Optimal positioning and capacities of new turnout facilities for the Surveillance and Protection Department in The Netherlands

Loog, Stan

Award date:
2024

Department of Industrial Engineering
Operations Management and Logistics

Optimal positioning and capacities of new turnout facilities for the Surveillance and Protection Department in The Netherlands

Master Thesis

S. (Stan) Loog
1336711

1st supervisor: L. Schlicher
2nd supervisor: L. Martin
DKDB supervisor: R. Zandbergen

Abstract

Our free society faces different kinds of threats, some of them with terrible consequences. Criminals put pressure on our society by intimidating or even attacking inhabitants of the Netherlands. In the Netherlands, more and more people are under threat. Because the capacity to handle this protection is lagging behind, the Dutch government stimulates initiatives to make better use of the resources available. This thesis investigates such an initiative at the DKDB, the organization responsible for the personal security of royalty's, politicians, diplomats, and other designated persons in the Netherlands. Currently, the DKDB operates from one turnout facility, which is a facility where agents have to show up before and after their shifts, in The Hague. This implies that DKDB agents spent a lot of working time traveling. This thesis investigates savings (in time) when multiple turnout facilities are opened using techniques from Facility Location Problems (FLP). The challenge in this thesis is that the number and location of threatened people are stochastic for the future. We developed a two-stage solution method to determine the optimal locations and capacity distribution among these turnout facilities. Numerical experiments demonstrate the robustness of these solutions. The Key Performance Indicators (KPIs) show that when opening one to six turnout facilities, up to 73.573% savings can be achieved in terms of total travel distances, resulting in a capacity increase where 111.7 more people can have one-on-one protection annually.

Management summary

Introduction

Our free society faces different kinds of threats, some of them with terrible consequences. Criminals put pressure on our society by intimidating or even attacking inhabitants of the Netherlands. In the Netherlands, the DKDB is responsible for the personal security of royalty's, politicians, diplomats, judges, and other designated persons. However, because of the increasing pressure due to the increasing threats of organized crime in the Netherlands, the DKDB lacks personnel for operational tasks.

Problem description

Currently, the DKDB operates with one turnout facility in the Netherlands which is located in The Hague. A turnout facility is a facility where DKDB agents have to arrive before and after their shifts. It seems that some agents drive a lot of kilometers every day. If there were multiple turnout facilities in the Netherlands, maybe, travel times would be reduced and capacity would increase. Capacity will increase because, during working times, agents drive fewer kilometers with more turnout facilities. Hence, is the objective function in this study to minimize the total travel times for DKDB agents. Not only the optimal locations of new turnout facilities but also, to get more insight into the capacities of these new turnout facilities, the optimal capacity (i.e., vehicles of bodyguards) distribution between these new facilities will be determined. Location Theory is well known for solving these kinds of problems. However, an additional challenge in this study is that, compared to traditional Location Theory, the number and location of demand nodes, which are represented by threatened people, are stochastic. This means that the goal is to develop a mathematical model that can deal with demand uncertainty and to find solutions that are robust for every possible scenario that unfolds in the future.

Research objectives

The overarching research objective (RO) is:

RO. Develop a model that minimizes the travel time for DKDB agents when opening a fixed number of turnout facilities and determines the optimal locations and capacity distributions for these turnout facilities.

To address this objective the following sub-objectives have to be addressed:

SO1. Execute a literature study on location theory with demand uncertainty.
SO2. Define a mathematical model capable of dealing with demand uncertainty while identifying the optimal turnout locations and capacity distributions.
SO3. Identify a solution method that can solve the mathematical model of SO2.
SO4. Specify the numerical experiment that will be executed in SO5.
SO5. Execute the numerical experiment of SO4 and delve managerial insights based on the optimal travel times, capacity distributions, and locations for placing varying numbers of turnout facilities.

Location theory

To address SO1, a literature study on Location theory was executed. Facility Location Theory is particularly well-suited to address the problem due to its inherent focus on optimizing the placement of facilities while ensuring a certain objective.

Because, the theory of demand uncertainty is less developed than, for example, price uncertainty and distance uncertainty, there is no mathematical model in existing literature that can be used immediately. In Location Theory, there are a few ways to handle demand uncertainty such as using an instance approach, a scenario approach, or introducing a capacity constraint.
A common method to solve a FLP with demand uncertainty is a two-stage solution method where the locations of facilities are the first-stage decision and the assignment of demand to facilities is the second-stage decision.

Mathematical model

In this study, first, nine instances will be formulated. These nine instances will represent different outcomes in the future related to the number and location of TBPs. The goal here is to find robust optimal locations and capacity distributions for all the instances. On top of that, instances will consist of different scenarios. Each scenario resembles a possible outcome in the future. The mathematical model must be able to optimize over all these scenarios and find the locations and capacity distributions that are optimal for the combination of these scenarios.
To use Location Theory, the assumption has been made that TBPs (Persons To Be Protected) will be clustered on geographical area. In this thesis, it has been decided to cluster TBPs on municipality level.
To address SO2, first, a mathematical model was formulated that ignores capacity. This model was referred to as the Scenario Turnout Model (STM) and combines aspects of the Uncapacitated FLP (UFLP) as introduced by Farhan & Murray (2006) and a scenario approach technique as introduced in SO1. On top of this, we also formulate the Capacity STM (CSTM) model to illustrate how capacity (i.e., vehicles of bodyguards) need to be pre-allocated over time.

Solution method

We developed a two-stage solution method for the CSTM because the computational time of the CSTM exceeds a reasonable computational time for the purpose of this study. In the first stage of this method, the fact that in the STM the assignment of TBPs to turnout facilities is only dependent on distance is used to rewrite the STM in an equivalent model in which the uncertain parameters are replaced by their means (Snyder, 2006). This equivalent model is referred to as the Average STM (ASTM), which ignores capacity and is computationally less complex than the STM. Subsequently, for the given allocation of the turnout facilities, the method solves the Simplified CSTM (SCSTM). The SCSTM uses the CSTM as a basis where constraints for placing the turnout facilities are left out and the potential facility location set is reduced to the solutions from the ASTM.

Numerical experiment

For the numerical experiment of SO4, the DKDB has been given the choice to either share historical data about TBPs or keep this information confidential. Due to the sensitive nature of this data, it has been determined that this data should be kept confidential. Therefore, only information available on the internet is used for this project. Because no information about TBPs is publicly available, the assumption is made that TBPs follow the same trends as the overall population of the Netherlands. In SO4, two trends in the Netherlands are examined. The first trend concerns the fact that people tend to move out of the populated areas in the Netherlands. The second trend concerns the growing number of TBPs. These trends are investigated and with the use of the Simple Average (SA) forecasting method of Blanc & Setzer (2016), different growth factors are determined.
With the use of the different growth factors, the instance approach as introduced in SO1 can be employed. The data input that is the same for each instance is the potential facility location set, the

maximum number of turnout facilities to open, and the distance matrix between the turnout facilities and TBPs. For each instance, the TBP data is different. The TBP data is represented by the different scenarios. The scenarios are based on gathering random values from a Poisson distribution. The expected value of the Poisson distribution is equal to the expected number of TBPs per municipality which in turn is dependent on the different growth factors.

Results

To address SO5, the two-stage method from SO3 with the specified numerical experiment from SO4 is used in Gurobi. To determine the optimal travel times, a weighted prediction result method is introduced that is based on the literature of Peterson et al. (2003). For opening one to five facilities, the experiments demonstrate that the outcomes are robust against various instance settings. For opening six facilities, robustness is shown in six out of the nine instances. Robustness is further substantiated by additional experiments on more demand fluctuations and using normal instead of Poisson distributed demand.

For the optimal capacity distributions, the nine instances all give approximately the same results with a maximum difference of 2%. Again, robustness is further substantiated by additional experiments on more demand fluctuations and using normal instead of Poisson distributed demand.

From the optimal locations, three things can be learned. At first, for a low number of facilities, optimal locations barely repeat because the model opts to locate turnout facilities such that the travel distance and the number of TBPs are balanced for every turnout facility. However, starting from four facilities, Amsterdam and Rotterdam will be present in all optimal solutions. At last, the optimal locations that are based in the south- and north-east sides of the Netherlands will tend to go further into the corners of the country when the number of turnout facilities increases. The optimal capacity distribution confirms the insights gained from the optimal locations that the model opts to locate turnout facilities such that the travel distance and the number of TBPs are balanced for every turnout facility.

Conclusion and recommendations

To address the RO, the study proposes the CSTM to improve the capacity of the DKDB. The SCTM is solved with a two-stage solution method, where the ASTM is used to determine the optimal locations, and the SCSTM is used to determine the optimal capacity distribution for these turnout locations. For opening one to six facilities, we found locations and capacity distributions that are robust for the future. That means that these solutions are optimal for whatever scenario unfolds in the future.

For the recommendation, the solutions were presented to the DKDB. The DKDB concluded that opting for four turnout facilities in Amsterdam, De Wolden, Maashorst, and Rotterdam seems to be the most logical choice taking into account the Key Performance Indicators (KPIs) which are the decrease in travel distance and the amount of Ful-Time Equivalents (FTEs) the DKDB could spare annually. The KPIs do not increase much from 4 facilities onwards. With this choice, the travel distances would decrease by 70.887% and the capacity would increase by 107.6 FTEs annually. However, to make a good decision, the DKDB should also take into account some other factors such as budget constraints, space availability, and political restrictions such as community relations or ideological differences.

The scientific contribution of this study lies in handling demand uncertainty in FLPs and finding solutions that are robust for the future.

Areas for future research include using a real TBP dataset and determining the optimal capacity distribution when the capacity is lower than the number of TBPs.

Because this project deals with confidential data, the decision has been made that some numbers are not correct on purpose.

Contents

List of Figures

List of Tables

1. Introduction

Our free society faces different kinds of threats, some of them with terrible consequences. One of the biggest threats comes from organized crime (Ministerie van Justitie en Veiligheid, 2022a). Criminals put pressure on our society by intimidating or even attacking someone. A good example of this is the assassination of the Dutch crime journalist Peter R. de Vries in 2021. De Vries got shot in Amsterdam after hosting a Dutch TV show. Next to organized crime, the Netherlands also still faces threats from different angels like jihadism and the right-extremist movement (Ministerie van Justitie en Veiligheid, 2022b). These right-extremists propagate that violence can be used as a legitimate tool to reach political goals. Also, the rise of belief in conspiracy theories in the Netherlands can lead to more personal attacks (Ministerie van Justitie en Veiligheid, 2022c). The Netherlands currently deals with a group of anti-government extremists who fundamentally distrust the government and other instances (Ministerie van Justitie en Veiligheid, 2022c).

The developments mentioned above cause that more persons need long-term intensive security (Ministerie van Justitie en Veiligheid, 2022ab). On the 14th of April 2022, this growing threat in the Netherlands was an agenda point in the Dutch Parliament (Ministerie van Justitie en Veiligheid, 2023). The Bewaken and Beveiligen (B&B) system provides protection for threatened individuals in the Netherlands. One of the main roles of the B&B is to be responsible for the personal security of individuals by bodyguarding these individuals but also by doing prior research. This could be, for example, screening the places where the individuals go and screening the people with whom the individuals come into contact. The Dutch parliament concluded that the B&B system needs more personnel for these operational tasks.

One of these operational aspects where there could be improvement lies at the DKDB. The DKDB is a specialist department of the KLPD (Corps National Police Services) which in turn has an advising role for the NCTV (National Coordinator of Counterterrorism and Security). The DKDB is responsible for the personal security of royalty's, politicians, diplomats, judges, and other designated persons (Calis, n.d). The main responsibility of the DKDB is to prevent, endure, and withstand attacks on persons and buildings. Since 2010, the focus has been more on the prevention part than on the endure and withstand part (Koeman, 2010). DKDB agents can have multiple roles and their jobs can be static and dynamic. In terms of dynamic, one can think of agents that scout a certain place before a TBP arrives and also agents that move alongside the TBP. In terms of a static job, agents stand, for example, next to buildings where TBPs are at that moment. Because the Dutch parliament knew that hiring Full-Time Equivalents (FTEs) is difficult, they also mentioned that more scientific knowledge on operational parts to improve the capacity of the B&B system was needed (Ministerie van Justitie en Veiligheid, 2023).

1.1 Problem definition

As described in the previous sections, there is a need for more capacity inside the whole organ of the B&B system. One of the main potential points of improvement concerns opening multiple turnout facilities for the DKDB.

1.1.1 Problem description

At this point, the DKDB operates with one turnout facility in the Netherlands which is located in The Hague. A turnout facility is a facility where DKDB agents have to arrive before and after their shifts. For example, an agent living in Groningen has to ride to The Hague for 2.5 hours every morning and then depart for a job that can, again, be in Groningen. Of course, this is an extreme scenario. However, it seems that some agents might drive a lot of kilometers every day. If there were multiple turnout facilities in the Netherlands, maybe, travel times would be reduced and capacity would increase.

Capacity will increase because, during working times, agents drive fewer kilometers with more turnout facilities. Therefore, the objective function of the mathematical model in this study will be to minimize the total travel time of DKDB agents. Because opening extra turnout facilities is costly and the DKDB will have budget constraints, it is not possible to open infinite turnout facilities.

In Literature, Location Theory is well known to solve location problems for facilities. Common models in Location Theory are Facility Location Problems (FLPs), cover models, and Police Districting problems (PDPs). For all these models, there is a set of demand nodes and a set of potential facilities. To illustrate this, Figure 1 is used. In Figure 1, the blue dots represent the customer/demand nodes and the red squares represent the potential facility nodes. In Location Theory, demand nodes are assigned to facilities based on a certain objective and a set of constraints. If all the facilities from Figure 1 are opened, the assignment of demand nodes to facilities could look like Figure 2 while if only two facilities were opened, the assignment of demand nodes to facilities could possibly look like Figure 3. In traditional location theory, demand nodes are fixed. In this project, demand nodes are represented by TBPs. The distribution of TBPs throughout the Netherlands is not known in the future. This means that it is not known where TBPs are located, and how many TBPs there will be in the future. This uncertainty in TBP data is what differs in this study compared to the standard Location Theory. On top of this, this project not only deals with demand uncertainty but also with finding a model that is robust against this uncertainty.

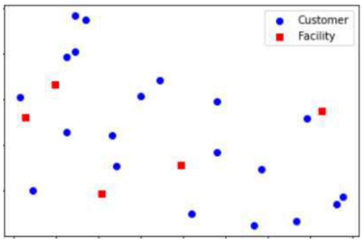

Figure 1: Location Theory example

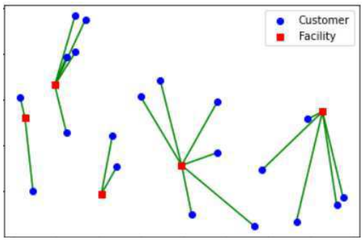

Figure 2: Location Theory assignment example 1

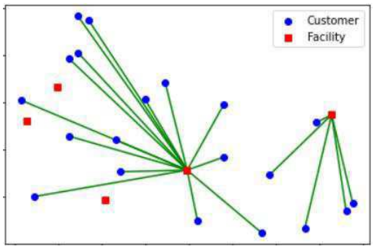

Figure 3: Location Theory assignment example 2

On a more generic level, one could say that the DKDB has a problem in the sense that they do not have a clear picture of the potential capacity increase of placing turnout facilities while minimizing travel times. An additional challenge in this problem is that, because of the uncertain distribution of TBPs in the future, there exist no mathematical models in the literature that can be used in this study.

1.1.2 Research Objectives

Based on the problem description, several research objectives are identified. The goal of this project is to gain insights for the DKDB into where to open turnout facilities and determine the capacity distribution of these facilities, and by doing so introducing and studying a new facility location Model. Therefore is the guiding research objective (RO) in this project formulated as:

RO. Develop a model that minimizes the travel time for DKDB agents when opening a fixed number of turnout facilities and determines the optimal locations and capacity distributions for these turnout facilities.

To address this objective, the following sub-objectives (SO) have to be addressed.
Because the problem has features of Location Theory, we execute a literature study on Location Theory with demand uncertainty (SO1). Based on the literature study, we define a mathematical model capable of dealing with demand uncertainty while identifying the optimal turnout facility

locations and capacity distributions (SO2). Because the computational time for the mathematical model from SO2 exceeds a reasonable computational time for the purpose of this study, we identify a solution method that can solve the mathematical model of SO2 (SO3). Next, to use the solution method from SO3, we specify the numerical experiment (SO4). Here, the parameters for the optimization process are set. All parameters that are used for one optimization are called an instance. One of the parameters that will be set is the distribution of TBPs. For this historical data about TBPs will be explored with the use of a forecasting method to prediction the future. The other parameters, which are the potential facility location set, the maximal number of facilities to open, and the distance matrix are fixed. At last, to delve managerial insights based on the optimized travel times, capacity distributions, and locations for placing varying numbers of turnout facilities, we execute the numerical experiment of SO4 (SO5).

1.2 Contribution

The scientific contributions lie in handling demand uncertainty. In facility location theory, the theory on demand uncertainty is less developed than, for example, price uncertainty and distance uncertainty. This thesis introduces nine different instances which represent different outcomes for the future. Subsequently, the mathematical model will be formulated to accommodate various scenarios. Again, these scenarios represent different demand distributions. Solutions that are robust for all these possible demand realizations have been found.

1.3 Outline

This report describes how the DKDB can increase its capacity by opening multiple turnout facilities in the Netherlands. To accomplish this, Chapter 2 provides a literature review on Location Theory thereby answering SO1. Then, Chapter 3 identifies a mathematical model that deals with demand uncertainty, thus dealing with SO2. In Chapter 4, we identify a two-stage solution method to solve the mathematical model form Chapter 3, and thereby answering SO3. Chapter 5 examines the numerical experiment, thus dealing with SO4. Finally, Chapter 6 states the results and managerial insights, thus answering SO5. This thesis concludes in Chapter 7.

2. Literature review

This chapter addresses SO1 where the objective is to execute a literature study on Location Theory with demand uncertainty.

2.1 Location Theory

The literature review in this section is aimed at providing an overview of relevant literature and different models in Facility Location Theory. Facility Location Theory is particularly well-suited to address the problem context due to its inherent focus on optimizing the placement of facilities while ensuring a certain objective function. In the case of the DKDB, the need to strategically position new turnout facilities throughout the Netherlands is crucial to reduce the travel time of agents. The problem definition aligns well with two particular dynamics within Facility Location Theory. At first, Facility Location Theory is designed to handle spatially distributed data and geographically dispersed facilities. In DKDB's scenario, the consideration of multiple potential turnout facilities across the country requires a spatial optimization approach to identify the optimal locations. By analyzing geographic data and distances between various locations, the theory can determine the optimal placement of turnout facilities while enhancing a certain objective. Secondly, facility location theory takes service coverage into account. The DKDB's responsibility to provide personal security extends to threatened individuals across the country. By taking service coverage into account, facility location theory ensures that the proposed facilities adequately cover regions with a high demand for security services. In the domain of facility location theory, categorization can be divided into the following areas.

2.1.1 Facility Location Problems (FLP)

The first model that has been examined is the Facility Location Problem (FLP) the FLP consists of a set of potential facilities ($L = \{1,2,..,n\}$). Secondly, there is a set of demand nodes *($D = \{1,2,..,m\}$)* that must be served. The objective is to open a subset of facilities in the set *L* that minimizes the distances between demand nodes and open facilities (Farhan & Murray, 2006). The distance between a demand node $i \in D$ and a facility $j \in L$ is stated as $d_{ij} \in \mathbb{R}_+ \ \forall\, i \in D, j \in L$. The facility location problem has two decision variables. The first is $y_j \in \{0,1\} \ \forall j \in L$ which is 1 if a facility is opened and 0 otherwise. The second decision variable is $x_{ij} \in \{0,1\} \ \forall\, i \in D, j \in L$ which is 1 if demand node *i* is served by facility *j*. At last, the Facility Location Problem requires a parameter that states how many facilities can be opened. This parameter is stated as *P*.

2.1.2 Covering models

The second model that has been examined are cover models. In literature, there exist many different types of covering models. Schilling, Jayaraman, and Barkhi (1993) defined two categories in cover models; the Set Covering Problem (SCP) and the Maximum Covering Location Problem (MCLP). These two models will be referred to as 'base models' in the rest of this literature review. Next to these base models, there exist many different extensions. These extensions take the base models as a base for their model and relax a certain assumption that these base models have.

Set Covering Problem (SCP)

In the SCP, there is a set of demand nodes *($D = \{1,2,..,m\}$)* and a set of potential facility nodes ($L = \{1,2,..,n\}$). In the SCP, every facility that will be opened will have a certain cost: $c_j \in \mathbb{R}_+ \ \forall\, j \in L$. The distance between demand node $i \in D$ and a facility $j \in L$ is stated as $d_{ij} \in \mathbb{R}_+ \ \forall\, i \in D, j \in L$. A demand node $i \in D$ can be served by a facility $j \in L$ if this facility is below a certain maximal distance. This maximal distance is stated as $S \in \mathbb{R}_+$. Consequently, a facility can thus only serve a demand node if $d_{ij} \leq S \; \forall \; i \in D, j \in L$. The binary parameter $a_{ij} \in \{0,1\} \; \forall \; i \; \in D, j \in L$, is equal to 1 if the distance between demand node *i* and facility *j* is lower or equal to *S* and 0 otherwise. The decision variable in the SCP is $x_j \in \{0,1\} \; \forall j \in L$. This decision variable is 1 if a facility is opened and 0 otherwise. The SCP tries to minimize the location cost of locating facilities while covering every demand node within the service distance *S.*

Maximal Covering Location Problem (MCLP)

In the MCLP, again, there is a set of demand nodes *(*$D = \{1,2,..,m\}$*)* and a set of potential facility nodes ($L = \{1,2,..,n\}$). The distance between a demand node and a facility is stated as $d_{ij} \in \mathbb{R}_{+} \forall \; i \in D, j \in L$. The maximal distance measure is stated as $S \in \mathbb{R}_{+}$. A facility *j* can only serve a demand node *i* if $d_{ij} \leq S$. The binary parameter $a_{ij} \in \{0,1\} \; \forall \; i \; \in D, j \in L$, will be equal to 1 if the distance between demand node *i* and facility *j* (d_{ij}) is lower or equal to *S* and 0 otherwise. The decision variable in the MCLP is $x_j \in \{0,1\} \; \forall j \in L$. This decision variable is 1 if a facility will be opened and 0 otherwise. In the MCLP, every demand node can have a different level of demand, this is stated as $h_i \; \in \mathbb{R}_{+} \forall i \in D$. The number of facilities to locate has to be prespecified in the MCLP which is indicated as *P*. Because the goal of the MCLP is to maximize the covered demand there is a need for a binary variable that indicates if a demand node is covered or not. This is indicated as z_i which is 1 if demand node *i* is covered and 0 otherwise ($\forall \; i \in D$).

2.1.3 Police Districting Problems (PDP)

The third model that has been examined is the Police Districting Problem (PDP). A PDP concerns the efficient and effective design of patrol sectors. This means that a police districting model is able to divide a large service area into smaller districts based on different attributes that are predefined. According to Liberatore, Camacho-Collados, and Vitoriano (2020), there exists no base model for the PDP. The only thing PDPs have in common is that the models try to divide large service areas into smaller balanced districts. The way that the districts are balanced is based on the different attributes that are predefined in the model. As seen in the literature, examples of these attributes could be area, demand, and diameter.

Upon examining the three previously explained models in Sections 2.1.1, 2.1.2, and 2.1.3, it becomes evident that the FLP is the most suitable match for the problem definition. Farhan & Murray (2006) introduce an Uncapacitated FLP (UFLP) which will be used as a basis for this project. The model dynamics are already explained in Section 2.1.1. Farhan & Murray (2006) introduce the following FLP:

$$\min \sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} * d_{ij} \qquad (1)$$

S.T

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall \; i \in D \qquad (2)$$

$$\sum_{j=1}^{n} y_j = p \qquad (3)$$

$$x_{ij} - y_j \leq 0 \qquad \forall i \in D, j \in L \qquad (4)$$

$$y_j \in \{0,1\} \qquad \forall j \in L \qquad (5)$$

$$x_{ij} \in \{0,1\} \qquad \forall \; i \in D, j \in L \qquad (6)$$

The objective (1) minimizes the total traveled distance. Constraint (2) requires that every demand node is assigned to one open facility. Constraint (3) specifies the number of facilities that are available to open. Constraint (4) limits assignment to open facilities only. Constraints (5) and (6) are the integrality constraints. This model will be referred to as the 'UFLP model' for the rest of this project.

2.2 Demand uncertainty

The UFLP model and other traditional FLPs treat demand nodes as fixed parameters in the model. However, for this project, the demand for the future is unknown. This is the case because of the two following trends that are visual in the Netherlands for the past few years. At first, as is concluded in Chapter 1, more people will need long-term intensive security in the Netherlands in the future. That is, the number of TBPs is expected to increase in the coming years. Secondly, since 2017, there has been a decline in the population of populated regions in the Netherlands (NOS, 2022). These two trends will play a vital role in this project. Consequently, in contrast to conventional FLPs, the distribution of the demand nodes (TBPs) in this project remains uncertain for the future.

Snyder (2006) researched how to handle uncertainty in FLPs. According to Snyder (2006), problems under uncertainty are known as 'robust optimization' models. The goal of robust optimization is to find a solution that will perform well under any possible realization of the random parameters. Snyder (2006) states that the random parameters can either be continuous or described instances. If probability information about the random parameters is known, uncertainty is described using a (continuous or discrete) probability distribution on the parameters. If no probability information is known, continuous parameters are generally restricted to lie in specific intervals. Because no information about the probabilities of TBPs is known, the instances approach will be used in this project. As mentioned by Peterson, Cumming, and Carpenter (2003), to make instances, growth factors have to be determined. First, a prediction growth factor will be determined based on a forecast method. Afterward, an above-prediction and below-prediction growth factor have to be employed. Peterson et al. (2003) mention that, for drawing conclusions, the 'prediction' growth factors are more important than the 'above' and 'below' growth factors.
Next to the instance approach, Snyder (2006) introduces a scenario approach for handling uncertainty in FLPs. These scenarios represent different distributions of the uncertain parameter (Snyder, 2006). In a scenario approach, a mathematical model must be able to optimize over all these scenarios. This means an extra dimension of some of the decision variables and or parameters will be added. An FLP using a scenario approach will choose the location that is optimal for the combination of these scenarios. This means that the objective will minimize the distance over all the scenarios. This is important because the Netherlands will have a lot of geographical areas that will host no TBPs. With the use of these scenarios, it could be possible that once in a while, such an area will host a TBP. When averaging the distance and choosing a facility location based on this average distance, the opportunity of such an area hosting a TBP is taken into account. This is in line with the theory of the *minisum location problems* (Snyder, 2006). The most common objective in Stochastic Programming (SP) is to optimize the mean outcome of the system; e.g., minimize the expected cost or maximize the expected profit (Snyder, 2006). For example, Cooper (1978) considers the Weber problem in which the locations of the demand points may be random. He assumes a bivariate normal distribution for these locations. The objective is to choose a point for the single facility location to minimize the expected demand-weighted distance to the customers. The first rigorous attempt to choose facility locations to minimize expected cost under scenario-based uncertainty was offered by Mirchandani and Oudjit (1980), who discuss the 2-median problem on a tree with stochastic edge lengths de-scribed by discrete scenarios. The objective is to minimize the expected demand-weighted distance.

Another way to handle demand uncertainty in FLPs is the introduction of a capacity constraint. Because the goal of this project is to gain capacity for the DKDB, it is also interesting to look at the literature about capacity in FLPs. Meraklı & Yaman (2017), introduce a hub location problem under demand uncertainty. The goal of the model of Meraklı & Yaman (2017) is to find the locations that minimize the worst-case total cost. For every facility, Meraklı & Yaman (2017) introduce a capacity constraint that states that the sum of the demand that is served from a certain facility cannot exceed a predefined benchmark. Louveaux, & Peeters (1992) solve a capacitated facility location problem with a two-stage stochastic program where the first-stage decisions are the location and the size of the facilities to be established and the second-stage decision is the optimal allocation of the available production to the most profitable demands. In their model, just as in the model of Meraklı & Yaman (2017) every potential facility has a capacity. With the use of the notation of the FCL model of Farhan & Murray (2006), a capacity constraint as introduced by Louveaux, & Peeters (1992) and Meraklı & Yaman (2017) would look as follows:

$$\sum_{i=1}^{m} z_i * x_{ij} \leq cap_j * y_j \qquad \forall\, i \in D, j \in L \;\; (7)$$

Here, cap_j stands for the capacity of facility *j* and z_i represents the demand at node *i* $\forall\, i \in D$. By using this constraint, a facility may reach its capacity and a TBP will be assigned to another open facility although that facility is not its closest open facility.

A common solution method for stochastic FLPs is a two-stage solution method. In stochastic location Modeling, locations are generally first-stage decisions whereas the assignment of customers to facilities is second-stage (Snyder, 2006). For instance, Ravi & Sinha (2004) developed an approximation algorithm for a stochastic version of the UFLP in which facilities may be opened in either the first or second stage, incurring different fixed costs in each. This means that in the second stage, the assignment of customers to a facility is based on fixed costs which is a stochastic variable in the case of Ravi & Sinha (2004). In this project, the allocation of TBPs to facilities is only based on distance which is a parameter. This means that both decisions occur in the first stage. If both decisions occur in the first stage, most problems can be reduced to deterministic problems in which uncertain parameters are replaced by their means (Snyder, 2006).

3. Mathematical Model

In this Chapter, SO2 will be addressed. This research objective pertains to determining a mathematical model capable of dealing with demand uncertainty while identifying the optimal turnout locations and capacity distributions. Because the goal of this project is to increase capacity for the DKDB, the mathematical model will minimize travel distances for DKDB agents. By minimizing these travel distances, agents will have more effective work time which will lead to a capacity increase. In Section 3.1 we first address a model that determines the optimal locations of new turnout facilities, ignoring capacity. Then, in Section 3.2, the model is extended and also the optimal capacities for these turnout facilities are determined.

3.1 Location model

3.1.1 Sets and parameters

We start by introducing a set $Q = (1,2, \dots, q)$ where q is equal to the number of scenarios that will be used. Also, we will introduce a set of potential turnout facility nodes $L = \{1,2,..,n\}$ and a set of demand nodes $D = \{1,2,..,m\}$ that must be served. In practice, one should interpret these nodes as a geographical area (e.g., of the size of a municipality). In this project, a demand node can host multiple TBPs. In other words, a geographical area can host multiple TBPs and we denote this number by $z_{iw} \in Z_+ \; \forall \; i \in D, w \in Q$. That is z_{iw} represent the number of TBPs at demand node *i* in scenario $w \in Q$. z_{iw} can only be a non-negative integer since an area can't host a half TBP or a negative number of TBPs. Note that it is still possible for a geographical area to host no TBP in a specific scenario. In traditional Location Theory, it is assumed that demand is not continuously distributed across a geographical area but concentrated on specific geographical locations. In Practice, a TBP can live in every single geographic location. One can see this as a continuous area where every single point in this area could be a possible home location of a TBP. However, to use a Facility Location Problem, the assumption has been made that TBPs will be clustered based on geographical area. In this project, the decision is made to cluster on municipality level. Why this decision is made is explained in Chapter 5. Lastly, The distance between a demand node $i \in D$ and a turnout facility $j \in L$ is $d_{ij} \in \mathbb{R}_+ \; \forall \; i \in D, j \in L$.

3.1.2 Decision variables

In this project, one of the two decision variables is $x_{ijw} \in \{0,1\} \; \forall \; i \in D, j \in L \; w \in Q$. This decision variable indicates if a demand node is served by a turnout facility node in a specific scenario. The last decision variable is $y_j \in \{0,1\} \; \forall j \in D$ which is 1 if a turnout facility is opened and 0 otherwise.

3.1.3 Objective function

For this project, the sum of the travel distance for all demand nodes, turnout facility nodes, and scenarios must be minimized. Also, a municipality can host multiple TBPs. This means that if a municipality hosts multiple TBPs, the distance is also driven multiple times a day. To account for this, the distance must be multiplied by the number of TBPs a municipality hosts. At last, the summation of the distances for all demand nodes, turnout facility nodes, and scenarios needs to be divided by the number of scenarios. By this, the outcome of the objective function will be the average distance of all scenarios. With these modifications, the objective function of the mathematical model becomes:

$$(\min \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{w=1}^{q} x_{ijw} * d_{ij} * z_{iw}) \; / \; q \qquad (8)$$

3.1.4 Constraints

First, there is a requirement that every demand node is assigned to one open turnout facility. Also for every scenario, every demand node has to be assigned to one open turnout facility. That is why the first constraint is:

$$\sum_{j=1}^{n} x_{ijw} = 1 \qquad \forall\, i \in D, w \in Q \qquad (9)$$

The next constraint ensures that the total amount of turnout facilities that will be opened is equal to a predefined number of turnout facilities (p):

$$\sum_{j=1}^{n} y_j = p \qquad (10)$$

The next constraint limits the assignment to open turnout facilities only. This means that x_{ijw} can only be 1 if y_j is 1. So, the third constraint becomes:

$$x_{ijw} - y_j \leq 0 \qquad \forall\, i \in D, j \in L, w \in Q \qquad (11)$$

An integrality constraint on the y_j decision variable must be introduced:

$$y_j \in \{0,1\} \qquad \forall j \in L \qquad (12)$$

The last constraint is the integrality constraint on the binary decision variable x_{ijw}. This constraint must also hold for all scenarios and thus this constraint becomes:

$$x_{ijw} \in \{0,1\} \qquad \forall\, i \in D, j \in L, w \in Q \qquad (13)$$

3.1.5 Mathematical model

In Summary, the mathematical model to determine the optimal turnout facility locations, ignoring capacity, is stated below. For the rest of this project, this model is referred to as the 'Scenario Turnout Model' (STM).

$$(\min \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{w=1}^{q} x_{ijw} * d_{ij} * z_{iw}) \,/\, q \qquad (14)$$

S.T

$$\sum_{j=1}^{n} x_{ijw} = 1 \qquad \forall\, i \in D, w \in Q \qquad (15)$$

$$\sum_{j=1}^{n} y_j = p \qquad (16)$$

$$x_{ijw} - y_j \leq 0 \qquad \forall\, i \in D, j \in L, w \in Q \qquad (17)$$

$$y_j \in \{0,1\} \qquad \forall j \in L \qquad (18)$$

$$x_{ijw} \in \{0,1\} \qquad \forall\, i \in D, j \in L, w \in Q \qquad (19)$$

With this mathematical model, uncertainty can be captured with the use of scenarios. These scenarios will have different numbers of TBPs per geographical area for each scenario. The model identifies the turnout facility or facilities with the lowest average distance among all these scenarios.

3.1.6 Example

To further clarify the STM, an example is given. For this example, the Parkstad region in the south of the Netherlands is used. Figure 4 represents this region and the municipalities belonging to the region.

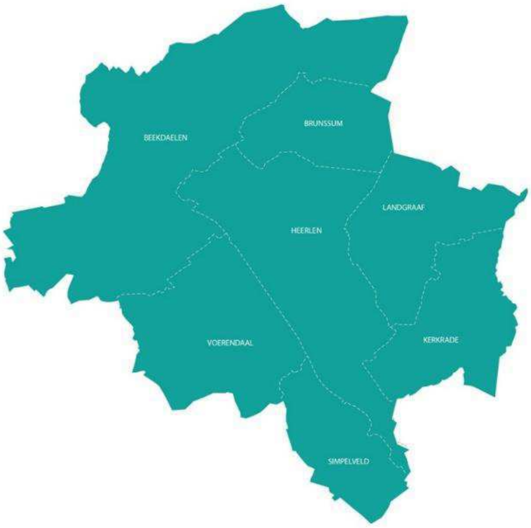

Figure 4: Municipalities in Parkstad region in the Netherlands

As of January 2023, this region consists of seven municipalities namely: Beekdealen, Brunssum, Landgraaf, Heerlen, Voerendaal, and Kerkrade. The borders of the municipalities are indicated by the white dashed line in Figure 4. For this example, two scenarios are made. These two scenarios are stated in Table 1.

Table 1. Two example Scenarios for Parkstad

Municipality	TBPs in Scenario 1	TBPs in Scenario 2
Beekdaelen	3	3
Brunssum	2	3
Heerlen	5	5
Landgraaf	1	1
Voerendaal	4	3
Kerkrade	3	4
Simpelveld	3	2

Table 2 shows the distance matrix between the seven municipalities in this region.

Table 2. Distance matrix between municipalities in Parkstad

	Beekdealen	Brunssum	Heerlen	Landgraaf	Voerendaal	Kerkrade	Simpelveld
Beekdealen	-	5	10	8	15	13	20
Brunssum		-	5	8	10	9	14
Heerlen			-	7	9	4	3
Landgraaf				-	8	3	10
Voerendaal					-	10	2
Kerkrade						-	11
Simpelveld							-

Note that the numbers in Tables 1 and 2 are random numbers and do not represent real life. For this example, the decision is made to open two facilities. The total traveled distance per scenario can be calculated by multiplying the distance between each municipality by the total number of TBPs residing in that municipality. The result of the optimization is stated in Table 3.

Table 3. Travel distance per scenario for Parkstad example

Locations	Total distance scenario 1	Total distance scenario 2	Average total distance
Heerlen, Voerendaal	65	72	**68.5**

After running this example in Gurobi, it becomes clear that Heerlen and Voerendaal are the optimal locations for a turnout facility. The values of the optimal solution of the decision variable y_j are stated in Table 4 below.

Table 4: Optimal values for decision variable y in the example

j	Beekdealen	Brunssum	Heerlen	Landgraaf	Voerendaal	Kerkrade	Simpelveld
y_j	0	0	1	0	1	0	0

When looking at the constraint (16) on the decision variable y_j in the STM, the sum of all the values of the y_j variables must be equal to the number of facilities we want to open. The sum of the variables in Table 4 is equal to two which is exactly the number of facilities we wanted to open.
Table 5 represents the optimal values of the x_{ijw} variables. Note that for both scenarios, the matrix looks the same. The vertical axis represents the demand set *D* while the horizontal axis represents the potential facility set *L*.

Table 5: Optimal values for the decision variable x in the example

	Beekdealen	Brunssum	Heerlen	Landgraaf	Voerendaal	Kerkrade	Simpelveld
Beekdealen	0	0	1	0	0	0	0
Brunssum	0	0	1	0	0	0	0
Heerlen	0	0	1	0	0	0	0
Landgraaf	0	0	1	0	0	0	0
Voerendaal	0	0	0	0	1	0	0
Kerkrade	0	0	1	0	0	0	0
Simpelveld	0	0	0	0	1	0	0

Looking at the constraint (15) on the x_{ijw} decision variable from the STM, it becomes clear that $\forall\, i \in D, w \in Q$ the sum of the x_{ijw} variables must be equal to one. Looking at Table 5, the sum of each row is equal to one. Since this matrix is the same $\forall\, w \in Q$, the solution adhered to the constraint.
At last, we take a look at constraint (17). This constraint mentions that $\forall\, i \in D, j \in L$, the x_{ijw} must be smaller or equal to y_j. When taking a look at Tables 4 and 5, this is true. This means that the solution to put a turnout facility in Heerlen and Voerendaal is a feasible and optimal solution.

3.2 Capacity model

In the STM model, we assumed that TBPs can always be allocated to the same turnout facility. This might not always be realistic, taking into account that also resources (i.e., vehicles) are needed to transfer to the TBPs. In this section, we will therefore introduce a generalized location model where capacity is included. However, the capacity constraint introduced by Meraklı & Yaman (2017) and Louveaux, & Peeters (1992) in the literature review does not make sense in this project. For this constraint, the capacity of each turnout location must be known beforehand which is not the case. Also, in this project, it would not make sense to use capacity constraints because now only one turnout facility in The Hague is used. This means that this turnout facility has an 'infinite' capacity. That is why in this project, the optimal capacities for the open turnout facilities will be determined. This will give the DKDB more insight into how to distribute resources among the open turnout facilities.
To model capacity, the STM will be used as a basis. To model this setting, we need to introduce $cap_j \in Z_+ \forall\, j \in L$ which is the maximal capacity of each turnout facility. We also introduce *T* which is the total number of resources available in the Netherlands. Also, two new constraints need to be added. At first, the sum of the demand coupled to a turnout facility needs to be lower or equal than the capacity of that specific turnout facility. This constraint becomes:

$$\sum_{i=1}^{m} x_{ijw} * z_{iw} \leq cap_j \quad \forall\, j \in L, w \in Q \qquad (20)$$

At last, the sum of all capacities needs to be equal to the total resources available which is ensured by the following constraint:

$$\sum_{j=1}^{n} cap_j = T \qquad (21)$$

The new mathematical model that handles this uncertainty is referred to as the Capacity Scenario Turnout Model (CSTM) and is stated below:

$$(\min \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{w=1}^{q} x_{ijw} * d_{ij} * z_{iw}) / q \qquad (22)$$

S.T

$$\sum_{j=1}^{n} x_{ijw} = 1 \qquad \forall\, i \in D, w \in Q \qquad (23)$$

$$\sum_{j=1}^{n} y_j = p \qquad (24)$$

$$x_{ijw} - y_j \leq 0 \qquad \forall\, i \in D, j \in L, w \in Q \qquad (25)$$

$$\sum_{i=1}^{m} x_{ijw} * z_{iw} \leq cap_j \quad \forall\, j \in L, w \in Q \qquad (26)$$

$$\sum_{j=1}^{n} cap_j = T \qquad (27)$$

$$y_j \in \{0,1\} \qquad \forall j \in L \qquad (28)$$

$$x_{ijw} \in \{0,1\} \qquad \forall\, i \in D, j \in L, w \in Q \qquad (29)$$

The objective function (22) is the same objective function as for the STM. The objective still is to minimize the travel distance while choosing the capacity per turnout facility. Also, constraints (23), (24), (25), (28), and (29) are directly copied from the STM. Constraint (26) makes sure that for every turnout facility and every scenario, the total number of TBPs coupled to that turnout facility is lower than the capacity of that turnout facility. Constraint (27) ensures that the sum of all capacities of the turnout facilities is exactly the same as the available resources.

The CSTM stated above will be the mathematical model that will be solved in this project. The model minimizes the travel distance for DKDB agents while determining the optimal locations for new turnout facilities and the optimal capacity distributions of these turnout facilities.

4. Solution method

This chapter addresses SO3 which concerns identifying a solution method to solve the mathematical model from SO2.

When using municipalities as geographical areas to cluster TBPs in the CSTM, there will be $342 * 342 * \mathrm{q} = 116{,}964 * \mathrm{q}\, x_{ijw}$ variables, where q stands for the number of scenarios that will be used. Because of these large numbers of decision variables, the computational time of the CSTM exceeds a reasonable computational time for the purpose of this study.

As explained in Section 2.2, a way to solve stochastic FLPs is a two-stage solution method. The decision on the locations are generally first-stage decisions whereas the assignment of TBPs to locations is second-stage (Snyder, 2006). This theory will be used in this project. In Section 4.1, it is described how the optimal turnout facility locations are determined. Section 4.2 explains how the optimal capacity distributions are determined for these turnout facilities.

4.1 Average model

In the first stage of the solution method, the optimal location for the new turnout facilities will be decided. This can be done with the STM from Section 3.1. However, just as for the CSTM, the computational time of the STM exceeds a reasonable computational time for the purpose of this study because of the large number of decision variables. When looking at the STM, it becomes clear that a TBP will always be assigned to its closest open turnout facility. In the example of Section 3.1.6, turnout facilities are opened in Heerlen and Voerendaal. Looking at Tables 2 and 5, the TBPs are assigned to the closest facility. For example, Table 2 states that the distance between Beekdealen and Heerlen is equal to 10 and the distance between Beekdealen and Voerendaal is equal to 15. Therefore, as stated in Table 5, Beekdealen is assigned to the turnout facility in Heerlen. This is the case because the objective is only dependent on the distance and no restriction on the capacity of the turnout locations is given. This means that $x_{ij1} = x_{ijw}\ \forall\ i \in D, j \in L, w \in Q$. Snyder (2006) states that if demand must be assigned to the same turnout facility in every scenario, the problem reduces to a deterministic problem in which the uncertain parameter can be replaced by their means. When combining the fact that in the Scenario Turnout Model, a TBP is always assigned to the same turnout facility and the theory of Snyder (2006), the Scenario Turnout Model can be rewritten as:

$$\min \sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} * d_{ij} * \overline{z_i} \qquad (30)$$

S.T

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall\ i \in D \qquad (31)$$

$$\sum_{j=1}^{n} y_j = p \qquad (32)$$

$$x_{ij} - y_j \leq 0 \qquad \forall\ i \in D, j \in L \qquad (33)$$

$$y_j \in \{0,1\} \qquad \forall j \in L \qquad (34)$$

$$x_{ij} \in \{0,1\} \qquad \forall\ i \in D, j \in L \qquad (35)$$

Where $\overline{z_i}$ stands for the average of the TBPs per geographical area for all scenarios ($\overline{z_i} = \frac{\sum_{w=1}^{q} z_{iw}}{q}$ $\forall\, i \in D$). For the rest of this project, the mathematical model above is referred to as the Average Scenario Turnout Model (ASTM). To further elaborate on the ASTM, the example of Section 3.1.6 is used again. For this, first the $\overline{z_i}$ is calculated. The $\overline{z_i}$ values (following from Table 1) are stated in Table 6 below.

Table 6: Calculation of average TBPs

i	TBPs in Scenario 1	TBPs in Scenario 2	$\overline{z_i}$
Beekdealen	3	3	3
Brunssum	2	3	2.5
Heerlen	5	5	5
Landgraaf	1	1	1
Voerendaal	4	3	3.5
Kerkrade	3	4	3.5
Simpelveld	3	2	2.5

Applying the theory of $x_{ij1} = x_{ijw}\ \forall\, i \in D, j \in L, w \in Q$ for Table 5, it is assumed that the optimal values of the decision variable x_{ij} in the ASTM will be the same as the values in Table 5. Now, the objective function of the ASTM can be calculated by multiplying the d_{ij} values for the combination of *i* and *j* values for which the x_{ij} = 1 with the $\overline{z_i}$ values. Table 7 represents this calculation. The closest facility for every municipality is based on the optimal x_{ijw} variable from Table 5.

Table 7: Calculation of objective function of the example for the ASTM

i	Closest facility	Distance to closest facility	$\overline{z_i}$	$\overline{z_i} * d_{ij}$
Beekdealen	Heerlen	10	3	30
Brunssum	Heerlen	5	2.5	12.5
Heerlen	Heerlen	0	5	0
Landgraaf	Heerlen	7	1	7
Voerendaal	Voerendaal	0	3.5	0
Kerkrade	Heerlen	4	3.5	14
Simpelveld	Voerendaal	2	2.5	5
				68.5

The value of the objective function again becomes 68.5 just as in the example of Section 3.6 (Table 3). This means that for the STM, the optimal x_{ijw} values are the same values as the optimal x_{ij} values in the ASTM ($\forall\, w \in Q$).

The fact that the Scenario Turnout Model will give the same results as the Average Scenario Turnout Model is also in line with the articles of Weaver and Church (1983) and Contreras, Cordeau, & Laporte (2011).

So, in the first stage of the solution method, the optimal locations of the turnout facilities will be determined by the ASTM because the ASTM is less computationally complex than the STM.

4.2 Simplified capacity model

For the CSTM, it is not possible to make an average model as is done in Section 4.1. This is the case because now when a facility has reached its capacity, the TBP will be served from another turnout facility. Looking at Table 6, when opening a facility in Heerlen and Voerendaal, the amount of TBPs assigned to the facility in Heerlen will be equal to fourteen and sixteen and the amount of TBPs assigned to Voerendaal will be equal to seven and five in scenario one and two respectively. If, for example, the capacities of Heerlen and Voerendaal are fifteen and six, a TBP originally assigned to Voerendaal will be served by the facility in Heerlen in Scenario One and a TBP originally assigned to Heerlen will be served by the facility in Voerendaal in the second scenario. However, when looking at the average TBPs in Table 6, the TBPs assigned to Heerlen and Voerendaal are equal to fifteen and six (which are equal to the capacity in this example). So for the average case, no TBP will be served from a facility that is not its closest open facility.

Therefore, we need to introduce a new model that determines the optimal capacity distributions. For the second stage of the two-stage solution method, the results of the ASTM will be input for a new model. This new model is a simplified version of the CSTM. In this new model, the dataset *L,* which represents the locations of potential turnout facility locations, will be reduced to only the turnout facilities that will be opened according to the ASTM. This makes the number of decision variables smaller. Also, constraints (24), (25), and (28) of the Capacity Scenario Turnout Model can be removed because the decision on where to place the turnout facilities is already determined by the Average Scenario Turnout Model. This new model is referred to as the Simplified Capacity Scenario Turnout Model (SCSTM) and is stated below:

$$(\min \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{w=1}^{q} x_{ijw} * d_{ij} * z_{iw})/q \qquad (36)$$

S.T

$$\sum_{j=1}^{n} x_{ijw} = 1 \qquad \forall\, i \in D, w \in Q \qquad (37)$$

$$\sum_{i=1}^{m} x_{ijw} * z_{iw} \leq cap_j \qquad \forall\, j \in L, w \in Q \qquad (38)$$

$$\sum_{j=1}^{n} cap_j = T \qquad (39)$$

$$x_{ijw} \in \{0,1\} \qquad \forall\, i \in D, j \in L, w \in Q \qquad (40)$$

$$cap_j \in Z_+ \qquad \forall\, j \in L \qquad (41)$$

In conclusion, the CSTM will be solved with a two-stage method where, first the ASTM is used to determine the optimal locations and secondly, these optimal locations are used in the SCSTM to determine the optimal capacity distribution. The models will be solved using the state-of-the-art solver 'Gurobi'. If, despite the changes made for the potential turnout facility location set, the computational time will still be too long, the decision is made to lower the number of scenarios until the computational time is acceptable. If the number of scenarios is insufficient to address the uncertainty, techniques such as row reduction can be used to decrease the computational time for more scenarios.
A numerical comparison between this two-stage method and the Capacity Scenario Turnout Model is made with a small dataset to see how the two-stage solution method performs. This comparison is stated in Appendix 1 and shows that the results of the two-stage solution method give approximately the same results.

5. Numerical design

This chapter specifies the numerical experiment (SO4) that will be executed in Chapter 6.

We create different instances and scenarios to account for the uncertainty in TBP data. An instance is a set of all parameters that are needed to optimize both the ASTM and the SCSTM once. To make these instances, a decision has to be made on which geographical level TBPs will be clustered. The decision is made to cluster TBPs based on municipalities. When clustering on municipalities, the Netherlands will be divided into 342 areas. For now, this seems enough areas to solve this problem. As discussed before, the TBPs residing in the same municipality will have their home address codes as a central location in this municipality. This project will use the geographic coordinates of city halls in all municipalities across the Netherlands as the representation of TBPs. Under this approach, a TBP residing in a specific municipality, such as Amsterdam, will have its 'home location' represented by the geographical coordinates of Amsterdam's city hall. If a municipality has multiple city halls, the most central one will be chosen. Using city halls as 'home locations' allows for straightforward calculations of distances and spatial relationships, thus simplifying the analysis. It reduces data preparation time since city hall coordinates are readily available in Google Maps, eliminating the need to individually geocode precise residential addresses for TBPs.

5.1 Exploration of TBP data

For the TBP data, the DKDB has been given the option to either choose to share information about the home addresses of TBPs throughout the years or keep this information confidential. Due to the sensitive nature of this data, it has been determined that this data should be kept confidential. Therefore, only information available on the internet is used for this project. As explained before, because not much information about TBPs is publicly available, the assumption is made that TBPs follow the same trends as the overall population of the Netherlands. As is concluded in Chapter 2.2, as of 2017, in the Netherlands, people tend to move out of the populated areas in the Netherlands. Because of this, only data from 2017 and later will be used for this study.

In sections 5.1.1 and 5.1.2, we examine and model two trends (see section 2.2): People moving out of the populated areas in the Netherlands and the increasing numbers of TBPs.

To make a prediction growth factor, a forecast method has to be chosen because historical data will be used to make a forecast for the future. For both trends, the Simple Average (SA) method is chosen as the forecast model. In SA, the forecast is equal to the average of historical data of N periods. Blanc and Setzer (2016) mention that SA is not systematically outperformed by other forecast methods when small datasets are used. Because, for the scope of this project, only data from 2017 and later is used, the SA is expected to be a useful forecast method. In both 5.1.1 and 5.1.2, this SA is used to make a prediction growth factor. In Appendix 2 this SA is compared to the Moving Average (MA) forecast method. In this appendix, the prediction growth factors of the two forecast methods are compared as well as the solutions these growth factors would result in.

5.1.1 People moving out of the populated areas in the Netherlands

First, the trend of people moving out of the populated areas in the Netherlands is examined. As of 2017, the number of people moving out of these areas is bigger than the other way around (Centraal Bureau voor de Statistiek, 2022). This trend can be assessed by utilizing Figure 5 which is retrieved from Centraal Bureau voor de Statistiek (2022).

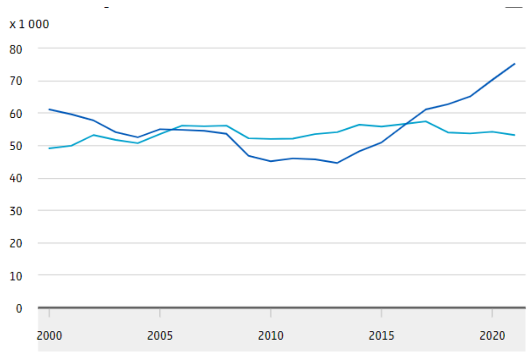

Figure 5: People moving in and out of the populated areas in the Netherlands

In this figure, the light blue line represents the number of individuals moving to populated areas in the Netherlands while the dark blue line represents the individuals that move the other way. As can be seen, in the last few years, more people have moved away from the populated areas. Because this trend is visual since 2017, only data starting from 2017 will be used. The data that is used to visualize Figure 5 is stated in Table 8 below.

Table 8. Numbers of people moving in and out of the populated areas in the Netherlands

Year	To populated areas (*1000)	Away from populated areas (*1000)
2017	57.4	61.1
2018	54.0	62.7
2019	53.7	65.1
2020	54.2	70.2
2021	53.2	75.1

When using the SA method, the prediction growth factor as introduced by Peterson et al. (2003) can be calculated. The SA is used to calculate the average incline of people moving away from populated areas and the average decline of people moving to these areas for two consecutive years. In the end, these SA values are added to the total number of the last year for which the data is available to determine the prediction growth factor. The in- and declines in emigration and immigration of populated areas are stated in Table 9 below.

Table 9. Difference in moving data for consecutive years

Year	To populated areas difference with the year before (*1000)	Away from populated areas difference with the year before (*1000)
2018	- 3.4	1.6
2019	- 0.3	2.4
2020	0.5	5.1
2021	- 1.0	4.9

Based on the data in Table 9, the SA concludes that immigration into populated areas drops by 1,050 persons each year and emigration rises by 3,500 persons each year. For 2021, 75,100 persons have left the populated regions while 53,200 immigrated to these regions. When adding the SA values to the number of 2021, the expectation is that 52,150 people will move into and 78,600 people will move out of the populated areas in the Netherlands. This is a difference of 26,450 individuals, which means that 26,450 more people will leave the populated areas. According to Wikipedia-bijdragers (2023), the Netherlands has 17,823,087 inhabitants. When dividing 26,450 by the inhabitants, it means that

0.148% of the whole population of the Netherlands moves out of the busy regions. This percentage will be used as the prediction growth percentage when making the instances. For this trend, a classification on which municipalities count as populated has to be determined. For this, Figure 6 is used as well as data from AlleCijfers.nl (2023) and Wikipedia-bijdragers (2023). Figure 6 graphically illustrates the distribution of municipal population across the Netherlands. The figure reveals that there are not a lot of municipalities in the Netherlands with more than 150,000 inhabitants. Consequently, for this project, all municipalities surpassing this threshold are designated as 'populated' municipalities. Compiled details, as of January 2023, are retrieved using Wikipedia-bijdragers (2023).

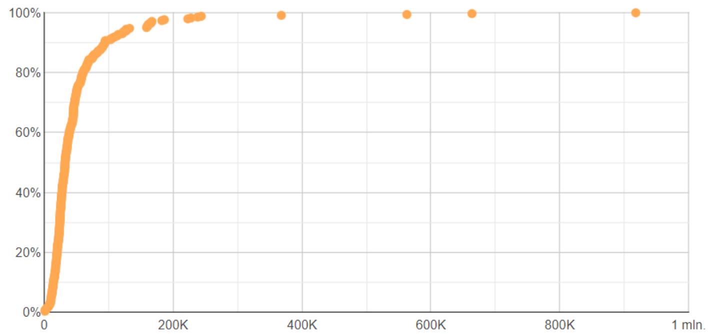

Figure 6: Distribution of inhabitants of municipalities in the Netherlands

5.1.2 Increasing numbers of TBPs

Secondly, the trend of the increasing number of TBPs is investigated. For this, Centraal Bureau voor de Statistiek (n.d.a) can be used. Again, SA is used to determine the average growth in population growth in two consecutive years. The data from Centraal Bureau voor de Statistiek (n.d.a) is stated in Table 10.

Table 10. Data about population growth

Year	Population growth	Growth compared to the year before
2017	99,577	
2018	101,079	1,502
2019	125,422	24,343
2020	67,830	-57,592
2021	115,257	47,427
2022	220,619	105,362

Using the SA, it is concluded that the average population grows with 24,208 persons every year. When adding this number to the number of 2022 that Centraal Bureau voor de Statistiek (n.d.a) provides us, the prediction is that the population will grow with 244,827 persons. When comparing this number to the total inhabitants of the Netherlands, the growth factor will be equal to 1.374%. Again, this growth factor is used as the prediction growth factor for the second trend when making the different instances.

5.2 Growth factors

As mentioned by Peterson et al. (2003), the construction of instances facilitates the exploration of uncertainties affecting future consequences. To make good instances, values below and above a certain prediction must be used (Peterson et al. 2003). These instances are based on the two trends that have been investigated in Section 5.1. In that section, it is shown that 0.148% of the whole population of the Netherlands moves out of the busy regions and that the population of the Netherlands will grow by 1.374%. These two growth percentages will be used as the prediction values for both trends. For this project, it has been decided that for both trends, an ‘above’ and ‘below’ prediction growth percentage is needed. The alternatives of the prediction values should be both plausible and relevant (Peterson et al. 2003). These alternatives will be based on the historical data that have led to the prediction values of both trends. This is done because the uncertainties chosen to define the alternatives should have differences that are directly related to the defining issue. They should imaginatively but plausibly push the boundaries of commonplace assumptions about the future. For the alternatives, reasonable instances can also be decided upon the past (Peterson et al. 2003). To determine the prediction growth factors in Chapter 3,1, the SA forecast method was used. This method calculates the average incline/decline in data. To determine the ‘above’ and ‘below’ growth factors, data from Tables 8, 9, and 10 is used. However, for the ‘above’ and ‘below’ growth factors the most optimistic and pessimistic data is used instead of the SA.

5.2.1 Growth factor for people moving out of the populated areas in the Netherlands.

To find the Alternative growth factors for the trend of people emigrating from populated areas in the Netherlands, again Centraal Bureau voor de Statistiek (2002) will be used. As Peterson et al. (2003) mention alternative instances can be derived from the past because this data is plausible and relevant. As concluded in Chapter 5.1, as of 2017, more people tend to move out of the populated regions in the Netherlands than vice versa. Since the ‘prediction’ growth factor calculated in Chapter 5.1.1 is based on average growth for consecutive years, it is now also important to look at the growths in consecutive years.

5.2.1.1 Above Prediction

First, the focus will be on the ‘above’ prediction growth factor for this trend. For this, the data from Tables 8 and 9 will be used. In 2019, 65,100 people moved out of the populated areas while in 2020 70,200 people moved in this direction. This is the biggest incline when comparing consecutive years. Since the focus is on the ‘above’ prediction instance, for the people that move to busy areas, we have to look for the biggest decrease in consecutive years. In 2018, there was a decrease of 3,400 persons moving towards the populated regions compared to 2017. This was the biggest decrease in the last few years. To make the ‘above’ instance, the same procedure is done for the growth percentage that is calculated in Chapter 5.1.1. This means that the numbers are added up to the numbers of 2022. This results in the fact that for the above prediction instance, it is assumed that 80,200 people will emigrate while 49,800 will immigrate to the populated areas. When making the same calculation for the growth percentage as done in Chapter 5.1.1, the growth percentage for the ‘above’ prediction will be 0.171%.

5.2.1.2 Below prediction

Again, the data from Tables 8 and 9 will be used to calculate the growth factor of the below growth factor for this trend. The same procedure as for the ‘above’ growth factor described above will be executed. However now, the lowest increase in the number of people that move out and the highest increase in the number of people that move in in two consecutive years will decide the growth factor. The lowest increase in the number of people who move out of the busy areas of the Netherlands was between the years 2017 and 2018. In 2017, 61,100 people moved out while in 2018 62,700 made the same move. This means an increase of 1,600 people. For moving into these regions, the biggest incline has been between the years 2016 and 2017 when 800 more people moved into the busy regions compared to the year before. To make the ‘below’ prediction, these differences will be added up to the numbers for 2022. When doing the calculation for the growth percentage, the 'below' prediction will be 0.131%.

5.2.2 Growth factors for the increasing numbers of TBPs

To calculate the growth factors for the above and below instances of this trend, the data in Table 10 will be used. The growth factor for both instances will be conducted in the same way as is done in Chapter 5.1.2. So again, the population increase of two consecutive years will be compared.

5.2.2.1 Above prediction

For the 'above' prediction instance, Centraal Bureau voor de Statistiek (n.d.a) is used to find the highest increase in two consecutive years. The highest increase was between 2021 and 2022 where in 2022 the population grew by 105,362 more than the year before. When adding this number to the number of 2022, as is done in Chapter 5.1.2, the expectation is that the population will grow to 325,981. When calculating the growth percentage in the same way as done before, the conclusion is that the 'above' growth factor for trend 2 is 1.823%.

5.2.2.2 Below prediction

For the 'below' prediction growth factor for trend 2, the lowest increase for the data in Table 10 needs to be known. In the period between 2017 and 2022, there has been a decline in growth percentages. In 2019 the population grew to 125,422 while in 2020 the population grew to 67,830. This means that there was a difference of -57,592 between these two consecutive years. Since this difference is negative, this number will be subtracted from the population growth in 2022. This means that the expectation for the 'below' growth factor for this trend is that the population will grow to 163,027 which results in a growth percentage of 0.915%.

Now, the growth factors for the instances can be calculated. Because there are three different growth factors for two trends, a total of nine instances will be created. The growth factors can be calculated by multiplying the factors from Sections 5.1.1, 5.1.2, 5.2.1.1, 5.2.1.2, 5.2.2.1, and 5.2.2.2 with each other. An overview of all the growth factors is stated in Table 11 below. Note that, as explained earlier, there are different growth factors for municipalities that are marked as 'populated' compared to the ones that are marked as 'unpopulated'. As explained earlier, the municipalities with less than 150,000 residents are indicated as 'unpopulated'. This means that the expected number of TBPs per municipality for these municipalities will be multiplied by the growth factor while the expected TBPs for the municipalities with more than 150,000 residents will be divided by the growth factor.

Table 11. Growth factors for the nine instances

			Trend 2		
			Below prediction	prediction	Above prediction
Trend 1	Below prediction	Unpopulated	1.01047	1.01507	1.01956
		Populated	1.00783	1.01241	1.01690
	Prediction	Unpopulated	1.01064	1.01524	1.01974
		Populated	1.00766	1.01224	1.01649
	Above prediction	Unpopulated	1.01088	1.01547	1.01997
		populated	1.007427	1.01201	1.01649

5.3 Scenario generation

To make the scenarios, first, an estimation of the number of TBPs per municipality has to be made. This is calculated with the use of the total residents of the Netherlands that is received from Wikipedia-bijdragers (2023) and the number of residents per municipality retrieved from Wikipedia-bijdragers (2023) respectively. With the use of these numbers, the inhabitants percentage for each municipality can be calculated by dividing the inhabitants by the total inhabitants of the Netherlands. For example, according to Wikipedia-bijdragers (2023), the municipality of Amsterdam has 922,679 inhabitants. When dividing this number by the 17,823,087 total inhabitants, the inhabitant percentage is equal to 0.05177. Willem Woelders, the director of the DKDB, disclosed that approximately 400 individuals necessitate protection in the Netherlands (Alledaagse Vragen, 2022). Since the assumption was made that the distribution of TBPs is equal to the distribution of inhabitants across the Netherlands, the number of TBPs per municipality can be calculated by multiplying the inhabitant percentage by 400. When taking the municipality of Amsterdam as an example again, when multiplying 0.05177 with 400, the expectation is that Amsterdam would host 20.7075 TBPs. For each of the 342 municipalities in the Netherlands, the expected number of TBPs is calculated.

Next, to get the expected number of TBPs for the nine instances, the previously calculated number of TBPs per municipality will be multiplied by the growth factors stated in Table 11. This means that, when using the 'above' growth factor for both trends, Amsterdam will host 21.0490 (20.7075 * 1.01649) TBPs. For each of the nine instances and the 342 municipalities, the expected TBPs per municipality per instance are calculated. These numbers will be input to generate the distinct scenarios.

The scenarios will consist of municipalities and the number of TBPs these municipalities will host. An example of two different scenarios is given in Table 1. When looking at the Average Scenario Turnout Model and the Simplified Capacity Scenario Turnout Model, one of the input parameters is the number of TBPs per municipality per scenario (z_{iw}). The scenarios will be generated based on gathering a random number from a Poisson distribution. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. The Poisson distribution is often used in facility location problems. Shavarani, Nejad, Rismanchian, & Izbirak (2018) used the Poisson distribution to generate demand for their distance-constrained mobile hierarchical facility location problem. The model of Shavarani et al. (2018) tries to locate charge stations for delivery drones. Like our model, the model of Shavarani et al. (2018) assumed that one demand is satisfied each trip and a demand is allocated to its nearest recharge station. Also, Jafari & Arkat (2016) use a Poisson distribution to generate demand for their network location problem. The Poisson distribution is often used for real-valued random variables whose distributions are not known.

To generate random values from a Poisson distribution, the numpy.random.poisson function in Python is used. This function has a lambda and size as input parameters and returns samples from a Poisson distribution. The size, also referred to as output shape, is equal to the number of samples one wants to draw. For example, if 1000 samples are needed, the size input parameter has to be equal to 1000. In this case, the size parameter is equal to one because only one random value has to be drawn for each municipality and scenario. The other input parameter, the lambda, is equal to the expected value and must be greater than or equal to zero. In this case, the lambda will be equal to the number of TBPs per municipality per instance. When taking the example of the municipality Amsterdam again, the lambda for the instance where both trends are in the 'above' growth factor will be equal to 21.0490. However, when using the 'below' growth factors for both trends, the lambda will be equal to 20.8696 (20.7075 * 1.00783). By choosing these numbers as the lambda, the expected number of TBPs per municipality per instance will still correspond to the instances that are made beforehand. The numpy.random.poisson function automatically rounds the unrounded lambda input parameters. By using this method, it is possible for each scenario that the total number of TBPs will be lower or higher than the 400 TBPs that Willem Woelders mentioned in Alledaagse Vragen (2022). However, because the lambda is equal to the expected value, the average number of TBPs will always be a bit higher than 400. This is the case because all the growth factors are higher than one and if these growth factors are multiplied by the number of TBPs, there will be always more TBPs than without these growth factors.

In conclusion, there will be nine instances all with their own z_{iw} values. These values are based on gathering random values from a Poisson distribution with the use of the numpy.random.poisson function in Python. The lambda input parameter for this Python function is equal to the number of TBPs per municipality. Because the numbers of TBPs per municipality are different for each of the nine instances, the z_{iw} values for each instance still correspond to their instance. For now, it is assumed that 1000 scenarios will be generated for each instance. With the use of nine instances and 1000 scenarios per instance, it is assumed that any possible distribution of TBPs is accounted for. If the computational time to solve the mathematical model for 1000 scenarios is too long, the number of scenarios will be reduced until an acceptable computational time is reached.

5.4 Instance generation

As mentioned, to account for the uncertainty in TBP data, nine different instances will be made. Each of the nine instances will be optimized individually. Below, the data that is used for each of the instances is explained. Note that for every instance, the TBP data is different while all other datasets are the same.

5.4.1 TBP data.

The TBP data represents the parameter z_{iw} in both the ASTM and the SCSTM. Note that in the Average Scenario Turnout Model the parameter $\overline{z_i}$ (from Section 3.6) represents the average of $z_{iw}\ \forall\ w\ \in Q$. How this z_{iw} parameter is generated is explained in Chapter 5.3.

5.4.2 Potential turnout facility data

For the potential turnout facility data, it was concluded that the same geographical data will be used as for the TBP data. This means that the geographic locations of every city hall across the Netherlands are stated as potential turnout facilities. Of course, in real life, a city hall cannot be a potential turnout facility. However, by choosing municipalities as potential turnout facility locations, we make sure that 342 different potential locations are evenly spread out across the Netherlands.

5.4.3 Maximal number of facilities to open

The decision on the number of turnout facilities to open will be crucial for the first phase of the two-stage method. It is decided that this will be an iterative process. This means that first, for each of the nine instances, the problem will be optimized for one facility, then for two, three, etc. To determine the stopping criterion, the objective of the current optimization is compared to the previous optimization. For example, if the objective function of the optimization for four turnout facilities is equal to 10,000 and for three turnout facilities it is equal to 16,500. Then, the objective increase is calculated as follows: $\frac{10{,}000-16{,}500}{16{,}500} * 100 = -39.39\%$. The decision is made that the iterative process is stopped when the objective increase is bigger than -10% for the second time.

5.4.4 Distance Matrix

At last, a distance matrix has to be made between the TBPs and potential turnout facilities. Since the same geographic locations are used for the TBPs and the turnout facilities, the diagonal of the distance matrix will be zero. Because no exact road distances are available, we use the great circle distance (retrieved from the pyproj.Geod package in Python) between each pair of locations.

6. Results

In this chapter, we determine the optimal locations and capacity distributions of various numbers of facilities with the objective of minimizing travel times for DKDB agents. Section 6.1 will show the results of the ASTM while Section 6.2 will present the results of the SCSTM. At last, Section 6.3 will show the calculation and results of some Key Performance Indicators (KPIs).

6.1 Optimal locations

In this section, we obtain the optimal turnout locations for different numbers of facilities. Since the ASTM and the STM return the same optimal solution (see Section 4.1), we use the computationally less complex ASTM. The ASTM is optimized in Gurobi over 1000 scenarios. Table 12 shows the results for nine different instances. Table 12 provides the range of the objective functions for the different instances.

Table 12. Optimal locations and objectives for nine instances

Facilities	instances	Solution	Objective
1	9	Utrecht	[35990.094 – 36409.566]
2	9	Arnhem, Waddinxveen	[25370.073 – 25684.045]
3	9	Alphen aan den Rijn, De Wolden, Maashorst	[20296.388 – 20532.293]
4	9	Amsterdam, De Wolden, Maashorst, Rotterdam	[16582.341 – 16774.181]
5	9	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	[14953.666 – 15117.470]
6	6	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	[13648.439 – 13834.334]
	2	Amsterdam, Deventer, Helmond, Noordenveld, Rotterdam, Zeist	[13754.875 – 13772.580]
	1	Amsterdam, Bunnik, Deventer, Helmond, Noordenveld, Rotterdam	13722.802
7	1	Almelo, Amsterdam, Arnhem, ‘s-Hertogenbosch, Oostwellingwerf, Rotterdam, Sittard-Geleen	12589.907

As Table 12 shows, all nine instances give the same solution for opening one to five turnout facilities. For opening six turnout facilities, the nine instances give three different solutions. The first solution is generated by six instances, the second by two out of nine instances, and the third solution is generated by only one instance. At last, only one instance generates a solution for opening seven turnout facilities. For the other eight instances, the stopping criterion of Section 5.4.3 is reached at six facilities.

When looking at the optimal solution in Table 12, there are a few notable things. At first, locations barely ever repeat. Appendix 3 showcases the geographic locations of the optimal locations from Table 12 and Figure 7 represents the number of inhabitants per square kilometer in the Netherlands.

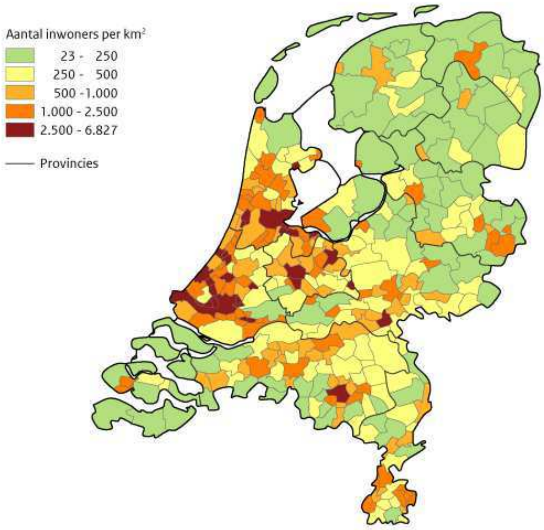

Figure 7: Inhabitants per squared kilometer in the Netherlands

Figure 8: Two different service areas when opening two turnout facilities

When opening two facilities, Waddinxveen is a municipality located in the west of the Netherlands while Arnhem is located in the east. As can be seen in Figure 7, the west side of the Netherlands is more densely populated than the east side. Assuming that TBPs will be served from the closest open facility when opening two turnout facilities, the Netherlands can be divided into two service areas. Figure 8 represents the approximate service areas of Waddinxveen (left yellow star) and Arnhem (right yellow star). As can be seen in Figure 8, the service area of Waddinxveen is smaller than the service area of Arnhem. This means that the agents working from the turnout facility in Arnhem tend to drive more kilometers to TBPs. However, taking into account Figure 7, the service area of Waddinxveen is more densely populated. This means that more agents are needed in this area. It looks like the model aims to find service areas where the total travel distance traveled by DKDB agents is balanced. For Waddinxveen in this example, this means that there are a lot of TBPs but the distances are small. For the service area of Arnhem, this means that the number of TBPs is lower but the distances are bigger. That results in the fact that optimal locations barely repeat when increasing the number of facilities. Secondly, Table 12 shows that, when opening four or more turnout facilities, Amsterdam and Rotterdam appear in every optimal solution although the conclusion made before that municipalities barely repeat. The fact that these two municipalities repeat lies in the assumption to cluster on municipality. Amsterdam and Rotterdam are by far the two municipalities with the biggest population in the country. This means that when a turnout facility is opened in these municipalities, the distances to the TBPs residing in these municipalities would be equal to zero. Also, another reason for the repetition is that both municipalities are close to some other big municipalities such as The Hague, Leiden, and Utrecht. Therefore, it is expected that even when opening 100 facilities, Amsterdam and Rotterdam will still be part of the solution. When adding more turnout facilities, the expectation is that more populated municipalities such as The Hague (which is the third populated municipality of the Netherlands) will become one of the optimal locations as well. At last, the optimal locations represented in Table 12 that are located on the east side of the Netherlands tend to go further into the corners of the country when the number of turnout facilities increases.

Figure 9: Locations of Maashorst, Eindhoven, and Someren

Figure 10: Locations of De Wolden, Midden-Drenthe, and Noordenveld

Figure 9 visualizes the geographic locations of Maashorst, Eindhoven, and Someren which are presented in the optimal solutions when opening four, five, and six turnout facilities. As Figure 9 reveals, the solution tends to go to the southeast corner of the country as the number of facilities increases. For the northeast side of the Netherlands, this is visual in Figure 10 where the locations of De Wolden, Midden-Drenthe, and Noordenveld, which are present as optimal locations for opening four, five, and six facilities, are visualized.

The result stated in Table 12 indicates that the location choice is robust against different realizations of future demand. As mentioned by Snyder (2006), the goal of robust optimization is to find a solution that will perform well under any possible realization of the random parameters. We further substantiate the robustness by additional experiments on more demand fluctuations (Appendix 4) and using normal instead of Poisson distributed demand (Appendix 5). To experiment with demand fluctuations, two additional instances are created. These instances are called the 'extreme' instances. One instance will represent an extreme version of the instance where both trends use the 'above' growth factor from Table 11. For this instance, both growth factors are doubled. The other instance will be an extreme version of the 'below' 'below' instance from Table 11 where both growth factors are halved. The ASTM is used to find the optimal locations for these extreme instances. The results of this experiment are stated in Appendix 4. Appendix 4 shows that for opening one to five facilities, the extreme instances give the same solution as the nine other instances in Table 12. For opening six turnout facilities, both extreme instances give the solution that is represented by six out of nine instances in Table 12. The objective values of the extreme instances in Appendix 4 are also close to the objective values for the normal instances in Table 12.

For the other experiment on robustness, scenarios will be built differently. In Section 5.3, it is explained how a Poisson distribution is used to generate scenarios. For this numerical experiment, the normal distribution is used to generate scenarios. How to generate scenarios based on a normal distribution and the results of this numerical experiment are stated in Appendix 5. Again, this numerical experiment uses the ASTM to get results. For opening one to five facilities, Appendix 5 shows the same solutions as in Table 12 and Appendix 4. For opening 6 solutions Appendix 5 presents the same solutions as in Appendix 4 and the solution that is generated by six instances in Table 12. For the numerical experiment where the scenarios in the instances are generated by a normal distribution, no instance generates a solution with seven facilities. This means that for all instances the stopping criterion explained in Section 5.4.3 is reached. The results of the experiments on robustness in Appendix 4 and 5 give the same results as the results in Table 12.

Table 12, and appendices 4 and 5 all show the same results for opening one to five facilities. Also, seventeen out of the twenty instances that are used to gather the results for Table 12 and appendices 4 and 5 give the same solution for opening six facilities. This means that the solution for opening six facilities is also considered robust. So the optimal location for opening one turnout facility will be Utrecht. Arhnem and Waddinxveen are the optimal locations for opening two turnout facilities, Alphen aan den Rijn, De Wolden, and Maashorst are the optimal locations for opening three turnout facilities. For opening 4 facilities, the municipalities of Amsterdam, De Wolden, Maashorst, and Rotterdam are optimal. The municipalities of Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, and Rotterdam are optimal for locating five turnout facilities. At last, Amsterdam, Bunnik, Deventer, Noordenveld, Someren, and Rotterdam are the optimal locations when opening six new turnout facilities. Because of the results in Table 12 and appendices 4 and 5, it does not matter how the growth factors and scenarios are generated, the solutions are fully robust. This means that, regardless of which scenario unfolds in the future, it can be concluded that the municipalities mentioned above are optimal locations. For determining the optimal capacity distributions, the optimal locations from Table 12 are used as input for the SCTSM.

6.2 Optimal capacity distribution

In this section, we obtain the optimal capacity distributions for the different numbers of turnout facilities. To get the optimal capacity distributions for the facilities, the SCSTM from Section 4.2 is used. For robustness purposes, the SCSTM will be used in different settings. In the first setting, it is assumed that capacity exceeds the number of TBPs. In the second setting, it is assumed that the capacity equals the number of TBPs. At last, just as for the ASTM, an average model is made. This average model is mathematically explained in Appendix 6 and is referred to as the Average Simplified Capacity Scenario Turnout Model (ASCSTM). This ASCSTM is studied for comparison reasons only. This means that, just as is done between the STM and the ASTM, an investigation is done on how the average solution performs compared to the SCSTM. Because of the theory mentioned in Section 4.2, it is expected that the ASCSTM will not give the same solution. This is the case because now when a facility has reached its capacity, a TBP has to be served from another turnout location. Note that there is also a possibility to investigate a setting with less capacity than TBPs. However, for such a setting, we consider the additional challenge of allocating capacity over these TBPs which is out of scope for this project.

In Appendix 1 it is numerically shown that the two-stage method has approximately the same performance as the CSTM. Next to the comparison of the performance of the solutions, it was also investigated to which extent using fewer scenarios results in similar results. This was investigated because solving the SCSTM over 1000 scenarios, the computational time to solve all instances would take approximately 100 days. For the optimization process of Appendix 1, the decision was made to optimize over 50 scenarios and see how these solutions performed. As can be seen in Appendix 1 the biggest difference in capacities between the two-stage method and the Capacity Scenario Turnout Model is 0.02. For all the other capacities, there is only 0.01 difference between the two solution methods. Because using 50 scenarios already proves that the two-stage method performs well in getting results for the CSTM and the fact that it would take approximately 10.5 days to solve for nine instances, the decision was made to use 100 scenarios for the SCSTM and the ASCSTM.

The results of the three settings mentioned above are stated in Tables 13, 14, and 15. Note that for these settings, the same scenarios are used.

Table 13. Optimal capacity distribution for the setting if the capacity exceeds the number of TBPs

	instances	Solution						
1	9	Utrecht						
		[1]						
2	9	Arnhem	Waddinxveen					
		[0.46-0.47]	[0.53-0.54]					
3	9	Alphen aan den Rijn	De Wolden	Maashorst				
		[0.50-0.51]	[0.21-0.22]	[0.28-0.29]				
4	9	Amsterdam	De Wolden	Maashorst	Rotterdam			
		[0.25-0.26]	[0.20-0.21]	[0.26-0.27]	[0.27-0.28]			
5	9	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam		
		[0.25-0.26]	[0.15-0.16]	[0.16-0.17]	[0.15-0.16]	[0.27]		
6	6	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren	
		[0.17]	[0.17-0.18]	[0.14-0.15]	[0.10-0.11]	[0.26-0.27]	[0.14-0.15]	
	2	Amsterdam	Deventer	Helmond	Noordenveld	Rotterdam	Zeist	
		[0.17]	[0.14]	[0.16-0.17]	[0.11]	[0.26]	[0.16]	
	1	Amsterdam	Bunnik	Deventer	Helmond	Noordenveld	Rotterdam	
		[0.17]	[0.17]	[0.14]	[0.17]	[0.10]	[0.26]	
7	1	Almelo	Amsterdam	Arnhem	's-Hertogenbosch	Oostwellingswerf	Rotterdam	Sittard-Geleen
		[0.07]	[0.23]	[0.13]	[0.15]	[0.10]	[0.25]	[0.07]

Table 14. Optimal capacity distribution for the setting if the capacity equals the number of TBPs

	instances	Solution					
1	9	Utrecht					
		[1]					
2	9	Arnhem	Waddinxveen				
		[0.46]	[0.54]				
3	9	Alphen aan den Rijn	De Wolden	Maashorst			
		[0.50-0.51]	[0.20-0.21]	[0.29-0.30]			
4	9	Amsterdam	De Wolden	Maashorst	Rotterdam		
		[0.26-0.27]	[0.20]	[0.27]	[0.27-0.28]		
5	9	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam	
		[0.26]	[0.15-0.16]	[0.16-0.17]	[0.15]	[0.26-0.27]	
6	6	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren

		[0.17-0.18]	[0.17]	[0.15]	[0.10]	[0.26]	[0.14]	
	2	Amsterdam	Deventer	Helmond	Noordenveld	Rotterdam	Zeist	
		[0.18]	[0.17]	[0.15]	[0.10]	[0.26]	[0.14]	
	1	Amsterdam	Bunnik	Deventer	Helmond	Noordenveld	Rotterdam	
		[0.18]	[0.17]	[0.15]	[0.09]	[0.10]	[0.26]	
7	1	Almelo	Amsterdam	Arnhem	‘s-Hertogenbosch	Oostwellingswerf	Rotterdam	Sittard-Geleen
		[0.08]	[0.23]	[0.13]	[0.15]	[0.09]	[0.25]	[0.07]

Table 15. Optimal capacity distribution for the ASCSTM

	instances	Solution						
1	9	Utrecht						
		[1]						
2	9	Arnhem	Waddinxveen					
		[0.62]	[0.38]					
3	9	Alphen aan den Rijn	De Wolden	Maashorst				
		[0.50-0.51]	[0.21]	[0.29]				
4	9	Amsterdam	De Wolden	Maashorst	Rotterdam			
		[0.34]	[0.20]	[0.27]	[0.19]			
5	9	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam		
		[0.34]	[0.16]	[0.17]	[0.14-0.15]	[0.18-0.19]		
6	6	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren	
		[0.27]	[0.16]	[0.14]	[0.09]	[0.18]	[0.15-0.16]	
	2	Amsterdam	Deventer	Helmond	Noordenveld	Rotterdam	Zeist	
		[0.27]	[0.16]	[0.14]	[0.09]	[0.18]	[0.15-0.16]	
	1	Amsterdam	Bunnik	Deventer	Helmond	Noordenveld	Rotterdam	
		[0.27]	[0.16]	[0.14]	[0.15]	[0.10]	[0.18]	
7	1	Almelo	Amsterdam	Arnhem	‘s-Hertogenbosch	Oostwellingswerf	Rotterdam	Sittard-Geleen
		[0.08]	[0.23]	[0.13]	[0.15]	[0.09]	[0.25]	[0.07]

The three tables above show that the results of the ASCSTM (Table 15) are different than the results of the settings with more capacity and equal capacity (Tables 13 and 14 respectively). Therefore, it is concluded that the results of Table 15 are only used for comparison purposes and not for drawing conclusions.

Looking at Tables 13 and 14, there are a few notable things. At first, the capacities are compared to the number of x_{ij} variables assigned to each of the open facilities in the ASTM. For this, the ASTM is optimized over 1000 scenarios. The x_{ij} variables assigned to each facility are divided by the total number of x_{ij} variables. This is done for opening two, three, and four facilities with the use of the ‘above’ ‘above’ instance from Table 11. For two facilities, the percentage x_{ij} variables are 45% and 55% for Arnhem and Waddinxveen respectively. For three facilities the x_{ij} distribution is equal to Alphen aan den Rijn: 41%, De Wolden: 25%, and Maashorst: 35%. At last, the x_{ij} distribution for four turnout facilities is 30%, 15%, 29%, and 26%. When comparing these results with the results of Tables 13 and 14 it can be concluded that there is too much difference between the results to conclude that the optimal capacities distributions coming from the SCSTM are equal to the x_{ij} distributions coming from the ASTM. Secondly, in Section 6.1 it was concluded that the turnout facilities that have a smaller cover area than the other facilities will be in dense areas. This is also in line with the capacities stated in Tables 13 and 14. As can be seen in these tables, the turnout facilities in the dense areas of the Netherlands (Figure 7) must have higher capacities than the ones outside of these areas. When looking at the optimal capacities for opening three facilities, Alphen aan den Rijn, which is located on the west side of the Netherlands (Appendix 3), has an optimal capacity that is equal to the other capacities combined. Also when looking at the capacities for Amsterdam and Rotterdam in Tables 13 and 14, it becomes clear that these turnout facilities will always have a relatively high capacity compared to the other turnout facilities.

Again, for the goal of robust optimization mentioned by Snyder (2006), the two extreme instances introduced in Section 6.1, are used to see if the solutions from Tables 13 and 14 are robust. The results of the optimal capacity distribution of the two extreme instances are stated in Appendix 7. This appendix shows that for all the numbers of turnout facilities, the two extreme instances give the same results with a maximal difference of 0.01.
Again, to further support the robustness of the optimal capacity distributions of the turnout facilities, the scenarios will be built differently. The Scenarios will be generated using a normal distribution as is explained in Section 6.1 and Appendix 5. The two settings that are used to get the results of Tables 13 and 14 are used for this numerical experiment. The results of these experiments are shown in Appendix 8.

Tables 13, and 14, as well as Appendix 7 and 8 show approximately the same results for the optimal capacity distribution for one to five facilities. For every turnout facility, there is a maximal difference of 0.02 when looking at the Tables and Appendix 6. Therefore, it is concluded that the results in Tables 13 and 14 are robust. This means it does not matter how the growth factors and scenarios are generated, the capacity distributions are fully robust and the solution is optimal for all the possible scenarios and instances that can unfold in the future. Also, for opening 6 facilities, Appendix 7 and Appendix 8 prove that the solutions presented in Tables 13 and 14 are robust solutions.
Now that we know that the solutions from Tables 13 and 14 are robust, the optimal capacity distribution can be determined. Because there are two settings and nine instances for each setting, a procedure has to be developed to get to one optimal capacity distribution. Peterson et al. (2003) mention that the 'prediction' instances are more important than the 'below' and 'above' instances when drawing conclusions. This theory will be used for making decisions in this project. For one instance, as can be seen in Table 11, one instance will use the 'prediction' growth factor for both trends. Four instances use one 'prediction' growth factor for one of the two trends. At last, four instances use no 'prediction' growth factor. Because of the theory of Peterson et al. (2003), the instances with a prediction growth factor should have a bigger weight in making decisions. That is why the results of the one instance that makes use of two 'prediction' growth factors will be multiplied by two, the results of the four instances where only one trend uses the 'prediction growth factor will be multiplied by 1.5 and the results of the four instances where no 'prediction growth factor is used will be multiplied by 1. In the end, the sum of these results will be divided by twelve. This method will be referred to as the 'weighted prediction result method'. This method is used for both settings which resulted in Tables 13 and 14. When applying this method, the optimal capacity distribution for opening one facility would be to assign 100% of the capacity to the capacity in Utrecht. For opening two turnout facilities, it is optimal to assign 54% and 46% to Waddinxveen and Arnhem respectively. Assigning 50%

to Alphen aan den Rijn, 21% to De Wolden, and 29% to Maashorst would be optimal for opening three turnout facilities. For four turnout facilities, it is optimal to assign 26%, 20%, 27%, and 27% to Amsterdam, De Wolden, Maashorst, and Rotterdam respectively. The optimal capacity distribution for opening five facilities would be Amsterdam: 26%, Arnhem: 15%, Eindhoven: 17%, Midden-Drenthe: 15%, and Rotterdam: 27%. Lastly, for opening six turnout facilities, it would be optimal to assign 17%, 17%, 15%, 10%, 26%, and 15% to Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, and Someren respectively. Because of the results in Tables 13, 14, and Appendix 6 and 7, it does not matter how the growth factors and scenarios are generated, the solutions are fully robust. This means that, regardless of which scenario unfolds in the future, it can be concluded that the capacity assignment mentioned above is the optimal capacity distribution for the number of facilities.

6.3 Key Performance Indicators

Because the goal of this project is to create additional capacity for the DKDB, two additional Key Performance Indicators (KPIs) will be calculated. For these KPIs, two additional results have to be determined. At first, the total traveled distance for the original setting has to be determined. This will be done in Section 6.3.1. Secondly, the distance per number of open facilities will be determined in Section 6.3.2. The first KPI that is calculated is the percentage decrease in distance compared to the original setting where there is only one turnout facility in The Hague. Secondly, the number of FTEs the DKDB could spare by opening a fixed number of turnout facilities is calculated.

6.3.1 Travel distance for the current situation

To calculate the distance for the current situation, the SCSTM is used with one fixed open turnout facility located in the municipality of The Hague. Just as is done in Section 6.2, to calculate the travel distance for the current situation, the two settings where capacity equals the number of TBPs and where capacity exceeds the number of TBPs are used. The total traveled distance for the current situation can be retrieved by acquiring the objective function of the SCSTM. However, because we have two settings and nine instances for each setting, a procedure has to be determined to get to one travel distance. The theory mentioned by Peterson et al. (2003) and the weighted prediction result method from Section 6.2 will be used to determine the total travel distance. The weighted prediction result method is used for both settings and from these two distances the average is taken. The outcome reveals that the total travel distance would be 51,926.13 Kilometers with only one facility in The Hague.

6.3.2 Distance per number of turnout facilities

To calculate the Distance per number of facilities, again, the SCSTM with the two settings where the capacity equals and where the capacity exceeds the number of TBPs is used. The distance per number of turnout facilities is retrieved by applying the weighted prediction result method that is introduced in Section 6.1. From both settings, the average is taken. The distance per number of turnout facilities is stated in Table 16.

Table 16: Travel distance for DKDB agents for opening a certain number of facilities

Number of facilities	distances
1	36409.566
2	25684.045
3	20532.293
4	15117.470
5	13834.334
6	13722.802
7	12589.907

6.3.3 KPI calculations

With the results presented in Sections 6.3.1 and 6.3.2, the first KPI, which is the percentage decrease in distance traveled by DKDB agents can be calculated. However, for calculating the number of FTEs the DKDB could spare per number of open turnout facilities, some more assumptions have to be made. At first, the average speed a car drives in the Netherlands is derived from Vekeerskunde (n.d). For the situation where there is only one turnout facility in The Hague as well as for the solutions, the average distance an agent drives can be calculated when it is assumed that one agent protects one TBP and has their own car. Based on these kilometers, and the pace of Verkeerskunde (n.d) the average drive time per agent can be calculated in both cases. This drive time per agent is multiplied by two because the agent also has to get back to the turnout facility at the end of the working day. Based on a nine-hour working day, the effective work time per agent can be calculated. Multiplied by the number of agents and an average of 208 working days each year, the effective work time per year for the DKDB can be calculated. The difference in the effective worktimes per year between the case where there is only one turnout facility in The Hague can be divided by 208 days per year times 9 hours a day which would result in the number of FTEs the DKDB could spare per solution. The assumptions for the calculation of the spare FTEs are discussed and approved by the DKDB. The results of both KPIs are stated in Table 17 below.

Table 17: KPIs

Facilities	Percentage decrease in kilometers	Spare FTE's
1	29.882%	45.3
2	50.537%	76.7
3	60.459%	91.8
4	70.887%	107.6
5	73.578%	111.4
6	73.573%	111.7
7	75.754%	115.0

7. Conclusion

This Chapter presents the findings of each of the research objectives, followed by the scientific contribution of this study, and closes with suggestions for future research.

7.1 Conclusions and recommendations

To present the findings on the research objective (RO) that deals with developing a model that minimizes the travel times for DKDB agents when opening a fixed number of turnout facilities, and determines the optimal locations, and capacities for these turnout locations, five different sub-objectives were defined. The findings of these objectives are presented first, followed by the overall conclusion and recommendations for the research objective.

SO1. Execute a literature study on location theory with demand uncertainty.

The literature study showed that Facility Location Theory is particularly well-suited to address the problem context described due to its inherent focus on optimizing the placement of facilities while ensuring a certain objective function. Of the three different models investigated in Chapter 2, it became evident that the Facility Location Problem (FLP) was the most suitable match for the problem definition from Section 1.2.
Snyder (2006) and Peterson et al. (2003) introduce a few methods to handle demand uncertainty in FLPs. At first, the instance approach from Peterson et al. (2003) was discussed. For this approach, the random variables are determined by the use of different growth factors. The goal is to find a robust solution that is optimal for all instances. Next to the instance approach, the scenario approach from Snyder (2006) is introduced. These scenarios represent different distributions of the uncertain variables. In a scenario approach, a mathematical model must be able to optimize over all these scenarios. An FLP that uses a scenario approach will generate a solution that is optimal for the combination of these scenarios. The last method that is introduced to include demand uncertainty is to introduce a capacity constraint. A capacity constraint states that the sum of the demand that is served from a certain facility cannot exceed a predefined benchmark.
A common solution method for FLPs that deal with demand uncertainty is a two-stage solution method where decisions on locations are generally first-stage decisions whereas the assignment of demand to facilities is second-stage.
After gathering literature about Location Theory, the next objective became:

SO2. Define a mathematical model capable of dealing with demand uncertainty while identifying the optimal turnout locations and capacity distributions.

For the mathematical model, the decision was made to include scenarios as introduced in SO1. Because a municipality can host multiple TBPs, the parameter z_{iw} was introduced. The objective of the mathematical model is the minimize the average distance traveled by DKDB agents over all the scenarios. The first mathematical model developed is stated as the Scenario Turnout Model (STM). Because in the STM, capacity is ignored, the decision was made to include capacity as a decision variable. This would give the DKDB more insight into how to divide the capacity optimally across the turnout facilities. This resulted in the Capacity Scenario Turnout Model (CSTM). Because the computational time to solve the CSTM for a large number of scenarios would not be reasonable for the purpose of this study a solution method must be defined in SO3:

SO3. Identify a solution method that can solve the mathematical model of SO2.

In SO3, a two-stage solution method, as explained in SO1 is introduced. In the first stage, the optimal locations are determined. In the STM, a TBP will always be assigned to its closest open facility. Therefore is the objective only dependent on the distance. Snyder (2006) stated that if demand must be assigned to the same turnout facility in every scenario, the problem reduces to a deterministic problem in which the uncertain parameter can be replaced by their means. Using this information and looking at the procedures used by Weaver and Church (1983) and Contreras et al. (2011) the Average Scenario Turnout Model (ASTM) was developed which will always give the same results as the STM. The results of the ASTM will be input for the model used for the second stage. This new model is a simplified version of the CSTM. This is because the potential facility set is reduced to the output of the ASTM and the constraints on opening facilities can be removed from the CSTM. This resulted in the development of the Simplified Capacity Scenario Turnout Model (SCSTM). So, in the second stage, the solutions of the ASTM are used as input for the SCSTM which, in turn, will define the optimal capacity distribution between the turnout facilities. Next, the numerical experiment will be specified resulting in SO4:

SO4. Specify the numerical experiment that will be executed in SO5.

The investigation of the two trends that are introduced in Section 2.2 and employing the Simple Average (SA) forecast method introduced by Blanc and Setzer (2016) resulted in the growth factors stated in Table 11. The expected number of TBPs per municipality for the nine instances could now be calculated by multiplying the inhabitant percentages by 400 (which is the approximate number of TBPs in the Netherlands) and by the according growth factors of Table 11.
The Scenarios are generated based on gathering random values from a Poisson distribution. The expected number of TBPs per municipality per instance is what differs in retrieving these random values.
After identifying the growth factors and the scenarios, the instances were generated. In total, nine different instances are all optimized individually. The TBP dataset of the instances is generated based on the scenarios. The potential turnout facility dataset is the geographical locations of the city hall of each municipalities in the Netherlands. For the maximal number of facilities to open, the decision was made that when the objective increase is bigger than -10% for the second time, to stop the iterative process. At last, the distance matrix is calculated by using the pyproj.geod function in Python. Now, the numerical experiment can be executed in SO5:

SO5. Execute the numerical experiment of SO4 and delve managerial insights based on the optimal travel times, capacity distributions, and locations for placing varying numbers of turnout facilities.

To determine the optimal locations for varying numbers of facilities, the ASTM is optimized over a thousand scenarios in Gurobi. The optimal capacity distribution is determined by using the solutions from the ASTM for the SCSTM in Gurobi. However, because of the large number of x_{ijw} variables the number of scenarios was reduced to one hundred. The optimized travel times are calculated by the weighted prediction result method that is based on the literature of Peterson et al. (2003). The optimal travel times, capacity distributions, and locations for placing varying numbers of turnout facilities are stated in Table 18 below.

Table 18: Optimal locations, capacity distributions, and travel distance for varying number of turnout facilities

Facilities	Locations	Capacity distribution	Travel distance
1	Utrecht	100%	36409.566
2	Arnhem, Waddinxveen	46%, 54%	25684.045
3	Alphen aan den Rijn, De Wolden, Maashorst	50%, 21%, 29%	20532.293
4	Amsterdam, de Wolden, Maashorst, Rotterdam	26%, 20%, 27%, 27%	15117.470
5	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	26%, 15%, 17%, 15%, 27%	13834.334
6	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	17%, 17%, 15%, 10%, 26%, 15%	12589.907

Based on the optimal locations, a few things can be learned. At first, optimal locations barely repeat because the model opts to locate turnout facilities such that the total travel distance for DKDB agents is balanced for each turnout facility. This is the case because a facility that has a smaller service area will have more TBPs assigned to it (higher capacity) than the other turnout facilities. However, starting from four facilities, Amsterdam and Rotterdam will present in all optimal solutions. This is because these are the two municipalities with the biggest population. Because TBPs are clustered based on municipality, the distance between the facilities and TBPs will be set to zero when opening a facility in these two municipalities. At last, the optimal locations that are based in the south- and north-east side of the Netherlands will tend to go further into the corners of the country as the number of turnout facilities increases. The optimal capacity distribution confirms the fact that the travel distance and the number of TBPs are balanced for every turnout facility.

RO. Develop a model that minimizes the optimal travel time for DKDB agents when opening a fixed number of turnout facilities and determines the optimal locations and capacity distributions for these turnout facilities.

This study proposes the CSTM to improve the capacity of the DKDB. The SCTM is solved in a two-stage solution method where the ASTM is used to determine the optimal locations and the SCSTM is used to determine the optimal capacity distribution for the turnout facilities. The results of this two-stage solution method are stated in Table 18. The solutions from Table 18 are optimal for whatever scenario unfolds in the future. This is proven based on the experiments that are stated in Appendices 2, 4, 5, 7, and 8.

To come up with a recommendation, the results from Chapter 6 were presented to the DKDB. Upon examining Table 17, the DKDB concluded that opting for four turnout facilities presented the most logical choice. This decision was primarily influenced by the realization that the addition of more turnout facilities does not improve the Key Performance Indicators (KPIs) in Table 17 much, whereas the disparities in KPIs for a reduced number of turnout facilities were notably smaller compared to opening four facilities. Another motivation factor for selecting four turnout facilities was the inclusion of Rotterdam as one of the turnout locations. According to the DKDB, this choice is strategic because Rotterdam is close to The Hague, where the Netherlands' parliament is located. However, before making the decision, some other variables should be taken into account. For example, the DKDB should determine if there is sufficient budget to open four turnout facilities. Also, this study uses the geographical locations of city halls as potential turnout facility locations. Therefore, the DKDB has to take space availability into account for making the decision. At last, maybe some other political restrictions must be taken into account. Two examples of these restrictions could be community relations or ideological differences.

7.2 Scientific contributions

The scientific contributions lie in handling demand uncertainty. In facility location theory, the theory on demand uncertainty is less developed than, for example, price uncertainty and distance uncertainty. In this project, an approach is introduced that handles demand uncertainty in two ways. At first, nine different instances are made that all have different sets of demand data. Subsequently, the mathematical model will be formulated to accommodate various scenarios. These scenarios represent different demand distributions that are based on gathering random values from a random distribution. The goal is to find robust solutions that are optimal for whatever scenario unfolds.

7.3 Future research

This section introduces two limitations which will be addressed with suggestions for future research.

At first, the DKDB has the option to conduct a comprehensive reevaluation using real data. While the created dataset appears to be representative, disparities in the solutions may arise. In the event that the DKDB decides to undertake research, a zip file containing all the necessary codes for generating the solutions has already been made. This allows the DKDB to focus solely on the data exploration phase.

Secondly, to determine the optimal capacity, two distinct cases were considered. In the first case, the available capacity exceeded the number of TBPs, while in the second case, the capacity equaled the number of TBPs. Both cases yielded identical optimal solutions. However as discussed in Section 6.2, it is plausible that the capacity could be less than the number of TBPs, given the increasing number of TBPs explained in Chapter 1. This presents a more complex case. When the capacity is lower than the number of TBPs, assumptions regarding how to allocate the capacity must be made, necessitating a completely new research methodology.

8. References

Alledaagse Vragen – Hoeveel mensen hebben persoonsbeveiliging? (2022, July 14). *NPO Radio 1*. https://www.nporadio1.nl/nieuws/podcast/f4c06b4f-0a64-423b-826e-40cf528ca568/alledaagse-vragen-hoeveel-mensen-hebben-persoonsbeveiliging

Betrán, A. P., Ye, J., Moller, A. B., Zhang, J., Gülmezoglu, A. M., & Torloni, M. R. (2016). The increasing trend in caesarean section rates: global, regional and national estimates: 1990-2014. *PloS one*, *11*(2), e0148343.

Blanc, S. M., & Setzer, T. (2016). When to choose the simple average in forecast combination. *Journal of Business Research*, *69*(10), 3951-3962.

Bobdenotter. (n.d.). *GitHub - bobdenotter/4pp: 4PP database - plaatsnamen, postcodes, geolocatie (latitude x longitude)*. GitHub. https://github.com/bobdenotter/4pp

Centraal Bureau voor de Statistiek. (n.d.b). *Inwoners per gemeente*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners

Centraal Bureau voor de Statistiek. (n.d.-c). *Kerncijfers per postcode*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode

Centraal Bureau voor de Statistiek. (2022, August 11). Trek uit de Randstad blijft toenemen. *Centraal Bureau Voor De Statistiek*. https://www.cbs.nl/nl-nl/nieuws/2022/32/trek-uit-de-randstad-blijft-toenemen

Centraal Bureau voor de Statistiek. (n.d.a). *Bevolkingsgroei*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingsgroei/groei#:~:text=De%20Nederlandse%20bevolking%20groeide%20in,worden%20dan%20er%20mensen%20overlijden.

Cheng, C., Adulyasak, Y., & Rousseau, L. M. (2021). Robust facility location under demand uncertainty And facility disruptions. *Omega*, *103*, 102429.

Contreras, I., Cordeau, J. F., & Laporte, G. (2011). Stochastic uncapacitated hub location. *European Journal of Operational Research*, *212*(3), 518-528.

Cooper, L. (1978) Bounds on the Weber problem solution under conditions of uncertainty. Journal of Regional Science,18(1), 87–93.

Dasci, A., & Laporte, G. (2005). An analytical approach to the facility location and capacity acquisition problem under demand uncertainty. *Journal of the Operational Research Society*, *56*(4), 397-405.

De Armas, J., Juan, A. A., Marquès, J. M., & Pedroso, J. P. (2017). Solving the deterministic and stochastic uncapacitated facility location problem: from a heuristic to a simheuristic. *Journal of the Operational Research Society*, *68*(10), 1161-1176.

Farahani, R. Z., Asgari, N., Heidari, N., Hosseininia, M., & Goh, M. (2012). Covering problems in facility location: A review. *Computers & Industrial Engineering*, *62*(1), 368-407

Farhan, B., & Murray, A. T. (2006). Distance decay and coverage in facility location planning. *The Annals of Regional Science*, *40*, 279-295.

Gemiddelde autorit is 19 kilometer en duurt een kwartier. (z.d.). Verkeerskunde. https://www.verkeerskunde.nl/artikel/gemiddelde-autorit-is-19-kilometer-en-duurt-een-kwartier#:~:text=Bij%20een%20gemiddelde%20autorit%20wordt,gehouden%20met%20'connected%20car'

Gutierrez GJ, Kouvelis P (1995) A robustness approach to international sourcing. Ann Oper Res 59:165-193

Hansun, S. (2013, November). A new approach of moving average method in time series analysis. In *2013 conference on new media studies (CoNMedia)* (pp. 1-4). IEEE.

Han, Y., Zhou, G., & Zhu, Y. (2016). A trend factor: Any economic gains from using information over investment horizons?. *Journal of Financial Economics*, *122*(2), 352-375.

Hodder JE, Dincer MC (1986) A multifactor model for international plant location and financing under uncertainty. Comput Oper Res 13:601–609

Hodder JE, Jucker JV (1985) A simple plant location model for quantity-setting firms subjected to price uncertainty. Eur J Oper Res 21:39–46

Internetbureau Antum [www.antum.nl]. (z.d.). *» Hoeveel werkdagen in een jaar | SnelVerder.nl*. https://www.snelverder.nl/v/wetenschap/tijd-en-meetkunde/hoeveel-werkdagen-in-een-jaar/

Jafari, R., & Arkat, J. (2016). Network location problem with stochastic and uniformly distributed demands. *International Journal of Engineering*, *29*(5), 654-662

Karasakal, O., & Karasakal, E. K. (2004). A maximal covering location model in the presence of partial coverage. *Computers & Operations Research*, *31*(9), 1515-1526.

Liberatore, F., Camacho-Collados, M., & Vitoriano, B. (2020). Police districting problem: literature review and annotated bibliography. *Optimal districting and territory design*, 9-29

Louveaux, F. V., & Peeters, D. (1992). A dual-based procedure for stochastic facility location. *Operations research*, *40*(3), 564-573.

Lynskey, J., Thar, K., Oo, T. Z., & Hong, C. S. (2019). Facility location problem approach for distributed drones. *Symmetry*, *11*(1), 118.

Marques, M. D. C., & Dias, J. M. (2018). Dynamic location problem under uncertainty with a regret-based measure of robustness. *International Transactions in Operational Research*, *25*(4), 1361-1381.

Meraklı, M., & Yaman, H. (2017). A capacitated hub location problem under hose demand uncertainty. *Computers & Operations Research*, *88*, 58-70.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2022, December 21). *Gemeentelijke herindeling*. Gemeenten | Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/gemeenten/gemeentelijke-herindeling#:~:text=Nederland%20heeft%20al%20lange%20tijd,gemeentelijke%20herindelingen%20afgenomen%20tot%20342.

Ministerie van Justitie en Veiligheid. (2022a, April 20). *Dreigingsbeeld Terrorisme Nederland*. Nationaal Coördinator Terrorismebestrijding En Veiligheid. https://www.nctv.nl/onderwerpen/dtn

Ministerie van Justitie en Veiligheid. (2022b, November 11). *Dreigingsbeeld NCTV: dreiging in en tegen Nederland veelzijdiger en meer diffuus*. Nieuwsbericht | Nationaal Coördinator Terrorismebestrijding En Veiligheid. https://www.nctv.nl/onderwerpen/dtn/nieuws/2022/11/07/dreigingsbeeld-nctv-dreiging-in-en-tegen-nederland-veelzijdiger-en-meer-diffuus

Ministerie van Justitie en Veiligheid. (2023, January 26). *Minister deelt update ontwikkelingen bewaken en beveiligen*. Nieuwsbericht | Nationaal Coördinator Terrorismebestrijding En Veiligheid. https://www.nctv.nl/actueel/nieuws/2022/11/01/minister-deelt-update-ontwikkelingen-bewaken-en-beveiligen

Mirchandani, P.B. and Oudjit, A. (1980) Localizing 2-medians on probabilistic and deterministic tree networks.Networks,10(4), 329–350.

Mulvey, J. M., Vanderbei, R. J., & Zenios, S. A. (1995). Robust optimization of large-scale systems. *Operations research*, *43*(2), 264-281

Narula, S. C., Ogbu, U. I., & Samuelsson, H. M. (1977). An algorithm for the p-median problem. *Operations Research*, *25*(4), 709-713.

NOS. (2022, August 11). Meer mensen verhuisden van Randstad naar rest van Nederland dan andersom. *NOS*. https://nos.nl/artikel/2440206-meer-mensen-verhuisden-van-randstad-naar-rest-van-nederland-dan-andersom

Peterson, G. D., Cumming, G. S., & Carpenter, S. R. (2003). Scenario planning: a tool for conservation in an uncertain world. *Conservation biology*, *17*(2), 358-366.

Ranglijst van de grootste en kleinste gemeenten in inwoners in Nederland (bijgewerkt 2023!) | AlleCijfers.nl. (2023, July 24). AlleCijfers.nl. https://allecijfers.nl/ranglijst/grootste-en-kleinste-gemeenten-in-inwoners-in-nederland/

Ravi, R., & Sinha, A. (2004, June). Hedging uncertainty: Approximation algorithms for stochastic optimization problems. In *International Conference on Integer Programming and Combinatorial Optimization* (pp. 101-115). Berlin, Heidelberg: Springer Berlin Heidelberg.

ReVelle, C., & Hogan, K. (1989). The maximum availability location problem. *Transportation science*, *23*(3), 192-200.

Schilling, D. A., Jayaraman, V., & Barkhi, R. (1993). A review of covering problem in facility location. Location Science, 1(1), 25–55

Shavarani, S. M., Nejad, M. G., Rismanchian, F., & Izbirak, G. (2018). Application of hierarchical facility location problem for optimization of a drone delivery system: a case study of Amazon prime air in the city of San Francisco. *The International Journal of Advanced Manufacturing Technology*, *95*, 3141-3153.

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE transactions*, *38*(7), 547-564.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, *15*(4), 101-115.

Weaver, J.R. and Church, R.L. (1983) Computational procedures for location problems on stochastic networks.Transportation Science,17(2),168–180.

Wijaya, C. Y. (2022, September 13). 3 Unique Python packages for Time Series Forecasting. *Medium*. https://towardsdatascience.com/3-unique-python-packages-for-time-series-forecasting-2926a09aaf5b

Wikipedia-bijdragers. (2023). Lijst van Nederlandse gemeenten. *Wikipedia*. https://nl.wikipedia.org/wiki/Lijst_van_Nederlandse_gemeenten

Wu, L. Y., Zhang, X. S., & Zhang, J. L. (2006). Capacitated facility location problem with general setup cost. *Computers & Operations Research*, *33*(5), 1226-1241.

9. Appendices

Appendix 1: Performance of two-stage solution method V.S CSTM

This comparison is important to compare the performance of the two-stage solution method compared to the Capacity Scenario Turnout Model. The dataset of municipalities of Wikipedia-bijdragers (2023) was cut off at the first 50 municipalities. This was done because otherwise, the solution generation of the CSTM would take too long. The comparison was done for two and three facilities, 50 scenarios, the 'above' growth factors for both trends, and the setting that there is exactly the same capacity as the number of TBPs. The results of this comparison are stated in Table 19. Note that the capacities are given in fractions.

Table 19: Solutions of two-stage method V.S SCSTM

facilities	Simplified Capacity Scenario Turnout Model		Capacity Scenario Turnout Model	
	municipalities	Capacities	Municipalities	Capacities
2	Amsterdam, Arnhem	0.65, 0.35	Amsterdam, Arnhem	0.66, 0.34
3	Amsterdam, Arnhem, Breda	0.51, 0.32, 0.17	Amsterdam, Arnhem, Breda	0.50, 0.34, 0.16

Appendix 2: Performance of SA forecast method V.S MA forecast method

When using the moving average forecast method in Excel on the data provided in Tables 8, 9, and 10, the prediction growth factors of the trends become 0.148% and 1.405%.

Table 20: Growth factor comparison for SA V.S MA

		SA	MA
Trend 1		0.148%	0.148%
Trend 2		1.374%	1.405%
Growth factor when both prediction growth factors are used	Unpopulated	1.01524	1.01555
	populated	1.01224	1.01255

With the growth factors of Table 20, the performance of the solutions of the SA and MA can be compared. Just as is done for Appendix 1, the dataset of municipalities from Wikipedia-bijdragers (2023) was cut off at the first 50 municipalities. The Average Scenario Turnout model is used to determine the optimal locations and the Simplified Capacity Scenario Turnout Model is used to determine the optimal capacity distribution. The comparison between the SA and MA forecast methods is stated in Table 21 below.

Table 21: Solutions comparison SA V.S MA

Facilities	SA			MA		
	facilities	objective	capacities	facilities	objective	capacities
2	Amsterdam Arnhem	3284.79	0.65, 0.35	Amsterdam Arnhem	3306.75	0.65, 0.35
3	Amsterdam Arnhem Breda	2613.14	0.51, 0.32, 017	Amsterdam Arnhem Breda	2630.60	0.52, 0.32, 0.16

Appendix 3: Geographical optimal locations for turnout facilities

1 facility: Utrecht

2 facilities: Arnhem, Waddinxveen

3 facilities: Alphen aan den Rijn, De Wolden, Maashorst

4 facilities: Amsterdam, De Wolden, Maashorst, Rotterdam

5 facilities: Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam

6 facilities: Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren

Appendix 4: Extreme cases for optimal locations

Table 22: Optimal locations for the extreme scenarios

Facilities	Instance	Solution	Objective
1	Extreme above	Utrecht	36964.943
	Extreme below	Utrecht	35868.887
2	Extreme above	Arnhem, Waddinxveen	26074.565
	Extreme below	Arnhem, Waddinxveen	25292.996
3	Extreme above	Alphen aan den Rijn, De Wolden, Maashorst	20872.597
	Extreme below	Alphen aan den Rijn, De Wolden, Maashorst	20247.020
4	Extreme above	Amsterdam, De Wolden, Maashorst, Rotterdam	17063.835
	Extreme below	Amsterdam, De Wolden, Maashorst, Rotterdam	16541.072
5	Extreme above	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	15364.297
	Extreme below	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	14892.763
6	Extreme above	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	14054.579
	Extreme below	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	13618.081

Appendix 5: Scenarios based on normal distribution

To generate z_{iw} values based on a normal distribution, the numpy.random.normal function in Python is used. This function has a 'loc', 'scale', and 'size' as input parameters. The size, also referred to as output shape, is equal to the number of samples one wants to draw. For example, if 1000 samples are needed, the size input parameters will be equal to 1000. In this case, the size parameter is equal to one because only one random value has to be drawn for each municipality and each scenario. The 'loc' parameter is the mean or center of the distributions. For the normal distributions this means that, if the mean for example is 10, if there would be 100,000 samples, the mean of these samples would be 10. In this project, the 'loc' input parameter is equal to the expected number of TBPs per municipality per instance. When taking the example of Amsterdam from Section 5.3, the 'loc' would be equal to 21.0490. At last, the 'scale' input parameter is equal to the standard deviation. For this project, the standard deviation is set to the 'loc' value divided by three. This decision is made because the normal distribution can draw a negative random value. When looking at Figure 11, which represents a standard normal distribution, there is a 99.7% chance of being 3 standard deviations away from the mean.

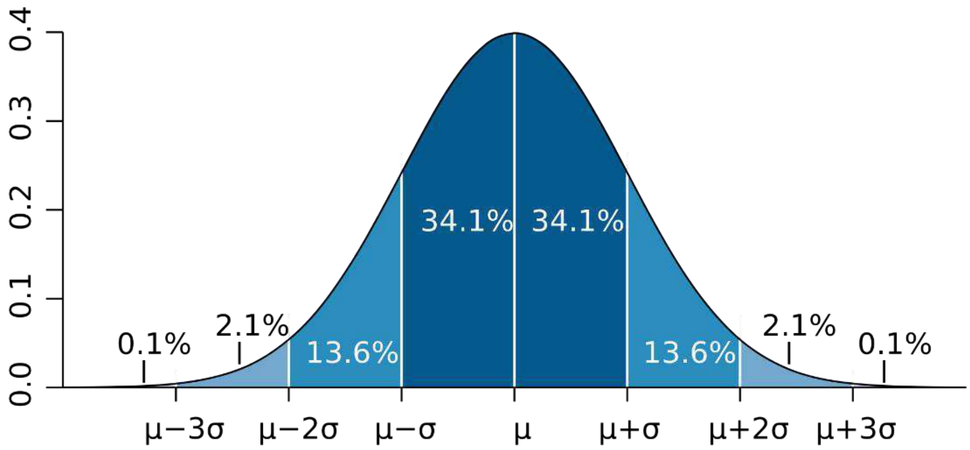

Figure 11: Normal distribution properties

So, by setting the standard deviation to one-third of the mean, the chance is 0.03% of retrieving negative numbers. By choosing these numbers as the 'loc' and 'scale' input parameters, the expected number of TBPs per municipality per instance will still correspond to the instances that are made beforehand. One disadvantage of using the numpy.random.normal function in Python is that this function does not automatically round the output as how the Poisson distribution does. Therefore, first, the random values will be rounded to the nearest integer. For every scenario, the sum of the TBPs after rounding is compared to the sum of expected TBPs per municipality per instance. If the difference is lower than 5%, the rounding is accepted. Otherwise, an iterative rounding procedure adds or subtracts (depending on the difference) 0.01 from the benchmark for rounding (which is 0.5 for rounding on the nearest integer) until the difference is lower than 5% between the rounded number of TBPs and the expected TBPs per municipality. The results of the numerical experiment are stated in Table 23.

Table 23: Results of the ASTM where scenarios are based on a normal distribution.

Facilities	instances	Solution	Objective
1	9	Utrecht	[35801.196 – 36259.896]
2	9	Arnhem, Waddinxveen	[25067.510 – 25398.338]
3	9	Alphen aan den Rijn, De Wolden, Maashorst	[19989.326 – 20253.839]

4	9	Amsterdam, De Wolden, Maashorst, Rotterdam	[16353.566 – 16564.327]
5	9	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	[14754.120 – 14944.229]
6	9	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	[13492.865 – 13666.698]

Appendix 6: Average Simplified Capacity Scenario Turnout Model (ASCSTM)

$$\min \sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} * d_{ij} * \overline{z_\iota}$$

S.T

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall\, i \in D$$

$$\sum_{i=1}^{m} x_{ij} * \overline{z_\iota} \le cap_j \qquad \forall\, j \in L$$

$$\sum_{j=1}^{n} cap_j = T$$

$$x_{ij} \in \{0,1\} \qquad \forall\, i \in D, j \in L$$

$$cap_j \in Z_+ \qquad \forall\, j \in L$$

Where $\overline{z_\iota}$ is average of z_{iw} $\forall\, w \in Q$ as is explained in Section 3.7.

Appendix 7: Extreme cases for the SCSTM

Table 24: Optimal capacity distribution for the extreme above instance

1	Solutions					
	Utrecht					
	[1]					
2	Arnhem	Waddinxveen				
	[0.45-0.46]	[0.54-0.55]				
3	Alphen aan den Rijn	De Wolden	Maashorst			
	[0.50-0.51]	[0.20-0.21]	[0.29]			
4	Amsterdam	De Wolden	Maashorst	Rotterdam		
	[0.26]	[0.20]	[0.26-0.27]	[0.27]		
5	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam	
	[0.26-0.27]	[0.15-0.16]	[0.16]	[0.15]	[0.27]	
6	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren
	[0.17-0.18]	[0.17-0.18]	[0.15]	[0.09-0.10]	[0.26]	[0.14]

Table 25: Optimal capacity distribution for the extreme below instance

1	Solutions					
	Utrecht					
	[1]					
2	Arnhem	Waddinxveen				
	[0.46-0.47]	[0.53-0.54]				
3	Alphen aan den Rijn	De Wolden	Maashorst			
	[0.50-0.51]	[0.20-0.22]	[0.29]			
4	Amsterdam	De Wolden	Maashorst	Rotterdam		
	[0.25-0.26]	[0.20-0.21]	[0.27]	[0.27]		
5	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam	
	[0.25-0.26]	[0.16]	[0.15-0.16]	[0.15-0.16]	[0.27]	
6	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren
	[0.18]	[0.17-0.18]	[0.15]	[0.09-0.10]	[0.26]	[0.14]

Appendix 8: SCSTM for scenarios based on normal distribution

Table 26: Optimal capacity distribution if the capacity exceeds the number of TBPs and the scenarios are based on a normal distribution

	instances	Solution					
1	9	Utrecht					
		[1]					
2	9	Arnhem	Waddinxveen				
		[0.46-0.47]	[0.53-0.54]				
3	9	Alphen aan den Rijn	De Wolden	Maashorst			
		[0.50-0.51]	[0.20-0.21]	[0.29-0.30]			
4	9	Amsterdam	De Wolden	Maashorst	Rotterdam		
		[0.25-0.26]	[0.20-0.21]	[0.26-0.27]	[0.27-0.28]		
5	9	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam	
		[0.24-0.25]	[0.15-0.16]	[0.16-0.17]	[0.15-0.16]	[0.27]	
6	9	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren
		[0.17]	[0.17-0.18]	[0.14-0.15]	[0.10-0.11]	[0.26-0.27]	[0.14-0.15]

Table 27: Optimal capacity distribution if the capacity equals the number of TBPs and the scenarios are based on a normal distribution

	instances	Solution					
1	9	Utrecht					
		[1]					
2	9	Arnhem	Waddinxveen				
		[0.46]	[0.54]				
3	9	Alphen aan den Rijn	De Wolden	Maashorst			
		[0.50-0.51]	[0.20-0.21]	[0.29-0.30]			
4	9	Amsterdam	De Wolden	Maashorst	Rotterdam		
		[0.26-0.27]	[0.20]	[0.26 - 0.27]	[0.27-0.28]		
5	9	Amsterdam	Arnhem	Eindhoven	Midden-Drenthe	Rotterdam	
		[0.26]	[0.15-0.16]	[0.16-0.17]	[0.15]	[0.26-0.27]	
6	9	Amsterdam	Bunnik	Deventer	Noordenveld	Rotterdam	Someren
		[0.16-0.17]	[0.17-0.18]	[0.15]	[0.10]	[0.26]	[0.14]

Optimal positioning and capacities of new turnout facilities for the Surveillance and Protection Department in The Netherlands

S. (Stan) Loog

Supervisors: Loe Sclicher & Layla Martin

Introduction

- Criminals put pressure on our society by intimidating or even attacking inhabitants of the Netherlands.
- Because of the increasing pressure on the Dutch Bewaken en Beveiligen (B&B) system, the system lacks personnel for operational tasks.

Problem description

- The DKDB is responsible for the protection of TBPs (Persons to be Protected).
- The DKDB operates from one turnout facility, which is the location where agents must arrive before and after shifts, in The Hague. This implies a lot of travel distance curing working times.
- Placing multiple facilities, travel times decrease and capacity increase.
- Location theory is well known to solve these kind of problems.
- However, the locations and numbers of TBPs are stochastic.
- This project deals with finding a model that is robust against stochasticity.
- Optimal locations and capacity distributions will be determined.

Research objective

Develop a mathematical model that minimizes the travel time for DKDB agents when opening a fixed number of turnout facilities and determines the optimal locations and capacity for these turnout facilities.

Mathematical model

- First, we formulate the Scenario Turnout Model (STM), which ignores capacity, to determine optimal locations.
- Next, the Capacity STM (CSTM) is formulated to determine the optimal locations and capacity distribution. The CSTM is stated on the right.
- Both models use a scenario approach to account for the demand uncertainty. The models can optimize over multiple scenarios. This is because one of the decision variables, as can be seen in the CSTM, is x_{ijw} which is one if demand node *i* is served by facility *j* in scenario *w*.

$$(\min \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{w=1}^{q} x_{ijw} * d_{ij} * z_{iw}) / q$$

S.T

$$\sum_{j=1}^{n} x_{ijw} = 1 \qquad \forall i \in D, w \in Q$$

$$\sum_{j=1}^{n} y_j = p$$

$$x_{ijw} - y_j \le 0 \qquad \forall i \in D, j \in L, w \in Q$$

$$\sum_{i=1}^{m} x_{ijw} * z_{iw} \le cap_j \qquad \forall j \in L, w \in Q$$

$$\sum_{j=1}^{n} cap_j = T$$

$$y_j \in \{0,1\} \qquad \forall j \in L$$

$$x_{ijw} \in \{0,1\} \qquad \forall i \in D, j \in L, w \in Q$$

Example

- The figure below visualizes Parkstad, a region in the south of the Netherlands.
- The figure represents a possible SCTM solution on how TBPs are assigned to facilities in Heerlen and Voerendaal in two different scenarios (green and blue nodes).
- The demand nodes closest to Heerlen and Voerendaal are black and red outlined respectively.
- In the green scenario, one TBP in Kerkrade is not served from its closest facility because of capacity restrictions.
- In the blue scenario, one TBP in Simpelveld is not served from its closest facility because of capacity restrictions.

Solution method

We developed a two-stage solution method for the CSTM because the computational time of the CSTM exceeds a reasonable computational time for the purpose of this study. In the first stage, the STM is rewritten in an equivalent model in which the stochastic parameters are replaced by their means. This equivalent model is referred to as the Average STM (ASTM) which ignores capacity. Subsequently, the optimal locations of the ASTM are input for the new Simplified CSTM (SCSTM) which determines the optimal capacity distributions of turnout facilities. In the SCSTM, constraints for placing facilities are removed.

Numerical experiment

Only publicly available data is used because data about TBPs should be kept confidential. The assumption is made that the TBPs follow the same two trends as the overall population which are:

1. People tend to move out of the populated areas in the Netherlands.
2. There is an increase of inhabitants in the Netherlands.

These trends are investigated and with the Simple Average (SA) forecasting method, three growth factors per trend are determined. The TBP dataset represents the scenarios which are based on gathering values from a Poisson distribution.

Results

A Weighted Prediction Result method is introduced to determine one result from the different instances. For opening one to six facilities, robust solutions are found. Robustness is further substantiated by additional experiments. The results of the two-stage method are stated in the Table below.

Facilities	Locations	Capacity Distributions	Travel distance
1	Utrecht	100%	36409.566
2	Arnhem, Waddinxveen	46%, 54%	25684.045
3	Alphen aan den Rijn, De Wolden, Maashorst	50%, 21%, 29%	20532.293
4	Amsterdam, De Wolden, Maashorst, Rotterdam	26%, 20%, 27%, 27%	15117.470
5	Amsterdam, Arnhem, Eindhoven, Midden-Drenthe, Rotterdam	26%, 15%, 17%, 15%, 27%	13834.334
6	Amsterdam, Bunnik, Deventer, Noordenveld, Rotterdam, Someren	17%, 17%, 15%, 10%, 26%, 15%	12589.907

From the optimal locations and capacity distribution three things can be learned:

1. For a low number of turnout facilities, optimal locations barely repeat.
2. Starting from four facilities, Amsterdam and Rotterdam will be present in all optimal solutions.
3. The optimal locations that are based in the south- and north-east sides of the country tend to go further into the corner of the country when the number of facilities increases.

With the results above, two Key Performance Indicators (KPIs) are calculated. The assumptions made to calculate these KPIs are approved by the DKDB, and the results are stated in the table on the right.

Facilities	Percentage decrease in kilometers	Spare FTEs
1	29.882%	45.3
2	50.537%	76.7
3	60.459%	91.8
4	70.887%	107.6
5	73.573%	111.4
6	73.578%	111.7

Conclusion and recommendation

This thesis proposes the CSTM to improve the capacity of the DKDB. The CSTM is solved with a two-stage method where the ASTM determines the optimal location and the SCSTM determines the optimal capacity distributions of different turnout facilities. For opening one to six facilities, robust solutions have been found. For the recommendation, the results in the two tables above were presented to the DKDB. Opting for four facilities (Figure on the right) seems the most logical choice because:

1. From four facilities onwards, the KPIs do not increase much.
2. Rotterdam is close to The Hague.

With this choice, the DKDB spares 70.877% travel distances and 107.6 FTEs. However, to make a good decision the DKDB should also consider other factors like budget constraints, space availability, and political restrictions.