The reality of Multi Depot Vehicle Routing models

Sjoerd Smink
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, Enschede, The Netherlands
sjoerdsmin@gmail.com

Abstract
The topic of this study is the Multi Depot Vehicle Routing Problem (MDVRP). Multiple models for this problem are pointed out and explained in some detail. The models have been implemented in a simulation program. Unique about this study is that the data used for the simulations comes from an existing Dutch transportation company. Results from simulations are that way based on real-life information and not randomly generated data. The most efficient model in this study is the random model with 2-opt, route first – cluster second and a limit of 1.00. Using these parameters to execute simulations for a transport company can help with cost calculations of specific customers and comparing ways of distributing goods.

Keywords
Mult depot vehicle routing problem (MDVRP), Travelling Salesman Problem (TSP), Simulations.

1. INTRODUCTION
Distribution companies face great challenges to organize their fleet efficiently [15]. Theoretical models have been developed to help good decision making. In this research, some of these models will be discussed and tested. Testing will be done with information from an existing Dutch transportation company.

1.1 Problem statement
The transportation business is very competitive. It is important to cut costs and make use of transportation material as efficient as possible. [3, 13 & 15] Making a good planning for the transportation fleet is imperative.

In this research the multi depot vehicle routing problem (MDVRP) will be discussed specifically. The problem is about how a transportation company decides to make a most efficient schedule for their trucks. The transportation company has multiple depots from which their trucks depart and arrive, and has multiple customers being served from the different depots. The challenge is to make a schedule for each truck individually so that the trucks drive in the most efficient way and the package arrives at the customer in time.

According to Baltz et al. [1] the MDVRP is defined as: “given several depots and a set of customers who must be served from these depots, the problem is to both (a) assign the customers to a depot and (b) find optimal service tours”.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2010, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

1.2 Practical relevance
Unique about this study is the simulation of models using real-life data and not randomly generated data. Using existing data also has the advantage that outcomes of simulations can be compared to what actually happened in reality when no models were used. Comparing that way results in outcomes of simulations implying more about the practical implications of the model. In other words: the results of the simulation are closer to reality because real-life data was used and not random data.

Applying models on an existing company has seldom been done. One of the conclusions in the research paper of Moonen [14] is that relatively little research is being done in real practice nor about implementation. In this case it is about Multi Agent Systems, which is a possible model or technique to solve the MDVRP.

1.3 Research purpose
A lot of models about the (multi depot) vehicle routing problem have been developed. Salhi & Sari [17] mention at least 15 publications about this problem. The purpose of this research is to study the MDVRP, simulate some models using data from an existing company and compare the results between the models.

The outcomes of this study can, off course, be used to improve the planning of the existing Dutch transport company. But more importantly the simulations can help making an estimate of the cost price of a certain customer. When a new customer adds delivery points to the routes the trucks have to drive, it is difficult to calculate the extra costs for the trucks to drive further. One way is to make a planning with and without the new customer, but that is time consuming. When simulations can make a schedule in a few seconds, it is possible to see the difference in the costs. It is plausible that relative cost differences with and without a certain customer in simulations can be applied to the real-life planning.

Applying the simulation techniques makes it possible for the transportation company to calculate which customers are the most expensive or profitable. Another practical application is that the transportation company currently delivers goods in certain areas of the country on specific days. The advantage of this method is that by clustering delivery points the trucks have to drive less far. Problem is that nowadays customers are requesting that delivery is possible every day. The consequence is that trucks have to drive every day throughout the entire country, which is less efficient. By using simulations it is possible to calculate the relative differences in costs between the two methods (cluster delivery vs. delivery of the entire country every day).

Besides usefulness for the company, this research has also scientific surplus value. Using results from a simulation and compare them to a real-life company has been done seldom in science. It can improve the understanding of how companies plan their transportation activities. According to dr. ir. M.R.K.
Mes - assistant professor in Operational Methods and Logistics at the University of Twente – recent contributions to solution methodologies for the vehicle routing problem merely focus on improvements of existing mathematical models. When using real-life data the efficiency of models can be different. A possible conclusion can be that although a model should be most effective because that is proven mathematically, in reality it is not the most efficient.

1.4 Research questions

In a nutshell, this research is about the MDVRP and simulation of different models to solve the problem. It is also about applying theoretical models in practice and comparing the outcomes to the reality. Of course, these elements should return in the research questions.

The main research question is formulated as follows: what is the most cost efficient model to solve a real-life multi depot vehicle routing problem (MDVRP)?

Two sub-questions are formulated to help answering the main research questions:

- What are possible models to solve the multi depot vehicle routing problem (MDVRP)?
- How cost efficient do the models solve a real-life multi depot vehicle routing problem (MDVRP)?

More information about how the research questions are going to be answered can be found in section 3 – methodology.

1.5 Structure of paper

This research paper first describes possible models to solve the MDVRP. After the literature review section a description is given about the methodology to answer the research questions. Because simulation is an important part of this research, section 4 is dedicated to simulations and describes the way simulations have been build. The next section gives some results of the simulations and section 6 discusses these results. Finally, conclusions are given to answer the research questions. This research paper closes with recommendations for future research, acknowledgements, references and appendices.

2. LITERATURE REVIEW

The multi depot location routing problem (MDVRP) is a specific variant of the vehicle routing problem (VRP). The literature about VRP is extensive [4]. A description of the VRP is that it has four criteria [4]:

- Each route starts and ends at the depot.
- Each customer belongs to exactly one route.
- The total demand of the route does not exceed a limit of a vehicle.
- The total duration of all routes is minimized.

For a vehicle routing problem, there is one distribution point, with multiple vehicles and a set of customers. It is common that the total demand of all customers exceeds the capacity of a vehicle, so multiple vehicles should be used. The routes for the vehicles start and finish at the distribution point. [8]

The VRP is similar to the Travelling Salesman Problem (TSP) [8]. Laporte [10] is recognized for his work about this topic. The problem is about a salesman who starts at point A, have to travel to points B, C, D and E (in random order) and in the end returns at point A. The problem is to make a route from A to A with visiting B, C and D in the most efficient way. So, for example, the model shows that A, D, C, E, B, A is the shortest path. The difference between the Travelling Salesman Problem and Vehicle Routing Problem is that VRP considers the capacity of the vehicles and the TSP does not. That way the VRP has multiple routes from a distribution point and the TSP has only one long route. [8]

A lot of calculations are necessary to solve the Vehicle Routing Problem or Travelling Salesman Problem the most efficient way. Adding only one extra delivery point to a route results in a lot more possible routes. The VRP is NP-hard, which means that every extra point lets the possibilities grow exponentially. Exact algorithms can come up with a solution until 50 points [4 & 16]. Because in this research there are more than 50 points that have to be visited, heuristic models are being used. Heuristics are also most often used for the MDVRP [16].

This research looks specifically to the multi depot vehicle routing problem. Compared to the VRP, there are multiple depots from which trucks depart. “The MDVRP is defined on a single day but vehicles operate from one of several depots instead of only one, and each vehicle route must start and end at the same depot. Otherwise, the statement of the problem is the same as that of the VRP.” [4] The MDVRP is more complex than the single-depot VRP and the available research about the MDVRP is less [8].

It is common that the goal of the MDVRP is to minimize the total travel cost of vehicles. The number of vehicles used does not affect the costs. [4] The cost of transportation between distribution points – so called long haul transport – is often fixed and that way not important when comparing possible routes.

The next sections will describe a few basic models to solve the TSP (subsection 2.1, 2.2 and 2.3) and VRP (2.5 and 2.6). These models can be used in different combinations to solve the MDVRP, as described in the last section 2.7. The discussed models were advised by dr. ir. M.R.K. Mes in an interview.

2.1 Nearest neighbour

The first model is the also the most simplistic model. The distribution point is chosen as starting point. Next point on the route (let’s call this point B) is the point closest to the distribution point. The point after that is the closest point to B. After all points are visited, the route returns to the home distribution point. Figure 1 shows how this model is build up.

The idea behind this model is that every time the distance between the points is minimized, which should result in a minimal total distance of the total journey. In practise the problem is that the last point before returning to the home distribution point is far-off, so the last connection in the route is long. The model often does not give the most optimal route, but the algorithm is fast. [12]
2.2 Nearest insertion

This model prevents the disadvantage of nearest neighbour – the last point added to the route is too far-off – by always having a complete route. The nearest insertion model (see figure 2) starts with a route from the home distribution point (point A) to the nearest point (point B) and back to the home distribution point; so the route is ABA. The second and iterating step is to find a point that is not already on the route and is closest to A or B. This point is added in the route, but in a way that the extra costs for adding this point to the route are minimal. The extra costs for adding C between AB are the connection AC plus connection CA minus connection AB. For example, when adding D to the route ABCA, it has to be checked whether it is cheaper to add D between AB, between BC or between CA. After point D is added to the route in the most cost efficient way, the next point closest to any point on the route is searched for and added to the route. That way, in the end, all points have been added to the route [9].

![Figure 2. Nearest insertion](image1)

2.3 Farthest insertion

Farthest insertion is almost the same as nearest insertion. The difference is that not every time the nearest point is searched for, but instead the farthest point is added to the route. The way this farthest point is added to the route is the same as with nearest insertion: the extra costs for adding connections are minimized by calculating the extra costs for adding the new point between two points on the route. [9] Figure 3 contains a graphical representation.

![Figure 3. Farthest insertion](image2)

2.4 2-opt

The first three models (nearest neighbour, nearest insertion and farthest insertion) are ways to create a route. After the initial path has been calculated, it can be improved by switching connections in the path. For example, if the initial route was ABCDEA, it can be calculated that ABDCEA is more efficient, so that two connections have to be switched. This example is shown in figure 4. Of course, it is also possible to switch three or more connections at the same time (called k-opt), but that requires more calculations. [11]

![Figure 4. 2-opt](image3)

2.5 Route first – cluster second

The models above solve the Travelling Salesman Problem by creating one long route. The assumption there is that the truck visiting the points on the route can carry infinite packages, but that is not the reality. There have to be multiple routes from the distribution center, to cope with the capacity of trucks. In other words: the route has to be split up in multiple sub-routes, in a way that all points on the sub-route can be visited without exceeding the capacity of the truck.

As the name of route first – cluster second model implies, a long tour to all points is first created and then the long tour is split up by clusters. Creating a long tour first can be done using nearest neighbour, nearest/farthest insertion or another model to solve the travelling salesman problem. It’s a possibility to improve the long route with 2-opt.

When splitting up the route the maximum capacity of the truck has to be taken into account. For example (see figure 5) there are 10 points on the long route and a truck can visit a maximum of 3 customers. In this case four trucks are needed and the route is split up after the 3rd, 6th, 9th and 10th point. But it is also an option to split up the route after the 2nd, 5th, 8th and 10th point on the route. This second option can be more cost efficient because the 2nd, 5th and 8th points are closer to the home distribution point than the 3rd, 6th and 9th points. The possible solutions have to be compared to each other calculating the sum of the four sub-routes. [2]

![Figure 5. Route first – cluster second](image4)

2.6 Cluster first – route second

Another way to solve the Vehicle Routing Problem is splitting up the points first and then create the routes [5]. A well-known model to do this is the sweep-algorithm from Gillett and Miller. This model looks for delivery points (for example) in the north of the distribution point and makes a cluster of all the points in the north. The number of points in a cluster depends on the capacity of the truck. [7]
Disadvantage of this model is that it doesn’t take the route the truck has to drive into account, so it could be more efficient to start making clusters in the south instead of the north. Another option is to assign points to a cluster counter-clockwise instead of clockwise. When starting in the north the clustering can be seen as a clock hand starting at 12 o’clock and splitting up at 5 o’clock and 9 o’clock. Rotating counter-clockwise could mean the clock hand clusters at 7 o’clock, 3 o’clock and 12 o’clock. This example is shown in figure 6.

After the clusters have been created, routes have to be made with a model to solve the travelling salesman problem. Possible models to create a route are nearest neighbour, nearest insertion or farthest insertion. After the routes have been created, it can be improved using the 2-opt model.

Different starting locations of the clock hand (north or south) and rotating clockwise or counter-clockwise has consequences for the total length of (sub)routes. The outcomes of these options have to be compared to know the most efficient route.

2.7 Assignment – sweep

The final model is a combination of all previous models. It is a model to solve the MDVRP. In contrast to previous discussed models, the assignment-sweep model handles multiple depots. The model originates from Gillett & Johnson [6].

The first step is to assign points to a specific distribution point; see figure 7. For some points this is immediately clear. Some other points are bit more difficult. For example point C is nearby distribution point A; distribution point B is a bit further but not much. It could be the case that adding this point to distribution point B leads to a lower total cost of routes. In this case the routes of B are already close to C, but the routes of A stay close to home and would make a detour if point C would have to be visited. In this case it would be smart to add point C to distribution point B, although this distribution point is a bit further away.

It is off course possible to check what the total cost of routes would be if point C would be added to distribution point A or B. Disadvantage of this method is that it is often clear that point C is far more closer to distribution point A than to B. Calculation of the cost when adding C to distribution point B is in that case a waste of processing time. To prevent this, the model looks at the relative difference of the distance between AC and between BC. When the relative difference is small – so AC/BC is close to 1 – both options should be calculated. A limit for the relative difference is set (e.g. 0.9) above which both options should be calculated and if the relative difference is below the limit the point is assigned to the closest distribution point.

After all points have been assigned to distribution points, it is a vehicle routing problem for each individual distribution point. Now the route first – cluster second or cluster first – route second models can be applied.

![Figure 6. Cluster first – route second](image)

![Figure 7. Assignment - sweep](image)

3. METHODOLOGY

As can be noted from the two research sub-questions, this research consists of two parts. First a small literature study has been conducted about multi depot vehicle routing problems. These models have already been mentioned in the previous section. The goal of this research is not to study the MDVRP in complete detail, but to simulate models and compare the results to a real-life company. That’s the reason the discussed models are not an exhausting enumeration of possible MDVRP models. The mentioned models are the most common and have been advised for the problem at hand by an expert in this area (dr. ir. M.R.K. Mes) in an interview.

The second part of the research is the implementation of the models for a simulation. Data from a Dutch transport company has been used to conduct this simulation. The results of the simulations have been compared with each other. For the comparison to be possible, it has to be clear how the Dutch transport company works and if the constraints of the model suffice.

The models that have been chosen to study and implement for a simulation are those that fit best in the situation at the Dutch transport company. This criterion for selecting models is necessary to compare the results of the theoretical models with data from the real-life company. If the model would not be capable to describe the situation at the company mathematically, it would be pointless to study it and build a simulation, because the results of the simulation cannot be compared to the real-life case. It is therefore important to know
how the Dutch transport company works and how they make their planning.

For this reason the transportation company has been visited twice. The first time two transport managers were interviewed. Questions about the way the company worked were asked. After this interview it could be concluded that it was a classic example of a multi depot vehicle routing problem. The second time the company was visited was to see how a planner created a schedule for one day. It was clear that making a schedule was handwork and few models were used. The reason for this is that models often have too many assumptions or do not account for specific events. It would therefore be extremely difficult to compare the exact outcomes of simulations with the real-life schedules. But relative differences between models or between situations (e.g. with or without a certain customer in a planning) should be applicable on the real-life schedule, as long as these are relative differences and not absolute differences.

In this research simulations are used to study the MDVRP models. An advantage of simulations is that it can give results for different parameters in different settings. Simulations give solutions for problems that are (too) difficult for analytical models, and require less simplifications then analytical models. This has a positive effect on the validity of the findings. A disadvantage of simulations is that they create an estimate, instead of an exact result as analytical models produce. [13] However, solving the MDVRP analytical is difficult because it is NP-hard, as discussed in the introduction of section 2.

4. SIMULATIONS

The program to run the simulations is written in PHP (version 5.2) with a MySQL 5.1 database back-end. Simulations are executed on an Apple MacBook laptop (2.1 GHz Core 2 Duo, 4GB DDR2 RAM) with Apache 2.0.

The first step for the simulations is to put the delivery points from the transportation company for one specific day in the database. The addresses of the points are obtained from the transportation company in an Excel file. This file is converted to a CSV file using Microsoft Excel. Then for every line in the CSV file (i.e. for each customer or address) the GPS coordinates of the address is looked up using Google Maps. Google Maps has an Application Programming Interface (API) with which can be communicated through JavaScript programming code.

After all delivery points have been put in the database, it is possible to use these points for different modelling techniques. The database has been structured in a way that multiple routes and multiple models can use the same delivery points. The design of the database can be found in appendix A.

All models discussed in the literature review have been implemented. As discussed before, the assignment – sweep algorithm solves the MDVRP. This model uses the cluster first – route second and route first – cluster second models. And these models use the nearest neighbour, nearest insertion and farthest insertion models. The 2-opt methodology can be applied anywhere in between the process. The simulation code has been build up in such a way that the results of one model can be the input of the next. This has been done using multiple classes and functions. That way it’s not necessary to copy code or have multiple code fragments scattered everywhere in the file for one specific model.

When execution a simulation, the program keeps a log of how many times a distance between points has been calculated, what the total distance is for routes and what the execution time was. For every step in creating a route, debugging information about the process and state of the simulation can be displayed on the screen. This information is useful to see how the models perform and check if the program executes the models in the right way.

Textual information on a screen is not very intuitive when it comes down to making a representation of the route. For this purpose, a user interface using Google Maps has been build. After a route has been selected, the interface shows this route in Google Maps. Google Maps draws a straight line between the points on the route. This straight line is useful information, because all models use a straight line to calculate the distance between two points. But the Google Maps interface also shows the route over roads between two points, using the direction planner from the API. This function makes it possible to see the distance in kilometres and minutes of the entire route in the interface. A screenshot of the interface can been found in appendix B.

Since a few months a new version of the Google Maps API is available. Version 3 makes it possible to add waypoints. A route can be created by assigning a start and stop point (in this research always the same distribution point) and also intermediate points (called waypoints) on the route can be assigned. With version 3 of the API it is possible to let Google Maps come up with a route that is most efficient, i.e. put the waypoints in the right order. Unfortunately this is limited to 23 points, but at the moment it’s not even possible to add more than 10 points because of a bug in the API. Because version 3 is in beta it can be expected than in a while the maximum number of additional points is 23, but this is still too little because a route for the transportation company sometimes has more than 100 points. Still, this technique can be useful in future research.

4.1 Assumptions and limitations

The goal of simulations is always to stay as close to reality as possible. Unfortunately, the time available for this research is limited, so when making the simulations the reality has to be simplified. A number of assumptions and limitations for the simulations and used models are mentioned below. This list has been composed by using assumptions from previous studied literature and by comparing the real-life situation at the transport company with the possibilities of the simulations.

• In the end of the day/route it sometimes happens that the truck has to bring back some goods from a customer to the distribution center. Because this is very seldom in practise, the model doesn’t account for it.

• The distance between two points is calculated using GPS coordinates and Pythagorean theorem.

• There is no preferred delivery time or time constraints for customers.

• Changing the start and finish distribution point location is not an option. Changing the location of the distribution point to a more strategic point in the country could make the routes more efficient. This is the topic of a different study, namely the Multi Depot Location Routing Problem.

• All trucks are the same and have the same maximum capacity. At the transportation company almost all of the trucks have indeed the same maximum capacity of 33 pallets.

• It is not possible to have half or quarter pallets, only whole numbers.

• All trucks start at the same time in the morning and cannot drive a second route in the afternoon. A planning is for one day.
• Uncertainty about delivery (stochastic events) is not part of this research. It is not possible to make a schedule for when the trucks are already on their way or an extra customer presents itself during the day.
• A distribution point has an unlimited amount of trucks available. In reality it is also possible to rent trucks externally.
• Traffic jams, maximum driving times or breaks for the truck driver are not used in the simulations.
• The loading and unloading times are not taken into account for the total transportation time from Google Maps.
• For the transportation between distribution points (long haul transport) is in reality only one truck used.

That results in a capacity of 33 pallets for the long haul transport between two distribution points. The simulations currently do not take this into account.

5. RESULTS
The goal of the simulations is to calculate how efficient the models solve the MDVRP with data from an existing company. As explained in subsection 2.7, the assignment – sweep model is suitable to solve the MDVRP. First this model looks for points that are close to a distribution point and assign these points to the distribution point. When points are close to multiple distribution points, the additional costs to assign the point to one of the distribution centres is calculated. After that the point is assigned to the distribution centre with the least additional costs. Whether calculations should be made if the point should be assigned to the one distribution point or the other, or whether the point can be assigned to the nearest distribution point automatically is based on a limit. This limit is calculated by dividing the distance between the point and the nearest distribution point to the point and the next nearest distribution point (see also subsection 2.7). When the outcome of this calculation is below the limit, the point is automatically assigned to the closest distribution point. When the outcome is above the limit, the possible options for multiple distribution centres are calculated.

After all points have been assigned to the distribution centres, the routes have to be made. This can be done using the model route first – cluster second (see subsection 2.5) or cluster first – route second (subsection 2.6).

Creating a route for route first – cluster second or cluster first – route second, can be done using one of the models nearest neighbour (subsection 2.1), nearest insertion (subsection 2.2) or farthest insertion (subsection 2.3). For the simulations also the sweep-algorithm from Gillett and Miller (subsection 2.6) and for comparison a random generator are used to create a route. All five models are also executed with the 2-opt (subsection 2.4) improvement model.

In total there are three parameters that can be set in different ways for the simulation. The first is the limit that has been set to 0.85, 0.90, 0.95 and 1.00. The second parameter is whether to use the route first – cluster second or cluster first – route second algorithm. The third parameter is the model used to create a long route, which can be random, Gillett & Miller, nearest neighbour, nearest insertion or farthest insertion and these five models have the option to be improved with 2-opt. So in total 4x2x10=80 simulations have been executed to create 80 different schedules for one day.

To compare the efficiency of the simulations, the length of the route calculated by the model is used. This length of a route is calculated using the Pythagorean theorem between two points.

Because this research is about real-life data, Google Maps has also been used to calculate the distance and time of the route over the roads. That way it is possible to compare the length calculated by the models to the Google Maps (real-life) distance and time.

5.1 Results route first – cluster second
As explained before, one of the parameters is the model that can be set to route first – cluster second. Four different limits have been tested. In table 1 are the most efficient models mentioned for each limit and measurement method. For example, the first row can be interpreted as: for the limit 0.85 the random model with 2-opt is the most efficient by measuring the length of the route using the model (Pythagorean theorem). For limit 0.95 the most efficient model when looking at the distance over the roads using Google Maps is farthest insertion with 2-opt; using this model means that trucks have to drive 4653.45 km on one day. The same applies for the time column, which is a measurement of minutes instead of kilometres also using Google Maps.

Table 1. Most efficient models and their outcomes for the route first – cluster second model

<table>
<thead>
<tr>
<th>Limit</th>
<th>Length</th>
<th>Distance (km.)</th>
<th>Time (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85 RD 2-opt 35.44</td>
<td>RD 2-opt 4594.10</td>
<td>RD 2-opt 4168.09</td>
<td></td>
</tr>
<tr>
<td>0.90 NI 2-opt 34.84</td>
<td>NI 2-opt 4581.48</td>
<td>NI 2-opt 4172.01</td>
<td></td>
</tr>
<tr>
<td>0.95 FI 2-opt 36.89</td>
<td>FI 2-opt 4653.45</td>
<td>FI 2-opt 4192.58</td>
<td></td>
</tr>
<tr>
<td>1.00 RD 2-opt 33.39</td>
<td>RD 2-opt 4288.95</td>
<td>RD 2-opt 4046.28</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations in table 1:
- FI: Farthest insertion
- NI: Nearest insertion
- RD: Random

The procedure for simulations of route first – cluster second is as follows. First step is assigning points to a distribution point, depending on the limit used. Second one of the models (nearest neighbour, nearest insertion, farthest insertion, Gillett & Miller or random) is used with or without the 2-opt improvement method to create one long route. Third step is splitting up the route.

5.2 Results cluster first – route second
The second half of the results are generated using the cluster first – route second model. The limits and measurement methods are the same as in subsection 5.1. The most efficient models when applying different limits and measurement methods can be found in table 2.

Table 2. Most efficient models and their outcomes for the cluster first – route second model

<table>
<thead>
<tr>
<th>Limit</th>
<th>Length</th>
<th>Distance (km.)</th>
<th>Time (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85 FI 2-opt 43.76</td>
<td>NI 2-opt 5826.28</td>
<td>FI 2-opt 5018.37</td>
<td></td>
</tr>
<tr>
<td>0.90 NN 2-opt 43.96</td>
<td>FI 2-opt 5800.75</td>
<td>NN 2-opt 4953.69</td>
<td></td>
</tr>
<tr>
<td>0.95 NN 2-opt 42.31</td>
<td>FI 2-opt 5482.93</td>
<td>NN 2-opt 4721.38</td>
<td></td>
</tr>
<tr>
<td>1.00 NN 2-opt 42.21</td>
<td>NN 2-opt 5424.96</td>
<td>FI 2-opt 4720.73</td>
<td></td>
</tr>
</tbody>
</table>
Creating a route using cluster first – route second is almost the same as for route first – cluster second. First the points are assigned to a distribution point depending on the limit used. Second the points assigned to a distribution point are clustered using the Gillett and Miller model. For each cluster a route is created using one of the five models (nearest neighbour, nearest insertion, farthest insertion, Gillett & Miller or random) with or without the 2-opt improvement method.

5.3 Pythagorean vs. Google Maps

All models use the Pythagorean theorem to calculate the distance between points. In reality the distance between points is different because the truck has to drive over roads. Therefore simulations have been executed using the models with Pythagorean or Google Maps as basis for cost calculations. Table 3 shows the results. In both cases the 107 points for which a route had to be created was the same and used the route first – cluster second method with a limit of 1.00 for only one distribution centre. In the most ideal case all simulations mentioned in the previous subsections would also be executed with Google Maps, but that would take months to execute.

<table>
<thead>
<tr>
<th>Model</th>
<th>Distance Pythagorean theorem (km.)</th>
<th>Distance Google Maps (km.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random with 2-opt</td>
<td>9290.85</td>
<td>9283.71</td>
</tr>
<tr>
<td>Cluster (Gillet &amp; Miller) with 2-opt</td>
<td>4671.59</td>
<td>4671.59</td>
</tr>
<tr>
<td>Nearest neighbour with 2-opt</td>
<td>2339.19</td>
<td>2292.15</td>
</tr>
<tr>
<td>Nearest insertion with 2-opt</td>
<td>2253.00</td>
<td>2188.43</td>
</tr>
<tr>
<td>Farthest insertion with 2-opt</td>
<td>1978.26</td>
<td>2046.11</td>
</tr>
</tbody>
</table>

6. DISCUSSION

When looking at the simulation results, a few observations can be made. First it seems that the 2-opt models always gives the best result both the model and the Google Maps calculations verify this.

Second it seems that the higher the limit, the more cost efficient the models are. This is very counter-intuitive. In the simulations with a low limit (e.g. 0.85) some points are not assigned to the closest distribution point, but to a distribution point further away. This is done because at that time it seems that it would be more cost efficient to assign the point to a distribution centre further away. Therefore a lower limit should give better cost efficient results, because not assigning points to the closest distribution centre only happens if this is more efficient. Apparently in the end, assigning points to a distribution centre further away is less cost efficient.

A reason why higher limits are more cost efficient than lower limits, can not be given. It also can’t be denied that there isn’t a difference between a high and a low limit. For both the route first – cluster second and cluster first – route second models, the most efficient model has the highest limit. A possible reason for this anomaly is that the cluster size is too small. With a low limit, points are assigned to a distribution centre far away; although another distribution centre would be closer to the point this is more efficient for the total route. The routes from the distribution centres are split up later because of the maximum capacity of trucks. This splitting up of routes can have the negative effect that a truck of a distribution centre further away now has to drive a longer route to first get to this point, because the split was applied just before or after this point. A larger cluster size – i.e. a higher capacity of a truck – could improve the efficiency of lower limits.

The most cost efficient model overall is the random 2-opt in combination with route first – cluster second and a limit of 1.00. Disadvantage of this model is the long execution time of 54 minutes. The second best cost efficient model is nearest insertion 2-opt with a limit of 0.90, which takes 7 minutes to execute. Difference of total distance between these two models is only 5%. For the cluster first – route second model, the most cost efficient model is nearest neighbour 2-opt with a limit of 1.00.

It is clear that the route first – cluster second performs much better than cluster first – route second. A possible reason for this is that the clustering of routes in the simulations is not very intelligent. When the clustering is not efficient, the entire model of cluster first – route second doesn’t perform well. Clustering in the route first – cluster second model plays a less significant role in creating an efficient route.

There are three measurement methods to calculate the efficiency of the routes: Pythagorean theorem, Google Maps distance (kilometres) or Google Maps time (minutes). When comparing the Google Maps distance and time, the observation can be made that the best model often matches. The few times the most efficient model for distance and time are distinct, the relative difference is small; between 0.1 and 0.8 percent. Another observation is that Google Maps calculated that in the Netherlands on average 1,19 kilometres can be driven every minute. Because for 80 simulations the distance and time of more than 1200 routes have been requested from the Google Maps API, the average of 1,19 is rather accurate.

The most cost efficient model measured using the Pythagorean theorem always matched the Google Maps distance or time model. Three times the Pythagorean distance was different than the Google Maps distance, although the relative difference was small; between the 0.0 and 0.5 percent. Differences between the efficiency of routes measured by the Pythagorean theorem compared to Google Maps are caused by e.g. road obstructions, bridges, highways, etc.. To get an idea of the effects using the Pythagorean theorem to create routes, simulations have been executed using Pythagorean or Google Maps (see subsection 5.3). When comparing the outcomes of these simulations it can be concluded that using Google Maps for distance calculations between points has almost no effect above using the Pythagorean theorem. In fact, the two most efficient models used the Pythagorean theorem. Using Google Maps resulted in 6 out of 10 times for more efficient routes, but the difference is only 0.1 to 4.5 percent. Using Google Maps to create routes is in practice not very useful; the execution time is in the range of hours instead of minutes. It also has to be noted that the simulations only used one specific set of points and the results therefore does not have to be representative.

As described in section 3 the data used for the simulations are from an existing company, but for one day. For another day or another company, the data would be different and results possibly as well. Because one simulation run creates 80 different routes, it is too much work to compare results of
simulations for multiple days. It can be expected that the results would be almost identical. This assumption can be made because one simulation day has over 200 points or delivery addresses, which should be enough to rule out the option that points are located on exceptional places in the country.

It has also been mentioned in section 3 that only relative differences between (outcomes of) models can be compared. Comparison against the planning of the company wouldn’t be fair because of the many assumptions and limitations of the simulation.

Using the most efficient model it is now possible for the transportation company to quickly see the results of different situations. As described in the purpose statement (subsection 1.3), cost calculation of specific customers and comparing ways of distributing goods is important for the transport company but very difficult. It is now possible to execute two simulations with the same model and data, but for the second simulation with an additional customer or without a specific customer. Calculating the additional cost for a new customer (or less costs for the leaving customer) can be done by comparing the outcomes of the two simulations. Without the simulations this has to be done by creating two different schedules, which is a lot of work. Now simulations can be executed in a few seconds.

7. CONCLUSIONS
The main research question is what the most cost efficient model is to solve a real-life MDVRP. The answer to this question is the random model with 2-opt, route first – cluster second. The first research question asked how cost efficient the farthest insertion and 2-opt. Their turn make use of nearest neighbour, nearest insertion, sweep. Assignment – sweep makes use of route first – cluster second, cluster first – route second and assignment – sweep. The simulations that are used have a lot of assumptions and limitations, as mentioned in subsection 4.1. The number of limitations can be reduced in further research. Especially time constraints for delivering customers and taking the capacity of long haul transport into account can have a large effect on the results of simulations.

As mentioned in section 4 the new Google Maps API offers a lot of options. One possibility is to let Google Maps come up with the most efficient route. Because of technical limitations this was not yet possible for this research, but seems promising in the future. Another option is to see what would happen if Google Maps would be used instead of the Pythagorean theorem for the calculation of distances between points. A possibility is that the results of simulations are more cost efficient that way. Some experimentation using Google Maps to calculate distances between points has been done in this research, but more simulations have to be executed to be able to draw more general conclusions.

The splitting up of routes can possibly be improved. The cluster first – route second models performs significantly worse than the route first – cluster second, presumably because of the splitting and clustering which can be improved. The clustering of points is now done by rotating clockwise or counter-clockwise and taking the distribution point as centre. A possible better solution is to look at the geographical characteristics (bridges, highways, islands, etc.) or create clusters for points far away and clusters for points closer to the distribution centre.

One of the conclusions is that 2-opt always improve the results. It can therefore be assumed that k-opt (switching more than two connections) improves the route even further. Disadvantage is that calculation times increases. Another improvement can be to use a kind of 2-opt model to switch points (instead of connections) between multiple routes or distribution centres.

9. ACKNOWLEDGEMENTS
For the realization of this research paper, a number of people have been very helpful for which I’m grateful. Alberdine van Velzen and Chris Beezemer are managers at the transport company Post-Kogeko and were kind enough to make time available to answer questions but were also willingly to give a lot of information that was very useful for this research. Secondly Martijn Mes answered some questions about my research in and pointed out useful models. And last but not least Hans Moonen for making this research project available in the first place, but also during the process directed the research to keep it on the right track.

10. REFERENCES

8. FUTURE RESEARCH
Because of time limitations, this research can not be complete. The simulations that are used have a lot of assumptions and limitations, as mentioned in subsection 4.1. The number of limitations can be reduced in further research. Especially time constraints for delivering customers and taking the capacity of long haul transport into account can have a large effect on the results of simulations.

The most efficient model is always the same for the program as Google Maps calculations. It can be concluded that the models are applicable in real-life.

- Route first – cluster second performs better than cluster first – route second.
- The most efficient model measured by Google Maps is often the same whether looking at the distance or time.
- It appears that using Google Maps for distance calculations between points has no positive effect compared to using the Pythagorean theorem.


APPENDICES

A. DATABASE STRUCTURE

The database is structured in a way that the orders of one day can be used by multiple routes. An order consists of the specific day the order has to be delivered, number of pallets, the starting distribution centre (which is a reference to the table points) and where the order should be delivered (also a reference to points). The created route is saved in the table routes, which has a solution number to keep track of the different routes (or models used) and a truck number to know which trucks were used for all the routes. The routes_orders table links multiple routes to the same orders.

A truck makes stops on his route. The first stop is the start distribution point, after that stops are the customers and the final stop is the distribution point again. Stops are saved in the stops table with a reference to points. A point is a GPS coordinate with the name and address of that point. Figure 8 is a graphical impression of the database structure.

B. USER INTERFACE GOOGLE MAPS

A user interface has been built to see in a very intuitive way how routes go. Figure 9 contains a screenshot of the user interface. This user interface makes use of the Google Maps API to show a map of the country with the selected route on it. The route is displayed with straight lines between points but also the route over roads is shown. The distance and time of the route is calculated and shown in the upper right corner. Other options on the top side of the interface are ways to select a route: by day, solution number, distribution centre or route number.