UNIVERSITY OF TWENTE.

Linehaul Forecasting at TNT

January, 2019

Author:
Joep Karis BSc.

Supervisors:
University of Twente
Dr. M.C. van der Heijden
Dr.Ir. M.R.K. Mes

TNT
Roelien Bijker MSc.
Mark Timmermans MSc.

University of Twente
Industrial Engineering and Management
Track Production and Logistics

Preface

I conducted this master assignment to complete my master Industrial Engineering and Management. I learned a lot during my time at TNT and I would like to thank some people that helped me with that. First of all, I want to thank my daily supervisor from TNT, Roelien Bijker, who always had time for me and with whom I had some very useful brainstorm sessions. Moreover, I want to thank Mark Timmermans who always provided critical feedback, which helped me improving my thesis. Furthermore I want to thank Wiepke Dijkstra, Mor Verbin (ORTEC) and Harm Peeters who helped me during different phases of my research. I would also like to thank the (other) colleagues of the Planning and Engineering department that made my time at TNT a lot more fun.

Furthermore, I want to thank my first supervisor from the University, Matthieu van der Heijden, who helped me a lot during this whole project. You gave constructive feedback and helped improving the quality of the thesis. In addition, I want to thank my second supervisor, Martijn Mes, who entered in a later stage but also contributed to the end result. You gave a lot of useful suggestions of how to improve the research and helped improving my academic writing.

January, 2019
Joep Karis

Management summary

This research is conducted at the Planning and Engineering department of TNT. TNT provides services for the delivery of parcels, documents and freight consignments, mainly in the business-to-business environment. TNT developed a model that is able to forecast the throughput in hubs. However, this model does not provide insight into the number of trucks that should be planned between those hubs. To assist the tactical planners, TNT wants to develop a forecast that provides the volume between locations. The goal of this forecast is to decrease the number of last-minute cancellations or ad hoc hired trucks. Moreover, such a forecast makes it easier to standardize planning procedures.

The current situation of TNT was analyzed. We found that 22% of total movements are carried out by additionally hired trucks, and that 5% of the movements are cancelled. A lot of these cancellations and extra arranged movements are done last-minute. This results in suboptimal planning decisions. Based on the current situation of TNT, we formulated the following research goal:

"To develop a model that is able to forecast the (range of the) total volume from one location to another location within the TNT European Road Network to improve the planning. The time horizon for the forecast should be 13 weeks ahead and for every day a forecast should be made."

To find out what models are suitable for this goal, the pattern of the volume over time was analyzed. For this research, five links were selected by TNT from the 369 links that were within scope. The links are between the Road Transit Hubs, and are representative for the other links. The five links have different patterns, but in general, we saw there was a trend over the years. All the links show seasonalities during the year and during a week. The models we selected can handle trends and multiple seasonalities, although some models required some alterations. We selected an adjusted method for Holt-Winters* (which TNT also uses for the Hub-forecast), Holt-Winters double seasonal, TBATS, Neural Network, Simple Exponential Smoothing, and combinations of these methods. The asterisk in Holt-Winters* indicates it is an adjusted version of the standard Holt-Winters method. The Simple Exponential Smoothing model is chosen as a simple method to benchmark the more sophisticated models.

An important step in the forecasting process is the cleaning of input data. Public holidays and anomalies were replaced with average values to smoothen the data. In a later stage however, these public holidays should also be forecasted. After the data cleaning process, the data was split in a training- and a test set. The models and different parameters were fitted to the training set. After this step, these fitted models were used to make a daily forecast with a forecast horizon of 13 weeks.

To validate the forecasts, the forecasted values were compared to the actual values of the test set. In total, per link 100 forecasts were constructed. Each forecast has a different forecast starting date, which is between the 1th of November and the 8th of February (100 days). Each of these 100 forecasts had a forecast horizon of 13 weeks. For every day of these 13 weeks a forecast value is created, resulting in 65 (13 weeks of 5 days) forecasted values per forecast. The Mean Absolute Percentage Error (MAPE) was used to calculate the accuracy of the forecasts. The MAPE is calculated for all 65 forecast values. After this, the average over these 65 MAPEs is calculated for the entire forecast. This therefore results in 100 MAPEs, one for every forecast. With these values, we were able to calculate the average MAPE, the

minimum MAPE, the maximum MAPE and the standard deviation of the MAPEs over all 100 forecasts. This methodology is used for all different methods in order to compare their performance. The results are shown in Table 1.

Table 1: Results forecasting models

Model	MAPE	MAPEMin	MAPEMax	MAPEStd
Combination Holt-Winters*, Neural Network, TBATS	11.5	7.2	17.0	9.9
Holt-Winters*	12.5	7.1	19.6	10.5
Neural Network	13.1	8.1	23.6	10.7
TBATS	13.2	7.1	43.5	10.7
Holt-Winters double seasonal	15.1	7.8	46.1	11.2
Simple Exponential Smoothing	15.4	10.3	31.9	12.4

The best performing combination of forecasts was the combination of Holt-Winters*, Neural Network and TBATS. Overall, this combination outperforms the other methods. Only the Holt-Winters* method has a slightly lower minimum MAPE. When the individual forecasting models are compared, the Holt-Winters* model performs best on all measures provided in Table 1. The Neural Network scores well on most metrics and could also be an interesting option. The TBATS, Simple Exponential Smoothing and Holt-Winters double seasonal all have a high value for the maximum MAPE, which could indicate that those forecast models are less robust. The Simple Exponential Smoothing method was included in the analysis as a benchmark for the other models. It shows that most of the sophisticated models perform better than the benchmark method.

The Holt-Winters* method is the easiest method to implement and to understand by TNT, because this method was also used for the hub-forecast. The Neural Network is hard to understand and it takes time to tweak the model. On the other hand, the Neural Network has the advantage that it is easy to add information to the model to improve the forecast. An example is the public holidays that can be easily added. As another example, TNT may want to add the country where the origin or destination location is in, to enable the model to find synergies between locations. The combination of forecasts is the most difficult method to implement. The disadvantage of a combination of forecasts is that people of TNT should understand multiple methods, and that a change or addition to the forecast possibly has to be done for multiple models.

Based on this research, we recommend to implement the Holt-Winters* method at TNT. The Holt-Winters* method showed the best results for the individual forecasts, is easy to implement and does not need much tweaking by TNT. Furthermore, we recommend conducting further research on the Neural Network. The Neural Network is a promising method and we think the results can still be improved by adding more links to the data set and experiment more with the input settings. In addition, the forecasting of public holidays should be added and we think that the Neural Network is a promising method to generate good public holiday forecasts.

We created a 95% one-sided prediction interval for the Holt-Winters* method to provide extra information to the tactical planners. A one-sided prediction interval is preferred, because the tactical planners want to know what the probability is that the actual values are below the lower limit. The costs for last-minute cancellations are higher than the additional costs when comparing an ad hoc movement to a masterplan movement. In addition, it is not beneficial for the relationship with the subcontractors when a movement is cancelled. For this reason, tactical planners prefer to make a planning based on a value between the lower limit and the point forecast and not above the point forecast.

By implementing one of the forecasting methods, TNT can provide valuable information to the tactical planners. With this information, the tactical planners can try to reduce the number of last-minute cancellations and ad hoc movements. The forecast cannot be translated directly into a planning, because there is a planning step conducted by the tactical planners in between. We created a model that represents the planning process of the tactical planners. This model is used to calculate the cost savings of the Holt-Winters* forecasts compared to the current situation. We calculated two scenarios, each having a different input value for the planning process. The first scenario is a representation of the current situation. The input value for this scenario is the average volume for that link and weekday of last year. In addition, a standard expected growth factor of 5% is added. Moreover, an increase or decrease is added for high- or low volume months. The second scenario has he forecasted cubic meters of the Holt-Winters* method as input. We found that the forecast of Holt-Winters* could save around €3.5 million per year assuming the assumptions that we made about the costs and planning process are correct. We also created a third scenario, with the lower limit of the 95% one-sided prediction interval of the Holt-Winters* as input. The third scenario is compared to the second scenario and resulted in a cost increase. When TNT wants to use a lower-limit of a prediction interval as input, we advise to use a more narrow prediction interval or conduct further research about the planning process.

Moreover, there are some important benefits for TNT of forecasting that cannot be expressed in cost-savings directly. The implementation of a forecast in the planning process is a next step in becoming more data-driven at TNT. A forecast can improve the standardization of the planning process across multiple regions in Europe and lowers the need for experienced planners. Therefore, it will be easier for new employees to start working as a tactical planner. When the planning process becomes more standardized, it is easier to improve the planning process and save costs. In addition, the relation with sub-contractors can be improved by informing them earlier of changes in the planning. Finally, a forecast can contribute to the ongoing efforts to improve the masterplan by providing extra information.

Based on this research, TNT already decided that it is going to implement one of the forecasting models presented in this research. Furthermore we recommend the following points:

- Use the Holt-Winters* method for the forecast.
- Conduct further research about Neural Networks.
- Provide extra information to the planners, such as the volume for a certain link and the type of link.
- Keep improving the data cleaning process to improve the forecast. The anomaly list can be extended and should be kept up to date. In addition, other methods to improve the data cleaning process by filtering out outliers can be investigated.
- Investigate how to include the forecast of public holidays in the model.
- Conduct further research about the implementation of the prediction interval.

Table of Contents

Glossary and abbreviations

Linehaul	The general definition of linehaul, used in logistics, is the transportation of freight over large distances. This can be done on the ground, in the air or trough waterways by using various transportation options such as a truck, plane or train (ORTEC, 2018). However, within TNT and within this research, linehaul only refers to transportation by road.
Swap point	A location where two trucks, of different origins, will swap there trailers and return to their origin with the trailer of the other truck.
Masterplan	The base, tactical or minimal planning of trucks. In the masterplan it is stated at what time, how many, from which subcontractor and from which origin to which destination the trucks are going. In this planning, it is stated that every Monday a truck will drive from A to B, until this is updated. The masterplan is published to the planners every week.
Movement	A route between two locations. A movement contains an origin and destination and is performed by a single vehicle.
Tour	A tour consists of one or multiple movements. For example, a tour from A to B to C contains the movements A-B and B-C.
Sector	A route between two locations, including the latest possible departure time.
Link	A link between two locations means that there are trucks driving between those locations and it has (a) sector(s). This link is not necessarily the origin or destination of the parcels.
ERN	European Road Network
P&E	Planning and Engineering
LPC	Linehaul Management and Planning center
OD	Origin/Destination

1 Introduction

In this chapter, the research will be introduced. Section 1.1 provides a general introduction about TNT and the most important departments of TNT for this research. Section 1.2 presents an overview of the Road Network within Europe. Chapter 1.3 explains the problem context of this research at TNT and chapter 1.4 provides the research goal and questions. Finally, the scope of this research is discussed in chapter 1.5.

1.1 TNT

TNT provides services for the delivery of parcels, documents and freight consignments, mainly in the business-to-business environment. The two main services that TNT offers are the express service and the economy express service. The express service delivers parcels the next day and the express economy service is meant for parcels that are less urgent. In 2015, TNT had a revenue of 6.9 billion euro and shipped one million packages daily to over 200 countries worldwide by road or plane.

Most of TNT services would be classified as a LTL carrier (=Less Than Truckload). LTL carriers carry smaller shipments of customers and they try to consolidate many shipments in one truck to make it economically feasible. LTL carriers often have multiple depots and hubs via a hub-and-spoke network they ship the freight. The parcels are collected with smaller vans from the customer and are distributed to the depot. After this, the parcels get distributed via the network and end at the destination depot. Finally, the parcels are again distributed with smaller vans to the end-destination. This process can be seen in Figure 1.

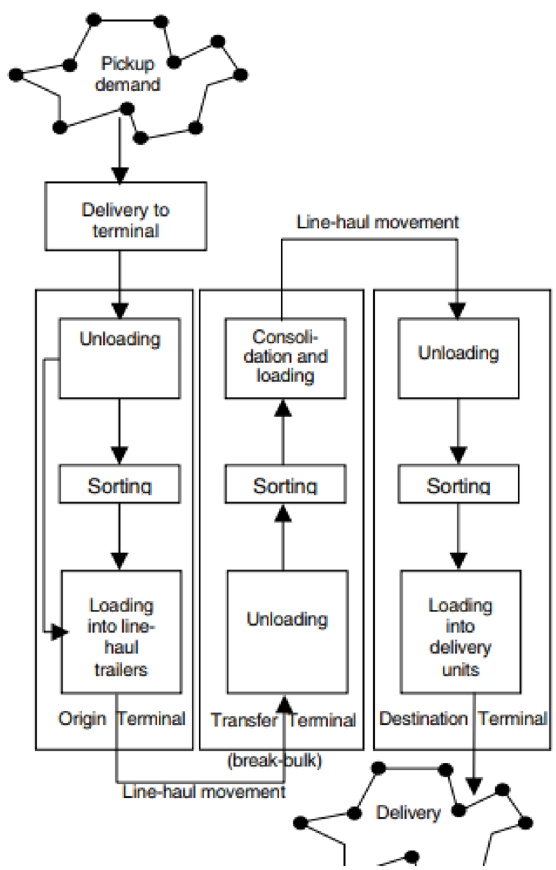

Figure 1: Hub and Spoke network

The advantages of a hub-and-spoke network are economies of scale and an improved customer service (Roy, 2001; Cunha & Silva, 2007).

In 2016 TNT was acquired by FedEx, which is an express distribution company from the United States. FedEx had a revenue of 47.5 billion dollars in 2015 and shipped 11.5 million packages daily to more than 220 countries. Currently the integration of FedEx and TNT is ongoing and more than 20 teams are working on this. The key factor for acquiring TNT for FedEx was the strong European network of TNT.

For this research, the Road Network P&E department and the Road Network Operations are the most important departments within TNT. The Road Network P&E department focusses on the optimization and different processes of the road network. The pick-up and delivery from and to the customer is not their responsibility. The activities of the P&E department are mainly focused on the longer term, and therefore more tactical and strategically orientated. Examples of different projects of this team are determining the location of new hubs, improving the network and quality monitoring of the network

Road Network Operations is responsible for the daily operation of all the hubs and linehauls. Within the linehaul department, there are nine Linehaul Management and Planning centers (LPC's). They are responsible for the movement of trucks throughout the hub-and-spoke network within their region. For example, they have the contact with subcontractors and plan how many trucks should leave at a certain time.

1.2 Road Network Europe

This section describes the Road Network of Europe. A more detailed version can be found in confidential Appendix A. The road network of TNT within Europe consists out of more than 500 hubs and depots, through which the volume is distributed to the customers. Figure 2 illustrates the 13 Road Transit Hubs within Europe. The Road Transit Hubs are the largest and most important hubs in Europe.

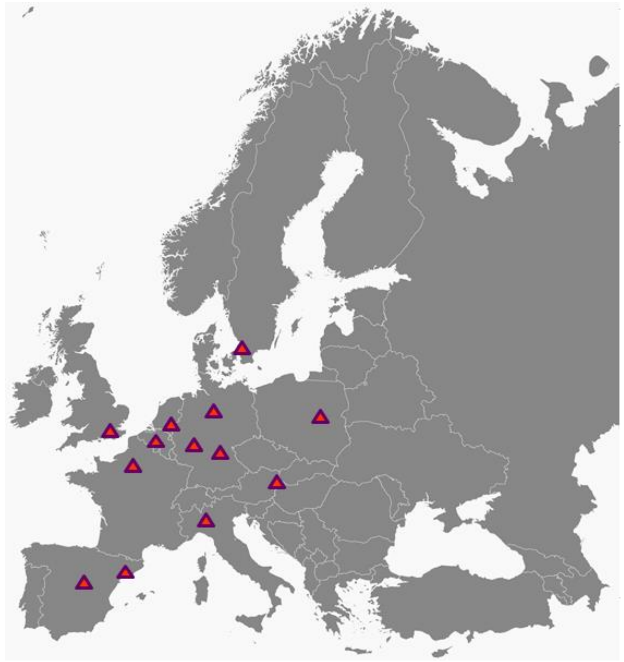

Figure 2: Road Transit Hubs of TNT

To explain how the routes from origin depot to destination depot are constructed, some terminology needs to be explained first. A link is a connection by road between two locations (similar terminology: “arc” or “leg”). There are different types of links specified by TNT.

Other terminology’s that needs explanation are a sector and a route. The definition of a sector and a route is comparable to respectively a link and path. A sector is a link that contains timing information, i.e., the scheduled departure and arrival times. A route is a set of consecutive sectors, defining the physical path implied with the sectors’ links, combined with the timing of the sectors. A movement is a vehicle moving between two locations, with specific details regarding departure location and departure time, arrival location and arrival time, and the vehicle type that is used. A tour describes a sequence of movements that a single driver will execute on one or multiple days.

Every OD depot combination has one main route per service type per weekday. Within TNT, a service type is called product. Customers can choose different products to deliver their goods. The two main products are the economy and express product. The express product has a faster delivery than the economy product and therefore, it is possible it has a different route. The main route is similar to the main path, but includes arrival and departure times. This main route is also the route that is scheduled by TNT. In practice, TNT has a few options to deviate from this route when volumes fluctuate but the rules and options for this are confidential.

These main routings are fixed and not adjusted operationally on a short term. If for example the volumes change over a long time, or there are improvement opportunities, these main routings can be changed by TNT.

1.3 Problem context

TNT has made a model to forecast the throughput of hubs, to see how many packages arrive daily. This forecast provides useful insights for the sorting hubs and can for example help with workforce scheduling. However, this forecast does not tell anything about the destination or origin of the packages. Therefore, this forecast is of no use to the tactical planners of the hub. These tactical planners are among other things responsible for the planning of trucks. They decide how many, when and to which destination trucks are going. In the current situation, this is primarily influenced by the experience of the planners. Because the number of trucks is a large cost driver and last-minute changes are stressful and costly, there is a need to have a forecast for the required load or number of trucks on a certain link.

If a truck is cancelled and it will not drive, but the subcontractor has to be paid, the direct costs are €135 on average. If an extra truck is scheduled last-minute, the truck will be €104 more expensive on average. However, closer to the departure, the risk of not being able to arrange an extra truck increases. Sometimes it is also possible to change the destination of the trucks when the alternative destination is close to the original destination, however this does not happen often.

Within TNT, there is a project to improve the quality of the base planning. This base planning is called the masterplan. This masterplan is the main input for the planners and could be described as the current forecast. In the masterplan, it is stated at what time, from which subcontractor and from which origin to which destination the trucks should be going. Based on the masterplan, TNT makes agreements with subcontractors. The number of trucks that are entered in the masterplan is most of the time the minimum number of trucks needed. For example, when the number of trucks fluctuates on a certain route between eight and twelve, the number of trucks in the masterplan will probably be eight. The reason behind this is that TNT does not want to cancel trucks because that is more expensive than the additional costs of arranging an extra truck and it is not beneficial for the relationship with the subcontractors. The masterplan does not take holidays or peak volumes during the year into account. TNT wants to improve the accuracy of the masterplan, in order to have less changes and the process will become smoother.

In the current situation, the process of making a planning is mainly based on the masterplan and the experience of the tactical planners. There is no standardized process in place. This makes it harder for new people that start working at TNT to learn the process and to make a good planning. It is also harder to contain knowledge within TNT when people leave. In addition, because the decisions are made mainly on experience, the decisions sometimes differ per tactical planner and it is hard to improve the process.

In the organization, there are different units to measure objects related to volume. For example, the forecast for the hub throughput is measured in the number of packages, but other people use cubic meters, kilograms or number of trucks. Therefore, an analysis that is made for a department can be useless for another department because of a different unit of measurement. This is important to keep in mind when making a forecast.

After the integration with FedEx, the focus will be more on planning, use of data and standardization. This means that they want to achieve a shift from doing more things on the tactical area compared to the operational area. They want to achieve less last-minute changes by getting a better planning beforehand for example.

Concluding, TNT wants to explore the possibilities to provide the planners with a linehaul transport forecast to support them in their decision-making. This way, costs can be saved on the number of trucks, less last-minute changes will be made and the standardization process can be improved.

1.4 Research goal and questions

The goal of the research is to develop a model that is able to forecast the volume from one location to another location within the European road network (ERN). The time horizon for the forecast is required to be 13 weeks ahead and this forecast will be made every four weeks. The forecast will have one output per 24 hours. A further explanation about why these decisions are made can be found in Section 2.4.

This results in the following research goal:

To develop a model that is able to forecast the (range of the) total volume from one location to another location within the TNT European Road Network to improve the planning. The time horizon for the forecast should be for 13 weeks ahead and for every day a forecast should be made.

To answer this research goal, different sub questions are formulated:

1.Current situation

-1a. What is the performance of the current masterplan forecast?

-1b. How long before a movement has departed are extra trucks or the cancellation of trucks being arranged?

-1c. What is the pattern of the historical volume data?

-1d. What are the requirements for the forecast?

To find out how the forecast should work and why a forecast is needed, it is important to investigate the current masterplan forecast. This has two reasons, namely to see if there is a need for improvement and to see if an improvement is even possible, and to compare the model in the end with the current situation. Examples of indicators about the current performance of the forecast are the number of cancelations and the number of extra trucks compared to the masterplan. To calculate this, data is needed that shows the number of movements that were in the masterplan for a particular day and the number of actual movements that were done that day. A very large dataset that contains all movements, together with the timestamps when these movements were updated, should be processed and aggregated to obtain the right data.

If the forecast would improve the current planning, it should also be easier to change the number of trucks compared to the masterplan by the tactical planners. This results in a shift from operational planning to tactical planning. For this reason, it is interesting to see how long before a departure the movements are added or cancelled in the current situation. This can be calculated by the different times stamps that are in the dataset we discussed previously.

To see what the historical pattern is of the volume data, the volume per link per day should be gathered. This can be collected from the Enterprise Data Warehouse system of TNT and this will probably take some time to obtain the right information. In addition, the requirements for the model should be gathered. One example is the aggregation level of the model, to see what the exact level of detail will be. For now, it is decided by TNT that a forecast will be made with certain detail levels, for example, the output will be in cubic meters and the forecast will be per link. But when the results show that with this aggregation the output of the forecast will not be reliable, another aggregation should be examined. To find out what an acceptable aggregation method is, interviews with stakeholders should be conducted. The stakeholders should for example indicate if a forecast per country or region is also acceptable compared to a forecast per link.

2. Literature

2a. What is known in literature about forecasting?
2b. What is known in literature about forecasting in the transport sector?
2c. Which forecasting techniques from literature are most suitable?

To come up with a solution and to see how related problems are handled in other companies and industries, it is good to do a literature study. In this literate study, the relevant literature about forecasting in general, forecasting in the LTL industry and the planning of trucks will be examined.

3. Design of the model

3a. How should the model be made?
3b. How can the validation of the model be done?

After the relevant literature is reviewed and the problem context is clear, these two should be combined and the most suitable forecasting technique(s) need to be customized for TNT. After this, the model should be developed and programmed. Finally, the validation of the model should be conducted, to see if the forecast realizes logical outputs and if it it has the required precision.

4. Model execution and evaluation

4a. What is the performance of the model?

When the model is developed and executed, the results should be collected and analyzed. The accuracy of the model should be investigated.

5. Conclusion and recommendations

5a. What are the conclusions and recommendations of this research?
5b. How can the solution be implemented?

After the results of the model are discussed, the conclusion and recommendations of the research should be presented. This is based on all the previous questions. Finally, an implementation plan for the solution should be developed.

1.5 Scope

After interviews with people from P&E, we decided not to take all movements within the TNT network into account. In this research, priority is given to the movements within the ERN, which have a Hub – Hub connection or have a Hub – Depot/Depot - Hub connection with at least a distance of 250 kilometer between the Hub and Depot. This means, that 369 links will be taken into account of the total of 2513 links. These links within scope are used to calculate the current situation in Section 2.2.

2 Current situation

In this chapter, answers will be provided to the first research question. Section 2.1 explains the current forecasting and planning process, and also describes the solution domain. When the current process is clear, Section 2.2 analyzes the performance of the planning. Finally, Section 2.4 explains the requirements of the forecast.

2.1 Current forecasting and planning process

As described in Section 1.3, TNT would like to develop a forecast. To understand what kind of forecast is needed, and how the forecast will be implemented, it is important to describe the current planning process and the forecast that is currently used. Section 2.1.1 explains the masterplan, which is the starting point of the planning and can be seen as the current forecast or input for the planners. In Section 2.1.2, the role of the tactical planners is described and finally in Section 2.1.3 the role of the operational planners is discussed.

2.1.1 Masterplan

The masterplan is a set of structural tours and movements, which is continuously updated. In the masterplan, it is stated at which day of the week, what time, from which subcontractor and from which origin to which destination the trucks are going. Based on the masterplan, TNT makes agreements with subcontractors. The number of trucks that are entered in the masterplan is most of the time the minimum number of trucks needed. When movements are cancelled or extra planned for a period of four or five weeks in a row, these movements often will be added or deleted to the masterplan. By adding or deleting these movements to the masterplan costs can be saved. The costs of a masterplan movement are lower than the costs of a movement that is planned in a later stage. In addition, when a masterplan movement is cancelled last-minute a cancellation fee should be paid. The masterplan is published every week for the tactical and operational planners.

2.1.2 Tactical planners

The tactical planners are normally looking four weeks ahead or longer when there are holidays that influence the network a lot. The tactical planners use the masterplan as a basis and try to see if there are changes needed. A week before execution, the tactical planners publish the plan and send it to the operational planners. Both the tactical planners and operational planners are situated in the Road Network operations department.

2.1.3 Operational planners

The operational planners are responsible for the daily planning. When there are for example more packages in the hub for a certain destination normal, the operational planners can arrange an extra truck for movements. Such changes that happened last minute are defined as “ad hoc”. There are a few disadvantages to arrange trucks in a later stadium. First, sometimes there are no subcontractors available anymore to arrange a movement. Second, if a subcontractor is willing to arrange a truck, the costs are on average €104 higher. Finally, when a truck needs to be arranged in a later stadium, it is harder to make the movement a round-trip instead of a single movement. A roundtrip is a trip where the truck contains volume back and forth. When the truck is arranged in an earlier stadium, it is possible to contact the destination location and try to find a synergy and make it a round trip. Round trips are preferred because fewer trucks are needed with more roundtrips. The operational planners provide daily feedback to the tactical planners about the planning they received.

2.2 Performance of current forecast

The masterplan, although it is very limited, can be seen as the current forecast. The masterplan is the most important input the planners receive to base their planning on. To study the quality of the masterplan, the number of trucks that is stated in the masterplan is compared to the number of trucks that actually drove. Data about all movements from January 2017 until March 2018 is collected, which results in a data file that contains 4.2 million rows. In this dataset, a lot of information is given per movement, for example the planned and actual departure/arrival times, the origin- and destination locations and what kind of load the movement had. For every movement, the data set also contains multiple rows, which contain information about the time this movement was updated in the system and what the status was on the moment of the update. When the data is for example sorted on the update time, the first row of a certain movement is the first time the movement is entered or updated in the system and often has the status “Missing Resources”, which means that not all resources such as the driver of that movement are yet entered in the system. Other statuses are, for example, “Ready for departure”, “Arrived” and “Completed”.

To obtain the dataset in the right format, we added some extra calculated columns, and performed validation checks. The data is edited and analyzed in R. To know the distance between locations and to assign a movement to a LPC, other tables with this information are merged with the initial data. A column is added that indicates if this row has the latest updated time, so when the number of movements is calculated, every movement is taken into account only once. Finally, the data is aggregated, to know per day, per OD what the number of planned masterplan movements and number of realized movements is. With these two numbers, the number of cancellations, extra movements and normal movements according to the masterplan can easily be calculated:

$$Normal\ movements = \mathrm{minimum}(Realised\ Movements, Planned\ Movements,)$$

$$Extra\ movements = \mathrm{maximum}(0, Realised\ Movements - Planned\ Movements)$$

$$Cancelled\ movements = \mathrm{maximum}(0, Planned\ Movements\ -\ Realised\ Movements)$$

On June 27th, 2017, TNT was hit by a cyber-attack, resulting in multiple IT-systems that were not operational anymore. This cyber-attack was costly but also decreased the data quality around that period. For this reason, the data between 2017-06-19 and 2017-07-16 are excluded from the data-analysis.

In Figure 3, the number of normal, extra and cancelled movements per LPC are displayed. Appendix B states the countries and cities of the abbreviations of the LPC. It can be seen in the figure that there are differences between the LPC, but on average, of all movements, 73% is normal, 22% is extra and 5% is cancelled. This is as expected since the costs of cancelations are higher than the costs of an extra movement.

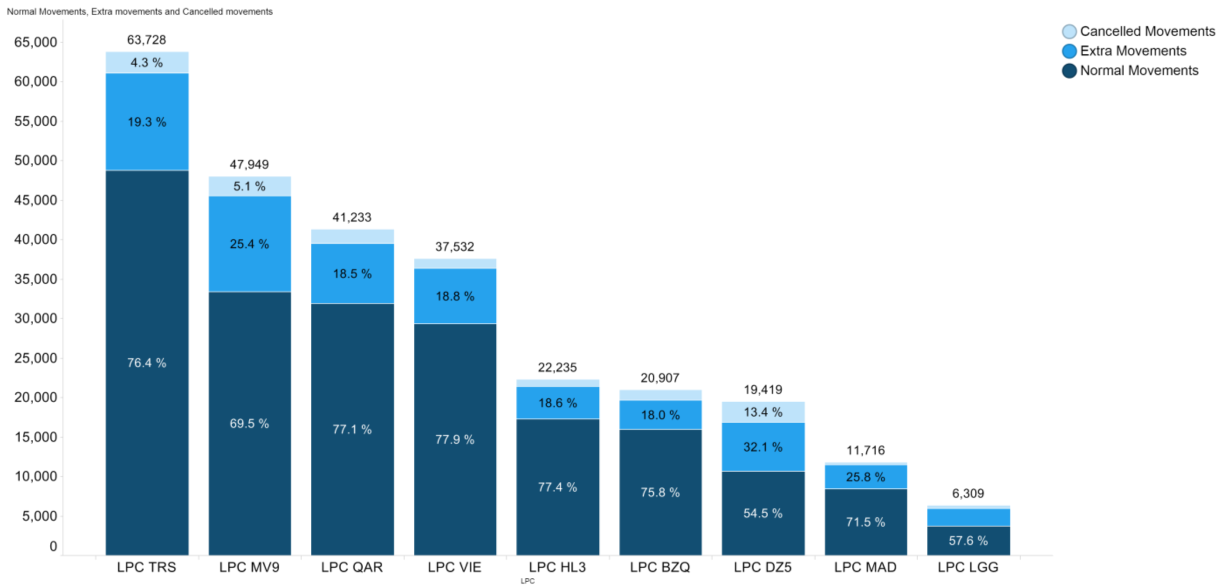

Figure 3: Number of different type of movements from 2017-01-01 until 2018-03-01

Figure 4 shows the performance of 2017 and the first part of 2018. It can be seen that the total number of movements (normal plus extra movements) is fluctuating throughout the year. In addition, the percentage of normal movements, which should be considered as the current performance and should be as high as possible, is fluctuating but does not show a real trend last year.

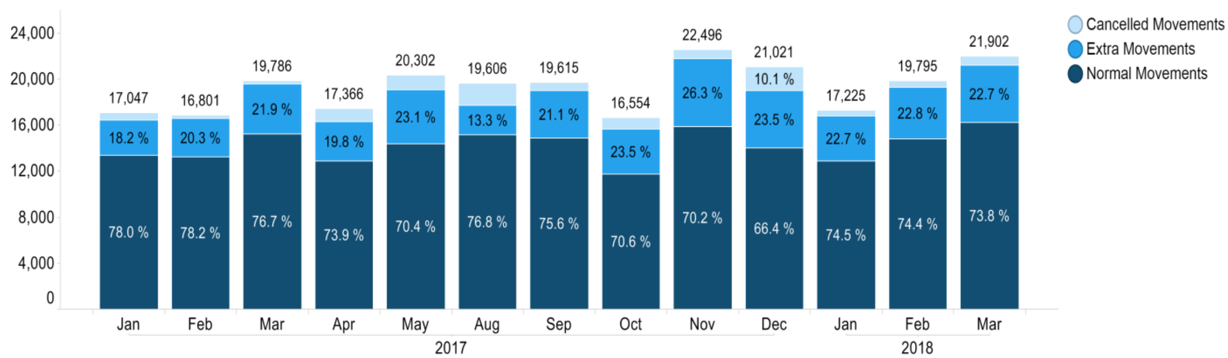

Figure 4: Number of different type of movements

Figure 5, shows the distribution of the percentage of normal movements per link. The links that have on average less than one movement per week are excluded. It can be seen that there are 30 outliers, these outliers are studied in detail and most of them have a logical reason. An example of an outlier is the new location Dartford; in the beginning there were only extra movements and no normal movements. The reason for this is that this location was probably not added to the masterplan initially. Furthermore, 75% of the links have a normal movement percentage between 66.4% and 99.4%, but the other 25% of the links have quite a low percentage of normal movements. It is interesting to see if these links can be improved by the use of a forecast.

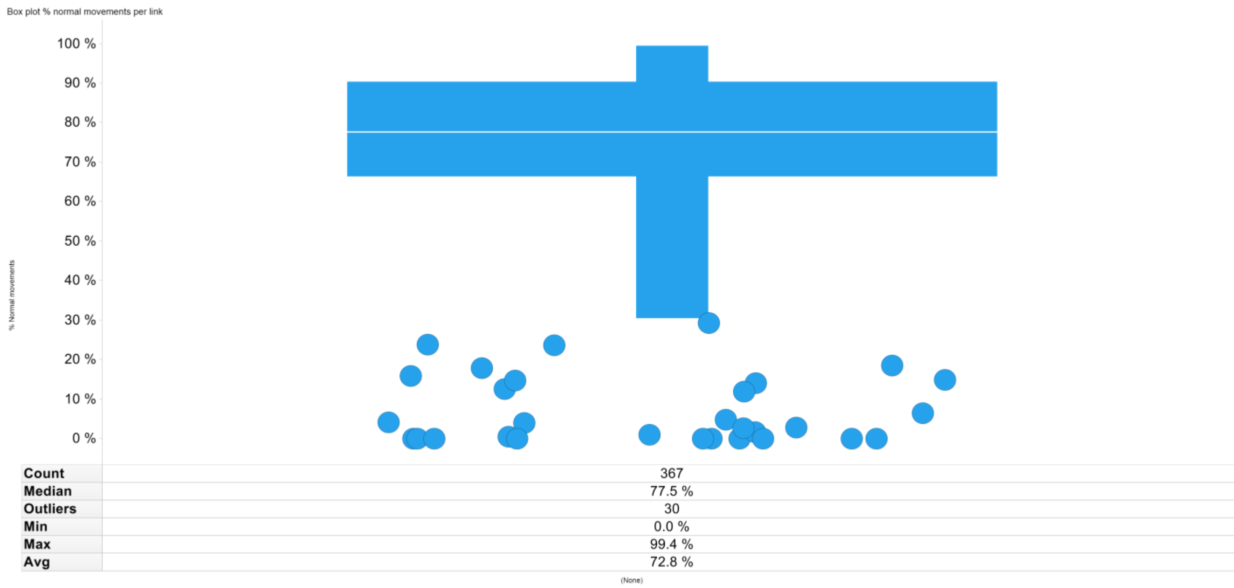

Figure 5: Percentage of normal movements per link

TNT thinks that a forecast can also help to arrange a shift of decisions from the operational area to the tactical area. Changes to the masterplan should be done preferably four weeks before departure instead of on the departure day itself. It is not known what the current distribution is between tactical or operational decisions and how long before departure changes are made. To see if the forecast actually realizes a shift from operational to tactical, it is important to analyze the current situation.

For the analysis of the timings of the extra and cancelled movements, the same movement dataset is used, without the aggregation step previously described and with some additional calculations. Two indicators are added. First, one that indicates if this row has the earliest update time. This will be used to compare the creation date of an extra movement and the planned departure time. Second, one that indicates if this row has the earliest update time, given that the status is "Cancelled". This indicates at what time the movement was cancelled.

The dataset contains columns that state the initial planned-, the actual planned- and actual departure and arrival date and time. The date and time when a cancelation is entered in the system is compared with the actual planned departure time. The actual planned departure time is used, because on the moment of cancelation this is probably the current planning. For the extra movements, the update time is compared to the initial scheduled departure, because when entering the extra movement, the initial scheduled departure was the current planning. The update time is subtracted from the departure time and date to see how long before departure a cancellation or extra movement is added.

Figure 6 shows that a large portion of the extra movements are planned 24 hours (the first two bars) before their initial departure time, namely 48% of the total extra movements. In addition, 99% of the extra movements are planned within ten days before initial scheduled departure. This implicates that there are not many extra movements planned long before the initial scheduled departure time and most of the extra movements are planned ad hoc.

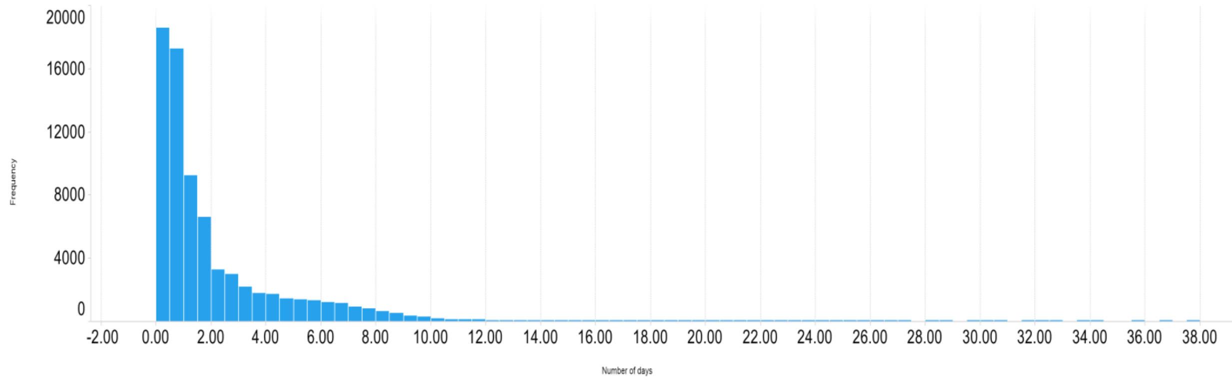

Figure 6: Frequency of number of days between update time and initial scheduled departure time for extra movements

Figure 7 is comparable with Figure 6, but in this figure the time difference for the cancellations between the update time and actual scheduled departure time is shown. Cancellations happen earlier before the scheduled departure time compared to the extra movements, namely 34% of the cancellations are cancelled 24 hours or less before the scheduled departure time. In addition, 99% of the cancelled movements are cancelled between 0 and 16 days before scheduled departure.

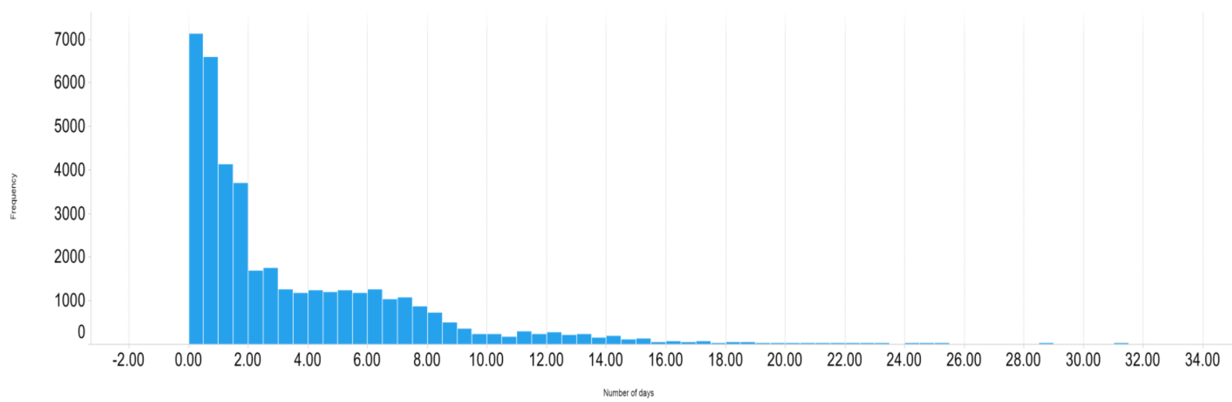

Figure 7: Frequency of number of days between update time and actual scheduled departure time for cancellations

Hence, cancellations are done earlier compared to adding extra movements, but in both cases, still a large quantity is done within 24 hours. With a forecast, a shift from operational to tactical can be achieved.

2.3 Pattern of the historical volume and movement data

In this section, the patterns of the historical volume are analyzed, to see if it is possible to make a reliable forecast and to see what kind of patterns exist in the data. This analysis will be done for two different sets of input data, which will be explained in Section 2.3.1 and Section 2.3.2.

2.3.1 Forecasting before taking routes into account

The first approach could be to make a forecast based on the Origin/Destination (OD) depot data and after this forecast, assign the volume to the corresponding links. In total, there are more than 520.000

OD combinations from 2014 until 2018. Some general remarks about the combined volume of all OD combinations are listed below:

- The volume is increasing over the years
- A seasonal pattern with 3 peaks and one drop during the year is visible (see appendix C)

Since there are more than 520.000 OD combinations, it is impossible to investigate all those combinations one by one to see if there is a possibility to make a reliable forecast for each individual OD combination. To make an estimation about the reliability, the number of observations (an observation is a day with volume) per OD combination is determined to find out if it has enough observations to obtain a reliable forecast. This showed that there are many OD combinations that do not have enough observations. The graph is shown in Appendix D. We analyzed the OD combinations that do not have enough observations, to find out how much volume is associated with these links. We decided that an OD combination should at least have volume in 80% of the weekdays. About 50% of the total volume belongs to OD combinations that have less than 80% of the times volume during weekdays. The other 50% have more than 80% of the times an observation during weekdays. When the requirement would be to account for 90% of the volume when choosing the links that are going to be forecasted, there will also be OD combinations that have only in 10% of the weekdays an observation. These links that have only in 10% of the weekdays an observation will be almost impossible to forecast, note that all links have at least one observation. We decided not to make a forecast based on OD combinations.

Another option is to aggregate the data more, to have better forecast results. To achieve this, the data can be aggregated based on country, all the volume of all the depots in the Netherlands is aggregated to a total volume for the Netherlands for example. This results in 8510 OD country combinations. About 1.7% of the volume belongs to OD country combinations that have less than 80% of the times an observation when this was possible during the weekdays. Based on this we assumed that a reliable forecast on OD country combinations is possible. After a forecast on country level, a translation has to be made from OD country pairs to the links. An option can be, to map the volume between two countries on the most important links, so for example of the total volume from the Netherlands to Spain, 10% will use the link from Eindhoven to Paris. This option is less preferable because this will need extra computing steps, will be time consuming to construct and only the larger lanes will have a reliable forecast. It is possible however that the accuracy of the larger lanes will increase compared to a separate forecast per link, since the aggregated data can have less variance, which makes it easier to forecast.

There are also other aggregations methods possible based on other characteristics of OD combinations. Examples can be a clustering of similar links, with the same pattern or an aggregation based on the same region of the origin or region of the destination, or based on both. A data analysis can be done, to find certain characteristics that influence that pattern of the volume. Other aggregations methods however still have the same disadvantage as an aggregation based on country level: it will result in a reliable forecast for the larger lanes, but the smaller lanes can have a low forecast accuracy.

2.3.2 Forecasting after taking routes into account

The second approach is to assign the volume from OD combinations to the scheduled main-routes and after this make a forecast per link. The scheduled routes for one week are downloaded, which contains for every weekday the corresponding route for an OD combination. This has the advantage that historical volume will be assigned to a current route and changed routes will not have an influence. This results in a total of 2513 links within Europe.

Together with TNT, five links are selected that are analyzed in more depth, to find the historical patterns. In general, the most important links are the links between the 13 Road Transit Hubs. The five links that are selected are between those 13 Road Transit Hubs, since these forecasts are the most important. The size of the links is different and there are two relatively large links, two medium links and one small link based on volume, compared to the other links between the 13 Road Transit Hubs. The main goal is first to find out if there is a trend, and if so, what kind of trend. Secondly, the goal is to find out if there are seasonal patterns.

In Appendix E, the five links are explored in more depth and per link the different patterns are given. In addition, some example graphs are given to provide a better understanding. Below the most important conclusions are presented:

- For all the five links, there is a yearly increase in volume between 2014-2016, but 2017 is different per link. In 2017, the cyber-attack took place and because of this it was a disappointing year for TNT. Concluding, there is a trend between 2014-2016 and looking to the period 2014-2017 it is either a damped trend or a linear trend. See Appendix F, for an example.
- The monthly pattern over the year is similar over the different years for the different links. There is a clear seasonality per link, but this seasonality is different per link. Concluding, there is a seasonal pattern during the year per link. See Appendix G, for an example.
- The week pattern during the year is for most links very weak over the years. See Appendix H, for an example.
- The distribution of volume between the weekdays is for some links the same over the months and for other links there is no or little visual pattern. This means that for some links, the distribution of volume between weekdays is the same for every month and in other cases this differs per month. See Appendix I, for an example.
- The daily pattern per weekday is for all links quite stable, except for some outliers, which can be explained by public holidays in most cases. See Appendix J, for an example.
- The distribution of volume between the weekdays per month per year is for most links quite the same. The distribution of volume between the weekdays in January is compared between the different years, for example. Therefore, per weekday there are only four or five data points with volume (4 or 5 Mondays in January per year). Since there are little data points, it was expected that it was not completely the same over the years.

In general, we see that the volume shows a trend with two seasonal patterns in every link. One seasonal pattern is the seasonal pattern during the year and the other seasonal pattern is during the week, with a distribution of volume between the weekdays. In addition, the different patterns can provide a useful indication if it is possible to make an aggregation based on time. In the end, the output of the forecast should be per day, but this does not mean that the input data for the forecast should be per day. An option is to forecast the volume per week and after this, translate it back per weekday based on historical distribution. With a higher aggregation level, the accuracy of the forecast will increase but aggregation is only feasible when a translation back to day-level is possible. Another option can be to develop a forecast model that can handle multiple seasonalities.

2.4 Requirements of the forecast

In this section, the requirements of the forecast are explained. Based on these requirements the research goal of Section 1.6 is formulated. For most requirements, there is balance between the aggregation level and statistical reliability, most of the time a more detailed and thus less aggregated forecast leads to lower statistical reliability. Section 2.4.1 explains the time horizon of the forecast and Section 2.4.2 explains the different options for the time buckets of the forecast. Section 2.4.3 describes the possibilities of the origin and destination of the forecast and finally Section 2.4.4 explains the output of the forecast.

2.4.1 Time horizon

The forecast will mainly be used by the tactical planners and since these planners look ahead four weeks, the forecast should at least have a horizon of four weeks. The forecast horizon also depends on the frequency with which the forecast is distributed. The hub-forecast is distributed every four weeks and this is also preferred by TNT for the linehaul forecast. This means that the forecast horizon should be at least eight weeks, but the same as for the hub-forecast the forecast-horizon is determined to be 13 weeks. The reason for a complete time horizon of 13 weeks, is that sometimes tactical planners look ahead longer, e.g., in case of upcoming holiday periods. In addition, when a forecast is not distributed in time to the tactical planners, the tactical planners can use the previous forecast, which also includes the upcoming four weeks.

2.4.2 Time buckets

There are two main options for the time buckets of the forecast, where a time bucket refers to the period a forecast is generated for, for example for a day, week, or month. The first option is to make a forecast for every 24 hours and the second option is to make a forecast per sector. A sector is defined as a connection between two locations including the latest departure time. For example, from hub A to hub B, during one week, every weekday at 11.00 am there is a sector (this does not necessarily mean that only one truck leaves) and on Wednesday there is an additional sector at 15.00 pm. On Wednesday there will be two connections between hub A and hub B with different latest departure times. Therefore, there will be one sector on a daily basis and two sectors on Wednesday. This means that for this week, for this route, if the forecasts were per sector, the model would generate six (4 + 2) forecasts.

We decided to use a time bucket of one day, because it would be more difficult and probably less reliable to have a forecast per sector. When taking sectors into account, it is hard to assign volumes to a certain sector. In addition, the sectors and corresponding timings change over time and will influence the results of the forecast. Only 18% of all links between locations have more than one sector per day. Generating a forecast per sector would be an option for further research.

2.4.3 Origin and Destination

For the origin and destination of the forecast, different options can be chosen. Origin does not necessarily mean the origin-depot of the package, with origin the origin of the movement is meant. A package can, for example, start at depot A and needs to be delivered to depot C. TNT uses the route A-B-C to deliver this package, then also hub B is the origin in the movement B to C. The first option is to take the actual location of a hub or depot, this means that no aggregation will be done based on location. The second option is to take a region and to aggregate the locations that are within this region. An example of this can be that the west of the Netherlands is one region and Hubs/Depots in the cities of Amsterdam, Utrecht and Den Haag are consolidated into one. When an option is chosen that contain regions instead of actual locations, it should be further investigated how these regions should be defined.

The option that is preferred by the tactical planners is a forecast from location to location on a link level, so without consolidation of hubs/depots in a region. This is also the option that is chosen in this research. Another option would be to make a forecast from a location to a region. This option is less preferred but could still help improving the planning. The option from region to region is not detailed enough for the tactical planners to be useful. Concluding, first a forecast is made from location to location and if this result is not good enough, than an option could be to make a forecast from a location to a region.

2.4.4 Output

There are several options for the unit of measure of the output of the forecast, namely in cubic meters, kilograms or the number of trucks. It is different per LPC what unit of measure they use to indicate how much load there is, but the number of trucks is the least favorable option for the LPC's. Based on interviews with people from P&E, it is decided that the output and main focus of the forecast will be in cubic meters to determine the number of trucks. A truck has two restrictions how many parcels can be loaded in the truck, namely the available volume of the truck and the weight restriction. However, in practice the weight is almost never the restricting factor, so volume is a better estimate of how many trucks are needed. In addition, the goal is that the tactical planners can use the forecast in combination with the utilization report, in which they can see what the utilization of trucks was in the past for certain movements. This utilization report is also in cubic meters, so it will be easier to combine with the forecast if the forecast is also in cubic meters. However, the planners now often use kilograms as a measurement unit, the reason for this is that in the past, the data of cubic meters was not accurate and now they are used to work with kilograms. So, when the forecast will be implemented it is important to change their way of thinking from kilograms to cubic meters. If the forecast is implemented, there will also be a translation from cubic meters to kilograms in the beginning, but the primary focus is on cubic meters.

It is also possible to use the cubic meters and add an extra level of detail. An option is to make a distinction between the type of load, so for example the distinction between non-stackable freight (called awkward freight), dangerous goods and normal freight. In case of awkward freight, less cubic meters fit into a truck. In the whole ERN network, there is about 89.9% normal freight, 3.9% awkward freight, 6.0% dangerous goods and 0.2% that is both awkward and dangerous. We decided not to make this distinction because this will probably result in an unstable prediction. If the results of this research are good enough, this could be something for further research.

The output of the model should preferably have a prediction interval and not only a point forecast. For the tactical planners it would be helpful if they receive a forecast with a prediction interval of 95%, for example. An option for the tactical planners can be to plan the trucks based on the lower limit and inform the operational planners there is a chance that extra trucks will be needed. This will result in fewer surprises for the operational planners, but the decision can be made in a later stadium. In addition, they can already see whether it is possible to create a round-trip.

2.5 Conclusion

Section 2 answered different research questions about the current situation of TNT. The main findings are listed below:

- With a new forecast, the number of last-minute cancelled movements and ad hoc movements can be reduced.
  - In the current situation 22% of the movements are arranged additionally by the tactical and operational planners and 5% is cancelled by the tactical and operational planners.
- Many extra and cancelled movements are arranged within 24 hours.
  - The analysis showed that 48% of the cancellations are cancelled within 24 hours of their departure and 34% of the extra movements are arranged within 24 hours of their departure. Almost no extra and cancelled movements were arranged longer than 2 weeks before departure. This confirms the feeling of TNT that a lot is done last-minute and the goal to carry out more beforehand is not yet achieved.
- The overall patterns showed that there are seasonal and trend components in the historical data.
  - There was a damped trend over the years (mostly 2017 was different for the different links, possibly due to the cyber-attack). During the year, there is a seasonal pattern and also during the week there is a seasonal pattern visible. The model that is chosen should be able to cope with multiple seasonalities and trends.
- The forecast have the following requirements:
  - Forecast the total volume in cubic meters.
  - Forecast the volume from origin to destination on a link level, with no further aggregation.
  - Make one forecast per four weeks with a forecast horizon of 13 weeks and a focus on the first four weeks. The forecast will contain one output value per day.

3 Literature

In this chapter, the third research question will be answered. In Section 3.1, different forecasting techniques with their (dis)-advantages are presented and Section 3.2 explains different methods to assess the performance of a forecasting model. Section 3.3 presents details about literature in the field of forecasting in the transport sector and Section 3.4 outlines the chosen forecasting models for this problem.

3.1 Forecasting techniques

Many forecasting techniques have been developed in history, each of them has its own advantages and disadvantages and it is important to choose the right method (Chambers, Mullick, & Smith, 1971). Makridakis et all. (1982) agree and add that the right choice is important both theoretically and practically. Often even small improvements of a forecast can have significant savings.

There are different methods available to forecast and these methods can be categorized as following (Chambers, Mullick, & Smith, 1971; Makridakis, et al., 1982):
1. Qualitative techniques, based on judgments and opinions of experts.
2. Time series analysis, when there is enough data and trends and relationships are clear.
3. Causal or explanatory models, for example regressions models.
4. Or a combination of the above.

A time series is a set of observations, in which each observation happened on a specific time t (Brockwell & Davis, 2016; Chambers, Mullick, & Smith, 1971). The problem and available data in this research can be classified as a time series analysis, since there is enough data and there are no external influences that can be taken into account. Although public holidays can be seen as an external influence, this will not be the main input and not the primary focus.

To offer a better understanding of all the formulas that are presented in this chapter, Table 2 provides the most common symbols with an explanation.

Table 2: Common symbols

Symbol	Explanation
y_t	Actual value of time period t
$\hat{y}_{t+h\|t}$	Forecast value for h steps ahead, given time period t
l_t	Level of time period t
b_t	Trend of time period t
s_t	Seasonality of time period t
${y_t}^{(w)}$	Transformed actual value of time period t
d_t	ARMA error of time period t
x_i	Independent variable
α, ε, β, ϕ, γ, ω, φ	(smoothing) Parameters

Section 3.1.1 until Section 3.1.6 explains different type of forecasts with the focus on time series.

3.1.1 Moving average

One of the most used methods of forecasting for time series is the moving average method. This method is easy to understand and easy to use. The two most frequently used versions of the moving average method are the Simple Moving Average (SMA) and Autoregressive Integrated Moving Average (ARIMA).

The SMA just takes the average of the last x periods. So, when a new data point is added, the last data point will not be taken into account anymore. ARIMA is seen as one of the most important and most used time series model in the field (Zhang, 2003). ARIMA can be used for a time-series that is not necessarily stationary. When the data is stationary, often an Autoregressive Moving Average (ARMA) can be used (Brockwell & Davis, 2016). Stationary data is data where the mean and co-variance are stable and its properties do not depend on the time at which the series is observed (Brockwell & Davis, 2016; Hyndman & Athanasopoulos, 2018). This means that there should be no trend or seasonality. If an ARIMA model is used to fit a non-stationary time series, often differencing is applied to convert the data to stationary data. Since the data of TNT shows a trend and seasonal behavior, the data should be preprocessed to make it stationary. The disadvantage of ARIMA models is that they are quite complex and there is a need for a skilled and experienced forecaster.

3.1.2 Exponential smoothing

Exponential smoothing is another forecasting model that is widely used in different industries (Gardner, 1985). Figure 8 shows different forecasting profiles, which can be forecasted by different exponential smoothing methods. Each square in Figure 8 has its own technique and formula, but in this chapter only the most common techniques are discussed.

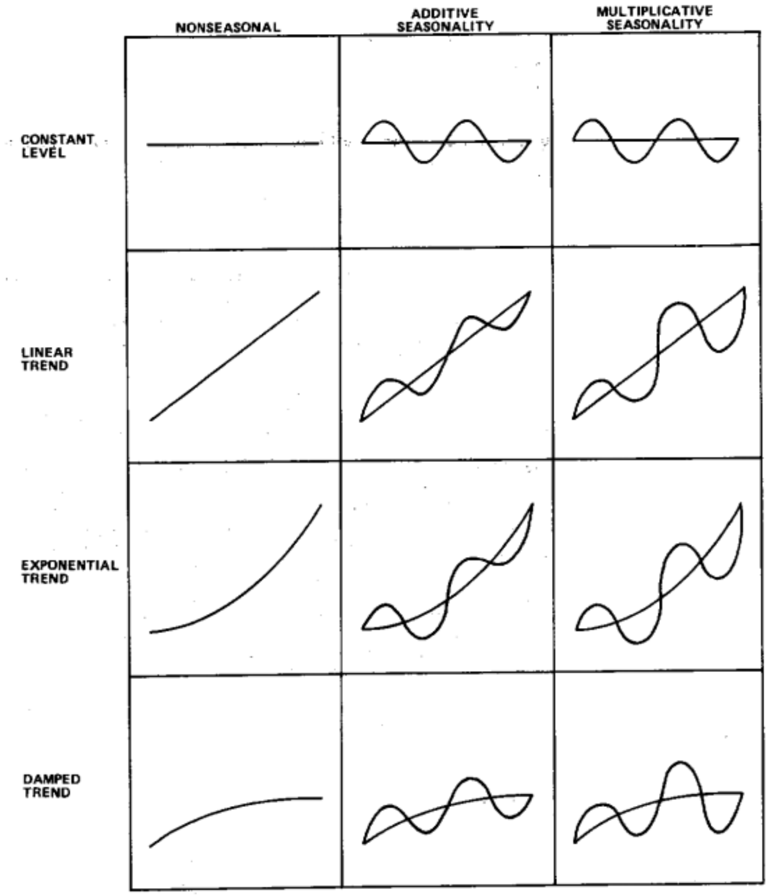

Figure 8: Forecast profiles (Gardner, 1985).

The simplest technique is the Simple Exponential Smoothing (SES) method. When a time series fluctuates around a base level and does not show a clear trend or seasonality, this method can be used to obtain a good forecast (Winston & Goldberg, 2004; Hyndman & Athanasopoulos, 2018). The formula for simple exponential smoothing is as following:

$$\hat{y}_{t+h|t} = l_t \tag{1}$$

$$l_t = \alpha y_t + (1 - \alpha) l_{t-1} \tag{2}$$

Where $\hat{y}_{t+h|t}$ is the forecasted value of time period *t* for *h* steps-ahead forecast, α is a number between 0 and 1 and y_t is the observation of time period t . When the value of α is closer to 1 it means that the smoothing factor is less and the formula gives more weight to recent changes.

When a data set has a linear trend (and no seasonality), a good option for a forecast is the Exponential smoothing method of Holt's, also called double exponential smoothing. The formula for double exponential smoothing is as following:

$$\hat{y}_{t+h|t} = l_t + h b_t \tag{3}$$

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{4}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{5}$$

Where l_t is the formula for the level of the series at time at time t and b_t is the trend component of the series at time t. Again, the smoothing parameters α and β should be between zero and one.

When the data is linear but also has seasonality, the Holt-Winters forecasting method can be useful, also called triple exponential smoothing. This method is an extension of the double exponential smoothing. There are two general forms of the Holt-Winters forecasting method, namely the multiplicative method and the additive method (Hyndman & Athanasopoulos, 2018). The multiplicative method is useful if the seasonality increases over time. The additive method should be used when this is not the case, this can also be seen in Figure 8. For Holt-Winters forecasting, equations (6) and (7) will be changed and s_t will be added which is the seasonal component. Below are equations for the Holt-Winters triple exponential smoothing for the multiplicative case (the formula for the additive case can be found in Appendix K):

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)} \tag{6}$$

$$l_t = \alpha(\frac{y_t}{s_{t-m}}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{7}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{8}$$

$$s_t = \gamma\left(\frac{y_t}{l_t}\right) + (1-\gamma)s_{t-m} \tag{9}$$

Where α, β and γ are between zero and one. m stands for the frequency of the seasonality, which means the number of seasons in one year. With a monthly seasonal pattern, the frequency would be 12 for example. k is the integer part of *(h-1)/m*, which makes sure that the estimation of the seasonal indices is coming from the final year of the training sample (Hyndman & Athanasopoulos, 2018).

Most exponential smoothing methods such as Holt-Winters are designed for small seasonal periods such as monthly or quarterly seasonal data. For daily or weekly data that contains about 365 or 52 periods the method is not that suitable, since it then needs to estimate 364 or 51 parameters and this will probably be too much (Hyndman & Athanasopoulos, 2018). The suitability of Holt-Winters also depends on the amount of data, if there is enough data, Holt-Winters can be a useful forecasting method. At TNT there is data for more than four years, which is assumed to be enough to perform a Holt-Winters analysis on a weekly basis.

The Holt-Winters method originally only accounts for one seasonal period, Taylor (2003) adapted the Holt-Winters model so that is possible to accommodate two seasonalities. The same approach can be used to incorporate more than two seasonalities. The original formulas of Taylor (2003) can be found below.

$$\hat{y}_{t+h|t} = (l_t + hb_t)s^1{}_{t+h-m_1(k+1)} * s^2{}_{t+h-m_2(k+1)} \tag{10}$$

$$l_t = \alpha(\frac{y_t}{s^1{}_{t-m_1} * s^2{}_{t-m_2}}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{11}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{12}$$

$$s_t{}^1 = \gamma\left(\frac{y_t}{(l_{t-1} + b_{t-1})s^2{}_{t-m_2}}\right) + (1-\gamma)s^1{}_{t-m_1} \tag{13}$$

$$s_t{}^2 = \omega\left(\frac{y_t}{(l_{t-1} + b_{t-1})s^1{}_{t-m_1}}\right) + (1-\omega)s^2{}_{t-m_2} \tag{14}$$

In the paper of Taylor (2003), the half-hourly electricity demand is forecasted, which shows a seasonal pattern over the day and a seasonal pattern over the week. In the example of the half-hourly electricity data, there are 48 half-hour periods within a day and 336 half-hour periods within a week. The seasonal periods in this example are nested, which means that the two periods are a multiplication of each other (48 multiplied with 7 is 336). Furthermore, by calculating the seasonality within a day and within a week, seasonal or calendar effects during the year are not taken into account in this example. The seasonal effects during the year are assumed to be the same.

The way Taylor (2003) defined his formulas, the two seasonal periods cannot be of different time measurements because there is only one parameter for the time (t). In the example of Taylor (2003), both seasonal periods were half-hour time periods. In our case, we want to take the seasonality within a week into account, for which we need the time period to be not larger than days. When the first seasonal period is chosen to be days, the second seasonal period should also be in days. To take the seasonality over the year into account, the second seasonal period will therefore result in 364 days. A frequency of 364 would result in many degrees of freedom, which is not ideal. Therefore, we altered the formulas of Taylor (2003) to facilitate multiple seasonalities without being restricted to using two seasonal periods of the same time measurement. The same logic and philosophy of Taylor (2003) is applied, but the mathematical formulation is different. The main change is the addition of a different time bucket for the second seasonal period. This makes it possible to have an aggregated volume for the second seasonal period, for example per month. This results in the calculation of the level, trend and seasonality separately for both periods. For example, the smaller period (related to $s\ short_t$) can be chosen to be a day and the larger period (related to $s\ long_r$) can be chosen to be a month. In this example, t represents the day and r the month, therefore r =2 means month two. There will be two input tables, one with the daily volume and one with the monthly volume. For both these tables, the level, trend and seasonality are calculated. The level and trend of the monthly data are not used in the daily forecast (formula (15)), but are only constructed to calculate the second seasonal factor per month (formula (21)). The altered formulas are presented below.

$$\hat{y}_{t+h|t} = (l\ short_t + hb\ short_t) * s\ short_{t+h-m_1(k+1)} * s\ long_{r+h-m_2(k+1)} \tag{15}$$

$$l\ short_t = \alpha(\frac{y_t}{s\ short_{t-m_1} * s\ long_{r-m_2}}) + (1-\alpha)(l\ short_{t-1} + b\ short_{t-1}) \quad (16)$$

$$l\ long_r = \varepsilon(\frac{x_r}{s^2{}_{r-m_2}}) + (1-\varepsilon)(l\ long_{r-1} + b\ long_{r-1}) \quad (17)$$

$$b\ short_t = \beta(l\ short_t - l\ short_{t-1}) + (1-\beta)b\ short_{t-1} \quad (18)$$

$$b\ long_r = \varphi(l\ long_r - l\ long_{r-1}) + (1-\varphi)b\ long_{r-1} \quad (19)$$

$$s\ short_t = \gamma\left(\frac{y_t}{l\ short_t * s\ long_{t-m_2}}\right) + (1-\gamma)s\ short_{t-m_1} \quad (20)$$

$$s\ long_r = \omega\left(\frac{x_r}{l\ long_r}\right) + (1-\omega)s\ long_{r-m_2} \quad (21)$$

The y_t represents the volume of the short period t and x_r represents the aggregated volume of period r. In formula (17) and (21), the seasonal effect of the short period is not taken into account. The reason for this is that it is not possible to take this into account, since the longer period contains multiple different shorter seasons.

3.1.3 TBATS model

The TBATS model is developed by De Livera (2011) and is comparable to the Holt-Winters model. TBATS uses a combination of Fourier terms, an exponential smoothing state space model and a Box-Cox transformation (Hyndman & Athanasopoulos, 2018). TBATS is an acronym for the most important elements of the model: Trigonometric, Box-Cox, ARMA errors, Trend and seasonality. The formula for TBATS is presented below, where $y_t^{(w)}$ represents the Box-Cox transformed observation with parameter w and y_t is the observation at time t. The Box-Cox transformation alters the input values, which are used by the rest of the formulas of TBATS. Formulas (22) and (23) represent the Box-Cox transformation, which is a power transformation technique. This technique is for example useful when the variance increases or decreases when the level increases or decreases (Hyndman & Athanasopoulos, 2018). A Box-Cox transformation tries to reduce the heteroscedasticity, non-normality and non-additivity (Sakia, 1992). When the parameter $w=1$, the shape of the pattern will not change, but the values will only shift a bit downwards. For the other values of w the shape of the pattern will change.

$$y_t{}^{(w)} = \frac{y_t{}^w - 1}{w} \qquad w \neq 0 \quad (22)$$

$$y_t{}^{(w)} = \log(y_t) \qquad w = 0 \quad (23)$$

$$y_t{}^{(w)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}{}^{(i)} + d_t \quad (24)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \tag{25}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t \tag{26}$$

$$s_t{}^{(i)} = \sum_{j=1}^{k_i} s_{j,t}{}^{(i)} \tag{27}$$

$$s_{j,t}{}^{(i)} = s_{j,t-1}{}^{(i)} \cos \lambda_j{}^i + s_{j,t-1}{}^{*(i)} \sin \lambda_j{}^i + \gamma_1{}^{(i)} d_t \tag{28}$$

$$s_{j,t}{}^{*(i)} = -s_{j,t-1}{}^{(i)} \sin \lambda_j{}^i + s_{j,t-1}{}^{*(i)} \cos \lambda_j{}^i + \gamma_2{}^{(i)} d_t \tag{29}$$

$$d_t = \sum_{i=1}^{p} \varphi_i \, d_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t \tag{30}$$

Formula (24) represents the measurement equation and consists of the level of last period*(l_{t-1}),* the trend of last period *(b_{t-1})*, the seasonal components *($\sum_{i=1}^{T} s_{t-m_i}{}^{(i)}$)* and an ARMA error process *(d_t)*. T represents the total number of seasons and i represents the *i*th seasonal component. The normal exponential smoothing methods assumes that the error process $\mathrm{d_t}$ is uncorrelated, however in practice this is not always the case and the forecast accuracy can be improved by explicitly modeling the errors (De Livera, Hyndman, & Snyder, 2011). The seasonal components of formula (28) and (29) are a trigonometric representation based on Fourier series. A Fourier series is a method that can model different patterns by the use of different harmonic sine waves. k_i is the number of harmonics for the *i*th seasonal component and *j* represents a specific harmonic. $s_{j,t}{}^{(i)}$ represents the stochastic growth in the level of the *i*th seasonal component with harmonic *j* in time t. This growth is needed to describe the change of the seasonal component over time by $s_{j,t}{}^{*(i)}$. $\gamma_1{}^{(i)}$, $\gamma_2{}^{(i)}$, α and β are the smoothing parameters, ϕ is the damped parameter and $\lambda_j^{(i)} = 2\pi j/m_i$ (De Livera, Hyndman, & Snyder, 2011).

Hyndman (2018) suggests using a TBATS model instead of Holt-Winters, when there is weekly or daily seasonal data. The advantage of using a TBATS model is that it allows the seasonality to change slowly over time and that it allows the period of seasonality to be non-integer, which is preferred for accommodating the seasonal period over the year. In addition, the model can accommodate multiple seasonalities and they do not necessarily have to be nested. A disadvantage is that it takes quite a lot of computational time and often the prediction intervals appear to be too wide (Hyndman & Athanasopoulos, 2018). For the data of this research, TBATS can be a good option, since there is daily data with a strong weekly pattern and seasonality over the year. TBATS can accommodate this multiple seasonality and because those two seasonal periods of the data are not nested, TBATS is a good option.

3.1.4 Artificial Neural Networks

The forecasting technique Artificial Neural Networks (NN) is based on mathematical models of the architecture of the biological neural networks in the brain (Mair, et al., 2000). NN can be used for time series forecasting as well as regression forecasting. NN can be applied to a wide range of forecasting problems and yield a high accuracy (Khashei & Bijari, 2011).

A NN can be seen as a network of neurons that are organized in different layers. In most cases, three layers can be distinguished. The first layer is the input layer, the second layer contains hidden nodes and the last layer is the output layer. The outputs for the nodes of one layer are the inputs for the nodes in the next layer (Hyndman & Athanasopoulos, 2018). Figure 9 provides a graphical representation of a neural network.

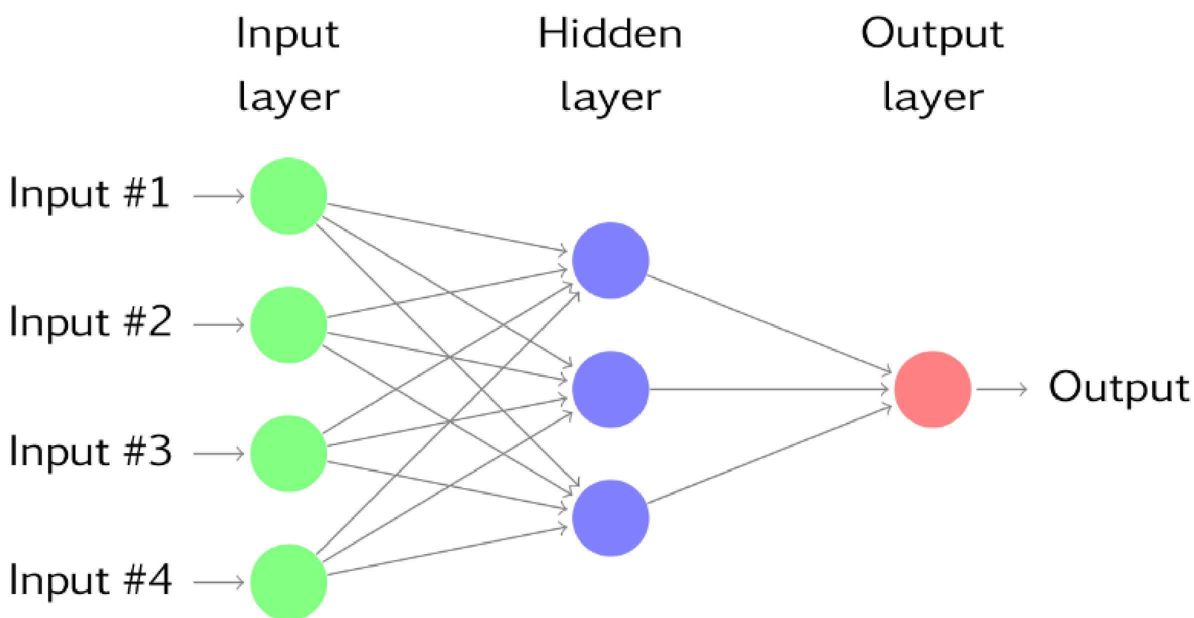

Figure 9: Example of a Neural network (Hyndman & Athanasopoulos, 2018).

Zhang and Qi (2005) conclude that in order to obtain good performances using NN for data that possesses seasonal and trend patterns, the data should be processed first. The NN cannot immediately capture the seasonal and trend characteristics. The data should obtain a detrending and deseasonality procedure, which means that the trend and seasonality components are removed from historical data according to Zhang and Qi (2005). However, other authors state that using NN is also possible without removing seasonality. Hamzaçebi (2008) shows that the opinions about the need for removing seasonality of the data are not in line. He found five articles that are in favor of removing seasonality and six articles that state that removing seasonality is not needed. In addition, NN often have the risk of overfitting and therefore it is important to have hold-out data to test this (Zhang & Qi, 2005). Overfitting means that the NN takes too much information into account and also learns from outliers and noise.

Makradis (2018) reviewed and compared statistical methods and machine learning methods. His conclusion based on literature was that there are many papers about machine learning and NN but an empirical comparison with more standard statistical methods seems to be lacking. The data set that Makradis (2018) investigated showed better accuracy of statistical models in comparison with machine learning models. Moreover, the paper expressed a concern that there is not enough research and no suitable method to construct a prediction interval around the point forecast. However, multiple studies showed ways of including prediction intervals. Khosravi (2011) reviewed four methods to include a prediction interval for NN: delta, Bayesian, bootstrap and Mean Variance-Estimation (MVE). The different methods each have their own advantages and disadvantages regarding quality, repeatability and computational load. A combination of the four methods results in the best results. Hyndman (2018) also showed how to implement a prediction interval with a simulation method based on the bootstrap model.

A disadvantage of NN is the black-box principle, which means that it is hard to know what the NN is doing and find out what the influence of the different input parameters and the relation between the input parameters is on the output. This results in none or little explanatory value of a model and it is difficult to improve the model since it is unknown what it is doing (Olden & Jackson, 2002).

A NN can be a good option to make a forecast for TNT, because it can find relationships between different input variables that another model is not able to find. The NN will take all data of all links into account, and might find a relation between the origin country of a hub and the weekly pattern. The other models that we discussed, fit a model on the data of one link, therefore possible valuable insights can be missed. In addition, in a later phase the prediction of public holidays can be easily added and can maybe offer better predictions than other methods. Finally, it is interesting to see how a NN performs in comparison to the more traditional models, since there is no consensus in literature about what performs best yet.

3.1.5 Linear regression analysis

Linear regression methods try to forecast a dependent variable based on one or more independent variables. An example of a dependent variable is the load on a certain movement, while independent variables may be the day of the week and the month of the year. Regression analysis often provides useful forecasts (Armstrong, 2011). Regression analysis is classified as a causal or explanatory model as described in Section 3.1.

Simple linear regression is the simplest method of linear regression. This method is displayed in the following equation:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{31}$$

Here x_i is the independent variable, β_0 and β_1 are regression coefficients and ε_i is an error term.

When there is more than one variable, the multiple regression method is used. This method is similar to the simple linear regression method. The formula for k independent variables is presented below.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_k x_{ki} + \varepsilon_i \tag{32}$$

Typically, regression is used when there is separate independent data available, for example when the weather or economic growth in a country is used to predict the number of trucks. As said, in this research there is no such external data available and to use a regression model, the current data should be divided in independent variables such as the year, week, day of week and origin country.

3.1.6 Combination of different forecasts

Finally, a good option to improve the result of a forecast is to combine different forecasts. Often two or more forecasts are executed and after this weights are attached to the forecast (Bates & Granger, 1969). The opinions of combining forecasts are nearly unanimous; it increases the forecast accuracy and even by averaging, the results can be improved dramatically (Clemen, 1989; Hyndman & Athanasopoulos, 2018). A disadvantage of combining multiple forecasts is that it will increase the computation time (Zhou, 2017). In addition, one condition to successfully combine forecasts is that the individual forecasts should be unbiased (Bates & Granger, 1969).

The reason why combining forecasts results in better performance is that the discarded forecasts typically contain some useful information. Discarded forecasts are the forecasts that otherwise would not be used. The discarded forecasts can contain information or variables that the other forecast(s) did not consider or the forecast makes a different assumption about the form, patterns or relationship between the variables. Different assumptions of the forecast do not necessarily lead to an improved forecast, but this is possible in some cases (Bates & Granger, 1969).

3.2 Forecast performance

Many different forecasting models can be chosen. In this research, the most promising models will be tested. To compare these models, it is important to find the right performance and accuracy measurements.

To assess the performance, it is generally agreed that the data set should be divided into two sets, i.e., the training set and the test set, also indicated by in-sample and out-sample (Tashman, 2000). The training set is used to determine the parameters of the model and to provide a first indication about the performance of the model. The test gives an indication how the forecast would perform in real-life, without already knowing how the future will look like. For this reason, it is important not to include the data of the test set in the training set. A test set should be large enough to obtain reliable output and it should be representative for the whole data set, therefore it is not wise to take for example one season of a year as a test set.

A lot is written about choosing the right forecast-error statistics (Tashman, 2000). First, the error measurement should be chosen. A forecast error (e_t) is the difference between the observed value and the forecasted value. The notation for this is as follows:

$$e_t = y_t - \hat{y}_t \tag{33}$$

In general, there are three options for the error measurement, namely scale-dependent errors, percentage errors and scale errors (Hyndman & Athanasopoulos, 2018). Scale-dependent error methods are used when the errors are on the same scale as the data, which means it is not a unit-free measurement. The disadvantage of not using a unit-free measurement is that when multiple time-series are compared, series with large numbers dominate the comparisons (S. & Collopy, 1992). Despite this disadvantage, scale dependent error methods are quite popular. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are widely used to assess the accuracy. These formulas are presented on the next page:

$$MAE = mean(|e_t|) \tag{34}$$

$$RMSE = \sqrt{mean(e_t^2)} \tag{35}$$

Percentage error measurements are unit-free and are calculated by the following formula:

$$p_t = 100e_t/y_t \tag{36}$$

The Mean Absolute Percentage error (MAPE) is similar to the MAE. The formula of the MAPE is presented below:

$$MAPE = mean(|p_t|) \tag{37}$$

The disadvantage of the MAPE is that it results in errors when y_t is zero or close to zero and it puts more weight to negative errors than it does on positive errors (Kim & Kim, 2016; Hyndman & Koehler, 2006). There are different alternatives that resolve these disadvantages, such as sMAPE. The Median Absolute Percentage Error (MdAPE) is recommended by Armstrong (1992) when having many different time series.

Scaled error is another possibility to measure the accuracy of different forecasts. It is ideal for measurement of different units or different scales. The method uses a naïve forecast, a training MAE, to calculate the accuracy. For seasonal data, the following formula for the Mean Absolute Scaled Error is used:

$$MASE = mean(\frac{e_t}{\frac{1}{T-m} * \sum_{t=m+1}^{T}|y_t - y_{t-m}|}) \tag{38}$$

The advantages of the MASE are that it can handle negative values or values close to zero and is scale independent. However, the method is also quite complex and can be hard to understand by others, so other methods might be preferred for simplicity.

3.3 Forecasting in the transport sector

The parcel industry has changed a lot during the last 15 years. The networks have expanded a lot nationally and the European networks improved regarding speed and costs. In addition, new services such as the express service have been added (Heitz & Beziat, 2016). The parcel industry is strategically orientated and very dynamic (Ducret, 2014). It is important to be able to cope with the changing environment. One of the reasons for the changing environment is the growth of e-commerce and globalization, which also results in the increase in air and freight transport (Hesse, 2002; Roy, 2001).

There is only little literature about forecasting in the transport sector. Most literature is about routing of trucks or setting up the Hub-and-Spoke Network. An interesting article is found about forecasting logistic demand for a LTL carrier by using Neural Networks of Zhou, Heimann and Clausen (2006). In the article, it is concluded that a NN with pre-data processing outperformed ARIMA and single-exponential smoothing (Zhou, Heimann, & Clausen, 2006). In another article of Zhou (2017), the combination between ARIMA, Neural Networks and single- exponential smoothing forecasting for a LTL carrier are researched. The combination of the three forecasting techniques results in the highest forecasting precision (Zhou, 2017). The two articles of Zhou (2006; 2017) are based on a simulation with sample data of a trucking company, however only little is specified about the data that is used. The forecast is probably made for the whole demand of the company and not related to certain links or regions. In addition, the forecast output is per month and not on a daily level. In this research, the forecast input and output will be different, so it is hard to say if the same conclusions hold for our research but these methods are interesting methods to investigate.

3.4 Selecting forecast models

We decided that we want to select one simple model, four more sophisticated models and create different combinations. The simple model is chosen as a benchmark, to study the added benefit of more sophisticated models.

Based on the literature and current situation of TNT, five models are chosen. We chose the Holt-Winters*, Holt-Winters double seasonal, TBATS, Neural Networks and Simple Exponential smoothing. The reasoning and development of the four models is described in Section 4.2.

The first sophisticated chosen forecast model is the Holt-Winters method. The Holt-Winters method is well known and different authors recommend using this method. An advantage of this method is that it can be easily adapted to different patterns. There are different formulas for many different cases regarding trend (damped or exponential) and seasonality (additive or multiplicative). The Holt-Winters method will be used but will have some added procedures to satisfy the needs of TNT. Since it is not only the Holt-Winters method, the asterisk is added to the Holt-Winters method, to indicate when we are discussing the combination of Holt-Winters and some added procedures. The Holt-Winters* method was also used in the hub forecast at TNT. This results in an easier knowledge transfer within the company; moreover TNT experienced satisfying results with this forecast. As explained before, the Holt-Winters method should not have too many degrees of freedom. To tackle this problem, TNT decided to first aggregate the data on a weekly basis. With this aggregated data a forecast is made with Holt-Winters. Finally, the weekly forecast is translated back to a daily forecast.

The second sophisticated model will be the Holt-Winters double seasonal model. As we explained before, we altered the formulas of Taylor (2003) to be able to model the seasonality within a week and over a year.

The third sophisticated model will be the TBATS model. This model is advised by Hyndman (2018) when using a weekly or daily pattern, but the computation time should be tested, to see whether it lies within an acceptable range. The biggest advantage is that it can cope with multiple non-nested seasonal periods and the seasonal periods can be non-integer.

The fourth sophisticated forecast model will be the NN model, since Zhou (2006) proved in his case that this outperformed the ARIMA and single exponential smoothing model. As explained, there are also some disadvantages of NN; Zhang (2003) identified that NN did not capture trend and seasonality well immediately, but other authors reject this statement. Since there are many different opinions about NN, it is interesting to incorporate a neural network in this research. In addition, there is much data available so the data set can be suitable for a NN. The other prediction methods will make one prediction per link based on the data of that link. The NN will make one prediction in total based on the data of all the links. This will increase the data size and the NN can learn from geographic (for example country) or time characteristics of all links.

Finally, the simple forecast model is the SES model. This model is easy to implement and can provide a good benchmark for the other more sophisticated models. As explained, the SES model is a simple variant of the Holt-Winters model. It will produce a forecast with a straight line.

For accuracy measurement, the MAPE performance indicator is chosen. Since the actual values are never negative, not close to zero and the MAPE is easy to understand, the MAPE is a good method to provide the accuracy of the different models (Hyndman & Koehler, 2006).

3.5 Conclusion

Chapter 3 answered different research questions about literature and below the main finding can be found:

- There is a lot of literature about different forecasting techniques.
    - The performance of a model depends on the data and the different patterns within it. It is preferred to select the most promising methods, test them and compare the accuracy to find out what the best forecasting technique is. Also a combination of different forecasting techniques often results in a higher accuracy. To compare the accuracy of the forecasts, different methods from literature are identified. Each of these methods has their own advantages and disadvantages and based on the characteristics of the data and forecast a suitable measure should be chosen.
- There is little literature available about forecasting in the transport sector.
    - There is much literature about the transport sector, but the combination of forecasting and transport, resulted in little results. Two studies that were useful were found, and the methods used in this research can offer a good indication about which method should be chosen. However, the type of input and output data that was used in this research is probably different, so results can be different for this research.
- The Holt-Winters*, Holt-Winters double seasonal, TBATS, Neural Networks and Simple Exponential smoothing are the most suitable models for TNT.
    - According to literature, these models are most promising and are also based on different theories and assumptions, which can result in higher accuracy when combining the different forecast methods. The Simple Exponential smoothing model will be mainly developed as a benchmark for the other more sophisticated models and to combine with the more sophisticated models. As accuracy measure, MAPE is chosen, since this provides a good indication about the performance of the forecast. The grouping of lanes can help to improve the forecasts and to better estimate seasonal factors. We chose to not include this is the current research, but this can be interesting for further research.

4 Model development and results

In this chapter, answers are provided to the fourth and fifth research question. Section 4.1 focuses on the cleaning of the data and Section 4.2 explains how the different forecasting techniques are developed. Finally, Section 4.3 compares the results of the different models.

4.1 Data cleaning

Public holidays have quite some influence on the volumes that travel through the network. Customers of TNT can be closed on a public holiday, resulting in less volume on that day and maybe a peak before the public holiday. In addition, some countries have driving restrictions during a public holiday, which means that trucks cannot drive in certain countries. Public holidays would normally not be classified as an outlier, since these can also be forecasted. But for now, we decided to classify public holidays as an outlier. By classifying public holidays as outliers, the results are not influenced by the volume of public holidays. For now, the focus is to first forecast normal days that are not a public day, but in the future also public holidays should be implemented to forecast. When analyzing the data, most outliers are related to public holidays.

A file that contains all important public holidays per hub is generated. Two days before the public holiday, the public holiday itself and two days after the public holiday are marked. All the marked dates are replaced by the average value of the last twelve weeks of the weekday that is replaced. We chose twelve weeks because this was also chosen in the hub-forecast. However, this can be an interesting topic for further research to see if there are better options. The altered values for public holidays resulted in a smoothened pattern and most outliers are excluded by this.

Anomalies can also have an impact on the volume through the network. Examples of anomalies are strikes, technical failures at a location or a natural disaster. Anomalies can normally not be predicted upfront. An anomaly that had much influence on the volume and data quality is the cyber-attack of 2017. From the historical data it can be seen that the pattern is completely off for many weeks during the cyber-attack. Ten weeks around the cyber-attack, the data is replaced by the average value for that weekday and month and multiplied with a trend factor. Other anomalies that influenced the volumes for a shorter period are replaced in the same way as the public holidays are replaced, but since this cyber-attack influenced such a long period this solution was not possible.

4.2 Development of forecasting techniques

This section describes the different forecast models that were chosen in Chapter 3, as well as the combination of the models. For each forecasting technique, the development and results of the technique are discussed.

4.2.1 Holt-Winters*

This section describes how the Holt-Winters model is adapted for the problem of TNT. The Holt-Winters* method as we developed it, is somewhat different from the other methods that are used. As said before, a statistical method performs worse with more degrees of freedom, therefore the daily volumes are first aggregated to a weekly volume. The other methods generate a forecast per day, this Holt-Winters* method makes a forecast per week and later translates it back to a daily forecast. The seasonal frequency for the Holt-Winters model is chosen to be 52, one for every week. This means that there is a parameter estimated for every week in the year and for example the first week of every year is compared to the first week of the other years. Section 4.2.2 describes a similar method, the Holt-Winters double Seasonal method and tests different settings for the seasonal frequency. The best performing setting is 52 weeks and because of that, we decided to keep the 52 weeks for the Holt-Winters* method.

The daily data is aggregated per week and afterwards split in a training- and test set. The training set consists of all weeks until the week before the week of the start-date of the forecast. The test set consists of all the weeks from the week of the start-date of the forecast.

With this weekly training data, the Holt-Winters model is constructed. Like all other models, this is done in R. The input data contains the aggregated volume per week per link. The model calculates the initialization values for the level, trend and seasonality by a moving average based on decomposition (Hyndman R. , Koehler, Ord, & Snyder, 2008). The four steps of the initialization are:

1. Calculate the trend, by using centered moving average of the first two years.
2. Remove the calculated trend from the data
3. Calculate the seasonal effects per time unit with centered moving averages and remove the seasonal component from the data.
4. Fit a linear trend to the seasonally adjusted data to find the initialization values for the level and trend.

After the initialization, the model calculates the level, trend and seasonality per week, for different parameters α, β and γ of formulas (7), (8) and (9). The model finds the best parameters for that link, by minimizing the squared prediction error for a one-step forecast. For every week, a forecast is made, based on the level and trend of last week and the seasonality of 52 weeks back, as in formula (6). Afterwards, this forecast is compared with the actual values to calculate the prediction error.

After the model is fitted to historical data, a forecast is made for the test set of the data, with the found parameters of the fitted model. In total 100 forecasts of 13 weeks ahead are made. From the 1th of November 2017 until the 8th of February 2018 (100 days), every day is the start-date of a forecast with a forecast horizon of 13 weeks. Every forecast provides weekly-forecasted values, resulting in 13 output values per forecast. Based on the level and trend component of the last week of the training data and the seasonality of 52 weeks back, a forecast for 13 weeks ahead is made. The smoothing parameters are again optimized for every forecast. This is needed, because the training set changes when the start-date changes.

After a forecast is made for the test set, the weekly-forecasted values need to be transformed in to daily data. First, the average distribution of volume between the different weekdays in the training set is calculated per month. Afterwards this volume is multiplied by the volume that was forecasted. This kind of aggregation on a weekly basis and then disaggregated back to daily level is called temporal aggregation (Dekker, Van Donselaar, & Ouwehand, 2004).

For each forecast a graph is made in which the forecasted values and actual values are shown. Figure 10 shows an example for the link Arnhem-Madrid from 2018-02-02 until 2018-05-03. Most links only have five days per week volume and two days with zero volume. In the analysis, the weekdays that never had volume are excluded, which results in 65 forecasted values in the 13 week horizon. As we can see, the pattern of the forecast is quite similar to the actual values. The predicted values however, are more around the average and the actual points sometimes reveal more extreme values.

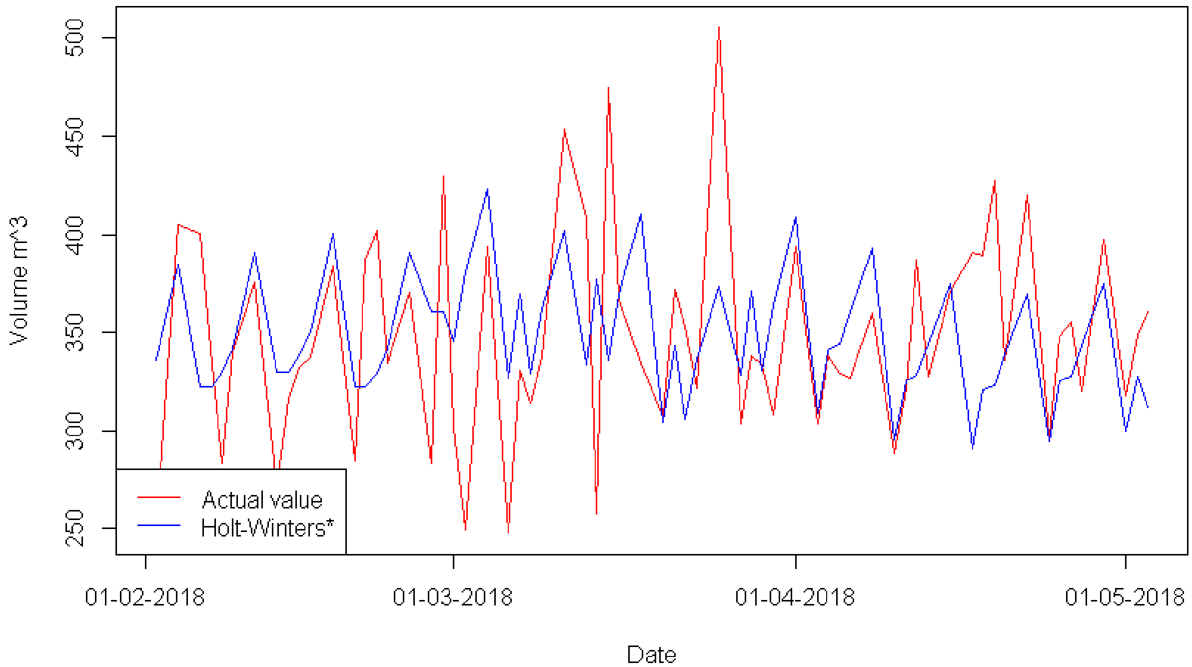

Figure 10: Actual and Holt-Winters* forecasted values for the link Arnhem-Madrid

Per forecast, the MAPE is calculated for every day of the 65 days in the 13 week horizon, using formula (37). Afterwards the average is calculated over these 65 MAPE values for every forecast, resulting in one MAPE value for each of the 100 forecasts. The average results for the 100 MAPE values of the five different links are presented in Table 3. The MAPE in Table 3 is the average MAPE of the 100 forecasting periods. It can be seen that the MAPE varies between 8.2 and 15.1 over the different links. The Holt-Winters* method has an average MAPE of 12.5 over all links and forecasts, which means that the forecast is on average 12.5% off from the actual values. The minimum and maximum MAPE are calculated by taking the minimum and maximum of the 100 MAPE values.

Table 3: Results Holt-Winters*

Link type	MAPE	MAPEMin	MAPEMax	MAPEStd.	MPE
Big	13.9	10.7	17.6	12.7	-7.0
Big	8.2	7.1	9.3	6.1	2.0
Small	15.1	10.6	17.8	13.0	3.7
Medium	13.6	10.0	19.6	10.3	-2.0
Medium	11.8	8.7	14.6	10.4	-2.7

To see if the errors are normally distributed, a histogram is created, which is shown in Figure 11. The bars on the x-axis compare the forecasted value with the actual value, where the actual value is increased or decreased with a certain percentage. If the actual value was for example 100 and the forecasted value 97, then this value belongs to the bin: "Forecast within (actual) and (actual-5%)". The bars left of the blue line cumulate for 51% of the forecast errors and the bars right of the blue line cumulate for 49% of the forecast errors. This indicates that the bias of Holt-Winters* is very small and there is almost no over- or under-forecasting. In addition, the distribution of errors seems to behave like a normal distribution.

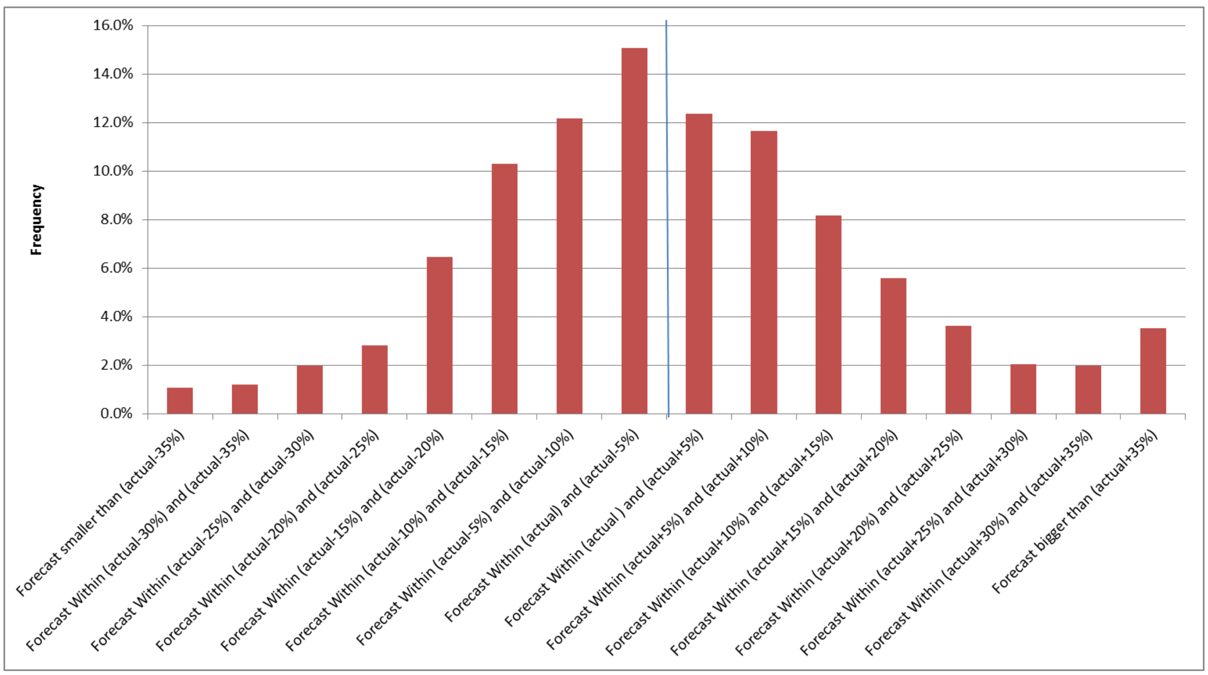

Figure 11: Distribution of forecast errors Holt-Winters*

4.2.2 Holt-Winters double Seasonal

The Holt-Winters double seasonal model, which is discussed in chapter 3.1.2, can handle two seasonalities with a different aggregation level. The two seasonalities that we want to capture are the seasonality over the year and the seasonality within a week. The seasonality within a week has a frequency of 7 (or less when some days never have volume). This seasonality, level and trend are calculated per day with formulas (16), (18) and (20). For the seasonality over the year, the data should be aggregated to reduce the degrees of freedom. We tested three different aggregation options, namely 52, 13 or 4 periods per year. The results will be discussed later; first the method is discussed, which is the same for the three different aggregation options. The seasonality, level and trend for the longer periods are calculated with formulas (17), (19) and (21).

The input data for the forecasting model contains two tables, one for each aggregation level. One table contains the volume per day and the other table contains the volume for the larger period. For the three options regarding different aggregation periods, this means the volume is aggregated per week, four weeks or 13 weeks (52, 13, 4 periods). The training set contains all the dates until the start date of the forecast. The test set contains all the dates from the start date onwards. The model is constructed only using the training set. The calculation of the initialization of the level, trends and seasonality is done according to one of the initialization methods of Hyndman (2010). This is a simple method that offers a quick initialization. The initialization value for the levels is calculated by averaging the values of the first year. The initialization value for the trends is calculated by determining the slope for each period between the first two cycles. Afterwards, the average is calculated over these slopes. For the daily trend, the slope between the first and second week per weekday is calculated. For the longer trend, the slope is calculated between the first and second year. The initialization values for the seasonalities are calculated by dividing the actual value of that period by the initialized level.

After calculating the initialization values, the model calculates the level, trend and seasonality per day and for the longer period, for the parameters α, ε, β, ϕ, γ and ω of formulas (16) until (21). These calculations are only done for the training set. The seasonal factors for the daily period and the longer period are normalized after each time they are updated. The normalization is done to make sure the seasonal factors add up to m_1 and m_2. The model finds the best smoothing parameters for that link, by minimizing the squared prediction error for a one-day ahead forecast in the training set. For every day, a forecast is constructed for one day ahead. This forecast is based on the level and trend of the last day and the seasonality of m_1 and m_2 periods back as in formula (15). Afterwards, this forecast is compared with the actual values to calculate the prediction error.

Again, 100 forecasts of 13 weeks ahead are made. From the 1th of November 2017 until the 8th of February 2018, a forecast is made daily with a horizon of 13 weeks. Each of the 100 forecast has a horizon of 13 weeks and provides a prediction for every weekday, resulting in 65(13 weeks * 5 weekdays) forecasted values per forecast. Because the runtime for the optimization for the smoothing parameters is very long, the number of iterations is restricted. The values for the smoothing parameters are calculated once per link for all the 100 forecasts. When the smoothing parameters would have been updated every forecast, the results can be slightly different. We think that this influence will not be substantial however, because the prediction error is calculated for the whole training set. The whole training set consists out of +- 1000 values and the increase or decrease with 50 data values will not make a big difference.

These 100 forecasts are executed three times, for the different aggregation levels. The results for 52, 13 or 4 periods per year are comparable and shown in Appendix L. The results of the aggregation of 52 periods was slightly better than the other models, so we chose 52 weeks as an aggregation method. 52 weeks results in more degrees of freedom than 13 or 4 periods per year. More degrees of freedom could lower the robustness of a forecast. When we compare 52 periods with 13- or 4 periods we see indeed that the Max MAPE is a bit higher. However, this is only a small difference and we decided that the average MAPE was of more importance. The results for 52 periods are presented in the remainder of this chapter.

Figure 12 shows an example of such a forecast with the actual values. The pattern of the forecast looks similar to the pattern of the actual values. The predicted values however, are more around the average and the actual points sometimes reveal more extreme values.

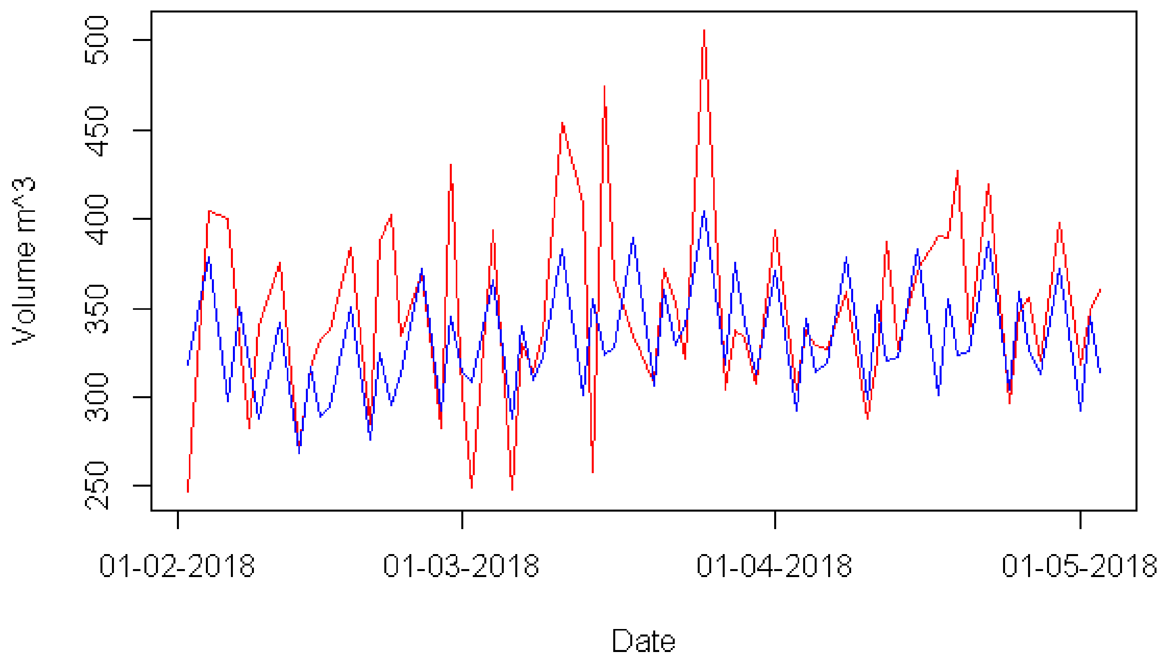

Figure 12: Actual and Holt-Winters double seasonal forecasted values for the link Arnhem-Madrid

The results of the five different links for 100 forecasts are presented in Table 4. The MAPE values are calculated in the same way as in Section 4.2.1. It can be seen that the average MAPE varies between 11.7 and 20.3. The Holt-Winters double seasonal method has an average MAPE of 15.1, which means that the forecast is on average 15.1% off. The maximum MAPE of the five links is 46.1, which is a big difference with the average of 15.1.

Table 4: Results Holt-Winters double seasonal

Link type	MAPE	MAPEMin	MAPEMax	MAPEStd.	MPE
Big	13.6	9.6	27.2	11.0	3.7
Big	11.7	7.8	21.4	7.9	4.0
Small	13.7	9.4	18.2	11.7	7.1
Medium	16.4	10.3	31.9	11.6	-3.0
Medium	20.3	9.8	46.1	13.9	-7.6

Figure 13 shows the distributions of the errors of the 100 forecasts. The last bar represents the forecast values that are bigger than the actual values plus 35%. This is a large portion of the total forecasts, namely 5.1%. The bars left of the blue line cumulate for 57% of the forecast errors and the bars right of the blue line cumulate for 43% of the forecast errors. This indicates the Holt-Winters double seasonal method has under-forecasting and the forecasted values are more often smaller than the actual values than larger. Under-forecasting is preferred over over-forecasting by TNT, because of the philosophy that it is better to plan too little trucks than too many trucks.

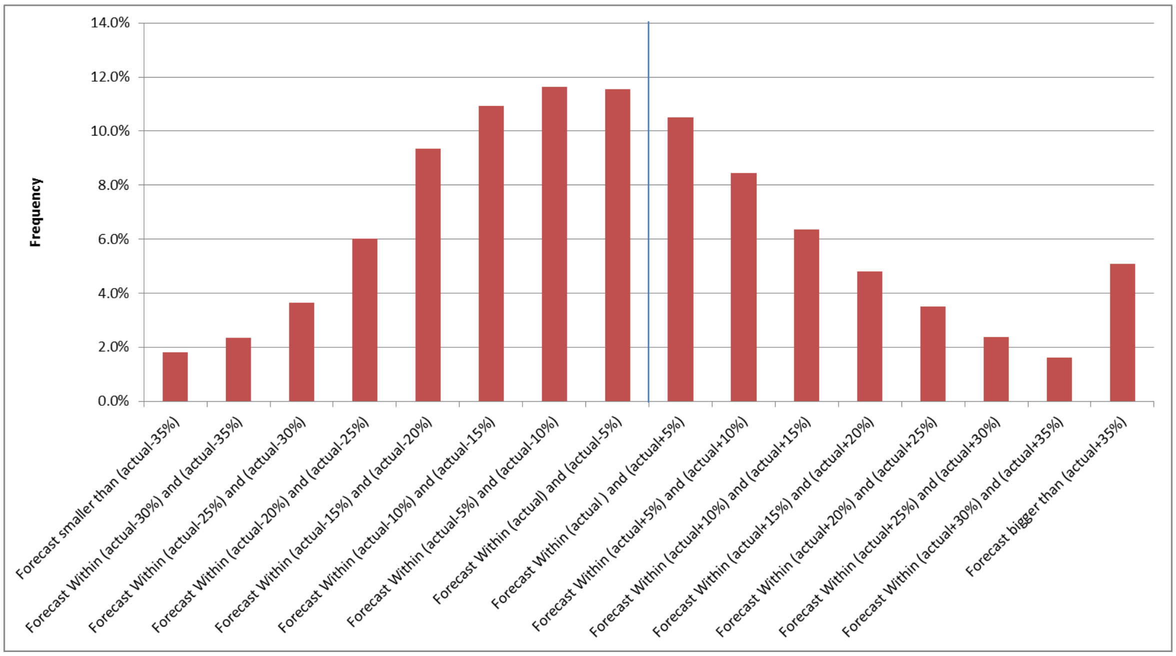

Figure 13: Distribution of forecast errors Holt-Winters double seasonal

4.2.3 TBATS

The TBATS model can handle multiple seasonalities and since the data reveals a yearly seasonal pattern and a weekly pattern, the seasonal periods that are chosen are 365.25 and 7. Because of an extra day every four years during a leap year, 365.25 is chosen instead of 365.

The TBATS model in R is used to obtain a forecast. In this model, it can be specified whether a Box-Cox transformation should be used, if a trend or damped trend should be applied and what the seasonal periods are. For now, these settings are determined by the model by testing the different settings and comparing the squared prediction error for a one-day ahead forecast in the training set. The input data consists of daily volume per link. The input data is split in training- and test data and the model is fitted to the training data.

After the model is constructed and the parameters are found, a forecast is constructed. The same number of forecasts is made as in Section 4.2.2. For each selected link, 100 forecasts are made with a forecasting horizon of 13 weeks ahead. Every forecast consists out of 65 forecasted values (13 weeks * 5 weekdays). For every forecast, the smoothing parameters are again optimized, because the training set is also different per forecast.

Figure 14 shows an example of a forecast from TBATS. As can be seen, the predicted values have a quite similar pattern as the actual values, but similar to the Holt-Winters Model, the more extreme points are poorly predicted.

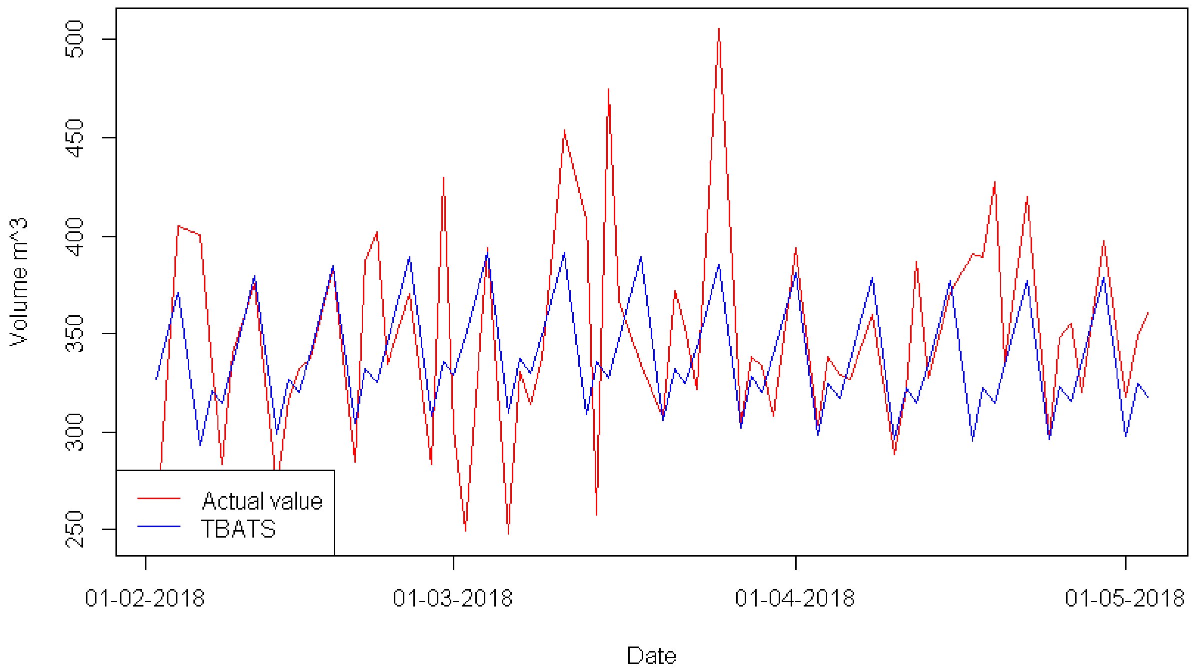

Figure 14: Actual and TBATS forecasted values for the link Arnhem-Madrid

For all days in every forecast the MAPE is calculated. Each forecast consists out of 65 MAPE values, and with these 65 values, the average MAPE for that forecast is calculated. This results in 100 MAPE values (one for each forecast) and of these 100 values, the average MAPE is calculated for that link.

The results are shown in Table 5. The maximum MAPE of the five links is 43.5, which is a big difference with the average of 13.2. In addition, there are three out of five links that have a maximum MAPE that is bigger than 37. A MAPE of 43.5 means that at least one forecast for that link was on average 43.5% off.

Table 5: Results TBATS

Link type	MAPE	MAPEMin	MAPEMax	MAPEStd.	MPE
Big	12.6	9.1	39.6	12.4	-1.7
Big	10.1	7.1	15.1	6.9	5.2
Small	13.4	9.2	16.8	11.6	6.7
Medium	15.1	9.5	37.5	11.3	5.8
Medium	15.0	8.5	43.5	11.5	0.3

Figure 15 shows the distributions of the errors of the 100 forecasts. The bars left of the blue line cumulate for 61% of the forecast errors and the bars right of the blue line cumulate for 39% of the forecast errors. This indicates the Holt-Winters double seasonal method has under-forecasting and the forecasted values are more often smaller than the actual values than larger. As said before, under-forecasting is preferred over over-forecasting by TNT.

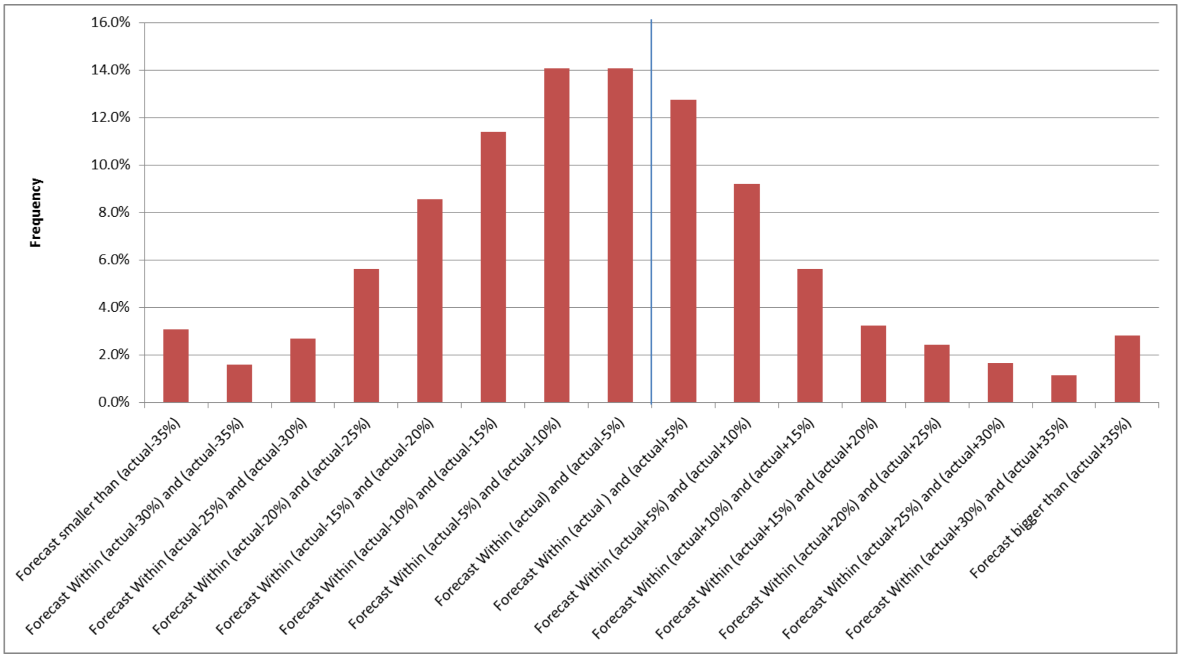

Figure 15: Distribution of forecast errors

4.2.4 Neural Network

To have a good result with NN, it is important to have as much data as possible. For this reason all the five links are used as input data and the NN will not be constructed per link but for all links. The advantage is that the NN can collect valuable information of other links in comparison with a separate NN for each link.

To construct the NN, the data first needs some processing. First, columns that contain information that is not numeric should be transformed to numeric values. For example, the two columns that contain the origin and destination location of a link are transformed in such a way that per origin and destination location a column with the location name is added. This is called a dummy variable and this column name is for example: ”OriginX” and contains a one if “X” is indeed the origin and otherwise a zero. This means that many extra columns are generated for all the different origin and destination locations. The creation of dummy variables can also be applied to columns that have numeric values, such as the weekdays (1 until 7) or months (1 until 12). These columns can also be seen as categorical data and this is tested later in this section.

The second step is to scale the numeric values between -1 and 1, which results in a better training outcome. This is done by multiplying the difference between a value and the mean by their standard deviation.

The third step is to determine the number of hidden layers and the number of nodes per layer. In general there are no strict rules to do so. However, there are some rules of thumb that can help in determining the number of hidden units. Hidden units are defined as the total number of hidden nodes over all hidden layers. Heaton (2008) formulated the following three most common rules of thumb:

- The number of hidden units should be between the size of the input layer and the size of the output layer.
- The number of hidden units should be 2/3 of the size of the input layer, plus the size of the output layer.
- The number of hidden units should be less than twice the size of the input layer.

These rules of thumb can help, but in the end the number of hidden units should be tested by trial and error. Later in this section, it is decided which columns should be in the input layer. Since the NN will be constructed for all links together, this process of finding the right amount of hidden units only has to be done once. It could be beneficial to reevaluate the number of hidden units after a certain time period. In general, NN should contain enough hidden units for a sufficient representation of the model. However, NN should not contain too much hidden units such that generalizations cannot be made. With too much hidden layers and nodes, there is a risk of overfitting (Swingler, 1996). Curry & Morgan (2006) state that sometimes one hidden layer can be enough to obtain good results, but often additional layers are required that will result in a better fit and a better generalization.

Besides the number of hidden layers and nodes, there are more settings that can be tested. Table 6 provides an overview of the different test settings. After each repetition, the NN gives the error, which is calculated by the sum of errors. In general, a lower error means a better fit, however this does not mean that the forecast will be better. A model with a low error could be overfitted.

Table 6: Test setting Neural Network

Test Setting	Explanation	Example
Data columns	Which columns of the data set are used as input parameters.	Origin Country, Week, Year, WeekDay
Dummy variables	Which columns of the input data should be transformed to dummy variables	Origin Country, WeekDay,
Hidden units	The number of hidden layers and nodes on these hidden layers.	(10,2,4) which means 3 hidden layers with respectively 10, 2 and 4 hidden nodes on them.
Threshold	The stopping criteria for the NN. A too high threshold will result in less accuracy and a too low threshold can result in the model being unable to achieve this threshold and may result in an error.	0.1
Repetition	The number of repetitions of the training of the NN. This can be useful to avoid a local minima.	3

The different test settings are related to each other. For example, when we choose more data columns, we probably need more hidden units. Therefore the test settings cannot be chosen individually. Since there are many settings that the user can define, and the running time for NN can be long, it was not possible to test all different combinations. For that reason, we sequentially tested different settings. The disadvantage of this is that different columns together influence the final result. It is not desirable to test one user setting when keeping all the other settings the same. We think the number of hidden units mainly influence the results when changing other settings. When the number of input columns increases, more hidden units probably give better results.

First we performed a preliminary analysis to see whether the exclusion of the column that indicates the year influenced the results. For different numbers of hidden units, the results showed that the MAPE was always higher when the column was included. This was also what we expected, since with the year, the model can apply a certain trend to the volumes. The results can be found in Appendix M.

After this, we examined whether the columns related to the date (weekday, week, month) should be transformed to a dummy variable or not. For example, when the weekday column is not transformed to a dummy variable, the model interpreter a Tuesday and Wednesday as very similar but Monday and Thursday are less similar. This is not necessarily the case, so it is interesting to test this. If it is the case that days closer to each other are more similar, there are also other transformation options possible to improve the planning. An example is to transform the values, with sine and cosine transformations to preserve the possible cyclical behavior. When the weekdays are from one until seven, one and seven are not as close to each other as one and two. In this research, we only tested the dummy variable transformation.

We considered three columns that can be transformed. This would result in 8 experiments, when keeping the other settings the same. But, as said, the number of hidden units would be a big influence and therefore we want to include four different settings for the hidden units. This would increase the number of experiments to 32. Since this would take too long, we decided to choose the four most promising dummy variable options, each with four different numbers of hidden units. Finally, this results in 16 experiments. We saw that the results for adding a dummy variable for the weekday resulted in similar results compared to not adding this dummy variable. Adding the weekday and month as a dummy variable resulted in worse results, and adding the weekday, week and month as a dummy variable in the worst results. We also expected that the weekday would be the most suitable column to transform to a dummy variable because a weekday is less related to days around it in this research. The volumes per month and week are more related to months or weeks close by.

Since the results with and without transformation of the weekday column to a dummy variable were close to each other, we decided to generate more forecasts for both options and to improve the settings for the number of hidden units further. For both options regarding the transformation of the weekday column, we tested multiple experiments with different numbers of hidden units for multiple forecasts. The results for the two best options are shown in Table 7. The other results can be found in Appendix N.

Table 7: Results experiments NN Dummy variable

HiddenUnits	DummyVariables	MAPE	MAPEMin	MAPEMax	MAPEStd	MPE	RunTime (min)
5,5	OriginLinkLocation,DestinationLinkLocation	11.1	8.6	13.9	9.1	2.7	74.0
6,6	OriginLinkLocation,DestinationLinkLocation,weekday	11.7	9.2	15.5	9.7	-1.5	35.0

The results show that the option without transforming the weekday to a dummy variable is 0.6 percentage point better regarding the MAPE. Also the other performance measures are better without transforming the weekday.

Based on these results we decided not to transform any other column than the origin- and destination-link. In addition, we know what the most promising number of hidden units is. With these findings, we tested different values for the threshold.

We could not find a direct link between the threshold and the MAPE. We expected that a lower threshold would result in a better MAPE, but this was not the case. As expected, the runtime increased by lowering the threshold. We chose a threshold of 0.2, which was the highest threshold we tested.

The number of hidden units with the best results overall is one layer with 10 hidden nodes. The model with two layers of 5 nodes each, performed slightly better regarding the MAPE (0.05 percentage point), but had a longer runtime.

All the experiments together result in the following final settings:

- Data Columns: Week, Year, Month, Weekday, Origin Link Location, Destination Link Location
- Dummy Transformation: Origin Link Location, Destination Link Location
- Hidden Units: (10), so one hidden layer with ten hidden nodes
- Threshold: 0.2
- Repetition: 5

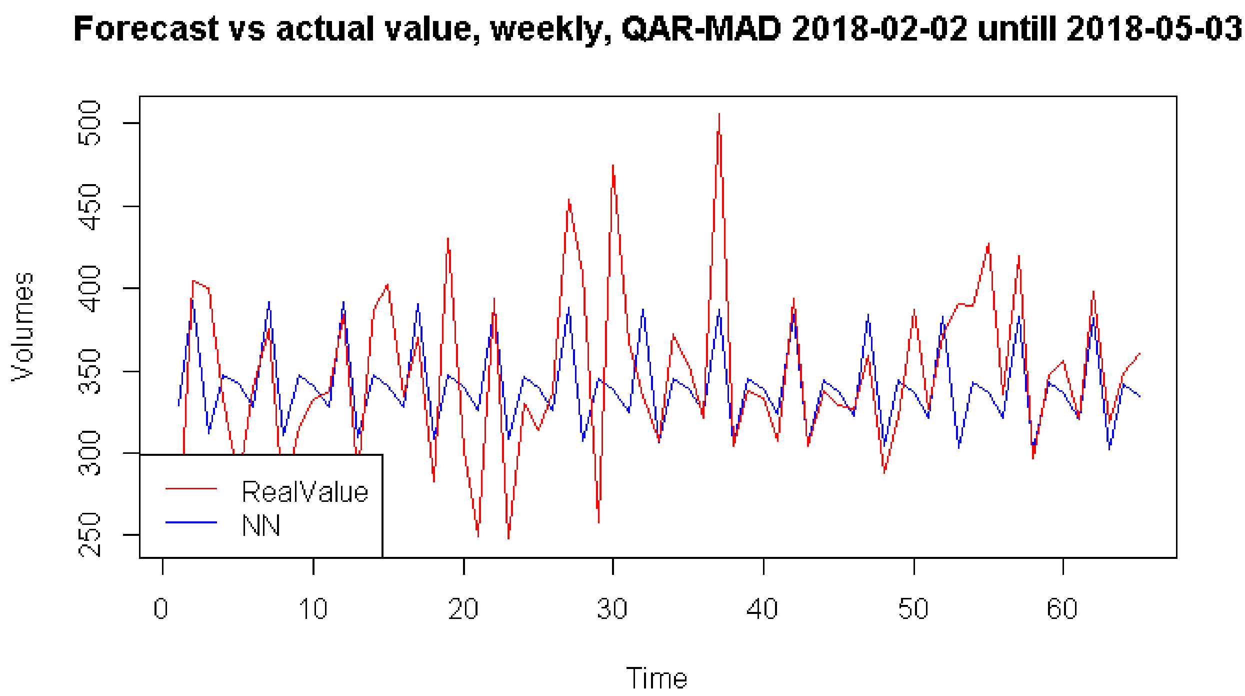

Figure 16: Forecast NN vs. actual value for QAR-MAD

With the test settings that are given before, ten forecasts are made for different dates. These ten dates are in the same period as the 100 forecasts in the previous chapters, but due to the long running time, we reduced the 100 forecasts to ten. Again, each forecast has a forecast horizon of 13 weeks and for every weekday a prediction is made, which results in 65 forecasted values per forecast.

The results are shown in Table 8. The MAPE values are calculated in the same way as in Section 4.2.1. The average MAPE varies between 10.0 and 16.0. The average MAPE over the five links is 13.1. The maximum MAPE of the five links is 23.6, which is close to the average MAPE of 13.1.

Table 8: Results Neural Network

Link type	MAPE	MAPEMin	MAPEMax	MAPEStd.	MPE
Big	12.6	10.1	14.9	12.2	-6.4
Big	10.0	8.1	12.9	6.7	6.7
Small	16.0	10.2	23.6	14.4	4.3
Medium	14.0	10.6	17.1	11.0	0.1
Medium	12.8	11.1	14.9	9.2	1.9

Figure 17 shows the distributions of the errors of the 100 forecasts. The bars left of the blue line cumulate for 56% of the forecast errors and the bars right of the blue line cumulate for 44% of the forecast errors. This indicates the NN has a bit under-forecasting and the forecasted values are more often smaller than the actual values than larger. As said before, under-forecasting is preferred over over-forecasting by TNT.

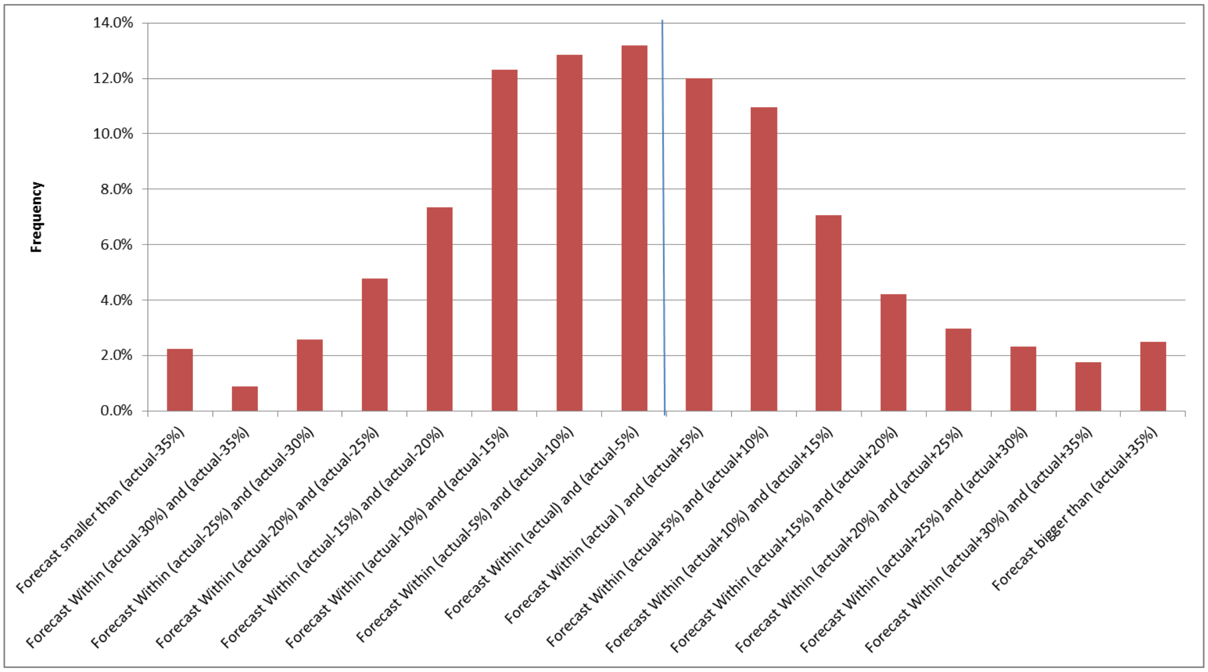

Figure 17: Distribution of forecast errors Neural Network

4.2.5 Simple Exponential Smoothing

The SES model is mainly used to benchmark the other models and to study the added benefit of more sophisticated models. For the SES, only the α should be specified upfront. The data is split in the training- and test set. With the training set the model is constructed. The α is found by minimizing the squared prediction error for the one-step ahead forecast.

When α is known, a forecast can be constructed. Since this model only takes the level into account, a forecast for 65 days ahead will result in 65 times the same value, and visually a straight line. We did not update it every day to get a more reliable forecast because this would not be a fair comparison with the other models. Again, 100 forecasts are made for all selected links with a forecast horizon of 13 weeks ahead. From the 1th of November 2017 until the 8th of February 2018 (100 days), every day is the start-date of a forecast with a forecast horizon of 13 weeks. For every forecast, the α is again optimized. Figure 18 shows the actual and forecasted values for one specific link. We can indeed see that it is just a straight line.

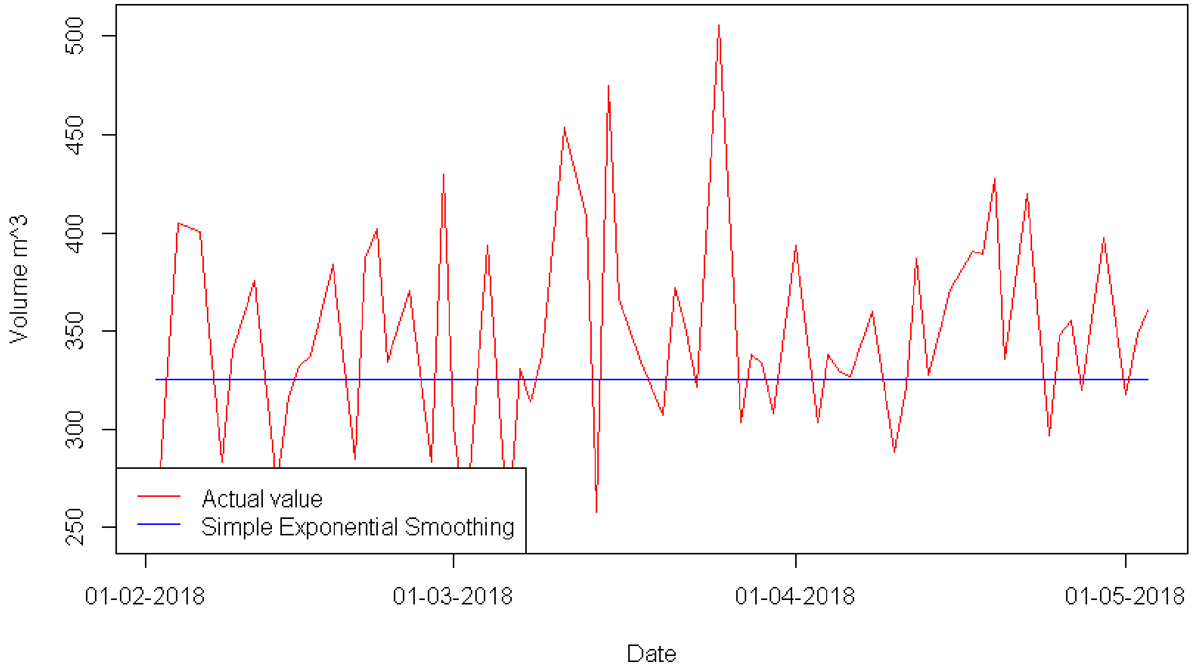

Figure 18: Actual and Simple Exponential Smoothing forecasted values for the link Arnhem-Madrid

4.3 Comparison of the results of different models

In this section the results of the different models are compared. Section 4.3.1 compares the performance of the individual forecast methods. Section 4.3.2 explains the results of the combination of forecasting models and 4.3.3 focusses on the runtime of the different models. Finally, Section 4.3.4 discusses the ease of implementation of the different forecasting models.

4.3.1 Performance

The results of the different forecasting models are shown in Table 9. It can be seen that the Holt-Winters* model performs best when looking at the MAPE. In addition, the standard deviation of the MAPE, the minimum MAPE and the maximum MAPE are the lowest of all models. We see that the MAPE values of the different models are close to each other and the difference between the best and worst performing model is only 2.9 percentage point. In addition, the standard deviations of the different models are close to each other. The maximum MAPE of the different models are quite different. With a difference of 26.5 percentage points between the best and worst performing model. Especially the Holt-Winters* and Neural Network models are performing better than the other models regarding the maximum MAPE. This could indicate that the Holt-Winters* and Neural Network model are more robust than the other models and that there is less change to generate a forecast that is completely off.

Table 9: Results of different forecast methods

Model	MAPE	MAPEMin	MAPEMax	MAPEStd
Holt-Winters*	12.5	7.1	19.6	10.5
Neural Network	13.1	8.1	23.6	10.7
TBATS	13.2	7.1	43.5	10.7
Holt-Winters double seasonal	15.1	7.8	46.1	11.2
Simple Exponential Smoothing	15.4	10.3	31.9	12.4

Another important aspect is to see how stable the forecast is within the forecast horizon. In general, it can be expected that a forecast with a horizon of one week is better than a forecast with a horizon of 13 weeks. Also for the creation of a prediction interval, it is important to know how the forecast performs over the forecast horizon. The average MAPE per week over the forecast horizon is shown in Figure 19.

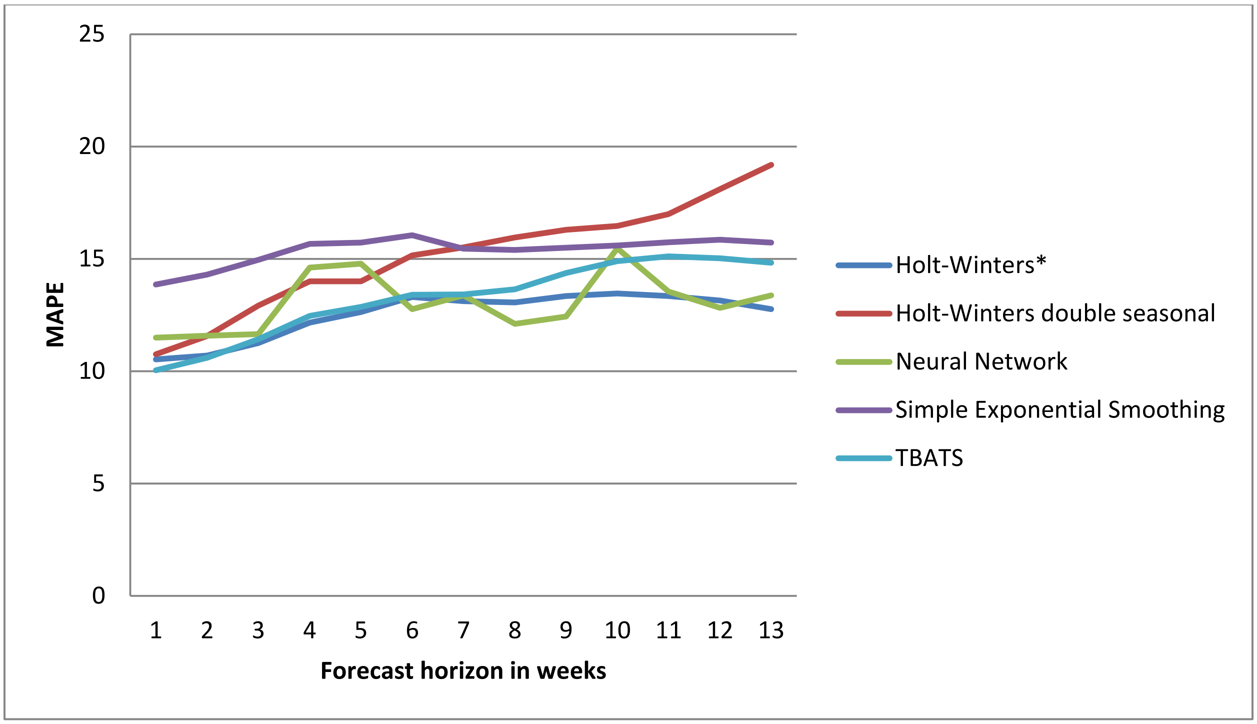

Figure 19: MAPE per week in forecast horizon

It can be seen that most models have an increase of the MAPE value during the first five or six weeks and after this the MAPE becomes stable or increases less than before. The Holt-Winters double seasonal method does not become stable and has a steady increase over 13 weeks. The expectation is that in practice, different users will use different weeks of the forecast. When the tactical planners would mainly use the first four weeks of the forecast for example, then the Holt-Winters double seasonal method is more preferable then the Simple Exponential smoothing method. The Holt-Winters double seasonal method has a lower MAPE the first four weeks and after this a higher MAPE then the SES method. For the other methods, it is not possible to draw comparable conclusions.

4.3.2 Combination of different forecasting techniques

The combination of different forecasting techniques is simply done by multiplying the forecasted values with a certain weight, which will result in a new forecast value. These weights can be distributed evenly or certain forecasting techniques can receive a higher weight. Based on historical performance, it is preferred to give better performing techniques a higher weight. Based on the results of Section 4.3.1, we decided to test different combinations of the Holt-Winters*, Neural Network and the TBATS model. In total 6 different combinations are tested. The distribution of weights for the first three combinations is based on the performance of the individual models. The last three combinations are in place to verify that we should give the best performing the individual methods the highest percentage. The exact percentage that is chosen is not based on something. When TNT wants to implement a combination of forecasts, further research should be conducted to find the best distribution of weights. We decided to make the same ten forecasts of Section 4.2.4 between the 1th of November and the 8th of February. Every forecast has a forecast horizon of 13 weeks.

The results of the six different combinations can be found in Table 10. All the six different combinations have a better MAPE than the best performing individual forecast (Holt-Winters*). The MAPE of the best performing combination is 0.9 percentage point lower than the MAPE of the Holt-Winters* model. The best performing combination is the combination of Holt-Winters*, Neural Network and TBATS with respectively the following weighs: 60%, 25% and 15%. Details of the different combinations can be found in Appendix M.

Table 10: Results combination of forecasts

Holt-Winters*	Neural Network	TBATS	MAPE	MAPEMin	MAPEMax	MAPEStd
60%	25%	15%	11.5	7.2	17.0	9.9
60%		40%	11.6	7.4	18.5	10.0
60%	40%		11.7	7.2	17.5	10.0
25%	60%	15%	11.8	7.5	19.1	10.0
40%		60%	11.9	7.2	19.3	10.1
	60%	40%	12.2	8.4	18.8	10.2

4.3.3 Runtime

Another important aspect when comparing the different models is the runtime. Table 11 shows the runtimes of the different models. It is hard to compare the numbers with each other one on one, because some settings of the models can influence the runtime. For example, when constructing the NN, the number of repetitions can be chosen by the user and a larger number of repetitions results in a higher runtime. Still, the runtime can give an indication about the feasibility in practice. The asterisk at the value of the runtime of the Neural Network and Holt-Winters double seasonal method indicates that this is an estimation and can be influenced by the settings of the user. To construct the Holt-Winters double seasonal method, it takes around 35 seconds per link. This means that the optimization of the different values for the smoothing parameters, also takes a long time. For this reason, we limited the optimization function to a certain number of iterations, to get results within acceptable time. The other models do not have this limit for the smoothing parameters. The largest part of the runtime of the NN is for the construction of the model. This model is constructed once for all the five links, but in practice more links will be added, so this can result in a in an increase or decrease of the runtime per link.

There are 369 links within the scope of this research. In the future, this could be increased to more links within Europe. In total there are 2513 links within Europe. The NN would run for 40 minutes for 369 links and more than eleven days for 2513 links. However, if a running time of eleven days is a problem, there are several solutions for this. An example of a solution is running the model on a faster computer or run the forecast parallel on multiple cores.

Table 11: Runtime different models

Model	Runtime (s)
Holt-Winters*	0.45
TBATS	36.19
Neural Network	390
Simple Exponential Smoothing	0.46
Holt-Winters double seasonal*	300

4.3.4 Ease of implementation

The Holt-Winters* method is easy to implement for TNT, mainly because they also use this method for the Hub-forecast. Some people within the Planning and Engineering department are therefore already familiar with this model and it will be easy for them to understand the linehaul forecast. In addition, some extra features are already developed for the Hub-forecasts that were not within the scope of this research. For example, the prediction of volume around public holidays is implemented in the Hub-forecast. Moreover the possibility to alter the trend if necessary is already developed. Another advantage is that the method does not require much tweaking of experts and can be run automatically. There are not a lot of settings that users have to define.

The Holt-Winters double seasonal method is comparable to the Holt-Winters* method and has the same philosophy. However, the approach is different, because the Holt-Winters double Seasonal forecast is made on a daily level and not on a weekly basis with aggregation to a daily basis. This means that it will take a bit more time for the people of P&E to understand. Again, the method does not require much tweaking by experts and can be run automatically. There are not a lot of settings that users have to define.

The TBATS model is difficult to understand, but it has the advantage that it is fully automated. This can also be a disadvantage however, since when someone wants to alter the model for a specific reason, this will be hard. For example, the prediction of public holidays cannot easily be implemented within the TBATS model. The prediction for the public holidays should be done separately. For the hub-forecast, this is currently done separately too, by multiplying the forecast with a certain public holiday factor.

The NN is together with the TBATS model the most difficult model to understand. The model itself is difficult, but it is also hard to assess the results of the model due to the “black-box” principle. The NN requires also some tweaking to get the best results, and for this reason, some knowledge about NN is required. An advantage is that it is easy to implement the prediction of public holidays in the NN and we expect that the NN can give good forecasts for this, since it can find relationships between different input-variables.

The combination of forecasts in general has the same advantages and disadvantages as the individual forecasts. However, there is a risk that when a change is needed, the change should be applied to multiple models and thus results in more work. An example could be calculation of public holidays, which can be done differently for the different models. In addition, the people of P&E should learn multiple forecasting methods instead of one.

4.4 Development of prediction interval

Section 2.4.4 discussed the need of a prediction interval for TNT. In this section, the prediction interval is constructed for the Holt-Winters* method, which is the best performing individual forecasting method. A prediction interval usually consists of upper- and lower limits with a certain probability. Formula (39) represents the most common formula used in practice to calculate the upper and lower limits for a 100 (1-α)% prediction interval.

$$\hat{y}_{t+h|t} \pm c\hat{\sigma}_h \tag{39}$$

The c denotes the appropriate percentage point of a normal distribution and $\hat{\sigma}_h$ is an estimate of the standard deviation of the h-step forecast distribution. Formula (39) assumes that the forecast is unbiased and the forecast errors are normally distributed. The main problem when creating a prediction interval is the determination of $\hat{\sigma}_h$. We chose to calculate the $\hat{\sigma}_h$ with an empirical procedure. An empirical based prediction interval results in more accurate results than approximate formulas (Chatfield, 1993). An example of an approximate formula that is widely used is multiplying the variance of the one step ahead forecast error with h. Another option would be to construct the prediction interval based on a theoretical formula for $\hat{\sigma}_h$. Yar and Chatfield (1990) created a formula for the $\hat{\sigma}_h$ for the Holt-Winters multiplicative model. However, this formula cannot be applied directly in our case because we use the adjusted Holt-Winters* method. In addition, because the Holt-Winters* method is applied to multiple time series, it is probably not the optimal model for all the time series and therefore it is probably not the best way to derive a prediction interval based on such a model (Chatfield, 1993).

We based our method to construct our prediction interval on the empirical procedures of Chatfield (1993). The forecast errors are collected at 1,2,3,…,65 days ahead (h) for every forecast that we constructed. In total we constructed 100 forecasts with different start dates, which means that each of the 65 days ahead (h) also has 100 forecast errors. The variance is calculated for every link for each h. The variance of each lead time is denoted as S_h and the prediction interval is calculated by replacing $\hat{\sigma}_h$ in formula (40) with S_h. In addition, TNT requested to create a one-sided prediction interval, where there is only a lower limit and no upper limit. This request is made because the tactical planners will never use the values above the point forecast and only want to know what the probability is that the actual value will be below the lower limit. The tactical planner are less interested in the upper limit, because an ad hoc movement is preferred compared to a cancellation.

We constructed prediction intervals for different probabilities and tested how often the actual values were within the prediction interval. We found that the prediction intervals are too narrow, this means that when we constructed a one-sided 95% prediction interval, more than 5% of the actual values were below the lower limit. Prediction intervals that are too narrow are a known phenomenon in forecasting (Chatfield, 1993). Chatfield (1993) stated different causes of a too narrow prediction interval. In our case, the too narrow interval can probably be explained by the errors not being normally distributed or the distribution of errors changing over the testing period.

Table 12 shows the theoretical probability of the prediction interval and the actual percentage that was above the lower limit. Together with TNT we decided what was the best balance between a high probability and a low average S_h (narrower prediction interval). We decided that we will use a 95% prediction interval, which will have a probability of around 89% in reality.

Table 12: Actual percentages within one-sided prediction interval

Probability one-sided prediction interval	85%	90%	92.50%	95%	97.50%	99.50%
Actual % within one-sided prediction interval	79%	84%	86%	89%	92%	96%
Average S_h in cubic meters	27	33	38	43	51	67

4.5 Conclusion

Chapter 4 answered different research questions about the development and results of the model. The main conclusions are:

- In the development of the different models, the same general steps are applied.
    - First, data should be cleaned and outliers should be removed. The values of the public holidays are also replaced by average values. In this research, the focus is on forecasting days that are public holidays and afterwards also public holidays should be forecasted. Second, the data is split in a training- and test set. The models and their parameters are fitted to the training set. Afterwards, a forecast is made for a forecast horizon of 13 weeks. Each forecast has a daily-forecast value and therefore consists out of 65 (13 weeks * 5 weekdays) forecasted values. In total, per link 100 forecasts were constructed. Each forecast has a different forecast starting date, which is between the 1th of November and the 8th of February (100 days).
- To validate the different models, the forecasted volumes are compared to the actual volumes of the test set.
    - To compare the forecasted values with the actual values, we use the MAPE indicator. The MAPE indicator shows how close the forecasted values are to the actual values, expressed in a percentage. Each of the 100 forecasts results in 65 MAPE values and the average over these values is calculated per forecast. Therefore, there are 100 MAPE values in total per link.
- The Holt-Winters method and Neural Network are the most promising methods.
    - First, the Holt-Winters* method performed best in terms of the average MAPE. Based on the maximum MAPE of the forecasts, the Holt-Winters* method and the NN

performs significantly better than the other models, which could indicate that these models are more robust. The accuracy of most models decreases during the first five or six weeks and become more stable in the nine weeks afterwards. Of all Holt-Winters* forecasts that were made, 51 % were within a range of -10% and +10% of the actual values and almost 70% of the forecasts that were made, were within a range of -15% and +15%. Finally, the Holt-Winters* method is the easiest method to implement and the NN has the advantage that it is easy to add information or incorporate new predictions to the model like the prediction of public holidays.

- A one-sided 95% prediction interval is created for the Holt-Winters* method.
    - A one-sided 95% prediction interval is developed with a lower limit. This one-sided prediction interval is tested and 89% of the actual values were within the prediction interval.

5 Implementation and cost savings

This chapter discusses the implementation of the forecasts and the savings that can be achieved with the implementation of the forecast. Section 5.1 describes how the forecast could be implemented and Section 5.2 explains what the cost-savings of this implementation are.

5.1 Implementation

In this research different forecasting methods were developed to improve the planning of TNT. Section 2.1 described the current planning process and we will now explain how the forecast can be implemented in this process.

The tactical planners will receive the forecasted values including the prediction interval for the next 13 weeks per link. The tactical planners will combine this information with other information sources to decide whether besides the masterplan movements, extra movements are needed. In addition, tactical planners decide whether masterplan movements should be cancelled. The other information sources that help the tactical planners are the reports about the loading performance of trucks of that link in the past months, the number of realized movements and loading methodology. The loading methodology can differ per link, for example a truck can be loose loaded or loaded with cages. The loading performance, together with the loading methodology, helps the tactical planners to translate the forecasted volume to the number of trucks. When the loading performance is decreasing, more trucks are needed for the same volume.

The decisions regarding changes of the masterplan are now mainly based on the experience of the tactical planners. The implementation of a forecast can help to provide the tactical planners with more information to base their decision upon. In addition, one of the biggest advantages of implementing the forecast is the possibility to standardize the planning process. It is easier to document how and why decisions related to extra and cancelled movements are made. The first step is to provide the forecast information to the tactical planners and collect their feedback on the forecast and the procedures of providing this extra information. After this, the planning process of the tactical planners should be documented and standard rules for determining extra and cancelled movements should be created. The experience of the tactical planners should be taken into account and based on this, different rules can be created to find the rules with the best results. By standardizing the planning process it is easier for new tactical planners that start working at TNT and the need for experienced planners decreases. The forecast information can also help in the ongoing efforts to improve the masterplan. With this extra information the process of improving the masterplan can be accelerated.

5.2 Cost savings

Chapter 4 developed and compared different forecasting methods. The best forecasting method is selected and this section explains the potential cost-savings when implementing the Holt-Winter* forecasting method. To analyze the cost-savings, the number of masterplan-, extra- and ad hoc movements, the number of last-minute cancellations and cancellations made by the tactical planners are determined for different scenarios. It is important to note that in order to calculate the cost-savings, several assumptions regarding the costs and planning process are made. The process of calculating these costs is visualized in Figure 20.

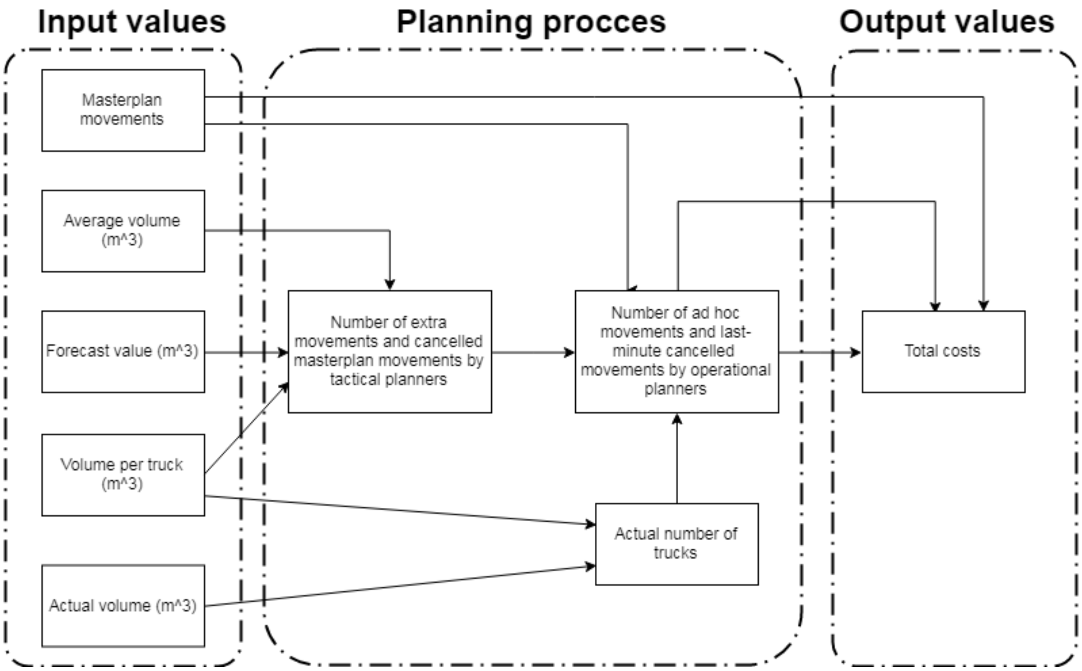

Figure 20: Process of calculating costs

The costs are calculated for two scenarios. The first scenario is a representation of the current situation with the current input of the tactical planners. The second scenario uses the forecasted values from Holt-Winters*. Only the forecast value is different for the two scenarios and the rest of the input values are the same. In addition, the same planning rules are applied for the two scenarios.

For the forecast value (input) of the first scenario, i.e. the current situation, it is assumed that the tactical planners know the average volume per weekday per link of last year. In addition, the tactical planners know if a month is a regular-, a low- or a high volume month for the whole network. The average volume per weekday is decreased or increased with a certain percentage if it is a low or high volume month. In addition, 5% is added because of the growth of TNT. For example, a link has on average 400 cubic meters on Monday and the tactical planners are making a planning for a high volume month. A high volume month has on average 10% more volume than the average volume of last year, which results in the tactical planner using 400 multiplied with 1.15 (1 + 0.05 + 0.1), resulting in 460 cubic meters as input (forecast value).

The rules for the planning process are based on interviews with people from TNT. In the planning process the forecast value is compared to the average volume per weekday. When the forecast value is at least one truck more than the average volume per weekday, the tactical planners plan one or more extra trucks. In the previous example a link had 460 cubic meters as forecast value, and the average for that link was 400 cubic meters which means there are 60 extra cubic meters. The average load of a truck at TNT is 38 cubic meters and as a consequence the tactical planners will plan one extra truck. The 60 cubic meters are divided by 38 and afterwards rounded down. It is rounded down, because an ad hoc movement is preferred over a last-minute cancellation. The cancellations performed by the tactical planners are determined in the same way as the extra movements, with the same restriction of at least one truck. Moreover, a cancellation of a masterplan movement is only done in low volume months.

The number of actual realized movements is the actual number of cubic meters divided by the average load of a truck of TNT and afterwards rounded up. The number of ad hoc movements arranged by the operational planners, are the actual realized number of movements minus the masterplan movements, minus the extra movements and plus the cancelled movements by the tactical planners. The number of last-minute cancelled movements is calculated similarly, but by multiplying the result of the ad hoc movements by minus one. When the number of ad hoc movements or last-minute cancellations is below zero, it is changed to zero. In addition, it is assumed only one of them can be bigger than zero per day.

TNT provided the costs per kilometer for the different types of movements. These normalized costs can be found in Table 13. We assumed that when an extra movement is cancelled last-minute, only the cancellations costs should be paid. Moreover, when the tactical planners cancel a masterplan movement, and it turns out this was an unjust cancellation, an ad hoc movement is arranged. When an unjust cancellation is performed by the tactical planners, both the costs for a masterplan cancellation and ad hoc movement should be paid. However, the costs for a masterplan cancellation by the tactical planner are zero.

Table 13: Normalised costs per movement type

Movement type	Normalised Costs
Masterplan	0.8
Extra	0.9
Ad hoc	1.0
Tactical cancellation	0.0
Last-minute cancellation	0.3

For every link of the five selected links, the average number of movements and average costs for a time period of 13 weeks are calculated for each of the two scenarios. The average over 100 forecasts is calculated for each link. After this, the average over these five links is calculated to obtain the average costs and movements per scenario. The results of the two scenarios and the average total costs can be found in Table 14.

Table 14: Average costs Scenario 1 and 2 per link for 13 weeks

	Current situation	Holt-Winters* point forecast
Forecasted movements	346.3	399.7
Masterplan movements	280.8	280.8
Extra movements	3.3	40.0
Masterplan cancellations tactical planners	0.0	0.0
Ad hoc movements	182.3	145.9
Cancellations last-minute	0.2	0.5
Cancelled extra movements	0.1	0.4
Unjust cancellations tactical planners	0.0	0.0
Realised masterplan movements	280.7	280.7
Total costs	€ 635,100.5	€ 632,744.5

We can see that the Holt-Winters* point forecast has less ad hoc movements and more extra movements compared to the current situation. This is also what we expected since the Holt-Winters* method can predict per week, while in the current situation there is only a distinction between a low-, medium- and high volume months. In addition, the Holt-Winters* calculates the trend per link and in the current situation a trend of 5% for all links is taken into account. This shift from ad hoc movements to extra movements in the Holt-Winters* point forecast results in a cost-saving. However, in the Holt-Winters* point forecast scenario the tactical planners arrange on average 0.4 extra movements that were not necessary compared to 0.1 in the current situation. These extra movements that were unnecessarily planned are more expensive because the last-minute cancellation fee should be paid. The total savings of the shift from ad hoc to extra movements are higher than the costs of the increased number of cancelled extra movements. The Holt-Winters* method saves on average €2.356 per link per 13 weeks compared to the current situation. This is a saving of 0.4% of the total costs in the current situation. The total saving of all 369 links within scope for a whole year would be around €3.5 million ($\frac{€2.356}{13\ weeks * 7\ weekdays} * 365 \text{ days per year} * 369 \text{ links}$).

We also created a third scenario, which uses the lower limit of the one-sided 95% prediction interval of the Holt-Winters* forecast as input. This scenario is created, to compare the Holt-Winters* point forecast with the lower limit of the Holt-Winters* forecast. As said before, TNT prefers ad hoc movements over cancellations and the tactical planners are also interested in the lower-limit of the prediction interval. Table 15 provides the results of the second and third scenario.

Table 15: Average costs Scenario 2 and 3 per link for 13 weeks

	Holt-Winters* point forecast	Holt-Winters 95% prediction interval lower bound
Forecasted movements	399.7	328.4
Masterplan movements	280.8	280.8
Extra movements	40.0	8.3
Masterplan cancellations tactical planners	0.0	3.6
Ad hoc movements	145.9	180.9
Cancellations last-minute	0.5	0.1
Cancelled extra movements	0.4	0.0
Unjust cancellations tactical planners	0.0	3.6
Realised masterplan movements	280.7	277.1
Total costs	€ 632,744.5	€ 636,185.8

The third scenario, i.e. the one-sided 95% prediction interval, has higher total costs compared to the Holt-Winters* point forecast. Scenario 3 has a lower amount of extra movements and a higher number of ad hoc movements. This is also expected because the input values are lower compared to scenario 2 and therefore the tactical planners will plan less extra trucks. In addition, the masterplan cancellations performed by the tactical planners were all unsuccessful, which is costly because the price per kilometer for a masterplan movement is lower than the price per kilometer for an ad hoc movement. However, the average number of cancelled extra movements is lower in the third scenario than in the second scenario. Overall the cost-savings of the third scenario by the reduced cancellations is lower than the cost-increase of the unjust cancellations and the increased number of ad hoc movements. We think that by taking the lowest point in the 95% prediction interval the predictions are on average too low. This can result in costly unnecessarily cancellations. In addition, we think that the desired shift from extra movements to ad hoc movements will not be achieved with the use of this lower limit. We advise TNT, if they want to use the lowest point of a prediction interval, to use a more narrow prediction interval or conduct further research about how the planning process can be improved.

5.3 Conclusion

Chapter 5 discussed the implementation of a forecast method. In addition the potential cost savings are calculated. The main conclusions are presented on the next page.

- The forecast combined with other information sources will help the planners reducing the number of last-minute cancellations and ad hoc movements.
    - To make the planning, the planners will use the forecast, together with the number of realized movements in the past weeks, loading performance and the loading methodology. With this combination a better planning can be made.
- Implementing the forecast will result in some important benefits that cannot be expressed in cost savings.
    - The forecast can help TNT standardize the planning process and the need for experienced planners becomes lower. In addition, for new employees it will be easier to become a tactical planner and the standardization makes it easier to improve the planning.
- Implementing the Holt-Winters* forecast will result in a cost saving of around €3.5 million per year, assuming the assumptions about the costs and tactical planning process are correct.
    - A model is created to represent the planning process of the tactical planners and to compare the results between the current situation and the Holt-Winters* forecast. The Holt-Winters* forecast resulted in less ad hoc movements and more extra movements compared to the current situation. This resulted in an average cost saving of €2.356 per link per 13 weeks.
- When TNT wants to implement a forecast and communicate the lower limit of a prediction interval to the tactical planner, further research should be conducted.
    - We created an extra scenario, where the lower limit of the one-sided 95% prediction interval of the Holt-Winters* method was used as input value for the tactical planners. When we compared this scenario with the scenario of the point forecast of the Holt-Winters*, the costs for this extra scenario were €3.441 higher. The lower limit was too low and resulted in costly unjust cancellations. If TNT wants to implement a forecast and communicate the lower limit to the tactical planners, the prediction interval should become narrower or the tactical planning process should be improved.

6 Conclusion and recommendations

This final chapter concludes the thesis and gives recommendations for TNT. Section 6.1 outlines the main conclusions and Section 6.2 provides the main recommendations for TNT and suggestions for further research.

6.1 Conclusion

The research goal of this research was:

To develop a model that is able to forecast the (range of the) total volume from one location to another location within the TNT European Road Network to improve the planning. The time horizon for the forecast should be for 13 weeks ahead and for every day a forecast should be made.

To achieve this goal, different suitable models are selected from literature and from a different forecasting project within TNT. These models are developed and compared with each other. The performance of the models is measured with the Mean Absolute Percentage Error (MAPE), which indicates how much the forecasts are off compared to the actual values in percentages. The main results are presented in Table 16.

Table 16: Results forecasting models

Model	MAPE	MAPEMin	MAPEMax	MAPEStd
Combination Holt-Winters*, Neural Network, TBATS	11.5	7.2	17.0	9.9
Holt-Winters*	12.5	7.1	19.6	10.5
Neural Network	13.1	8.1	23.6	10.7
TBATS	13.2	7.1	43.5	10.7
Holt-Winters double seasonal	15.1	7.8	46.1	11.2
Simple Exponential Smoothing	15.4	10.3	31.9	12.4

The MAPE of the combination of the forecasts is better than all the individual forecasts. When comparing the MAPE for the individual forecasts, the Holt-Winters* method performs best, the Neural Network and TBATS perform a bit less and the Holt-Winters double Seasonal and Simple Exponential Smoothing perform worst. In general, the performances are close to each other, where the best and worst performing model only have a difference of 3.9 percentage point. The Holt-Winters* method is able to forecast the total volume from one location to another location, where the forecast is on average 12.5% off for a forecast horizon of 13 weeks. During the first five or six weeks, the accuracy of the Holt-Winters*forecast decreases a bit and after this period the accuracy is stable. Of all Holt-Winters* forecasts that were made, 51 % were within a range of -10% and +10% of the actual values and almost 70% of the forecasts that were made, were within a range of -15% and +15%.

The Holt-Winters* method, Neural Network and the combination of forecasts have a low maximum MAPE, which could indicate that these models are more robust than the other models. The combination of forecasts performs best on all performance measures, except for the minimum MAPE, which is slightly higher. Compared to the other individual forecast, the Holt-Winters* method scores best on all factors of Table 16.

The Holt-Winters* method is easiest to implement within TNT, because this method is already used for a different forecasting project. In addition, some extra features are already developed for this model that were out of the scope for this research, but should be included in the future. Examples of extra features are the prediction of public holidays and the possibility to manually change the trend. The Neural Network has the advantage that it can accommodate the prediction of public holidays relatively easy, and we think it will generate good results. The Neural Network however is quite difficult due the "black box" principle and requires some knowledge to tweak the model.

The runtime of the Neural Network and Holt-Winters double seasonal are quite long. However, this can be improved in multiple ways. The runtime of the Holt-Winters double seasonal method can probably be reduced by improving the efficiency of the R code. In addition, TNT can decide to run the forecasts on a faster computer or a server with multiple cores.

Based on this research, we would advise to use the Holt-Winters* method because this has the best results compared to the other individual forecasts, is the easiest method to implement and has a short runtime. The combination of different forecasting methods has good performance, but the disadvantages regarding implementation and understandability are big disadvantages. An interesting alternative could be the Neural Network (NN). We think the NN eventually can outperform the Holt-Winters* but requires further research. The NN can still be improved by tweaking the input setting and by increasing the number of links. In addition, future research should be conducted to see if the NN can forecast public holidays better than the Holt-Winters*.

The forecast that we developed can help the tactical planners in their planning process. With the information of the forecast, the number of last-minute cancellations and ad hoc movements can be reduced. The forecast cannot be translated directly into a planning, because there is a planning step conducted by the tactical planners in between. We created a model that represents the planning process of the tactical planners. This model is used to calculate the cost savings of the Holt-Winters* forecasts compared to the current situation. The two scenarios have different input values for the tactical planners process. For the first scenario, i.e. the current situation, the input value is the average volume per weekday per link of last year. In addition, a standard expected growth factor of 5% is added to the input value. Moreover, an increase or decrease is added for high- or low volume months. The second scenario uses the point-forecast of the Holt-Winters* forecast as input. By using the Holt-Winters* forecast, on average €2.356 can be saved per link per 13 weeks compared to the current situation assuming the assumptions that we made about the costs and planning process are correct. The total saving would be around €3.5 million per year for all the 369 links.

Furthermore, by implementing a forecast, there are some important benefits for TNT of forecasting that cannot be expressed in cost-savings directly. The implementation of a forecast can improve the standardization of the planning process across multiple regions in Europe. The standardization lowers the need for experienced planners and it will be easier for new tactical planners that start working at TNT. In addition, a more standardized planning process makes it easier to improve this process and save costs. Moreover, a forecast can contribute to the ongoing efforts to improve the masterplan by providing extra information. Finally, the relation with sub-contractors can be improved by informing them earlier of changes in the planning.

6.2 Recommendations and further research

- Use the Holt-Winters* method for the forecast.
- Conduct further research about Neural Networks.
- Provide extra information to the planners, such as the volume for a certain link and the type of link.
- Keep improving the data cleaning process to improve the forecast. The anomaly list can be extended and should be kept up to date. In addition, other methods to improve the data cleaning process by filtering out outliers can be investigated.
- Investigate how to include the forecast of public holidays in the model.
- Conduct further research about the implementation of the prediction interval.

7 Bibliography

Armstrong, J. (2011). Illusions in Regression Analysis.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 451-468.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting.* Springer.

Chambers, J., Mullick, S., & Smith, D. (1971, July). How to Choose the Right Forecasting Technique. *Harvard Business Review*.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics,*, 121-135.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 559-583.

Cunha, C. B., & Silva, M. R. (2007). A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil. *Elsevier*, 747–758.

Curry, B., & Morgan, P. H. (2006). Model selection in neural networks: some difficulties. *European Journal of Operational Research*, 567-577.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 1513-1527.

Dekker, M., Van Donselaar, K., & Ouwehand, P. (2004). Dekker, M., Van Donselaar, K., & Ouwe How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, 151-167.

Ducret, R. (2014). Parcel deliveries and urban logistics: Changes and challenges in the courier express and parcel sector in Europe—The French case. *Elsevier*, 15-22.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 1-28.

Hamzaçebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, 4550-4559.

Heitz, A., & Beziat, A. (2016). The parcel industry in the spatial organization of logistics activities. *Elsevier*, 812-824.

Hesse, M. (2002). Shipping news: the implications of electronic commerce for logistics and freight transport. *Elsevier*, 211-240.

Hyndman, R. J. (2010, November 30). *Initializing the Holt-Winters method*. Retrieved November 13, 2018, from robjhyndman: https://robjhyndman.com/hyndsight/hw-initialization/

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice.* Melbourne: OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 679-688.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 2664-2675.

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *Transactions on neural networks*, 1341-1356.

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 669-679.

Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M., et al. (2000). An investigation of machine learning based prediction systems. *Journal of Systems and Software*, 23-29.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of forecasting*, 111-153.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*, 135-150.

ORTEC. (2018). *Line Haul Optimization*. Retrieved June 1, 2018, from ORTEC Corperation web site: https://ortec.com/dictionary/line-haul-optimization

Roy, J. (2001). Recent trends in logistics and the need for real-time decision tools in the trucking industry. *34th Hawaii International Conference on System Sciences* (pp. pp. 9-pp). IEEE.

S., A. J., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 69-80.

Sakia, R. M. (1992). The Box-Cox transformation technique: a review. *The statistician*, 169-178.

Swingler, K. (1996). *Applying neural networks: a practical guide.* Morgan Kaufmann.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 437-450.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 799-805.

Winston, W. L., & Goldberg, J. B. (2004). *Operations research: applications and algorithms (Vol. 3).* Thomson, Brooks & Cole.

Yar, M. &. (1990). *Prediction intervals for the Holt-Winters forecasting procedure.*

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 159-175.

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 501-514.

Zhou, L. (2017). Prediction of a service demand using combined forecasting approach. *Journal of Physics: Conference Series*, Vol. 887, No. 1, p. 012075.

Zhou, L., Heimann, B., & Clausen, U. (2006). Forecasting a logistic service demand based on neural network. *Service Systems and Service Management*, 530-534.

Appendix A: Detailed description Road Network Europe - Confidential

[Confidential]

Appendix B: Abbreviations of all LPC's

Abbreviation	Country	City
TRS	Germany	Troisdorf
MV9	France	Marly-la-Ville
QAR	Netherlands	Duiven
VIE	Austria	Vienna
HL3	Sweden	Helsingborg
BZQ	Belgium	Brussels
DZ5	England	Dartford
MAD	Spain	Madrid
LGG	Belgium	Liege

Appendix C: Monthly pattern during the year

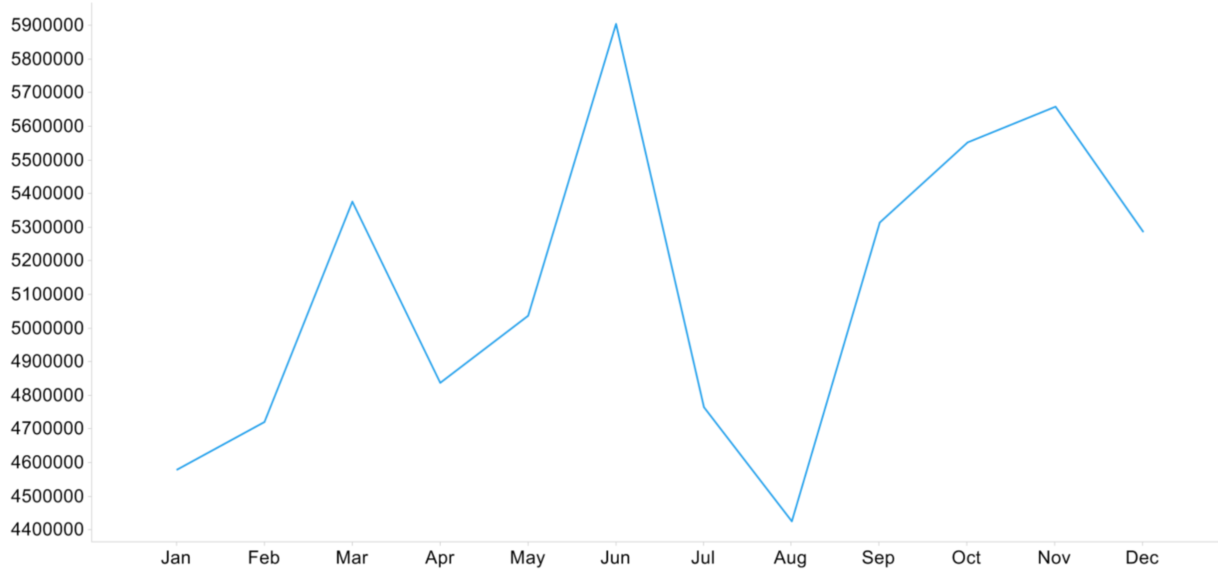

Figure 21: Total cubic meters per month

Appendix D: Number of days with volume per OD combination

In Figure 22, the number of days with volume per OD combinations is shown. The y-axis is the number of days with volume and the x-axis are all the OD combinations. Every OD combination has one bar with the number of days with volume, but because there are 520.000 OD combinations it is impossible to include all the OD combinations on the x-axis.

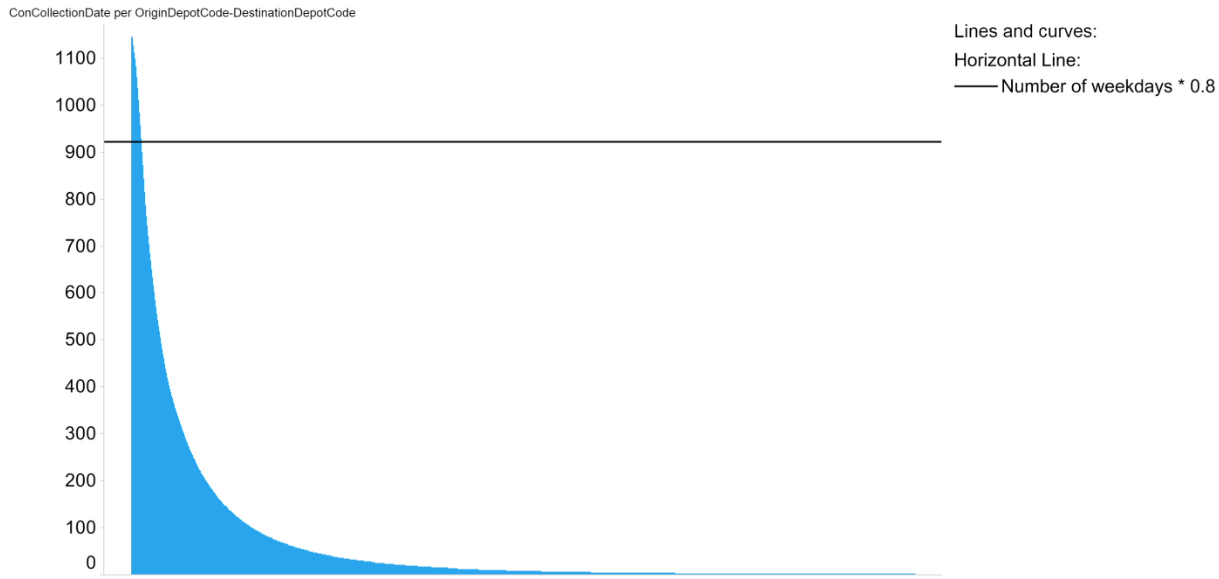

Figure 22: Number of days with volume per OD combination

Appendix E: Details seasonal patterns per link

Link	Size	Pattern over the years	Monthly pattern during the year	Week pattern during the year	Weekday pattern over the months	Daily pattern per weekday
Arnhem - Madrid	Large	Increase from 2014-2016, slight decrease 2017	Quite the same monthly pattern over the years, 2017 a bit more extreme	Weak pattern and an increase in the end of the year in most years.	Weak daily pattern during the weeks over the months	Quite stable pattern, except for outliers that are probably public holidays.
Arnhem - Marly-la-Ville	Large	Increase from 2014-2016, stable in 2017	From 2015-2018, quite the same monthly pattern over the years.	Very weak pattern and an increase in the end of the year in most years.	Same daily pattern during the weeks over the months	Quite stable pattern, except for outliers that are probably public holidays.
Darthford - Madrid	Small	Increase from 2014-2016, decrease 2017	Some monthly pattern over the years, but not always the same	Can't identify a visual pattern.	Very weak daily pattern during the weeks over the months	Quite stable pattern, except for outliers that are probably public holidays.
Hannover - Warschau	Medium	Increase from 2014-2016, strong increase 2017	Increase during the year, 2014-2016 quite stable, 2017-2018 more variance	Almost no visual pattern, except an increase in the end of the year in all years.	Can't identify a visual pattern, also not per year.	Quite stable pattern, except for outliers that are probably public holidays.
Neurenborg - Milan	Medium	Increase 2015, strong increase 2016, stable in 2017	Quite the same monthly pattern over the years	Almost no visual pattern, except an increase in the end of the year in all years.	Some daily pattern during the weeks over the months	Quite stable pattern, except for outliers that are probably public holidays.

Appendix F: Pattern over the years Arnhem-Madrid

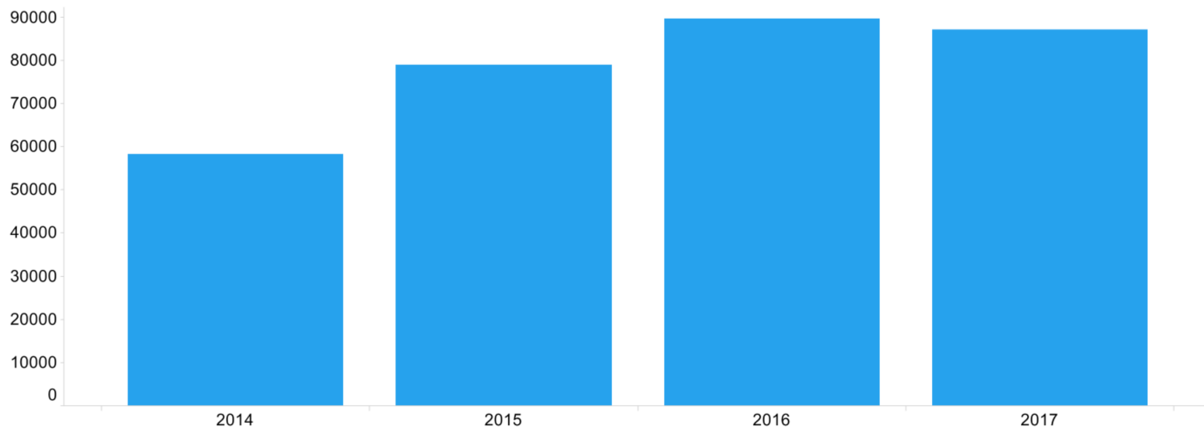

Figure 23: Total cubic meters per year Arnhem-Madrid

Appendix G: Monthly pattern during the year Arnhem-Madrid

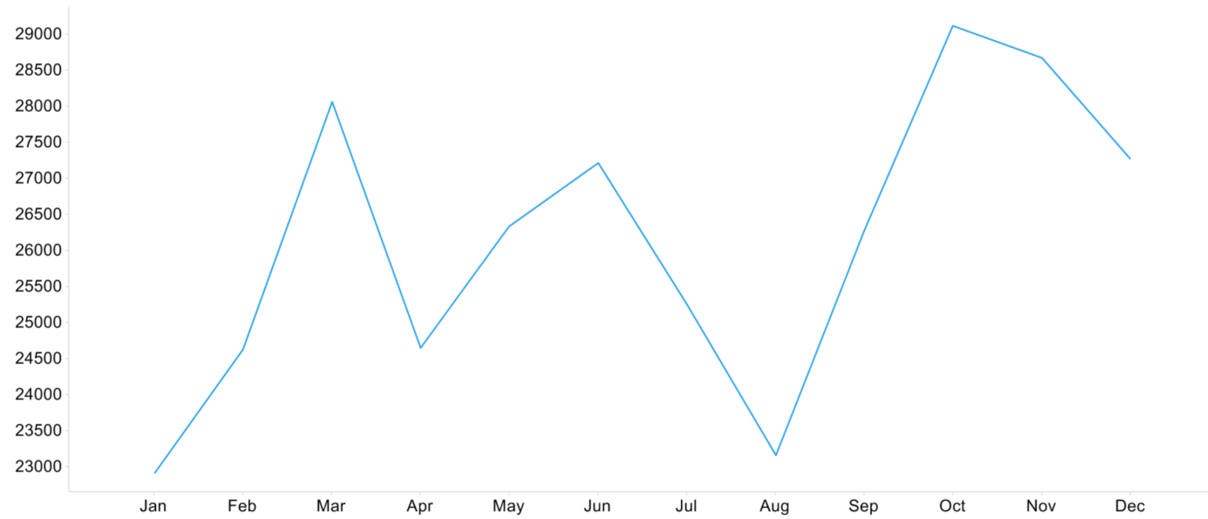

Figure 24: Total cubic meters per month Arnhem-Madrid

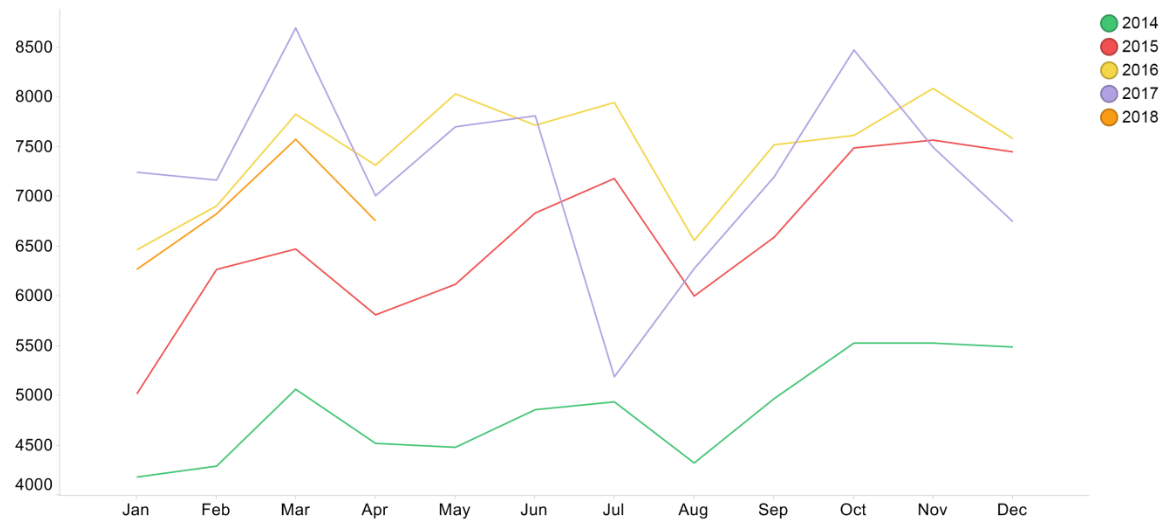

Figure 25: Total cubic meters per month and year Arnhem-Madrid

Appendix H: Week pattern during the year Arnhem-Madrid

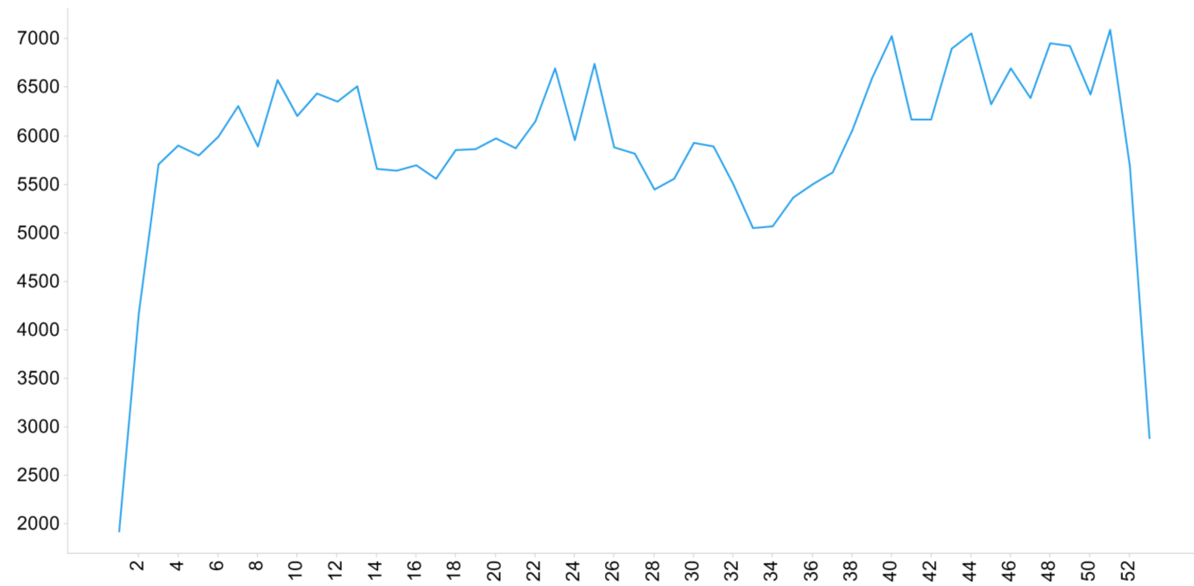

Figure 26: Total cubic meters per week Arnhem-Madrid

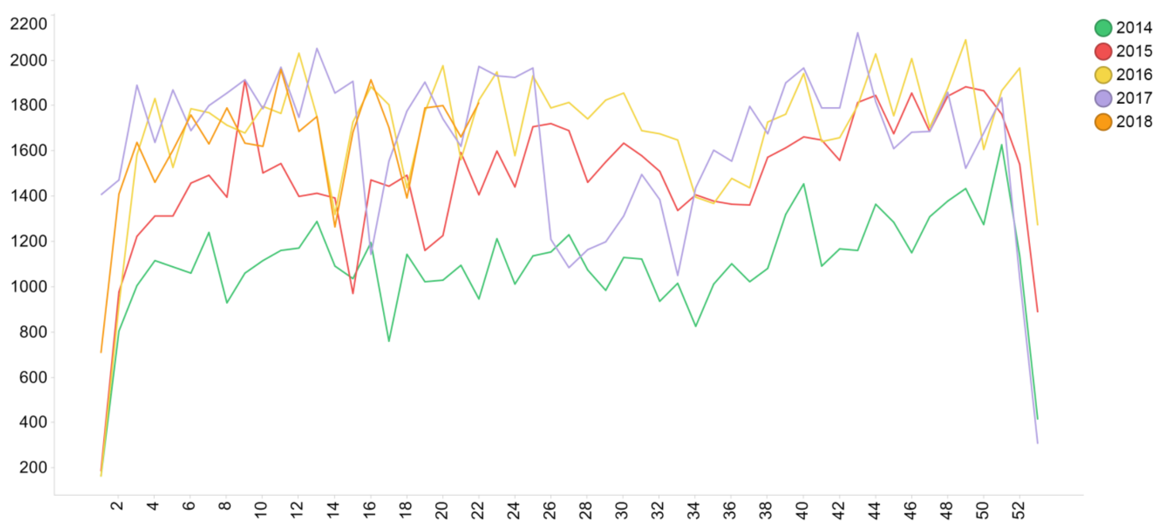

Figure 27:Total cubic meters per week and year Arnhem-Madrid

Appendix I: Weekday pattern over the months

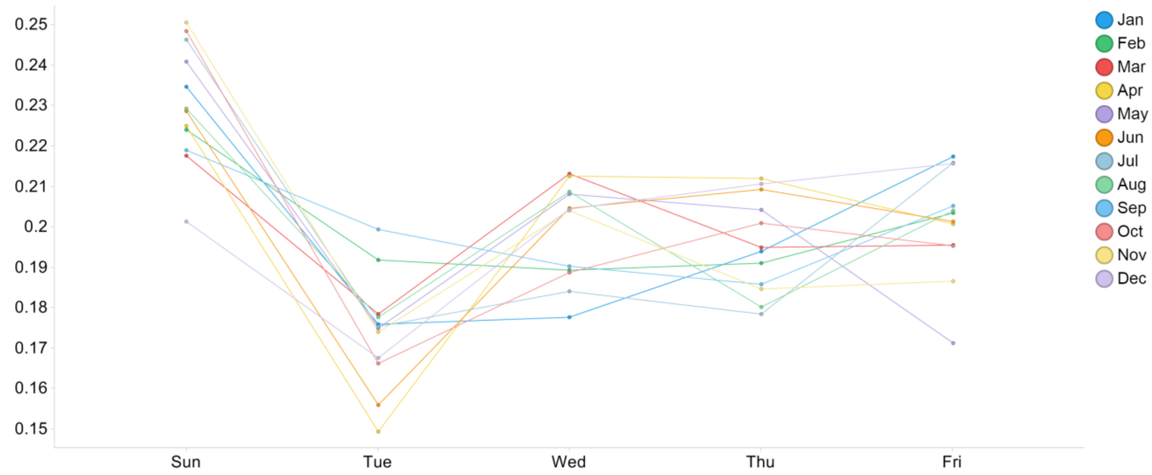

Figure 28: Distribution of volume per weekday and month Arnhem-Madrid

Appendix J: Daily pattern per weekday

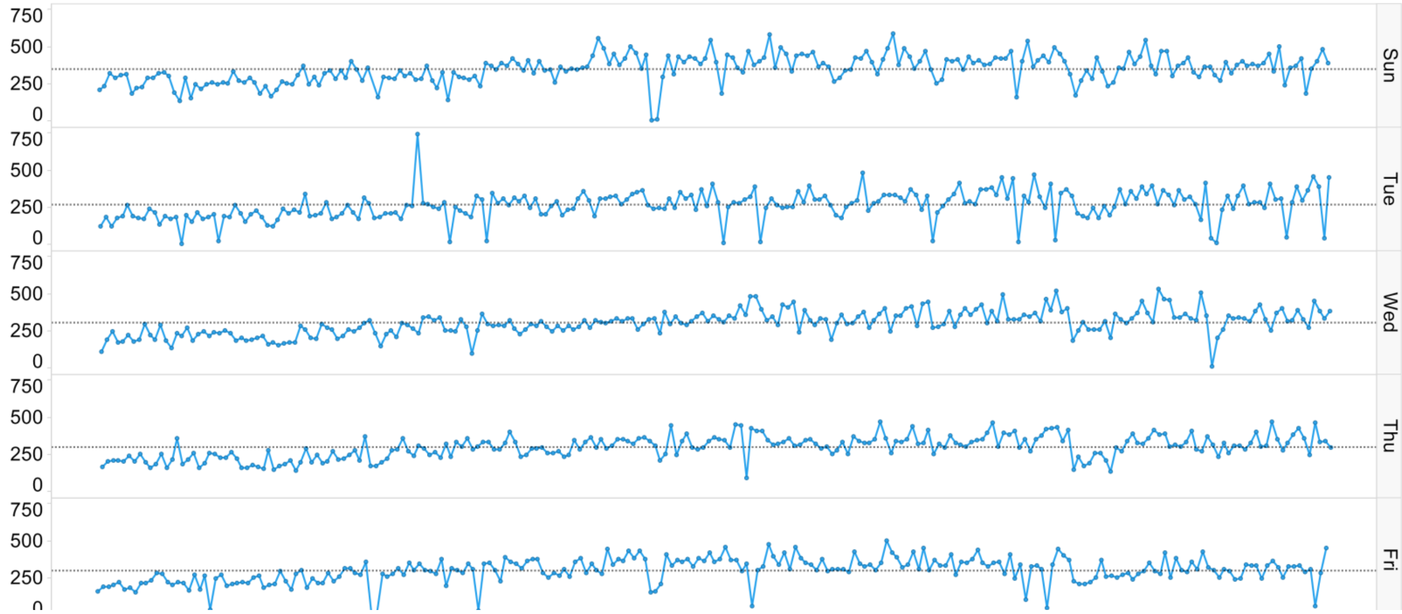

Figure 29: Daily total cubic meters per weekday

Appendix K: Additive Holt-Winters Formula

Below the formulas are given for the Holt-Winters additive variant. The additive variant is less used in practice, but is preferred when the seasonal variations are roughly constant during the year (Hyndman & Athanasopoulos, 2018).

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

$$s_t = \gamma(y_t - l_t) + (1-\gamma)s_{t-m}$$

Appendix L: Results comparison 52/13/4 periods

Table 17 provides the results for the Holt-Winters double seasonal method. The results are calculated for 52, 13 and 4 periods per year for all five selected links. The last three rows are the average results per period over the five links.

Table 17: Results Holt-Winters double seasonal different periods

OriginLinkLocation	DestinationLinkLocation	MAPE	MAPEMin	MAPEMax	MAPEStd	Periods
QAR	MAD	13.6	9.6	27.2	11.0	52
QAR	MV9	11.7	7.8	21.4	7.9	52
DZ5	MAD	13.7	9.4	18.2	11.7	52
HNJ	WA1	16.4	10.3	31.9	11.6	52
DNG	MI6	20.3	9.8	46.1	13.9	52
QAR	MAD	14.6	9.7	23.9	11.5	13
QAR	MV9	13.5	7.4	23.4	8.5	13
DZ5	MAD	12.3	9.3	14.3	11.8	13
HNJ	WA1	16.0	10.1	27.5	11.0	13
DNG	MI6	20.3	9.8	46.1	13.9	13
QAR	MAD	14.6	9.3	25.0	11.7	4
QAR	MV9	13.3	7.0	22.2	8.5	4
DZ5	MAD	12.7	9.6	15.3	11.6	4
HNJ	WA1	15.9	10.1	27.0	11.0	4
DNG	MI6	20.2	9.8	45.7	13.8	4
Average	**Average**	**15.1**	**9.4**	**29.0**	**11.2**	**52**
Average	**Average**	**15.3**	**9.3**	**27.0**	**11.3**	**13**
Average	**Average**	**15.3**	**9.1**	**27.0**	**11.3**	**4**

Appendix M: First test results NN

Table 18 shows the first test results of the NN. In this experiment it is tested whether "year" should be in the input layer and some first testing with the number of hidden units. We can see that the MAPE of the experiments including the year in the input layer are always higher compared to the experiments without year in the input layer.

Table 18: First results NN

Test	Columns	Threshold	Repetition	Algorithm	HiddenUnits	error	MAPE
1	filtered,TNTWeek,YearWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		5	374.6	21.7
2	filtered,TNTWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		5	1105.2	29.1
3	filtered,TNTWeek,YearWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		10	293.1	17.7
4	filtered,TNTWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		10	494.7	26.2
5	filtered,TNTWeek,YearWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		50	194.6	20.0
6	filtered,TNTWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		50	449.7	25.5
7	filtered,TNTWeek,YearWeek,Month,weekday,OriginLinkLocation,DestinationLinkLocation	0.2	3		20,20	145.6	19.3

Appendix N: Results Neural Network dummy variables

Table 19 gives the results for the Neural Network to test which columns should be transformed to a dummy variable. The second column indicates which columns are transformed to a dummy variable.

Table 19: Results NN dummy variables

HiddenUnits	DummyVariables	MAPE	MAPEMin	MAPEMax	MAPEStd
5	OriginLinkLocation,DestinationLinkLocation	17.7	7.8	28.2	12.5
8	OriginLinkLocation,DestinationLinkLocation	13.0	9.7	17.0	10.1
10	OriginLinkLocation,DestinationLinkLocation	11.4	7.8	14.9	9.2
12	OriginLinkLocation,DestinationLinkLocation	12.1	6.7	16.4	9.8
15	OriginLinkLocation,DestinationLinkLocation	12.3	7.5	15.7	9.7
4,4	OriginLinkLocation,DestinationLinkLocation	14.5	7.7	22.2	11.7
5,5	OriginLinkLocation,DestinationLinkLocation	11.1	8.6	13.9	9.1
6,6	OriginLinkLocation,DestinationLinkLocation	14.2	9.8	19.0	10.6
7	OriginLinkLocation,DestinationLinkLocation,weekday	14.4	9.2	20.4	11.4
11	OriginLinkLocation,DestinationLinkLocation,weekday	14.3	9.2	22.3	11.2
14	OriginLinkLocation,DestinationLinkLocation,weekday	13.7	8.6	18.4	10.7
17	OriginLinkLocation,DestinationLinkLocation,weekday	14.6	10.0	21.0	11.5
21	OriginLinkLocation,DestinationLinkLocation,weekday	13.8	8.8	20.7	11.7
6,6	OriginLinkLocation,DestinationLinkLocation,weekday	11.7	9.2	15.5	9.7
7,7	OriginLinkLocation,DestinationLinkLocation,weekday	13.1	9.5	15.8	10.4
8,8	OriginLinkLocation,DestinationLinkLocation,weekday	24.5	9.6	65.2	17.9
11	OriginLinkLocation,DestinationLinkLocation,weekday,Month	15.8	8.7	24.2	11.7
22	OriginLinkLocation,DestinationLinkLocation,weekday,Month	18.2	8.2	26.0	14.7
33	OriginLinkLocation,DestinationLinkLocation,weekday,Month	14.1	9.5	18.9	11.7
11,11	OriginLinkLocation,DestinationLinkLocation,weekday,Month	14.0	8.2	20.2	12.1
30	OriginLinkLocation,DestinationLinkLocation,weekday,Month,TNTWeek	21.0	11.7	26.2	19.5
60	OriginLinkLocation,DestinationLinkLocation,weekday,Month,TNTWeek	26.4	19.8	43.4	27.8