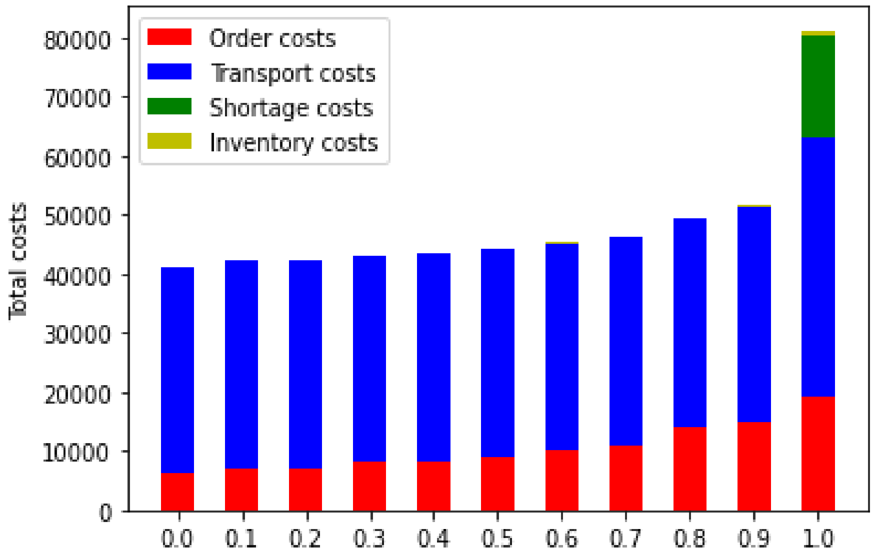


Figure 5-35: The costs of the different components of the distribution process for different values of Γ.

In Figure 5-36 the number of occurrences is indicated with black dots and the blue line is the length of the planning horizon. Similar to the cost increase shown in Figure 5- 35. The number of orders Increases stronger than the cost increase. It can be seen that in no Γ – solution orders occur in every period.

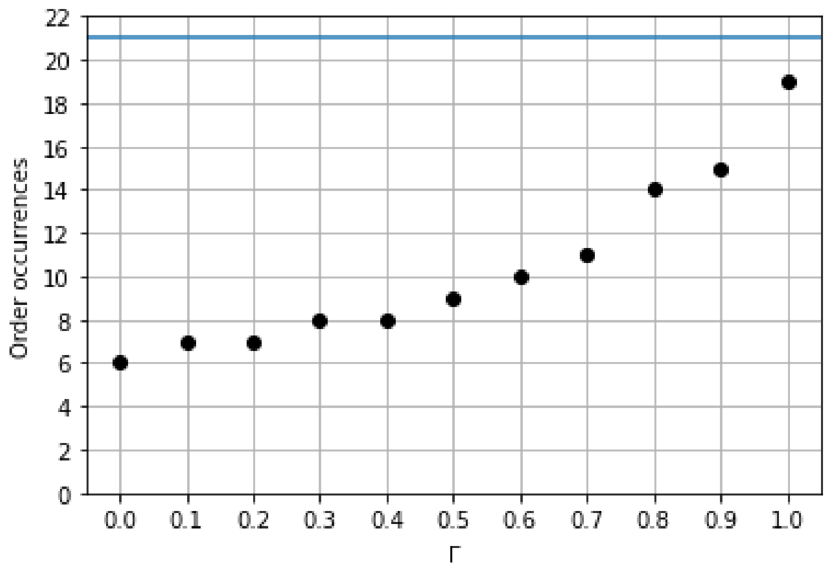


Figure 5-36 : Number of order occurrences in the planning horizon for different values of Γ where the blue line is the length of the planning horizon

The relative constant transport costs can be seen in Figure 5-37 where the number of total flights is depicted. Up to a Γ –value of 0.8 the total number of total flights are equal. A significant increase of total flights starts from a Γ –value of 0.9.

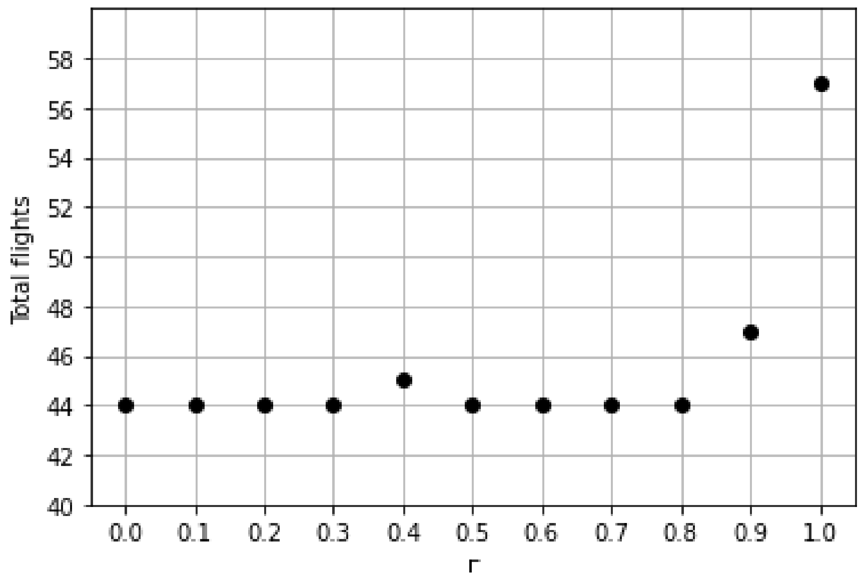


Figure 5-37: Total number of flights for different value of Γ

The solution for each Γ value provides the total costs, the order and the transport amounts to each destination combined with the number of flights and the periods in which an order occurs. The order moments and the number of flights can be gathered in a flight and order plan.

To test for which value of gamma the solutions are most protected against the supply uncertainty, simulations are performed using the rolling horizon method. In the simulation the available supply for each separate time period is uniformly drawn from the uncertainty interval. Note that if the order amounts and transport amounts from the robust solutions are combined with the random values of the available supply most probable only the most conservative robust solution leads to a feasible solution. Hence, a different approach is chosen to test the performance of the robust solutions. Neglecting the order and transport amounts and using the flight and order plan allow for more flexibility in testing the robust solutions with a simulation. The simulation determines the order and transport amounts based on the uniformly drawn available supply.

First it is determined if the simulation provides a feasible solution, if so the results from the simulations provide the order amount and transport amount to each destination. The total costs found with the simulations are compared with the total costs found in the robust solutions. The simulations are tested for 100 sequences of available supply. In Figure 5-38 the number of feasible solutions out of the 100 simulations is shown in dependence of Γ. With increasing conservative values of Γ, more feasible solutions can be found. For values which are less than Γ = 0.5 very few, much less than 15, feasible solutions are found.

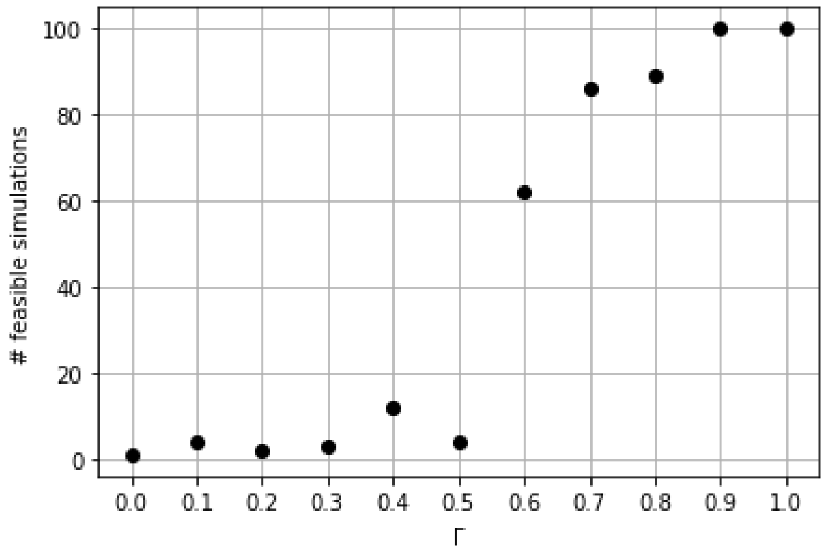


Figure 5-38: Number of feasible solutions resulting from the simulations with the flight and order plans of the respective Γ -values.

In Figure 5-39 for the range of values of Γ the total costs resulting from the robust solutions are compared with the average of the total costs of the feasible fraction of the 100 simulations.

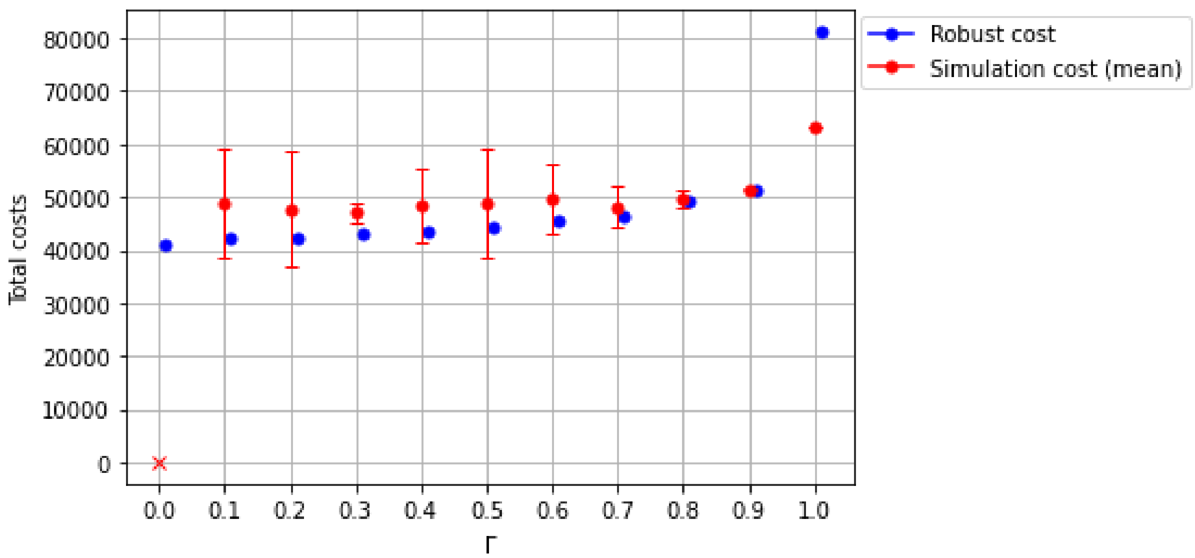


Figure 5-39: The total costs of the robust solutions versus the Γ -value. These are compared with the average total costs calculated from the concrete flight and order plans fitting the respective Γ -value. Only the feasible solutions out of the 100 runs are used, for the numbers of feasible solutions see Figure 5-38. The bars indicate two times the standard deviation. For a Γ -value of 0 (x) no feasible solution are obtained with the simulations.

The results of the total costs show a decreasing gap between the robust costs and costs resulting from the simulation with increasing value of Γ. Note that for lower Γ values the small fraction of feasible solutions result in a large scatter of the total costs found with the simulations, indicated in the graph with bars that represent two times the standard deviation. (Note that for a Γ of zero the simulation does not result feasible solutions at all). For a Γ in the range of 0.7 to 0.9 the costs from the simulations and the robust costs are nearly the same with only very little scatter, with slightly increasing total costs. The best robust solution is found in this range. 0.9 gives a feasible solution for all simulations. Lowering Γ to 0.7 results in a somewhat lower fraction of feasible solutions and somewhat more scatter on the total costs. This comes at the advantage of slightly lower total costs for the robust solution. For the worst-case scenario (Γ is equal to 1) the robust costs are higher than the simulated costs. This is caused by the constraints on the available supply in the robust solution.

6. Discussion

In this report a model for the distribution process for KLM has been designed and solved. The large computation time with the commercial solver requires a different solution approach. Two solution methods are studied. First the genetic algorithm is implemented and adapted to solve the model. The second solution method which is studied is the rolling horizon method.

6.1. Exact solution method

The exact solution method provides good quality results. For large size models, to reach the optimal solution, the computation time is excessive in line with findings in literature [35]. Given the quality of the solutions in terms of optimality, the results for the commercial solver will be used as a reference for the other solution methods.

6.2. Genetic algorithm:

The basic implementation of the genetic algorithm does provide results in a reasonable computation time, however the solutions are far away from the optimal solutions found with the commercial solver. Different parameter settings of the genetic operators prove not to affect this gap.

But the initial convergence rate can be significantly influenced by varying the parameter settings of the genetic operators. The multi point and single point strategies perform (by far) the best, so these are the strategies of choice. For these two strategies the genetic algorithm will be adapted to increase the quality of the solution.

Initially the convergence rate is high and can be maximised by choosing the right recombination strategy and the parameter setting. This is caused by the random generation of the initial population that offers the space to improve rapidly in the earlier generations.

The question is : why does this convergence rate diminish and why is the solution trapped in a local optimum? A qualitative analysis as well as literature provides a base for improvement. Mentioned as possible causes for premature convergence are: lack of diversity [26] and “blocking” of the algorithm [23]. In this work the focus is on the limitation of the used mutation operator to explore the search space [23]. The trapping of the genetic algorithm is investigated using a developed toolbox to analyse the diversity, convergence rate and the development of transaction-less genes.

The initial results of the basic implementation show that the genetic algorithm cannot adjust chromosomes in to the right direction. Due to the constraints which must be satisfied, the genetic algorithm gets trapped in a local optimum. The toolbox shows that valuable zeros in the chromosome are replaced with costly non-zero values.

To repair this phenomenon an improved mutation operator and a different diversity approach are proposed which allows for more flexibility. The functioning of the mutation operator is changed on the following aspects:

- A mutation vector is introduced which keeps track of the changes caused by the mutated gene(s) and defines the amount of value genes are allowed to change after the time period in which the mutation takes place.
- Impose a bias towards a zero with the help of monitoring the development of transaction less genes to prevent losing sparsity.
- Discretized interval to increase the chance for transaction less genes.
- Directing the mutation towards the genes that are involved with a large cost increase in a chromosome.
- Maintaining sufficient diversity based on the underlying chromosome structure.

None of these measures, individually, generates sufficient flexibility to avoid the local optimum. However, combining the measures creates an even higher initial convergence rate and provides results which are less than 5% off for a single destination.

Increasing the model size to four destinations the convergence of the solution shows a similar pattern to the convergence for a single destination. In the beginning the convergence is higher and gradually decreases. The difference is that there is no clear plateau formation yet as can be found for the single destination. The gap to the optimal solution is larger compared to the gap for a single destination. Here the gap after a 1000 generations is 11%. Observation of the chromosome evolution, shows that the genetic algorithm behaves more rigid compared to the single destination. The genetic algorithm has more difficulties in combining order occurrences it seems.

The accuracy can be increased by choosing for a larger number of the population size or extending the number of generations. The disadvantage of these measures is the significant increase in computation time.

The parameter which increases the convergence rate for the single destination the most is the directed mutation. Maybe in the future it is worthwhile to further explore the possibilities of improving this directed mutation for the multi-destination problem.

The large computation time and a too large gap to the optimal solution, asks for a different solution method.

6.3. Rolling horizon approaches

Three different applications of the rolling horizon method have been applied. The first application is the straightforward (myopic) approach. The second application, allows for adding a number of relaxed time periods in which the integrality constraints are relieved. The third application allows, besides adding relaxed time periods for a shift in time periods such that solved sub-problems are partly re-optimized with new information regarding future decisions. The straight forward approach(I) does not give a solution for the problem if the available supply is set to the minimum. The reason for this is the myopic approach of this method, where only the considered time window is optimized without considering future decisions. On the other hand, in the situations where the available supply is equal to the transport capacity, solutions are found. The quality of the solution increases with the increase of the time window. For all time windows the computation time is orders of magnitude lower compared to the commercial solver.

The rolling horizon approach with relaxation and without overlapping(II) is able to find a solution for the minimum available supply. If the solution is found, it is almost as good as or even better than the commercial solver. But to accomplish a feasible solution a large relaxed extension is required (a large value of mu). So for small and medium values of μ no solutions can be found. The main benefit is the reduced computation time.

Extending the rolling horizon approach with relaxation combined with overlapping(III) results in similar quality of the solution for both instances of available supply. The introduction of shifting broadens significantly the range of mu's for which a solution can be found. If the value of the shift is minimal (e.g. 2) the shortest extension of the time window is able to compute a feasible model for all values of mu. The reason for this is that limited shifts enable the algorithm to change the value of variables in subsequent iterations of the rolling horizon. This minimal overlap allows for a larger amount of variables to be reoptimized with the gained extra information.

The shift size also affects the computation time. An increase in shift decreases the computation time. This originates from the increase of sub problems which need to be solved. Notwithstanding this, for very limited instances (μ of 30 and step sizes larger than 2) this methods leads to infeasible solutions. It is unclear why.

Given the outstanding results on quality and computation time with the right choice of parameters this method will be used for the robust optimization in the next section.

6.4. Robust optimization:

For the robust optimization the method described by Bertsimas et al. [21] has been used. This method uses a Γ value to protect the solution of the model against the supply uncertainty up to a certain degree. The robust solutions for different values of Γ are determined. The robustness of each Γ value is tested by performing simulations using the resulting flight and order plans. The results show that a high value of Γ is required to protect the solution against supply uncertainty. Feasible solutions are obtained in the range of Γ values from 0.6 up to 1, while for lower values of Γ only small fractions of the simulations lead to feasible solutions. Based on the results of the robust optimization combined with the costs found with the simulations a Γ in range of 0.7 and 0.9 show a reasonable to good protection against the supply uncertainty. In this range the fraction of feasible solutions increases. This robustness does come at a price; The cost for the Γ value of 0.9 are 4% higher than the costs for the Γ value of 0.8 and 10% higher than the costs for Γ equal to 0.7. All total costs are higher than the nominal case without protection for uncertainty. Choosing for a robust approach with a Γ value of 0.8 results costs that are 16% higher than the nominal case. Needless to say that opting for this nominal case will expose the company to a very high probability of not satisfying the boundary conditions.

There is a high level of conservatism required. This might be influenced by the probability distribution used to generate the values of the available supply. Notwithstanding that the method developed by Bertsimas [21] is intended to be insensitive to the details of the probability distribution. Intuitively one might guess that robustness of the solutions could be sensitive to some details e.g. skewed probability distributions towards the transport capacity might allow for less conservative values of Γ. This influence of the probability distribution is not studied here, detailing on this aspect may require future attention.

7. Managerial insights and Conclusion

In this section first the managerial insights will be presented after which the report will be concluded with the conclusions from the experiments.

7.1. Managerial insights

This report provides an insight on the distribution process of multi-dose vaccines. Based on the conducted experiments and acquired knowledge the following points should be considered coping with supply uncertainty.

In extreme cases of order costs or available supply the deep freezers are used. However it should be noted that In the model formulation the acquisition for deep freezers is not taken into account. Hence, there is no need to use deep freezers at the KLM facility for the storage of vaccines.

Based on the assumption that there is no lead time for the vaccines which are ordered by the KLM facility, the larger part of the order amount is directly used for transport to reduce inventory costs. The size of the storage facility can be small based on the inventory cost but should be taken into account. Depending on the level of protection between a Γ- value of 0.7 to 0.9 orders occur in roughly 50 – 90% of the time periods.

First the uncertainty interval of the available supply should be determined. If the interval has a minimum value which is greater than the flight capacity per period there is no need to incorporate the uncertainty for the supply. In the case that there is no known uncertainty interval it is important to consider the supply uncertainty. To deal with supply uncertainty a flight and order plan is advised that is risk averse. The risk aversion comes at a price. Depending on the level of risk aversion, the increase of costs are up to 20% higher than the case where no supply uncertainty is taken into account. The experiments show that a protection against supply uncertainty from a Γ-value of 0.7 show increasing high levels of protection. Fully protecting the distribution process against supply uncertainty with a Γ of 1 shows that the estimated costs are higher than the cost realized with the simulations Not complying with the possibility of supply uncertainty can put the distribution process under large strain and has a high probability of not satisfying the boundary conditions.

To determine the optimal order and flight amount to each destination an agile method such as the rolling horizon method is advised. This allows for re-optimization of the order and flight amounts based on the chosen flight and order plan.

7.2. Conclusion

In this thesis a part of the vaccine supply chain is studied. A brief literature study shows the complexity of the vaccine supply chain. The distribution process considered here is the link of KLM between the vaccine supplier and national warehouses. This work addresses the influence of supply uncertainty on the distribution process of vaccines based on a mixed integer programming formulation. Two solution methods are developed, the genetic algorithm and the rolling horizon method. Significant attention has been paid to the analysis and development of the genetic algorithm. In the basic implementation of the genetic algorithm the solution is easy trapped in local optima. Analysis tools that were developed show that the mutation operator does not allow for proper search in the solution space and in specific has a bias towards too little transaction-less events. Methods to improve the flexibility are demonstrated. Individually the separate measures(improving mutation operator, sparsity, discretized interval and measurement of diversity, directed mutation) are effective in creating a higher initial convergence rate but are only limited capable in avoiding local optima. Combining all the measures creates an even higher initial convergence rate and provides more accurate results. For a single destination model the costs deviate less than 5% from the commercia solver but the computation time for the genetic algorithm is higher. It is known for heuristic methods that the computation time is larger for small sized models. Applying the modified genetic algorithm to a multi destination model creates results which deviate roughly 10% from the optimal solution. Also the convergence rate decreases. For multi-destination problem the genetic algorithm behaves more rigid compared to the single destination.

Although the genetic algorithm is strongly improved by the proposed measures it is not the method yet to be applied for robust optimization.

Three variations on the rolling horizon method are implemented. With stepwise improvements of the overlap strategy the quality of the results improves. The best results are obtained for the method that includes relaxation and shifting. This provides high quality solution at reasonably low computation times. And is therefore used for the analysis of robust optimization.

The results of the robust optimization show that a high level of conservatism is required to protect against uncertainty. The additional costs to obtain this level of protection are between 10-20 % compared to the nominal case.

Some recommendations for future work:

First of all it seems worthwhile to develop the algorithm further based on the proposed toolbox. It is believed that higher convergence rates and lower gaps to the optimal solution can be achieved. In specific exploring the possibilities of improving the directed mutation for the multi-destination problem may be fruitful in view of larger models to be solved.

The robust analysis is performed with a uniform distribution. Analysing the effect of different, in specific very skew, distributions might be worthwhile. Furthermore the results of the robust optimization show to be conservative. It would be interesting to develop a robust optimization method which does not require Γ for each period but where alternatively the Γ limits the sum of scaled deviations over all the time periods.

8. References

[1] Pharma.aero conducts review of Air Cargo in distributing Covid Vaccines (2021) International Airport Review. Russell Publishing Limited. Available at: https://www.internationalairportreview.com/news/161049/pharma-aero-review-air-cargo-covid-19-vaccines/ (Accessed: February 7, 2023).

[2] CDC Centers for disease and control and Prevention. (2021). Retrieved 9 March 2021, from https://www.cdc.gov/vaccines/vpd/vpd-vac-basics.html

[3] Chopra, S., & Meindl, P. (2013). *Supply chain management*. Essex, England: Pearson Education.

[4] Duijzer, L. E., Van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. *European Journal of Operational Research*, *268*(1), 174-192.

[5] Pfizer - Fact sheet. (2021). Retrieved March 7, 2021, from https://cdn.pfizer.com/pfizercom/Pfizer_PGS_COVID-19_Factsheet_071122.pdf

[6] Georgiadis, G. P., & Georgiadis, M. C. (2021). Optimal planning of the COVID-19 vaccine supply chain. *Vaccine*, *39*(37), 5302-5312.

[7] Tavana, M., Govindan, K., Nasr, A. K., Heidary, M. S., & Mina, H. (2021). A mathematical programming approach for equitable COVID-19 vaccine distribution in developing countries. *Annals of Operations Research*, 1-34.

[8] Zangwill, W. I. (1966). A deterministic multi-period production scheduling model with backlogging. *Management Science*, *13*(1), 105-119.

[9] Soysal, M., Bloemhof-Ruwaard, J. M., Haijema, R., & Van Der Vorst, J. G. (2015). Modeling an inventory routing problem for perishable products with environmental considerations and demand uncertainty. *International Journal of Production Economics*, *164*, 118-133.

[10] Chu, J., Huang, K., & Thiele, A. (2019). A robust optimization approach to model supply and demand uncertainties in inventory systems. *Journal of the Operational Research Society*, *70*(11), 1885-1899.

[11] Alipour, Z., Jolai, F., Monabbati, E., & Zaerpour, N. (2020). General lot-sizing and scheduling for perishable food products. *RAIRO-Operations Research*, *54*(3), 913-931.

[12] Gunpinar, S., & Centeno, G. (2015). Stochastic integer programming models for reducing wastages and shortages of blood products at hospitals. *Computers & Operations Research*, *54*, 129-141.

[13] Rohmer, S. U. K., Claassen, G. D. H., & Laporte, G. (2019). A two-echelon inventory routing problem for perishable products. *Computers & Operations Research*, *107*, 156-172.

[14] Coelho, L. C., & Laporte, G. (2014). Optimal joint replenishment, delivery and inventory management policies for perishable products. *Computers & Operations Research*, *47*, 42-52.

[15] Hsu, C. I., & Wang, C. C. (2013). Reliability analysis of network design for a hub-and-spoke air cargo network. *International Journal of Logistics Research and Applications*, *16*(4), 257-276.

[16] Woo, S. H., Kim, S. N., Kwak, D. W., Pettit, S., & Beresford, A. (2018). Multimodal route choice in maritime transportation: the case of Korean auto-parts exporters. *Maritime Policy & Management*, *45*(1), 19-33.

[17] Rinaldi, M., Bottani, E., Solari, F., & Montanari, R. (2022). The vehicle routing problem in the dairy sector: a case study. *International Journal of Food Engineering*, *18*(3), 239-252.

[18] Ben-Tal, A., & Nemirovski, A. (2002). Robust optimization–methodology and applications. *Mathematical programming*, *92*(3), 453-480.

[19] Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, *21*(5), 1154-1157.

[20] Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization* (Vol. 28). Princeton university press.

[21] Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, *52*(1), 35-53.

[22] Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations research*, *54*(1), 150-168.

[23] Holland, J. H. (1973). Genetic algorithms and the optimal allocation of trials. *SIAM journal on computing*, *2*(2), 88-105.

[24] Tabassum, M., & Mathew, K. (2014). A genetic algorithm analysis towards optimization solutions. *International Journal of Digital Information and Wireless Communications (IJDIWC)*, *4*(1), 124-142.

[25] Goldberg, D. E. (1990). *Real-coded genetic algorithms, virtual alphabets and blocking*. Champaign: University of Illinois at Urbana Champaign

[26] Mohammadi, M., Musa, S. N., & Omar, M. B. (2020). Optimization of multi-plant capacitated lot-sizing problems in an integrated supply chain network using calibrated metaheuristic algorithms. *International Journal ofOperational Research*, *39*(3), 325-363. [39(3)].

[27] Baker, B. M., & Ayechew, M. (2003). A genetic algorithm for the vehicle routing problem. *Computers & Operations Research*, *30*(5), 787-800.

[28] Burke, E. K., Gustafson, S., & Kendall, G. (2004). Diversity in genetic programming: An analysis of measures and correlation with fitness. *IEEE Transactions on Evolutionary Computation*, *8*(1), 47-62.

[29] Tamer F. Abdelmaguid & Maged M. Dessouky (2006) A genetic algorithm approach to the integrated inventory-distribution problem, International Journal of Production Research, 44:21, 4445-4464, DOI: 10.1080/00207540600597138

[30] Dolgui, A., Tiwari, M. K., Sinjana, Y., Kumar, S. K., & Son, Y. J. (2018). Optimising integrated inventory policy for perishable items in a multi-stage supply chain. *International Journal of Production Research*, *56*(1-2), 902-925.

[31] Glomb, L., Liers, F., & Rösel, F. (2022). A rolling-horizon approach for multi-period optimization. *European Journal of Operational Research*, *300*(1), 189-206.

[32] Hartleb, J. M. D. L., & Schmidt, M. (2020). A rolling horizon heuristic with optimality guarantee for an on-demand vehicle scheduling problem.

[33] Duinkerken, M., & Altasoy, B. (2018). *Integer programming*. Presentation, TU Delft.

[34] Gurobi Optimization. (2021). Retrieved April 2021, from https://www.gurobi.com/resource/mip-basics/

[35] Maaranen, H., Miettinen, K., & Penttinen, A. (2007). On initial populations of a genetic algorithm for continuous optimization problems. *Journal of Global Optimization*, *37*(3), 405-436.

Appendix A: Scientific paper

A robust approach to the distribution of double dose vaccines:

J. Bulters, dr. N. Pourmohammad-Zia, dr. F. Schulte, prof.dr. R. Negenborn

Department Maritime and Transport Technology,
Delft University of Technology, Delft, The Netherlands

Abstract. The distribution chain of two-dose vaccines by an air carrier (KLM cargo) with its practical features is modelled. An efficient solution approach for the model is applied that provides good quality solutions in reasonable computation time to be able to account for supply uncertainty. To circumvent the high computation time required for a commercial solver, the rolling horizon method is applied of which three approaches are compared. The best quality results in terms of computation time and gap to the optimal solution are found with the approach (i) using shifting and relaxation compared to (ii)the straight forward (myopic) approach and the approach (iii) with only relaxation. To study the effect of the supply uncertainty, where the exact probability distribution is unknown, robust optimization is applied. The uncertainty resides in a polyhedral uncertainty set. The influence of supply uncertainty is studied over a specified practical interval of the available supply. Over this interval the total costs proves to be sensitive to variation in the ordering amount. The shifting rolling horizon method is used to compute a flight and order plan given this uncertainty in the supply amount. To protect the solution against supply uncertainty a high level of conservatism proves to be required. This protection is obtained using Γ values ranging from 0.7 to 0.9. The level of protection is described as the fraction of feasible solutions out of 100 simulation runs. For a Γ of 0.7 approximately 85 % of the runs were feasible, for a Γ of 0.9 all runs were feasible. To obtain the protection level of a Γ value of 0.7 the costs are 12% higher compared to the nominal case while for a Γ of 0.9. the costs are increased 25%. The increase of protection comes at a price of higher cost.

Keywords: Vaccine supply chain, supply uncertainty, double dose vaccines, perishable, robust optimization, rolling horizon method, air carrier

1 Introduction:

Immunization is considered to be an effective way to prevent or decrease the spreading of contagious and transferable diseases such as the current SARS-Cov-2 virus. A controlled method for achieving immunity is by means of actively vaccinating the population. This requires a vaccine supply chain. The supply chain is the network of a vaccine manufacturer and its suppliers as well as other parties involved to produce, store, distribute and administer the product to the final user [1]. In the distribution network the product is transported from the production facility via the storage facilities eventually to the end consumer. What makes the supply chain of vaccines unique, is the set of stringent requirements which must be fulfilled to enable a safe administration at the point of use.

In the supply chain KLM forms the distributing link between the production facility and the national warehouses of several countries. The distribution process has the following structure: At the production facility the vaccines are filled in vials that contain multiple vaccination doses. Multiple vials are packaged in designated boxes. Once the vaccines have been produced and packaged accordingly the vaccine units will first be stored in a warehouse at the production facility. KLM can demand these vaccine units when KLM has the available resources to distribute or to store in KLM's own facility. This procedure is referred to as “the ordering of vaccines”. In the next step the vaccine units will be distributed by KLM to national warehouses at the airports of several locations. KLM has a fleet of several planes to transport the vaccines to the required destination and offers a storage facility.

The type of vaccine in this model is based on the characteristics of the Pfizer\BioNtech vaccine[2]. The vaccine is a two dose vaccine where the prescribed time between the first and the second dose is fixed to three weeks. This implies that it is required that the quantity of vaccines used to fulfil the demand of the first dose in a certain period at a certain destination also needs to be available three weeks later.

To ensure the efficacy of the vaccine, the vaccine should be maintained at a temperature of – 70 degrees Celsius. During transport and storage this is accomplished by using dry ice combined with a passive cool box (PCB) having a certain capacity for vaccine units that must be utilized. The dry ice of the passive cool box must be replaced every ten days, with a maximum of three dry ice (re)placements. This puts a maximum limit on the shelf life of 30 days coinciding with 3 periods in the model.

An extensive overview of the vaccine supply chain is presented in the literature review of Duijzer et al [3], comparing the supply chains in general and the vaccine supply chain in specific. Unique to the vaccine supply chain is the aspect of mass distribution under time pressure and accounting for perishability.

Optimization of the distribution of vaccines often focuses on the location of the distribution centres in different countries. The location of the distribution centres and points of use are based on routing decisions. An example of optimizing this part of the supply chain for vaccines is given by Georgiadis et al.[4]. Using a decomposition method based on distances of the distribution centres and points of use a mixed integer linear programming model is solved for a multi-echelon vaccine supply chain with a short planning horizon. A similar approach is followed by Tavana et al. [5] for equitable distribution of COVID-19 vaccines in developing countries, in specific in India. Both use short planning horizons of respectively 14 and 8 time periods, treat single dose vaccines and do not take uncertainty explicitly into account.

Decisions based on deterministic optimization models can be very sensitive to the input parameters and thus may result in wrong outcomes, see Ben-tal et al. [6]. To provide a better ground for decision-making and to reduce the probability of a severe outcome, uncertainty should be taken into account. The two main approaches used for optimization of uncertain models are robust optimization and stochastic optimization. They differ in the way the probability distribution of the uncertain data is treated. Stochastic optimization has two major drawbacks[6]. The first drawback is the need for knowledge of the exact probability distribution. The second drawback is that by taking the details from the probability distribution into account the solution can become computational intractable [6]. Due to the nature of the problem considered in this work the probability distributions of the considered parameters are assumed to be unknown. Therefore, from the two main approaches, the robust optimization method will be applied.

Robust optimization is a method developed to protect the schemes of distribution against uncertainties. Soyster [7] started exploring the field of robust optimization based on the worst-case scenarios. Ben-tal and Nemirovski[6] and Soyster[7] himself point out that this early method produces ultra-conservative solutions. Continuing the work of Soyster two less conservative main approaches are developed by the groups of respectively Ben-tal[7] and Bertsimas[8]. Ben-tal et al. use an ellipsoidal uncertainty set. This results in a non-linear set of equations. While for the method developed and Bertsimas et al., using a polyhedral uncertainty set, the set of equations remain linear. Bertsimas and Thiele [9] extend this robust optimization to a broad variety of networks to protect the distribution process against demand uncertainty. In [10] the method is extended to supply uncertainty. This approach will be applied here in combination with the rolling horizon method [11]

In this paper first the assumptions of the distribution process will be discussed. Based on the description of the distribution process and the assumptions the model will be formulated as a mixed integer programming model. For solving the model a solution method is selected, resulting in the rolling horizon method. After specifying the solution method, robust optimization will be applied to study the effect of the supply uncertainty on the distribution process.

2 Model assumptions and formulation:

The model is a multi-period inventory control problem with a specified planning horizon. Here it will be formulated in the form of a mixed integer program (MIP) with the goal to minimize the total costs, subject to the constraints resulting from the description in the introduction and the assumptions described below.

The assumptions and boundary conditions underlying the model are:

- The amounts of vaccines the number of vaccine units each containing a fixed amount of vaccines.
- The demand for the first dose is prescribed
- The demand for both doses must be fulfilled
- Shortages at the destination can only occur for the first dose.
- The second dose must be available in a prescribed time interval
- No shortages can occur at the KLM facility.
- Shortages must be fulfilled in the following period
- Shortages can only occur if there is no inventory at hand.
- The vaccines have a fixed shelf life
- Ordered vaccines have no lead time and arrive fresh at the KLM facility.
- Products are depleted on a FIFO basis from the respective inventories
- The dry ice in the passive cool boxes needs to be replaced at the end of each period.
- The dry ice can be replaced three times, thus the maximum stay of a box is four periods.
- All planes have a fixed capacity in terms of vaccine units.
- The transportation time is negligible compared to the period length.

The goal of the model is to minimize the cost involved for KLM.
Furthermore the model will determine:

- The total cost for the distribution process
- The amount of vaccines which need to be ordered in each period
- The amount of transport in each period to each destination

Results will be generated for the amount of:

- inventory at KLM and at different locations in each period.
- vaccine units transported each period
- shortage for the first dose at each destination location in each period
- demand for the second dose for each destination in each period
- dry ice required each period
- flights in each period to each destination

In the presentation of the model the indexes of the variables belong to the following sets:

Set notations:

$\boldsymbol{t} \in \boldsymbol{T}$	t is a time period of the planning horizon set $\boldsymbol{T}$. $\boldsymbol{T} = \{\mathbf{1}, \dots, \boldsymbol{T}\}$
$\boldsymbol{l} \in \boldsymbol{L}$	l is a location belonging to the set of all locations $\boldsymbol{L}$. $\boldsymbol{L} = \{\mathbf{0}, \mathbf{1}, \dots, \boldsymbol{L}\}$
$\boldsymbol{p_d} \in \boldsymbol{\mathcal{P}d}$	p_d is the product age of the vaccine, $\boldsymbol{\mathcal{P}d}$ is the set of product ages.
$\mathbb{Z}^{+}$	Set of all positive real integers including 0

The assumptions and the description are formalized by the following set of equations.

Total costs:

To calculate the total costs of the model a cost function is defined in equation (1). This cost function consist of the elements: cost of respectively: ordering, inventory at the KLM facility, (penalty for) shortage at the destinations, wasted vaccines and transport.

$$\text{Total cost} = \overbrace{\sum_{t \in H} v_t * K}^{Order\ cost} + \overbrace{\sum_{t \in H} \left((h^A + re + c_{di}) * I_t^{A,KLM} \right)}^{Inventory\ cost} + \overbrace{\varphi * \sum_{t \in H} \sum_{l \in L} S_t^l}^{Shortage\ cost}$$
$$+ \overbrace{\omega * \sum_{t \in H} (C_{cb} * Waste_{A.t} + \sum_{l \in L} Waste_{t,l})}^{Wastage\ cost} + \overbrace{\sum_{t \in H} \sum_{l \in L} y_{t,l} * d_l * tc}^{Transport\ cost} \quad (74)$$

The total costs are subject to the following constraints:

Constraints on ordering:

$$u_t - M * v_t \le 0, \; M \gg u_t, \forall t \in H \quad (75)$$
$$u_t \le Q_t^{ext}, \forall t \in H \quad (76)$$

Inventory constraints:

$$I_{t,p_d}^A = u_t^A - \sum_{l \in L} w_{t,l,pd}^A, \qquad for\ p_d = 0 \quad (77)$$

$$I_{t,p_d}^A = I_{t-1,p_{d-1}}^A - \sum_{l \in L} w_{t,l,p_d}^A, \quad (78)$$
$$for\ 1 < p_d < p_{d,max}$$

$$I_{t,p_d}^A = I_{t-1,p_{d-1}}^A - \sum_{l \in L} w_{t,l,p_d}^A - Waste_{A,t} = 0, \quad (79)$$
$$for\ p_d = p_{d,max},$$

$$I_{t,p_d}^l = C_{cb} * w_{t,l,pd}^A - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2}, \quad (80)$$
$$for\ p_{di} = 0$$

$$I_{t,p_d}^l = I_{t-1,p_d-1}^l + C_{cb} * w_{t,l,pd}^A - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2} \quad (81)$$
$$, for\ 0 < p_d < p_{d.max}$$

$$I_{t,p_d}^l = I_{t-1,p_d-1}^l + C_{cb} * w_{t,l,pd}^A - d_{t,p_d}^{l,1} - d_{t,p_d}^{l,2} - Waste_t^l = 0 \quad (82)$$
$$, for\ p_d = p_{d.max}$$

$$\sum_{p_d \in \mathcal{P}_d} I_{t,pd}^A, \qquad \le C_A \quad (83)$$

$$\sum_{p_{di}=0}^{p_{di,max}} I_{t,p_d}^l, \qquad \le C_{L,l} \quad (84)$$

$$DI_{t,required} = \sum_{p_d=0}^{p_{d,max}} (DIF_A * I_{t,p_d}^A) \quad (85)$$

$$DI_{t,required} \le DI_{t,available} \quad (86)$$

Constraints on the demand:

$$D_t^l = D_t^{l,1} + D_t^{l,2} + S_{t-1}^l \quad (87)$$
$$D_t^{l,2} = D_{t-r_t}^{l,1} - S_{t-r_t}^l, t \le r_t, \forall t \in H, \forall l \in L \quad (88)$$
$$D_t^{l,1} - S_t^l = \sum_{p_{di}=0}^{p_{d,max}} d_{t,p_d}^{l,1}, \quad (89)$$
$$D_t^{l,2} = \sum_{p_{di}=0}^{p_{d,max}} d_{t,p_d}^{l,2} \quad (90)$$
$$S_t^l - q_t * M \le 0 \quad (91)$$

$$\sum_{p_{di}=0}^{p_{di,max}} I_{t,p_d}^l - (1 - q_t) * M \le 0 \quad (92)$$

Transport constraints:

$$T_{t,l} = \sum_{pd \in Pd} C_{cb} * w_{t,l,pd}^A \quad (93)$$
$$0 \le y_{t,l} \le f_{flight,l}, \forall t \in H \text{ and } \forall l \in L \quad (94)$$

$$T_{t,l} \le y_{t,l} * C_{pf} \quad (95)$$

Elaborating briefly on the constraints: The total costs for the distribution of vaccines are presented in equation (1). The choice of the model to order in a period, is determined with the Big-M method[12] given by equation (2). The order amount cannot exceed the available supply at the external supplier is stated by equation (3). To account for the perishability of the vaccines the inventory levels for different product ages are traced by equation (4-6) for the KLM facility and (7-9) for the destinations. In (7-9) the age-specific demand fulfilment for both doses is taken into account. The storage facilities cannot store more than the capacity allows as shown in equation (10) and (11). The amount of dry ice required for storage at the KLM facility is determined in equation (12) and cannot exceed the upper bound imposed by equation (13). Based on the double dose requirement of the vaccines and the allowed shortage for one period, the periodic demand for a destination is given by equation (14). The fixed period between the doses and the possible shortage determines the required amount for the second dose (15). The demand for both doses can be fulfilled with vaccines independent of the product age (16) and (17). Equation (16) shows the relation between the fulfilment of the first dose and the shortage. To compensate for the possibility in equation (16) to have inventory at hand while incurring a shortage, equation (18) and (19) are imposed. Equation (20) determines the transport amount from KLM to the respective destination in a given period taking the product ages of the vaccines into account. The number of flights available to a destination is bounded by the flight frequency as shown in equation (21). The transport capacity based on the available flights and the capacity of a flight bounds the transport amount (22).

This set of equations constitutes the model used to study the influence of supply uncertainty on this distribution process.

3 Methodology:

To study the influence of supply uncertainty, the formalism of robust optimization[8] will be presented and applied to the distribution process.

Robust optimization

The generic optimization model consists of a cost function subject to a set of constraints and boundary conditions. The model can be formulated as:

$$\text{Min } \mathbf{c'x} \tag{96}$$

$$subject\ to:$$

$$\boldsymbol{Ax} \leq \boldsymbol{b} \tag{97}$$

For robust optimization the uncertainty resides in the matrix **A** that consists of elements a_{ij}.

Robust optimization has to find a solution which is feasible for all values of the uncertain parameters. With robust optimization an unknown uncertainty probability distribution can be taken into account. This is accomplished by specifying an interval of the uncertainty that is given as: $[\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$ in which the uncertain parameter a_{ij} resides. $\bar{a}_{ij}$ is used to denote the nominal value and $\hat{a}_{ij}$ is the ‘half length’ value.

Less conservative approaches [6][8] use the uncertainty interval to determine the scaled deviation of the uncertain parameter. The scaled deviation is determined as:

$$z_{ij} = \frac{a_{ij} - \bar{a}_{ij}}{\hat{a}_{ij}} \tag{98}$$

The value of the scaled deviation lies in the interval of [-1,1]. By imposing a boundary condition on the scaled deviation conservative solutions can be avoided. In this work the polyhedral uncertainty set as described by Bertsimas et al.[8] is used.

The generic application of the polyhedral uncertainty set works as follows:

For each row of the product **Ax** that contains uncertainty it should hold that:

$$\sum_j \bar{a}_{ij} x_{ij} + \sum_j z_{ij} \hat{a}_{ij} x_{ij} \leq b_i \,, \forall i \tag{99}$$

The idea of robust optimization is to maximize the uncertainty term $\sum_j z_{ij} \hat{a}_{ij} x_{ij}$ under the conditions formulated by Bertsimas et al [8] in equation (27).

$$\sum_j \bar{a}_{ij} x_{ij} + \max\,(\sum_j z_{ij} \hat{a}_{ij} x_{ij}) \leq b_i \,, \sum_j z_{ij} \leq \Gamma_i, \forall i \tag{100}$$

The total allowed sum of scaled deviations of all uncertain parameters for each row i is limited by the parameter Γ_i, this Γ_i is called the ‘budget of uncertainty’.

This constraint cannot be solved directly due to the maximization term. Using the principles of duality theory, proven by Bertsimas et al.[8] the maximization term can be converted into a minimization term.

The implementation of the dual formulation is implemented in the original uncertain equation (27) and results in the robust counter part of the constraint:

$$\sum_j \bar{a}_{ij} x_{ij} + q_i * \Gamma_i + \sum_j p_{ij} \ \leq b_i \,, \forall i \tag{101}$$

$$q_i + p_{ij} \geq \hat{a}_{ij} y_i \tag{102}$$

$$-y \leq x \leq y \tag{103}$$

$$q_i \geq 0, p_{ij} \geq 0, y \geq 0 \tag{104}$$

The $q_i * \Gamma_i + \sum_j p_{ij}$ -term protects the solution against variation of the uncertain parameter a_{ij} in a defined interval.

Application of robust optimization:

The available production capacity at the supplier, and thus the available supply is allowed to fluctuate. This uncertainty will have an effect on the total costs. The implementation of robust optimization starts with the constraints on the uncertain parameter. From the model formulation the constraint which requires a robust counterpart is equation (3):

$$u_t \leq Q_t^{ext} \tag{3}$$

As Q_t^{ext} is uncertain, Q_t^{ext} can be written as:

$$Q_t^{ext} = \bar{Q}_t^{ext} + z_t * \hat{Q}_t^{ext} \tag{105}$$

Where $\bar{Q}_t^{ext}$ is the nominal value of Q_t^{ext} and $\hat{Q}_t^{ext}$ is the half length of the interval in which Q_t^{ext} resides.

From initial experiments it can be seen that using a production capacity higher than the total transport capacity per period has no influence on the total costs. Hence, the interval in which the uncertainty resides is one sided [10] and defined as:

$[\bar{Q}_t^{ext} - \hat{Q}_t^{ext}, \bar{Q}_t^{ext}]$.

From the theory of Bertsimas et al. [8] the constraints regarding the order amount (u_t) can be written as:

$$u_t - \bar{Q}_t^{ext} + q_t * \Gamma_t + \sum_j p_{tj} \le 0 \quad (106)$$

$$q_t + p_{tj} \ge \hat{Q}_t^{ext} \quad (107)$$

$$q_t \ge 0, p_{tj} \ge 0, \quad (108)$$

Implementing the robust counterpart completes the model. The size of the model requires a solution method that provides good quality results in reasonable computation time.

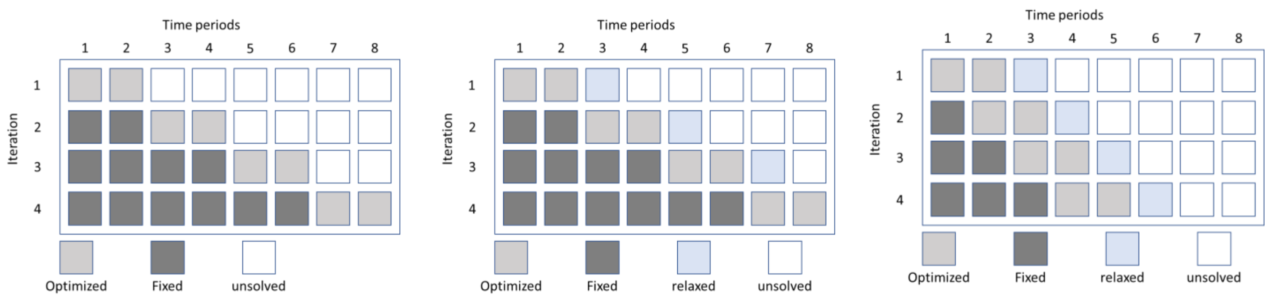


Figure A-1: from left to right: Straight forward RH-approach, considering future periods in the RH-approach and overlapping the consecutive solved time periods.

4 Solution method

First the genetic algorithm was modified [13] creating alternative mutation operators and selection process. Although promising, the speed and accuracy of this method are not on the level yet to perform the robust analysis. Instead as an alternative the rolling horizon method is used here.

Rolling horizon is a method which solves the model by dividing the considered planning horizon into smaller time windows. This way the computation time can be decreased by consecutively solving the reduced planning horizons, using the solution of the previous time window as the input for the following time window. The method runs until the total planning horizon is reached.

Three approaches of the rolling horizon method will be discussed. All three approaches share the decomposition of the total planning horizon. The differences between the approaches are on how the decomposed time windows are solved. The straight forward method solves the model by using the subsequent time windows for the given decomposition size. The second method considers relaxed future periods in solving the decomposed problem. In this method in each iteration an additional time period is added, but for this time period the integrality constraint is relieved. The integrality for this time period is imposed in the next iteration of this method. The third method is an extended version of the second method. The difference is in the addition of overlapping of the time periods subject to integrality constraints. Examples of the three different approaches are depicted in figure 1.

The variables involved for solving the rolling horizon approaches are the number of time periods which are optimized with the integrality constraints (η), the extension of this time window for the number of periods which are relaxed (μ) and the shift amount for overlap (step).

Evaluating the rolling horizon methods shows that the approach with relaxation and shifting performs superior to the first two approaches [13]. Therefore this method will be used.

5 Results:

Robust optimization is applied on the situation with four destinations, a constant demand rate of 100 for all destinations and a planning horizon of 21 weeks. The model will be solved using the rolling horizon method including relaxation and shifting. The time window of 12 and a shift amount of 2 are used. When the available supply is larger than the transport capacity the total costs are not affected, see fig 2. But the total costs are largely influenced by the available supply in the range of interest i.e. from the minimum amount required up to the transport capacity. Consequently the supply uncertainty in this range affects the total cost strongly, so this interval of the available supply, will be used as the uncertainty interval for the robust optimization.

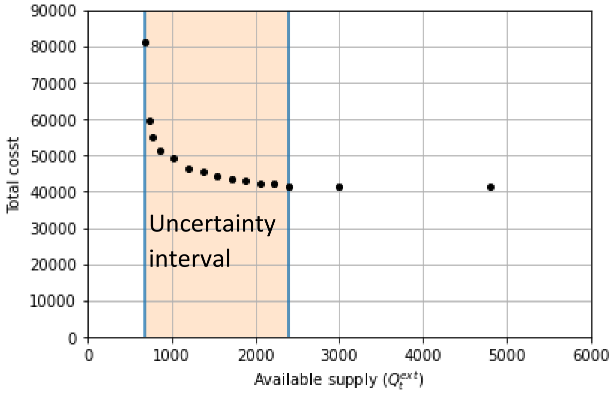


Figure A-2: Influence of the available supply on the total costs. The interval ranging from the minimum amount required up to the transport capacity is used as the uncertainty interval for the robust optimization.

The optimal robust solutions for different values of Γ are determined. The values of Γ range from 0 to 1. The solution for each Γ value provides the total costs, the order and the transport amounts to each destination combined with the number of flights and the periods in which an order occurs. The order moments and the number of flights can be gathered in a flight and order plan.

To test for which value of gamma the solutions are most protected against the supply uncertainty, simulations are performed. The rolling horizon method is used for this. In the simulation the available supply for each separate time period is randomly generated from the uncertainty interval assuming a uniform probability distribution. The flight and order plans from the solutions from the robust solutions will be used as input for the simulation. If the order amounts and transport amounts from the robust solutions would be combined with the random values of the available supply, most probable only the most conservative robust solution leads to a feasible solution. The order and transport amounts are determined from the simulation.

The simulations are tested for 100 sequences of available supply. The total costs found with the simulations are compared with the total costs found in the robust solutions.

In figure 3 the number of feasible solutions out of the 100 simulations is shown in dependence of Γ. With increasing conservative values of Γ, more feasible solutions can be found. For values which are less than Γ = 0.5 very few, much less than 15, feasible solutions are found.

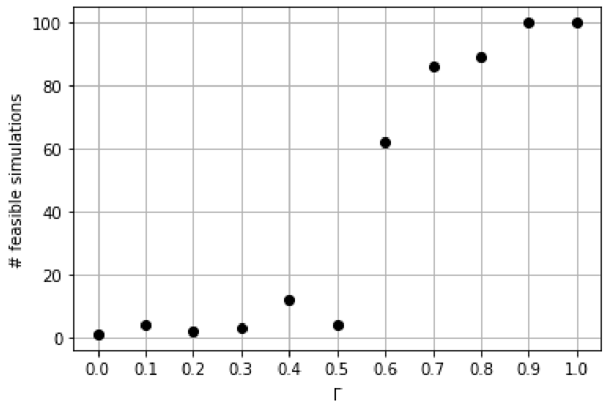


Figure A-3: Number of feasible solutions resulting from the 100 simulations with the flight and order plans of the respective Γ -values.

In Figure A-4 for the range of values of Γ the total costs resulting from the robust solutions are compared with the average of the total costs of the feasible fraction of the 100 simulations.

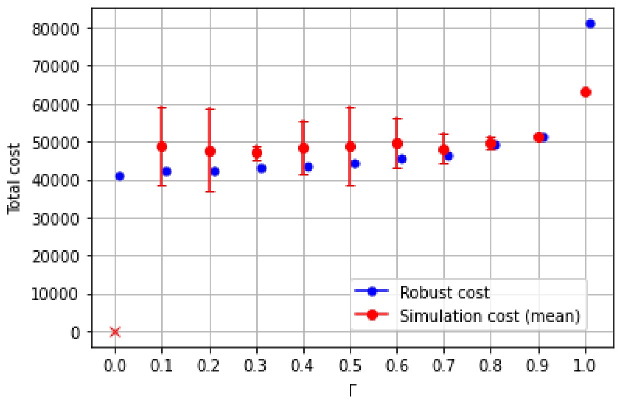


Figure A-4: The total costs of the robust solutions versus the Γ -value. These are compared with the average total costs calculated from the concrete flight and order plans fitting the respective Γ -value. Only the feasible solutions out of the 100 runs are used, for the numbers of feasible solutions see Figure A-3. The bars indicate two times the standard deviation. For a Γ -value of 0 (x) no feasible solution are obtained with the simulations. (The blue symbols are shifted 0.01 to the right for readability)

The results of the total costs show a decreasing gap between the robust costs and costs resulting from the simulation with increasing value of Γ. Note that for lower Γ values the small fraction of feasible solutions result in a large scatter of the total costs found with the simulations, indicated in the graph with bars that represent two times the standard deviation. For a Γ in the range of 0.7 to 0.9 the costs from the simulations and the robust costs are nearly the same with only very little scatter, with slightly increasing total costs. The best robust solution is found in this range. 0.9 gives a feasible solution for all simulations. Lowering Γ to 0.7 results in a somewhat lower fraction of feasible solutions and somewhat more scatter on the total costs. This comes at the advantage of slightly lower total costs for the robust solution. For the worst-case scenario (Γ is equal to 1) the robust costs are higher than the simulated costs. This is caused by the constraints on the available supply in the robust solution.

6 Discussion:

For the robust optimization the method described by Bertsimas et al. [8] has been used. This method uses a Γ value to protect the solution of the model against the supply uncertainty up to a certain degree. The robust solutions for different values of Γ are determined. The robustness of each /gamma value is tested by performing simulations using the resulting flight and order plans. The results show that a high value of Γ is required to protect the solution against supply uncertainty. Feasible solutions are obtained in the range of Γ values from 0.6 up to 1, while for lower values of Γ only small fractions of the simulations lead to feasible solutions. Based on the results of the robust optimization combined with the costs found with the simulations a Γ in range of 0.7 and 0.9 show a reasonable to good protection against the supply uncertainty. In this range the fraction of feasible solutions increases. This robustness does come at a price; The cost for the Γ value of 0.9 are 4% higher than the costs for the Γ value of 0.8 and 10% higher than the costs for /gamma equal to 0.7. All total costs are higher than the nominal case without protection for uncertainty. Choosing for a robust approach with a Γ value of 0.8 results costs that are 16% higher than the nominal case. Needless to say that opting for this nominal case will expose the company to a very high probability of not satisfying the boundary conditions.

There is a high level of conservatism required. This might be influenced by the probability distribution used to generate the values of the available supply. Notwithstanding that the method developed by Bertsimas [8] is intended to be insensitive to the details of the probability distribution. Intuitively one might guess that robustness of the solutions could be sensitive to some details e.g. skewed probability distributions towards the transport capacity might allow for less conservative values of Γ. This influence of the probability distribution is not studied here, detailing on this aspect may require future attention.

7 Conclusion and further research:

In this paper the distribution process of the vaccine supply chain is studied. A literature study shows the complexity of the vaccine supply chain. The influence of supply uncertainty on the distribution process of vaccines based on a mixed integer programming formulation is addressed. Of the three rolling horizon method approaches, the one that includes relaxation and shifting performs the best and results high quality solution at reasonably low computation times. This method is therefore used for the analysis of robust optimization that takes supply uncertainty into account following the approach of Bertsimas et al.[8].

The results of the robust optimization show that a high level of conservatism is required to protect against uncertainty. Expressed in Γ values this ranges from 0.7 to 0.9 showing increasing protection. The protection is described as the fraction of feasible solutions of 100 simulation runs. For a Γ of 0.7 85 % of the runs were feasible, for a Γ of 0.9 100% of the runs were feasible. The increase of protection comes at a price of higher cost. To obtain the protection level of a Γ value of 0.7 the costs are 12% higher compared to the nominal case and 25% for a Γ of 0.9.

Some recommendations for future work: The robust analysis is performed with a uniform distribution for the uncertain parameter. Analysing the effect of different, in specific very skew, distributions might be worthwhile.

The results of the robust optimization show to be conservative. It would be interesting to develop a robust optimization method which does not require Γ for each period but where alternatively the Γ limits the sum of scaled deviations over all the time periods.

References:

[1] Chopra, S., & Meindl, P. (2013). *Supply chain management*. Essex, England: Pearson Education.

[2] Pfizer - Fact sheet. (2021). Retrieved March 7, 2021, from https://cdn.pfizer.com/pfizercom/Pfizer_PGS_COVID-19_Factsheet_071122.pdf

[3] Duijzer, L. E., Van Jaarsveld, W., & Dekker, R. (2018). Literature review: The vaccine supply chain. *European Journal of Operational Research*, *268*(1), 174-192.

[4] Georgiadis, G. P., & Georgiadis, M. C. (2021). Optimal planning of the COVID-19 vaccine supply chain. *Vaccine*, *39*(37), 5302-5312.

[5] Tavana, M., Govindan, K., Nasr, A. K., Heidary, M. S., & Mina, H. (2021). A mathematical programming approach for equitable COVID-19 vaccine distribution in developing countries. *Annals of Operations Research*, 1-34.

[6] Ben-Tal, A., & Nemirovski, A. (2002). Robust optimization–methodology and applications. *Mathematical programming*, *92*(3), 453-480.

[7] Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, *21*(5), 1154-1157.

[8] Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, *52*(1), 35-53.

[9] Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations research*, *54*(1), 150-168.

[10] Chu, J., Huang, K., & Thiele, A. (2019). A robust optimization approach to model supply and demand uncertainties in inventory systems. *Journal of the Operational Research Society*, *70*(11), 1885-1899.

[11] Glomb, L., Liers, F., & Rösel, F. (2022). A rolling-horizon approach for multi-period optimization. *European Journal of Operational Research*, *300*(1), 189-206.

[12] Duinkerken, M., & Altasoy, B. (2018). *Integer programming*. Presentation, TU Delft.

[13] Bulters, J., Pourmohammad-Zia N., Schulte & F., Negenborn N. (2022). The distribution process of multi-dose vaccines, *Master thesis*

Appendix B: Additional intermediate support results on genetic operators

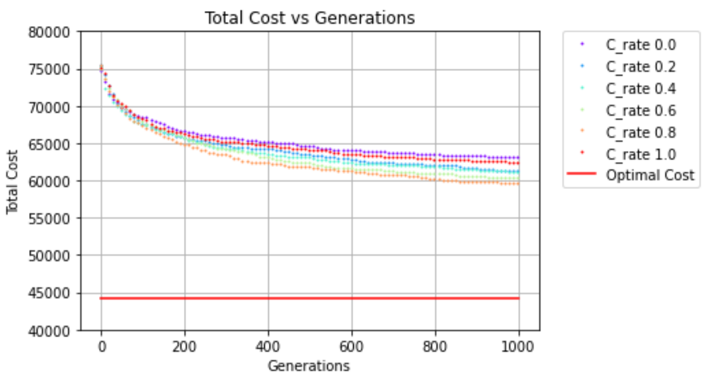
The parameters for the cross-over rate, mutation rate, elitism and diversity will be varied. The population size and the number of generations are the same for all experiments. The population size is 50 chromosomes and the number of generations is 1000.

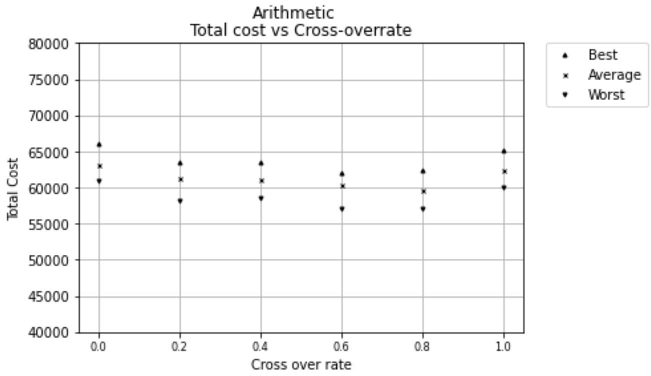
Recombination strategies:

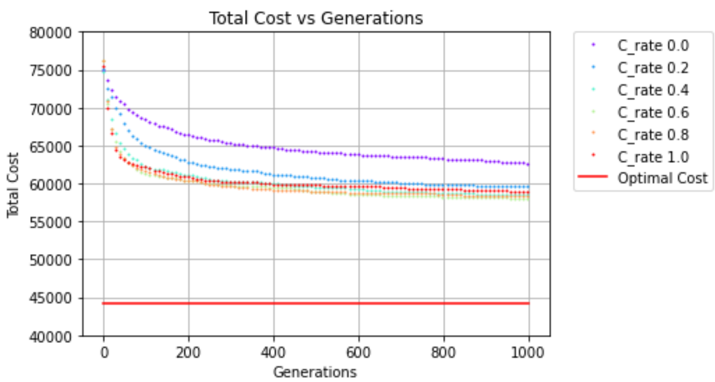
For each recombination strategy the cross-over rate is studied. The figures below show the average total cost for 10 runs for each cross-over rate (C_rate). The cross-over rate is varied between 0 and 1 with increments of 0.2. The results will be presented in two figures for each recombination strategy. The first figure shows the average cost of the ten runs per generation and compared with the optimal solution. The second figure shows the found minimal cost for each cross-over rate.

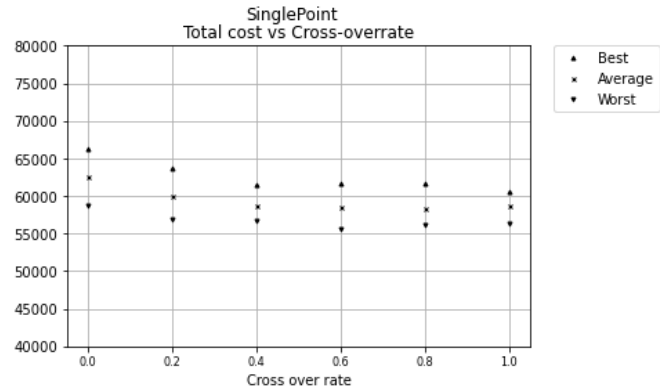
Arithmetic:

Single Point:

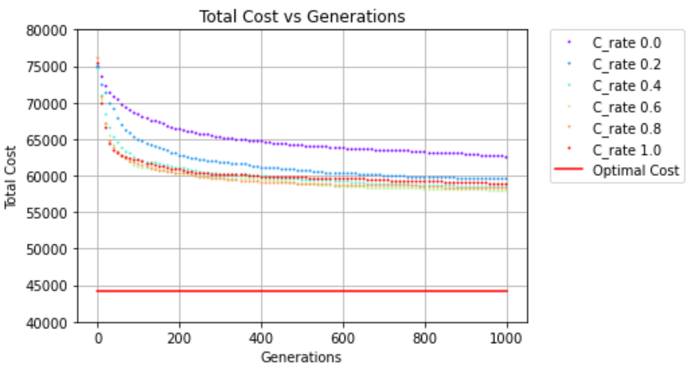


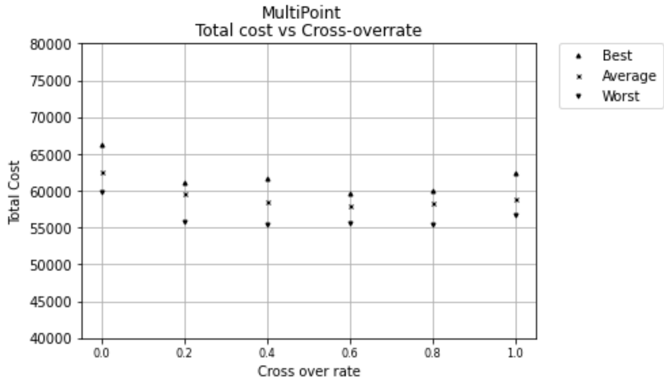






Multi point:

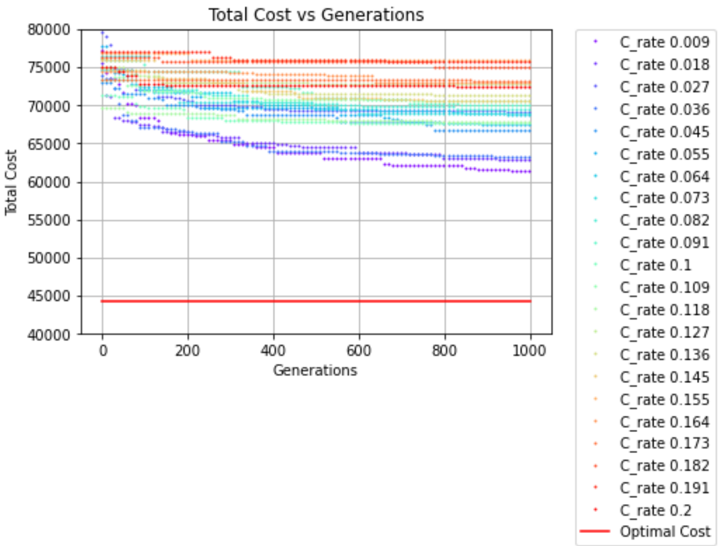


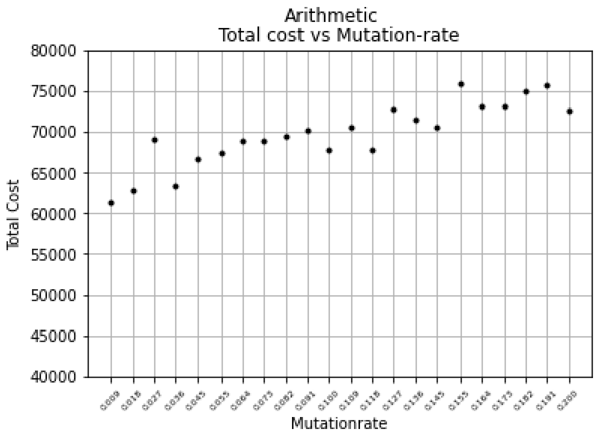


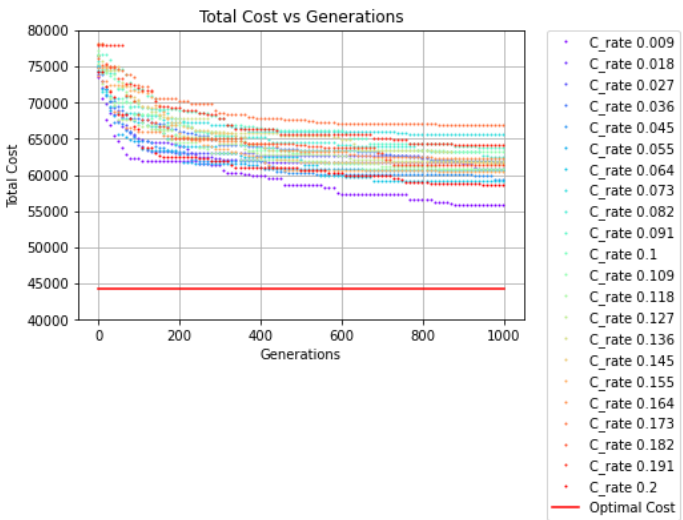
The results show typical behaviour of the GA where for all strategies the GA gets stuck in a local optimum. The results show a higher convergence speed for multi-point and single point cross-over

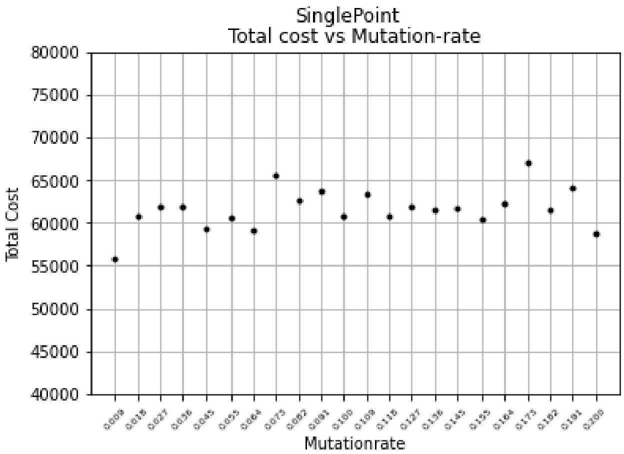
<u>Mutation rate combined with the different cross-over strategies:</u>

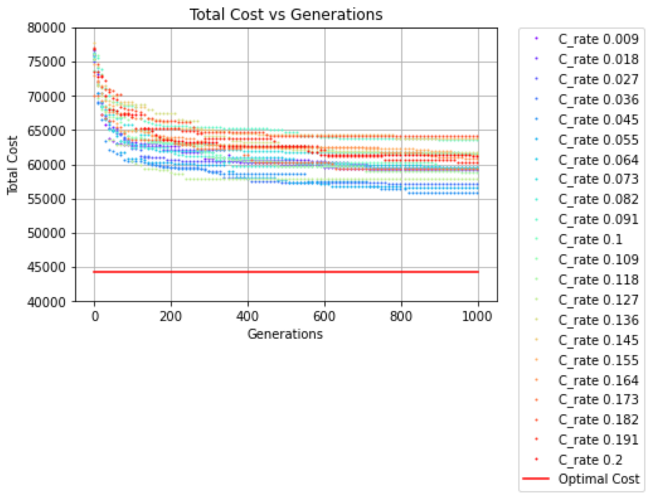
For each recombination strategy the mutation rate is studied. The figures below show the average total cost for 10 runs for mutation rate. Often used mutation rates found in literature vary between [1/#genes – 0.2]

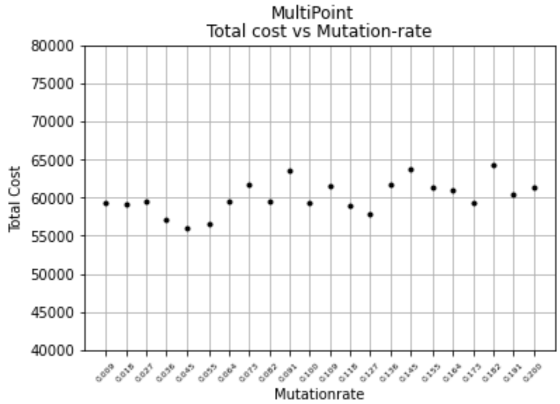








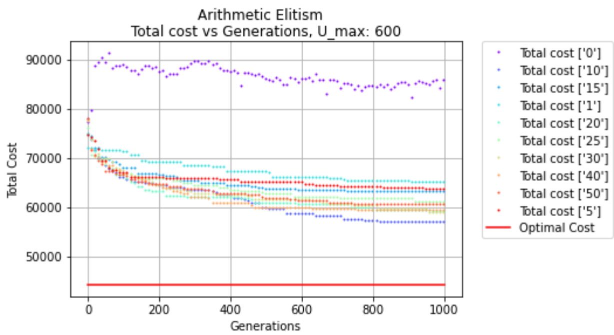


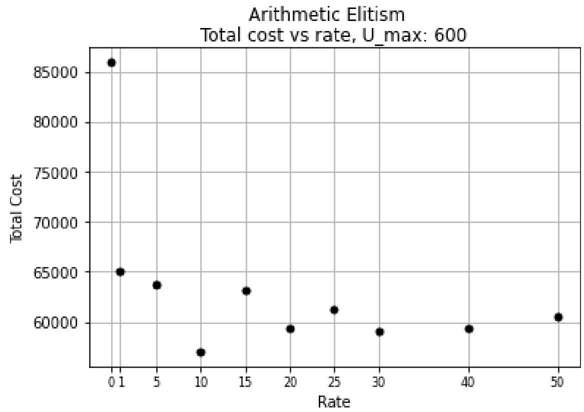


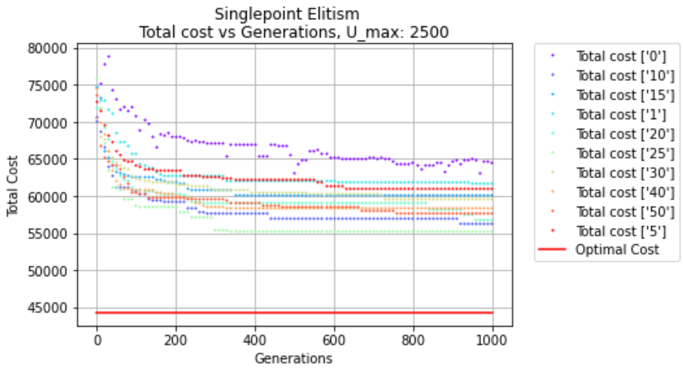
The results with varying the mutation rate show non typical behaviour for the arithmetic cross-over and the single point cross-over and to less extend for the multipoint. The cost increase with the increase of mutation rate.

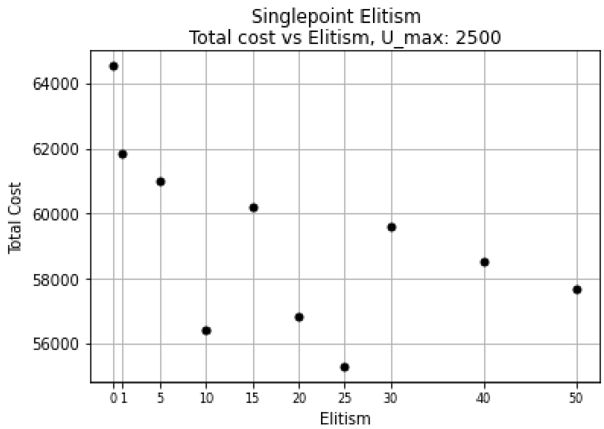
<u>Elitism:</u>

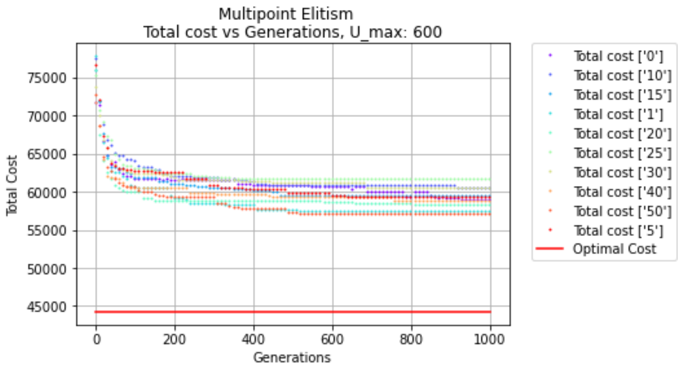
What is the influence of elitism (How much good chromosomes are taken to the next generation)

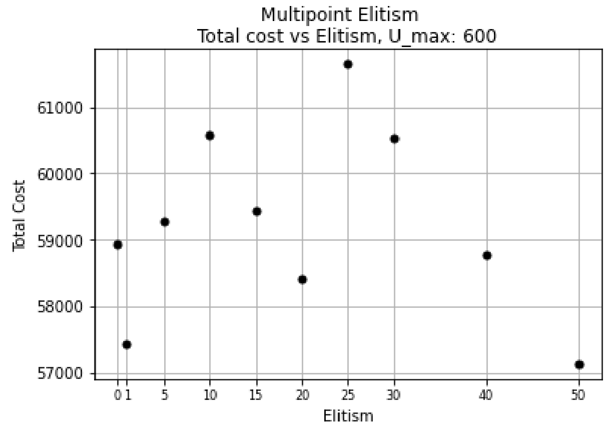








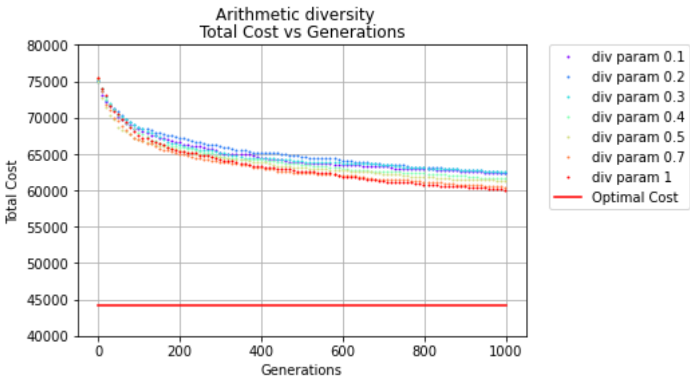


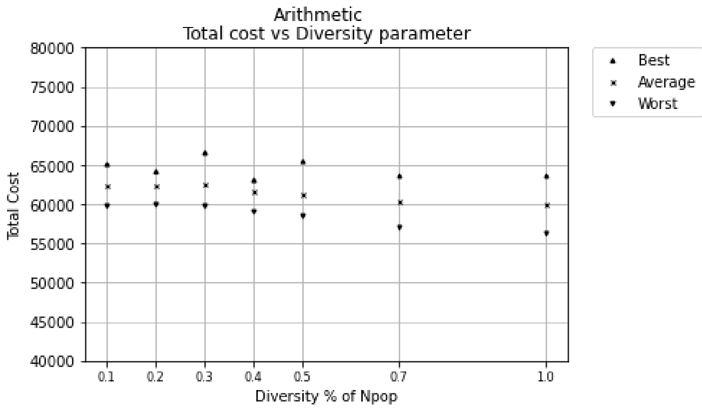


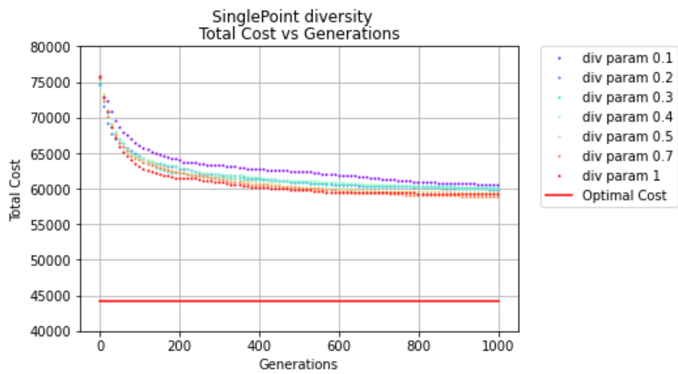
The need for elitism is evident from the figures except for the multipoint cross-over. This cross-over strategy is less influenced by the elitism.

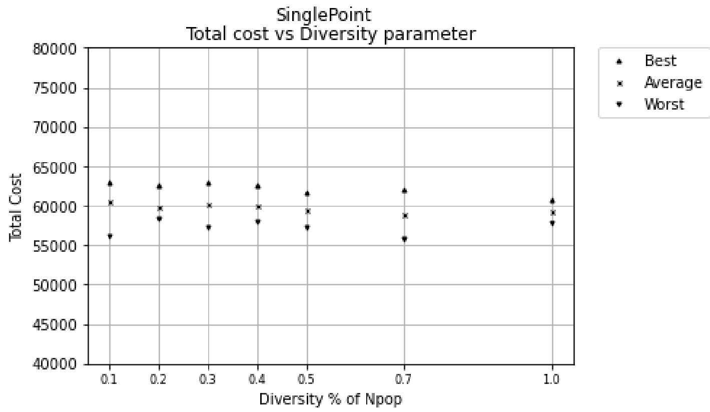
Diversity:

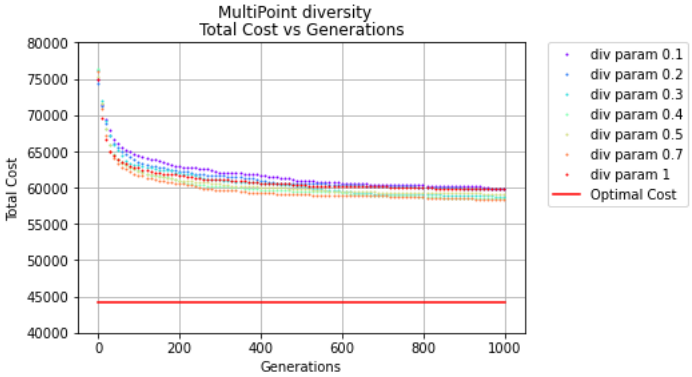
Adding diversity to the population is a method to aid the GA escaping local optima [21]. Diversity initially is based on the cost. The following figures show the percentage of best performing individuals of the current population. Diversity is implemented as the X% best unique chromosomes and (100-X)% unique chromosomes are selected at random from the current population which do not belong to the X% best unique chromosomes.

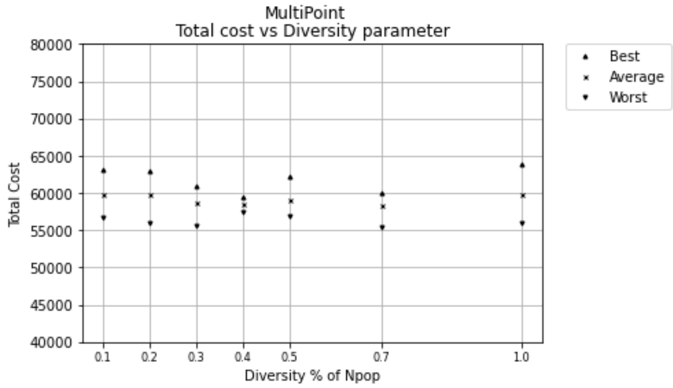










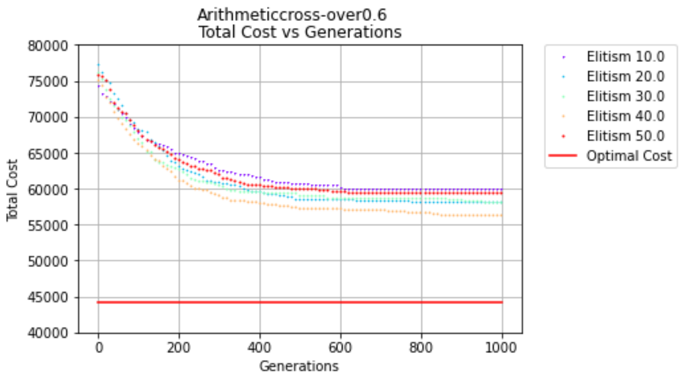


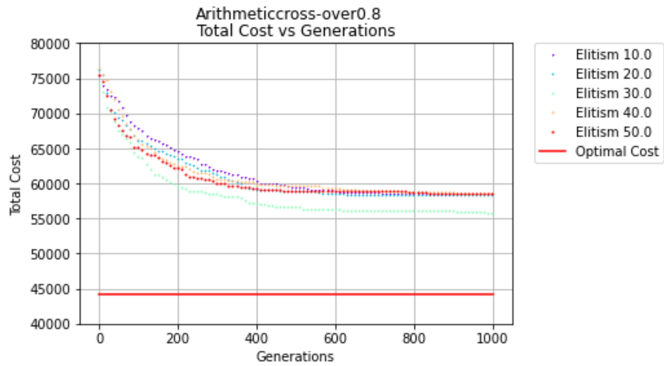
For the arithmetic cross-over adding diversity has a worse effect. Where increasing the number of best solutions leads to better results. For single point and multi-point cross-over adding diversity from 0.4 has a positive influence on the convergence

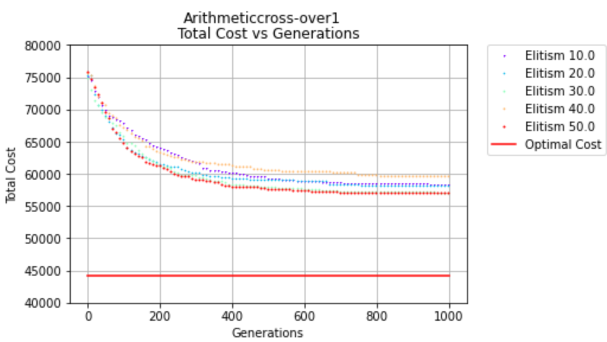
Results of the improved mutation operator

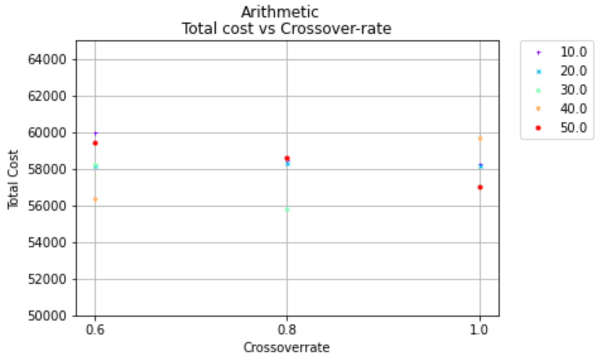
As only a single gene is selected for mutation, the mutation rate has no influence. From this the parameters which will be varied are the cross-over rate and the elitism parameter. This is done as the functioning of the algorithm cannot be assigned to a single parameter or operator. The cross-over rate is varied between 0.6 and 1, where the elitism is set in steps of 10 % starting at 10% to 100%.

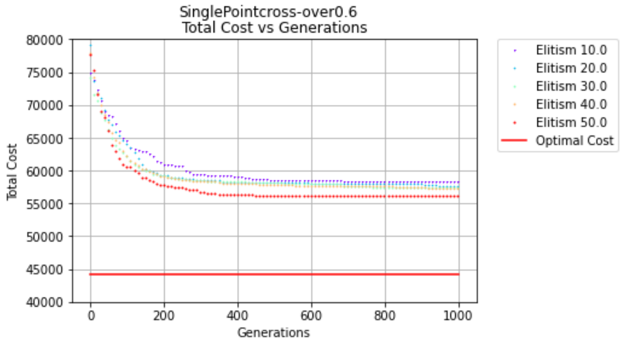
Starting with the arithmetic cross-over followed by the single point and multi-point cross-over:

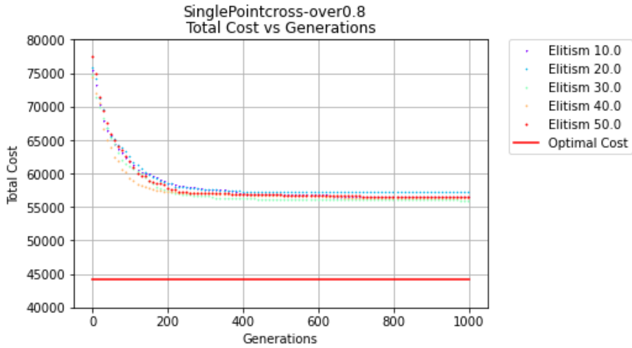


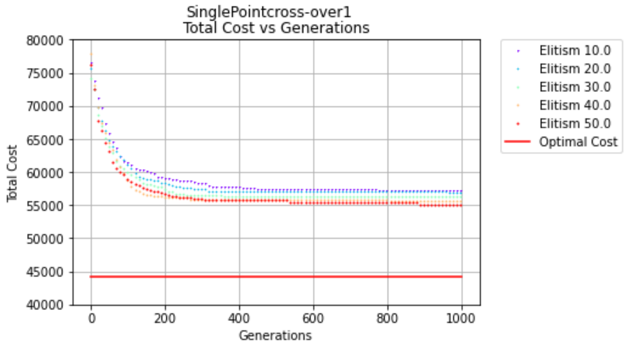


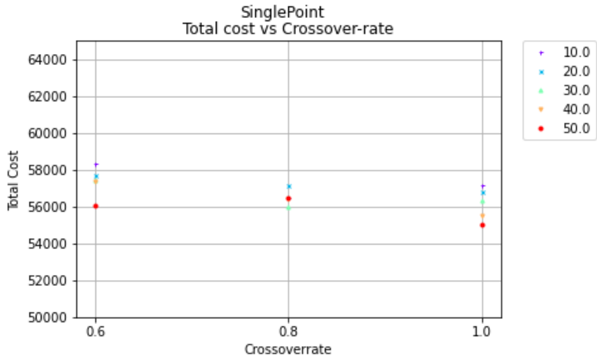


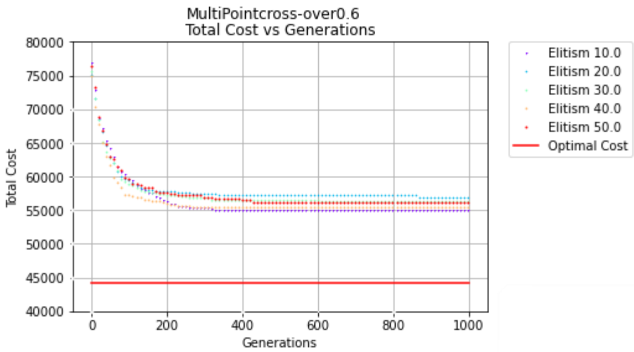


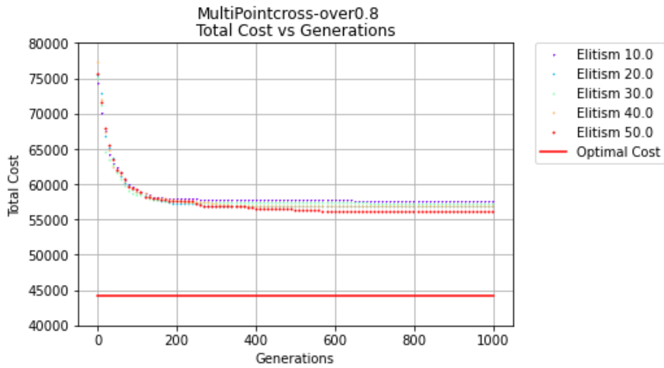

SinglePointcross-over0.6
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost

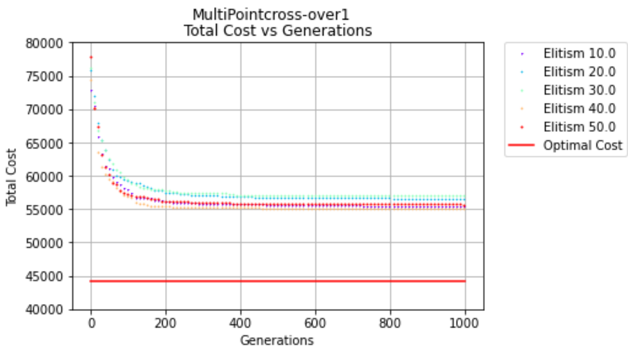

SinglePointcross-over0.8
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost

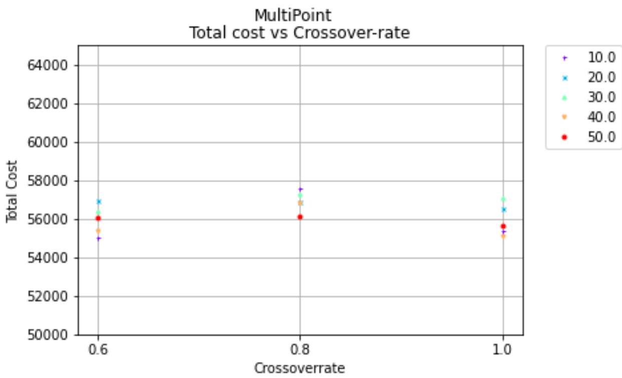

SinglePointcross-over1
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost


SinglePoint
Total cost vs Crossover-rate
Total Cost
Crossoverrate
10.0
20.0
30.0
40.0
50.0


MultiPointcross-over0.6
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost


MultiPointcross-over0.8
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost


MultiPointcross-over1
Total Cost vs Generations
Total Cost
Generations
Elitism 10.0
Elitism 20.0
Elitism 30.0
Elitism 40.0
Elitism 50.0
Optimal Cost


MultiPoint
Total cost vs Crossover-rate
Total Cost
Crossoverrate
10.0
20.0
30.0
40.0
50.0