Average speed prediction using Artificial Neural Networks

Cynthia Jacobs
Delft University of Technology
Faculty of Information Technology and Systems
August 2003

Graduation Committee:

drs. dr. L.J.M.Rothkrantz
prof. dr. ir. E.J.H. Kerckhoffs
prof. dr. H. Koppelaar

Jacobs, Cynthia
Master thesis, August 2003
Average speed prediction using Artificial Neural Networks

Delft University of Technology, The Netherlands
Faculty of Information Technology and Systems
Data and Knowledge Engineering Group

Acknowledgements

First of all, I would like to thank Leon Rothkrantz for his feedback, comments and advice during the project. I have really enjoyed interesting discussions about travel time and average speed prediction and many other non-relating topics.
Furthermore I would like to express my gratitude towards Pascal Wiggers for the conversations leading to new ideas, the constructive criticism on my report and his friendship.
I also want to thank my friends Frans Flippo, Michiel Konstapel and Otto Visser for making most of my time at the university very enjoyable.
And at last I would like to thank my parents, my sister and especially Lars van Leeuwen for their moral support and more.

Cynthia Jacobs
Delft, August 2003

Abstract

Nowadays cars have navigation systems guiding the traveler from departure to destination. The navigation system uses route planners to determine the best route for the traveler. However these route planners do not use dynamic traffic information to decide what route is best. These route planners use the maximum speed allowed on a road to determine the travel time. So it is possible that the traveler is guided over congested routes while another route will lead the traveler faster to its destination.

In this thesis at first it is explored what influences the average speed on the road and it is tried to predict this average speed using a feed forward neural network. If this average speed is known the real travel time can be determined for parts of the road and the best route can be discovered using this travel time.

Analyzing the data showed that time, day of the week, month, weather, events, holidays and special events like accidents are factors that influence the average speed on the road. These influencing factors are used to predict the average speed on a specific point on a road. Therefore three models are developed. Two of these models are tested using JavaNNS developed at the Wilhelm-Schickard-Institute for Computer Science in Tübingen, Germany.

The research on the architecture of neural networks and tests showed that a feed forward neural network with one hidden layer, from which the number of hidden neurons is determined by the formula of Fletcher and Goss (between $2N_i+1$ and $2\sqrt{N_i}+N_o$), is capable of predicting the average speed using only the time, month, day of the week and the average speed of 40 minutes before. At least 76% of the predictions have a difference smaller or equal to 10% if the prediction is made 40 minutes ahead. From these results it can be concluded that Artificial Neural Networks are capable of making good approximations for problems like predicting the average speed.

Table of contents

Chapter 1

Introduction

1.1 Problem setting

In The Netherlands the traffic on the roads suffers heavily from congestion. In the rush hours most of the time a total of 200 kilometers of traffic jam can be found. These traffic jams have serious financial consequences. The direct costs of the delays caused by the congestion have been estimated to be 1.7 billion guilders in 1998. The indirect cost may even be higher. These cost include follow-up costs because of arriving early or late at the destination and prevention costs of trying to be in time [Eggenkamp 2001].
Consequently it is tried to solve these traffic problems. Research is done into automatic car following systems. These systems may be used to reduce the minimum distance between the cars and therefore the capacity of the road increases. Teleworking at home can be another way to reduce the traffic problems. This will reduce the number of movements on the road. Also there is invested in public transport and the road network. During this time monitoring systems are installed to track the traffic on the motorway network.
Nowadays new cars are often outfitted with navigational systems that guide the driver from the place of departure to his destination. Such a route planner or navigation system has to know the best route from a certain point A to another point B. When determining the best route from A to B one first should define what best means. In real life there are several roads or combination of roads that lead from A to B and the traffic density differs from time to time. But which route has to be taken on some given moment.

The best route from A to B can be determined depending on different criteria:

- The shortest route
- The cheapest route
- The most convenient route
- The fastest route

The shortest route depends on the number of kilometers. If this criterion is used to determine the best route, at first the distance along the route is measured for every combination of roads. After that the route with the least amount of kilometers is taken.

Some people like to pay as less as possible to drive a car. So the route, which is the cheapest, has to be determined. This can be very complicated, because it has to be estimate what the cost are of a certain route. This is a combination of the cost for toll roads and the consumption of fuel on a certain road. The consumption of fuel is depending on the type of car, the age of the car and the state of maintenance of the car. These costs are hard to determine. Further driver traveling for business, e.g. transportation firms, may also consider time to be valuable. So the time it takes to get from A to B is also important to consider.

Moreover the most convenient route can be taken. Avoiding for example traffic lights and crossings by taken the highway may be a longer route, but the driver can experience this as more comfortable.

The fourth method to calculate the best route is to take the fastest route. Most route planners nowadays take this criterion to determine the best route. These route planners are called 'static' route planners. They determine the minimum travel times on a road by taking the length of the road and the maximum speed allowed on that road. This seems a very good way to give an optimal route. But in reality this isn't a very reliable route. Due to traffic congestion the maximum speed cannot be reached. The travel times can be much higher than the minimum travel times.

To reduce the travel time for individual travelers a route planner can be developed that searches the fastest route using traffic information and public transport possibilities. If the normal route is congested this planner should be able to find the best alternative route. It is possible that this alternative is leading the traveler along the highways, but also an advice may be given to use the public transport. Such a route planner is not yet available.

The existing route planners have two major drawbacks. At first they do not take any dynamic traffic information, like information about traffic jams, into account. The so-called 'static' route planners are capable of giving an optimal route if no congestion is available. But this route may not always be optimal due to traffic jams. Above it is stated that an average of at least 200 kilometers of traffic jams can be found during working hours. Therefore these route planners only give optimal routes during a few hours of the day. If the dynamic route information is used a planner is able to search for another route when the route found by the 'static' route planner is congested. Another advantage of using dynamic traffic information is that it is possible to estimate a realistic travel time.

The second major drawback is that current route planners do not incorporate information about other kinds of transport than to go by car. With these planners it is not possible to plan a route using different kinds of transport. Besides a lot of route planners for cars there is a route planner for the public transport. However there is no route planner that searches for a possible combination of those kinds of transport. In [Eggenkamp 2001] a prototype of such route planner is introduced.

It is stated above that current route planners do not use any dynamic traffic information and that using this information leads to an optimal planned route even when traffic congestion is present. This will also make it possible to give an estimation of a realistic travel time. But this route will not be optimal if a traffic jam arises when the route is already planned. In [Eggenkamp 2001] using the PITA the route will be planned again during the trip using new information. If a traffic jam did arise a new optimal route is planned from the point where the traveler is at that moment. However most of the times it would be advisable to take a completely different route if this information is known in advance. For example if a traveler wants to go from The Hague to Groningen by car, there are three major routes leading to it. The first route goes along the Afsluitdijk. Another possibility is to go through Flevoland, without traffic congestion this will be the fastest route, or the traveler can go via Amersfoort. If it is known in advance that the A9 and the A10 around Amsterdam are not yet congested but will be before the traveler gets there, it can be advised to take the route via Amersfoort. The reason for that decision is that the first two routes both go around Amsterdam. Without this information the traveler was send in the direction of Amsterdam and is not able to take the third route after all unless a detour is taken.
If the speed on a road can be predicted for a certain period ahead it will be possible to give more reliable travel times for a part of a route. By combining these different travel times the real fastest route can be given. This route can be different over time. For example:

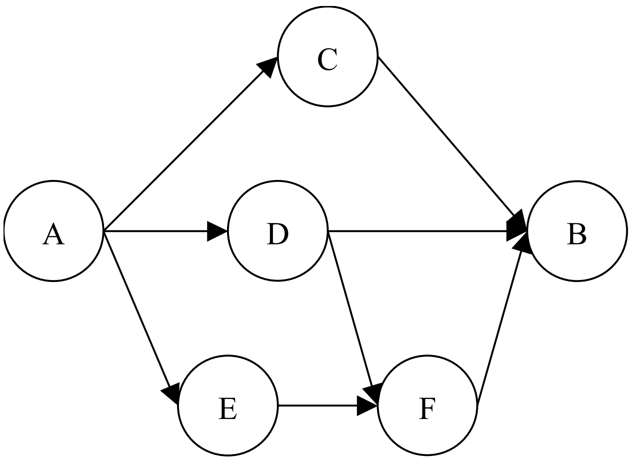


Figure 1.1 Example of a traffic network

In Figure 1.1 four different routes can be seen from A to B.

- A → C → B
- A → D → B
- A → D → F → B
- A → E → F → B

Every arrow corresponds to a certain travel time. The travel times are depending on the time of the day. In Table 1.1 the travel times for two possible situations are given, one for 7 to 8 o'clock on a Monday morning in January and one for 3 to 4 o'clock on a Saturday afternoon in April.

Table 1.1 Travel times on Monday 7.00 - 8.00 AM and Saturday 3.00 - 4.00 PM

	Monday 7.00 - 8.00 AM	Saturday 3.00 - 4.00 PM
A → C	13 minutes	9 minutes
A → D	5 minutes	2 minutes
A → E	6 minutes	4 minutes
C → B	22 minutes	10 minutes
D → B	27 minutes	15 minutes
D → F	7 minutes	5 minutes
E → F	11 minutes	8 minutes
F → B	14 minutes	12 minutes

Table 1.2 Route times on Monday 7.00 - 8.00 AM and Saturday 3.00 - 4.00 PM

	Monday 7.00 - 8.00 AM	Saturday 3.00 - 4.00 PM
A → C → B	13 + 22 = 35 minutes	9 + 10 = 19 minutes
A → D → B	5 + 27 = 32 minutes	2 + 15 = 17 minutes
A → D → F → B	5 + 7 + 14 = 26 minutes	2 + 5 + 12 = 19 minutes
A → E → F → B	6 + 11 + 14 = 31 minutes	4 + 8 + 12 = 24 minutes

On Monday morning the fastest route is A → D → F → B under the assumption that the travel times will not change dramatically the next hour, as can be seen in Table 1.2. This is not the same as on the Saturday afternoon. On that time the best route is A → D → B.
In [Van Lint 2000] is shown that it is possible to predict the travel times accurate up to 10 to 25 minutes in the future, but this is not enough if a long route has to be given.
The dynamic route information should therefore be expanded with a prediction of travel times on a road over a certain time period longer than 25 minutes. Using the travel times of the prediction the route planner is able to give an optimal route whether or not traffic congestion does or will occur.

1.2 Goals

The goals can be divided into three topics:

- Research on traffic
- Research on Artificial Neural Networks
- Research on travel time prediction

1.2.1 Research on traffic

This part of the study consists of two points:

- The literature survey
- The traffic data

The literature survey
In the literature research can be found about the influences of the travel times on the economy and the prediction of these times. At first this literature has to be studied.

Traffic data
The second step of this part is investigating which information about traffic is available and where this information can be collected.

1.2.2 Research on Artificial Neural Networks

In the literature about traffic neural networks are used to predict travel times. Therefore research on neural networks should be done. The research on artificial neural networks is also divided into two parts:

- The architecture of a neural network
- Experiments

The architecture of a neural network
First of all a research in literature is necessary to determine the architecture of a neural network. This means it must be decided what kind of neural network should be used. Also the number of hidden layers and the number of neurons in these layers should be determined.

Experiments
Secondly the architectures that conclude from the previous research should be tested.

1.2.3 Research on travel time prediction

In dynamic routing, the average travel time on the links is needed. In literature systems can be found that uses the speed measurements on different locations along a link to predict the travel times. These measurements are taken every minute. When the speed measurement fails, the system fails completely, because there is no actual input. In this research the possibility to predict the speed measurements is investigated.
This research is divided in three parts:

- Research on the influences on the speed
- Developing a model
- Experiments

Research on the influences on the speed
Before a model can be developed it has to be investigated which factors influence the speed on the road and from which location this data can be collected.

Developing a model
Using the influences on the road a model has to be developed capable of predicting the average speed on a given place x and time t. Basically the function F(x, t) on every link is considered. This function can be approximated by the average values. But a computing device is wanted. An analytic representation of F is one way to do it. Also an implicit representation of F by Artificial Neural Networks can be used. This is one of the goals of our research.
To be able to use this average speed for predicting the travel times into the future the following goal should be reached:

Study the possibilities of predicting the average speed on a certain place over 25 minutes in the future with an acceptable accuracy. An acceptable accuracy of the predicted speed is defined as a difference of at most 10 percent of the real speed.

Experiments
Last topic of the research on travel time prediction is testing the model(s) and checking the model(s) on accuracy.

1.3 Structure

The report will be structured as follows. In the second chapter the considered literature is summarized. The third chapter is about factors that possibly influence the travel times and will also contain a description of the structure in which the information about these influences is available. The analysis of influence of this data is discussed in Chapter 4. In Chapter 5 a general overview of neural network is given and it also contains a discussion about possible architectures of neural networks. The tool that is used to simulate neural networks is presented in Chapter 6. Chapter 7 consists of a description of the preparation of the data necessary to be able to feed it to the simulation program. The models developed are presented in Chapter 8 and in Chapter 9 results of the prediction using the model are given. Chapter 10, 11 and 12 respectively contain the conclusion, recommendations and references.

Chapter 2

Literature survey

2.1 Introduction

Every day people have to travel. The travel times on the same part of the road can be different every day due to traffic congestion or other reasons. If the travel times on a given day and place would be known in advance another route or another departure time can be chosen. In that case the travel times have to be predicted in advance. Also route planners are depending on travel times in order to determine the fastest route. The route planners now use the minimum travel time of that route. If these travel times can be predicted, more realistic fastest routes can be given. But is it possible to predict these times? In the past years there already has been some research in relation to predicting travel times. In this literature survey an overview of some representative papers in the area of travel time prediction is given.
At first a list of papers will be given and these paper will be categorized. After this categorization a summary of every paper will be given followed by comment. Section 2.3 contains the conclusion.

2.2 Literature

The subject in this bibliography is predicting travel times. All the papers are about traffic, but not all about predicting travel times. The papers can belong to one or more of the following categories:

- Specific articles
- Overview articles
- Scientific articles
- General articles

First a numbered list with al the literature is given. After this list a table will be presented, categorizing the literature in the categories. In the third section the summaries of the papers will be given with some comments.

2.2.1 Literature list

1. Dynamisch verkeersmanagement, H.J.van Zuylen.
2. Robust freeway travel time prediction with state space neural networks, J.W.C. van Lint.
3. Robust and adaptive travel time prediction with neural networks, J.W.C. van Lint.
4. Short-term prediction of Traffic Flow Conditions in a Multilane Multiclass Network, S.P. Hoogendoorn.
5. Freeway Travel Time Prediction with State-Space Neural Networks, J.W.C. van Lint.
6. State Space Neural Networks for Freeway Travel Time Prediction, J.W.C. van Lint.
7. Determination and evaluation of traffic congestion costs, I. Hansen.
8. Different Perspectives on the Global Development of Transport, B. Ubbels.
9. Using stated preference data for studying the effect of advanced traffic information on drivers' route choice, M.A. Abdel-Aty.
10. Spectral Basis Neural Networks For Real-Time Travel Time Forecasting, D. Park.

2.2.2 Categorization

In Table 2.1 a categorization is stated of the literature in the categories 'specific', 'overview', 'scientific' and 'general'. If a paper belongs to the category 'specific' the paper describes one aspect of a subject into detail. A paper in the category 'overview' will give a broad view of a subject. To the third category, 'scientific', belong the papers that contain a systematic approach of the subject or stick to the exact science. The papers that do not belong to this category are labeled as a 'general' paper. All the papers are sorted into two of the categories.

Table 2.1 Literature categorization

Literature number	Specific	Overview	Scientific	General
1	+			+
2	+		+	
3		+	+	
4	+		+	
5	+		+	
6	+		+	
7	+		+	
8	+		+	
9	+		+	
10	+		+	

2.2.3 Summaries

Dynamisch verkeersmanagement

H.J.van Zuylen
Pao Cursus Model- en Evaluatiestudies Verkeersbeheersingsmaatregelen, februari 2000

Summary
In The Netherlands there will be more traffic every day. In the cities it is attempted to control traffic with traffic lights and other kinds of controllers. Dynamic traffic management is meant to control the traffic. By improving the accessibility, there will be more possibilities for the economy. If traffic congestion disappears and the travel times will become lower, the demand for transport will change. There are many different dynamic traffic management systems.
If the dynamic traffic management arrangements will be effective, the congestions will be less or even disappear. However, this effect is not noticeable. If the situation on the road gets better, the traffic grows and the situation will be the same as before. At first sight there seems to be no use of dynamic traffic management. But if this traffic management results in equally amount of traffic congestion, while there is more traffic, this can be seen as an improvement.

Comment
This paper is about traffic management. Predicting travel times can be very useful in traffic management. If it is known that traffic congestion will occur, then with traffic management traffic can be rerouted.

Robust freeway travel time prediction with state space neural networks

Ir. J.W.C. van Lint, Dr. ir. S.P. Hoogendoorn, Prof. ir. H.J. van Zuylen
Dept. of traffic Engineering and transportation analysis, Faculty of Civil Engineering and Geosciences, Delft University of Technology
Accepted for presentation at the 13th Mini-EURO Conference "Handling Uncertainty in the Analysis of Traffic and Transportation Systems" June 10 to 13, 2002, BARI, Italy.

Summary
The need for systems that provide accurate travel information is increasing. The travel time prediction model has to be robust. In this paper the sensitivity of faulty or missing input data on a state space neural network is studied. The model tells that the state of a system at time t+1 is uniquely defined by its state in the previous time period t and the inputs to the system at time t with a time step of one minute. This model produces with correct and complete data a result that only differences 10% of the real travel times. But there can be corrupt or missing data.
The travel time prediction framework consists of a preprocessing layer and the actual travel time prediction model. Besides the instantaneous data along the link or route of interest also the information on traffic conditions on up- or downstream links is required. The state space neural network model is a data driven model that is capable of learning the non-linearity of congested traffic from data. The state space neural network is derived from state spaced formulation of the travel time prediction problem, and is capable of incorporating the spatio-temporal dynamics of traffic flow implicitly, yielding not only accurate travel time predictions, but also plausible internal states,

which are related to the traffic processes at hand. The hidden nodes each hidden neuron represents a link along the route. On a route of K links,
there are K hidden neurons. Neural network have the advantage not to be very sensitive to noise as some bias in its inputs.
The data comes from a simulation network (FOSIM) to control the (non-)robustness of the data. The first and one of the most straightforward strategy in handling faults and incompleteness in the input data involves only the preprocessing layer, such that failures are 'repaired' or smoothed. The second strategy is to feed the network with a sufficient amount of corrupted data, but this increases the non-linearity of the problem to be learned by the model. The third strategy is to make several models for various types of failures in the input data, but this yields a lot of models. The study is about a combination of the first two strategies.
This combination produces a fairly robust system. But if the failures occur on the significant links (on the ramps and surrounding the bottleneck) the system isn't that robust anymore. More enhanced preprocessing may be a solution.

Comment
This paper is mainly about robustness of a system on missing and corrupt data. This system can predict travel times using data of one minute before with an accuracy of 10% using robust data. It only uses data measured along the highways and on the up- and downstream links. All other information that may influence travel times, like the kind of weather, is ignored. With non-robust data it is stated that a combination of pre-processing (input-repairing) and a SSNN model for travel time prediction seems a promising approach toward robust travel time prediction, but the accuracy is never mentioned.
However nothing is said about predicting the data measured along the highways and on the up- and downstream links. If that data can be predicted a certain time in advance it is possible to determine the travel times before the driver gets to that part of a road and can use the information to take another route if the delay is to large.

Robust and adaptive travel time prediction with neural networks

Ir. J.W.C. van Lint, Dr. ir. S.P. Hoogendoorn, Prof. ir. H.J. van Zuylen
Proceedings of the 6th annual TRAIL Congress (part 2), December 12th 2000.

Summary
Much research is now done to the potential effect of an advanced traveler information systems (ATIS) on route choice. The provided information can have a positive effect on the experienced traffic conditions and can lead to a more efficient use of the available network infrastructure. But improved conditions can lead to an increase in travel-demand and thus traffic load. Traffic information has an effect on driver behavior if it consists of clear and unambiguous messages based on accurate and reliable predictions.
The information provided must be direct information. For travelers this means travel times instead of queue lengths. A robust prediction should be made to make it useful and it is assumed that this can be achieved by efficient combination of different data sources, ranging from inductive loop data, floating car data to weather forecasts, using data fusion techniques. There has been some investigation to the influences of weather on travel times, but further research is necessary.

There are many possible sources of data such as the data from inductive loops, traffic control, and ramp metering systems. There are two main approaches to travel time predictions that can be distinguished. Travel time prediction using statistical or inductive models, like ANN's (Artificial neural networks) and travel time prediction on the basis of analytical or deductive models. ANN's are a class of mathematical models capable of approximating any nonlinear mapping. ANN's are related to modern methods in statistics. Multi-layer perceptions used for classification or prediction can be interpreted as generalized non-linear classifiers or regression models. Neural networks are, however, more suited for modeling complex nonlinear problems. There are many optimization techniques for training a neural network. But there are two phenomena that yield poor generalization if not properly dealt with (curse of dimensionality and bias-variance tradeoff). Feature extraction reduces the complexity of the model and early stopping and cross-validation forces better generalization for neural network and statistical models. Considering the requirements for travel time prediction, ANN's seem a natural modeling choice, because they are adaptive systems and they are robust to missing or faulty input data.
There have been studies to predict traffic flow related parameters using ANN's. But there are only a few studies to predict traffic time or link speeds using ANN's. These studies look however very promising, so further study is justified. A distinction can be made between direct and indirect travel time predictions. Direct prediction with ANN's predicts the travel time directly from the available data. Indirect prediction with ANN's first predicts traffic flow conditions and calculates subsequently the travel times.
By predicting traffic conditions at first a model is designed. Different models were tried always predicting the traffic flow conditions using the traffic flow conditions one or more time steps before. After that a neural network and a training algorithm is chosen. The longest time ahead for which a prediction was made was ten minutes. The prediction had an accuracy of 14%. Direct prediction requires the same approach. A model must be made and a neural network and training algorithm must be chosen. The models of the direct approach predict the travel times based on the current speed and intensity measurements one or more time steps before or on the previous travel times. It has been tried to predict travel times this way up to 5 time intervals ahead. The best performing modular design was a Fuzzy C means clustering algorithm with 12 clusters feeding a feedforward ANN.

Comment

In this paper an overview of research to travel time prediction using neural networks is presented. In this paper weather conditions are mentioned as a factor to make a more robust prediction. But further research should be done. The most successful system described, is capable of predicting the link travel times up to 5 time intervals of five minutes ahead with a feedforward ANN feeded by a Fuzzy C-means clustering algorithm. It is not stated how accurate this prediction was and whether this system is capable of predicting the link travel times over more than five time intervals. The data used, was only data measured along the links. Other factors that may influence the situation on the road are not taken into account.

Short-term prediction of Traffic Flow Conditions in a Multilane Multiclass Network

Serge P. Hoogendoorn, Hans van Lint
Dept. of Transportation and Traffic Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology
Full paper to be presented at the TRISTAN IV conference, June 2001, Ponta del Gado (The Azores, Portugal)

Summary

There are multiple user classes in traffic, like vehicle-types, traveler types, paying and non-paying traffic etc. These different classes are important for the traffic control options. Because of the many control options there is a need for a decision support system or an automated traffic control system. But the existing dynamic network models do not distinguish between user-classes and roadway lanes or are not suitable for decision support or model-based control.

HELENA is a multilane-multiclass traffic network simulation model that can forecast traffic operations in multi-route networks. HELENA is capable of distinguishing different lanes and user-classes. Nodes describe which lanes of the incoming links connect to which lanes of the outgoing links. The traffic dynamics on these links are described by multi class multilane difference equations.

The multi-class flow model for network links consist of the multi-class conservation of vehicle equation and a dynamic equation describing the dynamics of the multiclass heterogeneous traffic flow. There are a few possible processes. For all these processes there is an equation.

It is necessary to distinguish (at least) three user classes. The flow conditions of the roadway lanes result from interactions between the different vehicles.

In HELENA, nodes involve merging or diverging. Two multilane links merge into a single multilane link. It will involve dropping one or more lanes. In HELENA this is modeled by using virtual vehicles. Because of these virtual vehicles, the average speed on the dropped lane will decrease and the vehicles are changing lanes towards the remaining lane. HELENA uses class-specific predictions of the fraction of the total traffic demand of the ingoing link for node *n* en class *c* that chooses outgoing link *a* at instant *t*. In practice, a historic database of destination-unspecific splitting rates or route-choice models are used to forecast the route-choice behavior. Diverge nodes describe how the lanes of the incoming link connect to the lanes of the outgoing links. The virtual vehicles will again be used to get the traffic on the correct lane of the incoming link.

Various traffic control measures will be modeled in HELENA. The reaction to information provision will be included.

HELENA can be used as an offline simulation tool and as an online tool for multiclass traffic forecasting. In HELENA, a road network is modeled as a directed graph. The edges consist of a set of lanes and divided in cells. The modeling approach used is the object oriented modeling approach. This allows a maximum flexibility and ensures consistency in both attributes and behavior of all network components. In HELENA, not a single relation of the inheritance type is included. This makes the storage of the network in a relational database simpler.

Comment
Predicting the actions of traffic is the subject of this paper, not predicting travel times. It is about how lane- and class-specific traffic will behave, change lanes and so on. Also route-choice behavior, with or without additional information, like traffic congestion, is forecasted. This is done with a multilane-multiclass traffic network simulation model. No neural network is used.

Freeway Travel Time Prediction with State-Space Neural Networks

ir. J.W.C. van Lint, dr. ir. S.P. Hoogendoorn, prof. dr. H.J. van Zuylen
Transportation Planning and Traffic Engineering Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology
Accepted for presentation & publication at TRB 2002

Summary
In this paper a new approach to freeway travel time prediction is presented, based on recurrent neural networks. Travel time prediction requires a modeling approach that is capable of dealing with complex non-linear spatio-temporal relationships between flows, speeds and densities. Based on literature, feed-forward neural networks are a class of mathematical models that are well suited to solve this problem. A drawback of the feed-forward approach is that size and composition of the input time-series are inherently design choices and thus fixed for all input. This may lead to unnecessary large models. Moreover, for different traffic conditions different sizes and compositions of input time-series may be required, a requirement that is not satisfied by any feed forward data-driven method.
The recurrent neural network topology presented in this paper is capable of dealing with the spatio-temporal relationships implicitly. This neural network consists of two trainable layers of neurons (hidden- and output-layer), and two layers which simply distribute their activations to the hidden layer (input and context layer). The topology of this neural net is derived from a state-space formulation of the travel time prediction problem, which is in line with traffic flow theory. The performance of several versions of this state-space neural network is tested on synthetic data from a densely used highway stretch in the Netherlands. The neural network models tell that the state of a system at time t+1 is uniquely defined by its state in the previous time period t and the inputs to the system at time t with a time step of one minute. They were capable of accurately predicting experienced travel times, producing approximately zero mean normally distributed residuals, rarely outside a range of 10% of the real expected travel times. Moreover, analyses of the internal states and the weight configurations revealed that the neural networks were able to develop an internal model that is linked to the underlying traffic processes.

Comment
A recurrent neural network is used to predict the travel times. The predictions are accurate for travel times using data of one minute before. But it isn't possible to predict travel times before arriving to the part of the road for which the travel time is predicted. So it is not possible to use the travel times in a route planner to lead the driver over the fastest route. However it can be useful to use the data produced by this system to recalculate the travel times along the way.

State Space Neural Networks for Freeway Travel Time Prediction

ir. J.W.C. van Lint, dr. ir. S.P. Hoogendoorn, prof. dr.ir. H.J. van Zuylen
Traffic Engineering and Transportation Planning Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology
Accepted for presentation at the INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, ICANN 2002, August 27-30, 2002, Madrid, Spain

Summary
The highly non-linear characteristics of the freeway travel time prediction problem require a modeling approach that is capable of dealing with complex non-linear spatio-temporal relationships between the observable traffic quantities. Based on a state-space formulation of the travel time prediction problem, we derived a recurrent state-space neural network (SSNN) topology. The SSNN model is capable of accurately predicting experienced travel times, producing approximately zero mean normally distributed residuals, generally not outside a range of 10% of the real expected travel times. The hidden nodes each hidden neuron represents a link along the route. On a route of K links, there are K hidden neurons. This model is capable of accurately predicting experienced travel times, producing approximately zero mean normally distributed residuals, generally not outside a range of 10% of the real expected travel times. Furthermore, analyses of the internal states and the weight configurations revealed that the SSNN developed an internal models closely related to the underlying traffic processes. This allowed us to rationally eliminate the insignificant parameters, resulting in a Reduced SSNN topology, with just 63 adjustable weights, yielding a 72% reduction in model size, without loss of predictive performance.

Comment
This paper is similar to the paper above. It states that using the SSNN model the experienced travel times can accurately predicted giving a result generally not outside the range of 10% of the real expected travel times. But again it is not possible to predict these travel times a certain time ahead. Because in that case the input parameters of the model should be known that certain time in advance instead of only one minute. That means that the input parameter must be predicted.

Determination and evaluation of traffic congestion costs

Ingo Hansen
Traffic Engineering and Transportation Planning Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology
EJTIR, 1, no.1 (2001), pp. 61-72

Summary
The economic loss by traffic congestion is very large. The people want the government to do something about it by creating new motorway or increasing road capacity. But the actions to reduce congestions have negative impacts. The time loss of the vehicles on the road by congestion isn’t very accurate. The calculation of the cost is done by applying different values of time for the different categories of traffic. But it is not very easy to split the traffic in the different categories. The overall costs and benefits for suitable measurements for improvement of the traffic conditions on a regional scale are often determined unsatisfactorily. The long-term investment cost for increasing the capacity seems

underestimated and the long-term benefits of new roads are overestimated to get the project approved.
Time loss due to traffic congestion is determined by comparing the average trip time with the trip time under free flow conditions. In the city it isn't possible to determine the trip time under free flow conditions because there is always congestion. On motorways congestion is monitored regularly. Observing the queue length does this. The observation of queue length by human beings is not very accurate because of the limited frequency. Nowadays traffic congestion is determined by computing the flow and speed data. The estimated loss of time during traffic congestion is very sensitive to the difference in volume and speed realized.
There will be an investment in the road transport net if the cost for the investment is lower than the total social benefits. But the social benefits do not have a monetary value. Congestion can lead to loss of work time, but in most cases the work times are flexible and the time loss due to congestion reduces a persons free time. So the economic loss isn't very large. But the congestion cost of freight transport is much more. Most of the time loss is due to infrastructure bottlenecks and occur daily at peak times. But besides the costs of congestion there will be other costs when building new motorways. These other costs are due to air pollution, accidents and space occupation.
Increasing the capacity of the congested roads is not the solution. The need for more capacity is growing faster then the capacity of the road can grow. Other solutions must be searched. Road pricing and congestion pricing can be introduced.

Comment

In this paper not the traffic congestion is predicted, but the costs of traffic congestion are determined. It is an interesting paper because also a few options to diminish traffic congestion are mentioned.

Different Perspectives on the Global Development of Transport

Barry Ubbels, Caroline Rodenburg and Peter Nijkamp
Department of Spatial Economics, Free University Amsterdam, The Netherlands
EJTIR, 1, no. 1 (2001), pp. 9 – 28

Summary

The future of the transport sector is clearly fraught with uncertainties, as the system is influenced by many factors that can develop in various ways. The aim of this paper is to get insight into the future development of the transportation sector seen from a worldwide perspective. This is done by applying a scenario approach and by designing four possible development paths for transport. These future developments are presented here by sketching four global contrast images based on outcomes of earlier research. The outcomes for the transport sector, expressed in transported volumes for both passenger transport and freight transport, are first described qualitatively, based on expected developments of several indicators. Next, based on those descriptions, quantitative numbers of transported volumes are calculated with the year 2020 as the time horizon. It appears that all scenarios foresee a worldwide growth in transport volumes (for both passengers and freight). The achievement of sustainable mobility based on the outcomes in this paper may seem difficult and

will be confronted with several hurdles. But policy changes and (unexpected) technology developments offer possibilities to realize this objective. Thus policy makers are faced with formidable policy challenges to achieve the Kyoto objectives in the next 20 years.

Comment
This paper is not about predicting travel times, but it describes the development of transport. Different scenario's where used to show the possibilities. These scenarios are combined to come to a prediction of the transport volumes in 2020. If the transport volume will grow more traffic congestion can be expected if nothing will change.

Using stated preference data for studying the effect of advanced traffic information on drivers' route choice

Mohamed A. Abdel-Aty, Ryuichi Kitamura, Paul P. Jovanis
Transportation Research Part C: Emerging Technologies v 5C n 1 Feb. 1997, p 39-50

Summary
In 1992 a computer-aided telephone interview was conducted to investigate how much information drivers have about their routes, their awareness of alternate routes, their awareness of traffic conditions, which could effect their route choices, and their use of available traffic information either en route or pre-trip or both. A second interview in 1993 was meant to probe further into drivers' route choice behavior, to measure any changes within the last year, to investigate commuters' attitudes and perceptions regarding several commute characteristics, and to understand the effect of travel time variation on route choice. There was also a mail-back follow-up survey. This was about which route attributes are considered important by the individual in the decision process that leads to the choice of a route, commuters' familiarity with their highway/streets networks, and its potential effect on route choice, commuters' willingness to use advanced traveler information systems (ATIS), and the effect of advanced traffic information systems on route choice. Each respondent was provided with three scenarios, in each, he had to choice between two routes. Two route choice models are estimated. The first model uses the five hypothetical binary choice sets collected in the telephone interviews. The objective of the model is to determine how travel time variation affects route choice and the potential interplay among travel time variation, traffic information acquisition and route choice. The second model uses the data from the mail survey. The objective of the model is to investigate the potential effect of advanced traveler information systems on route choice. The correlation among error components in repeated measurement data is addressed with individual-specific random error components in a binary logit model with normal mixing distribution.
There seems to be a strong indication of the importance of travel time reliability for route choice. So travel time is not the dominant route choice criterion, ATIS should avoid advising commuters to use uncertain routes even if on the minimum path, there is a need to develop reliability measures as part of ATIS, and there is a need to investigate whether to display reliability information to drivers.

Comment
Here is investigated how people determine there route, what factors play a role in this and does advanced traffic information influence the choice. Predicted travel times, if accurate and reliable, can be a part of advanced traffic information.

Spectral Basis Neural Networks For Real-Time Travel Time Forecasting

D. Park, L.R. Rilett, G. Han
Journal of Transportation Engineering, 1999/11. 125(6), p. 515-523

Summary
Artificial neural networks (ANN) and in particular multilayer perceptron neural networks that utilize a back-propagation algorithm have been applied successfully for forecasting link travel time and other traffic parameters. Spectral basis neural networks combine a pre-transformation of the input features and a conventional ANN for forecasting multiple-periods link travel times into the future. The pre-transformation is based on sinusoidal functions and is performed on the input features to obtain linearly separable input features. In literature is said that a conventional ANN works efficiently when the function that is approximated is relatively monotonic with only a few dimensions of the input features and when the features are linearly separable. There are two techniques to reduce these problems. The first is input-partitioning and the second input-transformation.
The spectral basis neural network (SNN) incorporates an input-transformation. With this network classification errors can be reduced and the performance of the ANN can be improved. The hard vectors are transformed using a nonlinear transformation, such that the input features become simpler to classify. Here, the transformation is based on a sinusoidal transformation, but any orthonormal (orthogonal) basis function can be used.
The average travel times of the target link in the preceding time periods are important parameters for identifying future travel time patterns. Also the link travel times experienced on the upstream and downstream links during the preceding time periods are important.
There was a SNN with one hidden layer tested and one with two hidden layers. More hidden layers mean more ANN capacity, but the network will also be slower. The SNN with two hidden layers produced the best results. Better than other travel time forecasting methods like Kalman filtering. This method gave the best results for one period ahead, but the SNN gave good results for two to five periods ahead.

Comment
In this paper spectral basis neural networks are used to predict travel times less than 25 minutes in the future. The data used to train and test the network, was the data measured along the way. No additional information was used. This network was not capable of predicting travel times two hours in the future.

2.3 Conclusion

The literature points out that there is need for travel time prediction and especially for reliable predictions.
In this same literature there is a lot of information about predicting travel times, but they give a solution for a prediction for only 10-25 minutes in the future. To make this prediction only the traffic measurements along the way and on ramps are used. But it is also mentioned that, for example, the weather will influence the values of these measurements. There has been some research of these influences, but it is also stated that further research on that area is necessary. If these influences are noticeable then by using the traffic measurements along the road the conditions on the road, for example the weather, are indirectly taken into account while predicting travel times. What will happen if these conditions are used as real, hard parameters? Maybe it will make it possible to predict travel times more than 25 minutes in the future.
In some of the papers a system is described predicting travel times using the average speed of one minute before. If the average speed can be predicted a certain time t ahead than this system can be useful to predict the travel times time t +1 minute ahead.
In Table 2.2 the degree of connection to predicting travel times of the literature is stated. The literature can be about predicting, about travel times or about predicting travel times. The literature with a plus in the last table is most relevant to the subject. The literature numbers are the numbers mentioned in Section 2.2.

Table 2.2 Degree of connection to predicting travel times

Literature number	Predicting	Travel times	Predicting travel times
1	-	-	-
2	+	+	+
3	+	+	+
4	+	-	-
5	+	+	+
6	+	+	+
7	+	+	-
8	+	-	-
9	-	-	-
10	+	+	+

The papers with number 1, 5, 6 are very related to each other. They have the same author and are about the same study, but in every paper the point of view is a little bit different. Paper number 2 is an overview of a few papers about travel time prediction. Prediction of travel times is also the subject in paper 10. In this paper another kind of neural network is used.
From this literature can be concluded that the problem of predicting travel times can be solved using different model types:

1. Analytical model
2. Basic feedforward neural network
3. Special neural network, like a recurrent neural network, a spectral basis neural networks or a state space neural networks

The analytical model requires explicit modeling of the process, which becomes even more difficult, when information like the weather conditions is used, because of the uncertainty of this data. The different neural networks all have there advantages and disadvantages. But from the papers can be concluded that most research is based on the use of ANN's for this kind of problems. A special neural network, like a recurrent neural network, will probably give the best results, but it will make the model somewhat more complicated. To have a relatively simple model just to try out a new theory a basic feedforward neural network can be sufficient.

Chapter 3

Possible influences and their data description

As pointed out in the introduction the goal is to predict the average speed on a certain time and date on a special location of a road. This is not a straightforward problem as may be clear from the literature in the previous chapter. In order to make a prediction, information about factors that influence the average speed is necessary. But what are these factors? One can expect that the following factors have an impact on the estimated average speed. The most common are:

- Time
- Date
- The day of the week
- The weather
- Events
- Holidays
- Accidents
- Incidents

Of course also the information about the situation on the road and the place on the road itself is needed.
In the next chapter the influence of these features on the situation on the road is analyzed. In this chapter the sources the data was obtained from and the format of the data of the time, the date, the day of the week, the weather, the events and the holidays are discussed. Also the structure of the data about the situation on the road is given here. Accidents do influence the travel times too but information about them is not known in advance. If they would be, then maybe they can be prevented. Incidents will contain roadblocks, pruning activities on the edge of a road and drawing lines on the road. Incidents do not occur very often and therefore not much data samples are available for this situation. So we do not include incidents in our research. Accidents will not be discussed in a separate section, but it will come forward in the other sections.

3.1 The date and time

Besides the fact that the date and time are factors on which the situation on the road itself depends, it is also clear that all the influencing factors summed up in the section above are correlating with date and time. So the date and time are important information. The date can consist of the following information:

- The year
- The month
- The day

Time can be split up in the next three parameters:

- The hour
- The minute
- The second

In Figure 3.1 an example of date and time is given.

```
01-07-2002 7:00:00
```

Figure 3.1 Date and time

3.2 The day of the week

Using the date the day of the week can be found in two ways. To obtain the day of the week one can calculate it using the current day and day of the week. Another method is to use a calendar. For the example in Figure 3.2 the calendar is shown in Figure 3.2.

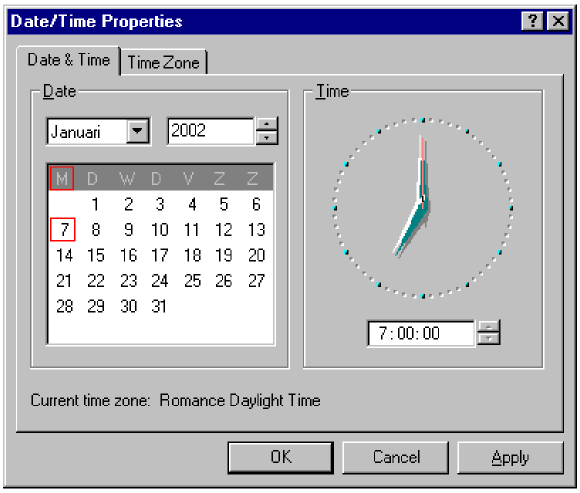


Figure 3.2 Calendar of the 7th of January 2002

From this calendar can be read that the day of the week on the 7th of January 2002 was a Monday.

3.3 The weather

Weather circumstances can also influence the average speed, like fog, rain and black ice. Fog and rain will reduce the visibility and black ice will make the road slippery. This results in slower driving and even accidents. Therefore weather information can be useful by predicting travel times. There are several websites that give information about the weather nowadays. In this section two websites that give this kind of information are described. The first is [Weeronline] and the second is [KNMI]. The websites both give a weather prediction and information about the past weather. [Weeronline] also gives information about the current weather.

3.3.1 The past weather

Weeronline

The information about the past weather on [Weeronline] consists of:

- The place
- The temperature
- The type of weather.

The information is available for every half an hour starting on the 31st of January 2000 at 0.30 AM. The place is telling where the measurement had taken place. Every half an hour the information of a number of places is given, but not all of the places. Therefore it is possible that the measurements of a certain place are not available every half an hour. At every place two measurements are done, the temperature and the type of weather. The type of weather is not really a measurement, but more somewhat like a classification. In plain text is stated what the weather looked like. What kind of types can be possible is shown in Appendix A. Table 3.1 gives an example of types of weather. These types of weather are practically the same weather. The temperature is given in degrees Celsius. An example of the information about the past weather of [Weeronline] is stated in Table 3.2.

Table 3.1 Types of weather

Weinig / geen bewolking
Vrijwel onbewolkt
Vrijwel onbewolkt, rook
Vrijwel onbewolkt, nevel

Table 3.2 The information of the weather on the 1st of July 2002 7:00 AM from [Weeronline]

actueel weer : ma, 01-07 07:00 MEZT		
	temp.	**weer**
Amsterdam-Schiphol	14°	motregen
De Bilt	12°	-
Deelen	12°	geheel bewolkt
Den Helder	14°	bui
Eindhoven	12°	lichte regen
Ell	13°	-
Euro Platform	14°	-
F3 (Platform)	13°	-
Gilze-Rijen	12°	lichte regen
Groningen-Eelde	12°	bewolkt
Hoogeveen	13°	geheel bewolkt
Hupsel	13°	-
K13-A	14°	-
Lauwersoog	13°	-
Leeuwarden	13°	-
Lelystad	13°	-
Maastricht-Beek	14°	lichte regen
Meetpost Noordwijk	14°	-
Rotterdam Luchthaven	13°	motregen
Soesterberg	12°	motregen
Terschelling Hoorn	14°	-
Twenthe	13°	bewolkt
Valkenburg	14°	lichte regen
Vlieland	14°	-
Vlissingen	13°	lichte regen
Woensdrecht	12°	-

Table 3.3 The information about the weather in July 2002 from [KNMI]

STN	YYYYMMDD	DDVEC	FG	FHX	FX	TG	TN	TX	SQ	SP	DR	RH	PG	VVN	NG
260	20020701	215	55	70	150	135	120	150	5	3	54	95	10034	19	8
260	20020702	217	53	70	160	141	121	186	41	25	27	28	10015	59	8
260	20020703	202	46	60	120	141	121	172	4	2	166	144	9988	47	8
260	20020704	275	39	50	120	156	123	194	105	63	29	27	10133	56	6
260	20020705	203	32	50	90	157	132	204	7	4	0	-1	10127	56	7
260	20020706	249	24	40	90	161	116	198	11	7	0	-1	10142	23	8
260	20020707	209	27	50	90	157	116	193	7	4	6	4	10165	30	8
260	20020708	194	27	50	90	187	110	243	78	47	0	0	10118	21	4
260	20020709	159	26	50	100	190	143	256	46	28	36	134	10075	3	6
260	20020710	211	24	40	70	146	109	172	2	1	81	35	10111	40	7
260	20020711	216	35	60	110	151	100	208	111	68	0	-1	10181	30	3
260	20020712	85	18	30	60	166	85	215	44	27	0	-1	10204	3	6
260	20020713	341	19	30	60	174	131	211	28	17	27	168	10174	23	7
260	20020714	341	35	50	90	188	157	230	63	39	0	0	10229	38	6
260	20020715	8	35	50	90	200	163	244	114	70	0	0	10218	50	4
260	20020716	15	41	60	120	198	144	249	106	65	0	0	10179	58	3
260	20020717	5	39	60	110	184	141	230	87	54	0	0	10175	58	5
260	20020718	338	41	50	110	160	97	202	93	57	0	0	10187	67	5
260	20020719	325	19	30	60	149	89	192	4	2	0	0	10191	3	7
260	20020720	156	22	40	100	157	69	226	77	48	18	64	10157	5	6
260	20020721	279	38	60	130	154	134	182	65	40	32	52	10135	14	6
260	20020722	269	35	50	100	164	128	201	65	41	0	0	10176	66	7
260	20020723	232	44	60	130	172	154	207	1	1	43	42	10115	37	8
260	20020724	303	31	50	90	166	120	200	41	26	1	1	10133	50	7
260	20020725	258	28	40	90	161	95	206	42	26	10	4	10153	3	6
260	20020726	256	31	40	90	191	140	243	76	48	28	11	10163	7	4
260	20020727	177	20	30	70	207	131	272	102	65	0	0	10190	1	3
260	20020728	136	18	30	60	230	157	297	130	83	0	0	10188	5	0
260	20020729	53	21	40	70	246	161	310	133	85	0	0	10176	43	1
260	20020730	106	21	70	170	239	160	333	118	75	26	159	10108	2	3
260	20020731	279	19	30	60	213	172	271	32	21	2	1	10085	16	6

KNMI

The KNMI (Royal Netherlands Meteorological Institute) gives other information about the past weather than [Weeronline]. This information has a frequency of a day starting on the first of January 1991. The following measurements are done:

- Station number (STN)
  - (235 = De Kooy, 260 = De Bilt, 280 = Eelde, 290 = Twenthe, 310 = Vlissingen, 380 = Maastricht)
- Date (YYYYMMDD)
  - YYYY means the year, MM gives the month and DD gives the number of the day
- Prevailing wind direction in degrees (DDVEC)
  - (360 degrees means north, south is indicated by 180 degrees, west by 270 degrees and east by 90 degrees. 0 means calm/variable wind)
- Daily mean wind speed in 0.1 m/s (FG)
- Maximum hourly mean wind speed in 0.1 m/s (FHX)
- Maximum wind gust in 0.1 m/s (FX)
- Daily mean temperature in 0.1 degrees Celsius (TG)
- Minimum temperature in 0.1 degrees Celsius (TN)
- Maximum temperature in 0.1 degrees Celsius (TX)
- Sunshine duration in 0.1 hour (SQ)
  - (-1 if its less than 0.05 hour)
- Percentage of maximum possible sunshine duration (SP)
- Precipitation duration in 0.1 hour (DR)
- Daily precipitation amount in 0.1 mm (RH)
  - (-1 if its less than 0.05 mm)
- Daily mean surface air pressure in 0.1 hPa (PG)
- Minimum visibility (VVN)
  - (If VVN is 0 than the minimum visibility is less than 100m. The number 1 means a minimum visibility of 100-200m, 2 of 200-300m, ..., 49 of 4900-5000m, 50 of 5-6km, 56 of 6-7km, 57 of 7-8km,..., 79 of 29-30km, 80 of 30-35km, 81 of 35-40km,..., 89 of more than 70km)
- Cloud cover in octants (NG)
  - (The number 9 means that the sky is invisible)

In Table 3.3 an example is given of this information.

3.3.2 The weather forecast

Weeronline

[Weeronline] gives a forecast for three days for most places in Holland and also a lot of places in the rest of the world. Besides the weather forecast also a trend for the three days after that is given. In this section only the prediction for the next three days will be considered. This weather forecast knows the following parameters:

- The day for which the forecast is
- The place for which the forecast is
- The maximum temperature in degrees Celsius for each day
- The minimum temperature in degrees Celsius for each day
- The weather in the morning for each day in plain text
  - For example clouded, rain, mostly sunny
- The weather in the afternoon for each day in plain text
  - For example clouded, rain, mostly sunny
- The weather in the evening for each day in plain text
  - For example clouded, rain, mostly sunny

This information is given in Table 3.4 for the 11th of June 2003 for the location Gouda.

Table 3.4 The weather forecast from [Weeronline] given on the 11th of June 2003

de verwachting voor de regio/plaats: Gouda			
	Wo, 11-06	do, 12-06	vr, 13-06
minimum temperatuur	15°C	12°C	10°C
maximum temperatuur	21°C	21°C	20°C
ochtend	overwegend zoning	overwegend zoning	overwegend zonnig
middag	overwegend zonnig	half bewolkt	overwegend zonnig
'avonds	Geen bewolking	geen bewolking	geen bewolking

Besides the weather forecast also the wind prospect is given. This prospect is also for the next three days. It consist of the following information:

- The day for which the prospect is
- The place for which the prospect is
- The wind in the morning for each day
  - The wind direction (N (north), S (south), W (west) or O(east)) and the wind speed in Beaufort
- The weather in the afternoon for each day
  - The wind direction (N (north), S (south), W (west) or O(east)) and the wind speed in Beaufort
- The weather in the evening for each day
  - The wind direction (N (north), S (south), W (west) or O(east)) and the wind speed in Beaufort

In Table 3.5 an example of the wind prospect is given for the 11th of June 2003 for the place Gouda.

Table 3.5 The wind prospect from [Weeronline] given on the 11th of June 2003

de windverwachting voor de regio/plaats: Gouda			
lokale tijd	wo, 11-06	do, 12-06	vr, 13-06
's morgens	ZW 4	Z 2	ZW 2
's middags	W 3-4	W 3	W 3-4
's avonds	NW 2	NW 2-3	NW 3

KNMI

[KNMI] only gives a weather forecast in plain text instead of a table like [Weeronline]. Therefore the elements of which the forecast consist of must be subtracted from the text. The following elements can be found in the text:

- The minimum temperature in daylight of the day from which the prediction was
- The maximum temperature in daylight of the day from which the prediction was
- The wind speed of the day from which the prediction was
- The wind direction of the day from which the prediction was
- The weather of the day from which the prediction was
    - The weather as it was that day (for example clouded, rain, mostly sunny)
    - The weather to come that day (for example clouded, rain, mostly sunny)
- The weather of the evening and night of the day from which the prediction was
- The minimum temperature at night of the day from which the prediction was
- The weather of the next day
    - For example clouded, rain, mostly sunny
- The maximum temperature in daylight of the next day
- The wind direction of the next day
- The wind speed of the next day
- The weather for the days after the next day
    - For example clouded, rain, mostly sunny
- The temperature in the afternoon for the days after the next day

An example of the text from which these elements can be extracted can be seen in Figure 3.3.

```
PERIODEN MET ZON, MORGEN EEN ENKELE BUI

De dag begon vandaag fraai met veel zon, maar in de loop van de
ochtend is er stapelbewolking ontstaan die soms de zon
afschermt. Verder waait het stevig uit zuidwestelijke richting,
boven land meest matig, kracht 4, langs de kust vrij krachtig,
kracht 5. Vanmiddag wordt het geleidelijk aan beter. Er komt
steeds meer ruimte voor de zon doordat de bewolking oplost.
Bovendien neemt de wind in kracht af. De middagtemperatuur
loopt uiteen van 18 graden langs de kust tot plaatselijk 23
graden in het oosten van het land.

Vanavond en vannacht is er weinig bewolking, alleen in het
zuiden drijven wat wolkenvelden over. Het koelt bij weinig wind
tot ongeveer 12 graden.

Morgen zijn er perioden met zon en vooral in de oosten en
zuidoosten van het land is er in de loop van de dag kans op een
regen- of onweersbui. De maximumtemperatuur loopt uiteen van 19
graden vlak aan zee tot 27 graden in het zuidoosten van het
land. De wind komt eerst uit uiteenlopende richtingen, maar
wordt in de loop van de dag westelijk en matig van kracht.

De dagen daarna zijn er flinke perioden met zon en blijft het
op de meeste plaatsen droog. De middagtemperatuur ligt met 22
graden op een aangenaam niveau.
```

Figure 3.3 The weather forecast on the 11th of June 2003 from [KNMI]

3.3.3 The current weather

Weeronline

On the website of [Weeronline] also the current weather situation is given. There are two different ways of displaying the weather, every hour or every six hours. With the option 'every hour' the results as given in Table 3.6 are given.

Table 3.6 The current weather every hour on the 11th of June 2003 2:25 PM on Rotterdam Luchthaven from [Weeronline]

actueel weer: Rotterdam Luchthaven		
	Temperatuur	weer
wo, 11-06 14:25	20°C	licht bewolkt
wo, 11-06 13:55	20°C	licht bewolkt
wo, 11-06 13:25	19°C	licht bewolkt
wo, 11-06 12:55	19°C	half bewolkt

Of every half hour from the past two hours the temperature and the kind of weather are given. This information is the same information that [Weeronline] gives for the current weather.
The option 'every six hours' gives the same information as the option 'every hour', but not from every half hour in the past two hours. The information goes back to four times six hours. So the information is given of a whole day in time intervals of six hours. An example is shown in Table 3.7.

Table 3.7 The current weather every six hours on the 11th of June 2003 2:00 PM on Rotterdam Luchthaven from [Weeronline]

actueel weer: Rotterdam Luchthaven		
	temperatuur	weer
wo, 11-06 14:00	19°C	half bewolkt
wo, 11-06 08:00	17°C	licht bewolkt
wo, 11-06 02:00	17°C	bewolkt
di, 10-06 20:00	20°C	vrijwel onbewolkt

KNMI

[KNMI] gives the past weather for the current day in its weather forecast. It doesn't really say what kind of weather it is on the moment of watching. The forecast can be from hours before that moment and therefore also the current weather is the weather from hours ago.

3.4 The events

Events attract a lot of people. Some will go by train or other means of public transport, but also a lot of people will take the car to visit events. Therefore it may be helpful to know which events occur on a certain day and time. There are two ways to find out which events will take place. The first one is to check the websites of all large places where events could take place, like the Jaarbeurs in Utrecht, the RAI in Amsterdam, the Amsterdam Arena in Amsterdam, the Heineken Music Hall in Amsterdam and the Gelredome in Arnhem. There are also a few more general websites, which announce large events or sell tickets for those events. Checking these websites is a second option.

The websites of large places

Only two examples of websites will be given here. The [Amsterdam Arena] will be discussed and the [Heineken Music Hall].

After starting the website [Amsterdam Arena] a link becomes visible that shows the events in the upcoming period. The list of events displayed after using the link is visible in Figure 3.4.

14 June 2003	Panorama Mega Barbeque and Concert Blof, 16.00 hours More
17 June 2003	Eminem, 19.00 hours More
18 June 2003	Eminem, 19.00 hours More
21 June 2003	André Hazes, 20.30 hours More
23 June 2003	André Hazes, 20.30 hours More
28 June 2003	Kane, 19.30 hours More
5 July 2003	Sensation The White Edition, sold out! More
12 July 2003	Sensation The Black Edition More
17 July 2003	Robbie Williams, SOLD OUT! More
18 July 2003	Robbie Williams, SOLD OUT! More
1 August 2003	The Sony Amsterdam Tournament , Galatasaray - Internazionale, 19.00 hours More
1 August 2003	The Sony Amsterdam Tournament, Ajax - Liverpool, 21.15 hours More
3 August 2003	The Sony Amsterdam Tournament, Galatasaray - Liverpool, 18.30 hours More
3 August 2003	The Sony Amsterdam Tournament, Ajax - Internazionale, 20.45 hours More
19 August 2003	Rolling Stones SOLD OUT More
20 August 2003	Rolling Stones More

Figure 3.4 The events on [Amsterdam Arena]

An overview of a certain event can be obtained by using the link ‘more’. The overview given by doing that for the event of 18th July 2003, the concert of Robbie Williams, is shown in Figure 3.4.

This overview contains the following information:

- The place where the event takes place
- The date on which the event takes place
- The time on which the event starts
- The costs of the tickets for the event
- The place where tickets can be obtained
- The announcement of the concert itself

Robbie Williams, 18.30 uur, UITVERKOCHT! (18 juli 2003)

Robbie Williams geeft volgend jaar een concert in de Amsterdam ArenA. Dit heeft hij zojuist tijdens een persconferentie in Berlijn bekend gemaakt. Op donderdag 17 juli zal hij voor het eerst een stadion in Nederland veroveren. De kaartverkoop voor dit concert start aanstaande zaterdag 23 november.
In maart 2001 gaf het 'enfant terrible' van de hedendaagse popmuziek twee uitverkochte concerten in Ahoy' Rotterdam. Deze concerten waren zeer overtuigend en Williams maakte de tekst van zijn openingsnummer 'Let Me Entertain You' meer dan waar. Naast een goede zanger bewees Williams ook een top-entertainer te zijn met zijn one-liners, flitsende show en ijzersterke songs.
Na het vertrek van Robbie Williams uit Take That, scoorde hij zijn eerste solo hit met een cover van George Michael's nummer 'Freedom'. Van zijn eerste solo album kwamen de megahits Let Me Entertain You en Angels. Het tweede en derde album,'I've Been Expecting You' en 'Sing When You're Winning' waren nog grotere successen, artistiek en commercieel gezien. De singles 'Millenium', 'She's The One', 'Rock DJ', 'Supreme' en 'Kids', het duet met Kylie Minogue, werden stuk voor stuk hits. Inmiddels is de 'lolbroek' uit Engeland uitgegroeid tot een ware rockster en heeft hij ook de waardering en het respect van de 'serieuze' media. Vorig jaar verraste Williams met zijn optreden in The Royal Albert Hall, 'Swing When You're Winning', waar hij met groot orkest nummers van Frank Sinatra ten gehore bracht. Dit concert is ook op dvd en cd verschenen.
Vandaag is het nieuwe album van Robbie Williams uitgekomen, 'Escapology' genaamd. De eerste single van dat album 'Feel' is al wel te horen en zien op radio en tv maar ligt pas maandag 2 december in de winkels.

Robbie Williams + support act
Amsterdam ArenA
Vrijdag 18 juli 2003
Aanvang: 18.30 uur
Entree: € 42, € 39, € 36, € 34 (exclusief service kosten).

Kaartverkoop door Ticket Service via de grotere postkantoren, VVV's en GWK's, 0900-300.1250 (0,45 ct.p.m), www.ticketservice.nl, en de bekende voorverkoopadressen.

Figure 3.5 The overview of the concert of Robbie Williams on the 18th July 2003 on [Amsterdam Arena]

The link 'program' on [Heineken Music Hall] gives an overview of all the events. This overview is shown in Figure 3.6. The events without a name are business events and because of privacy they are not named and they do not have a start time. These events shall be not as large as a concert and therefore will they not give a lot of trouble on the road. The named events contain a link to information about the event, but this information isn't useful because the start time, the costs and the date are mentioned the overview.
All the other websites shall contain the same kind of information. This information will tell which event will take place on what date and what time and will also give the price to enter the event.

EVENEMENT	DATUM	AANVANG	PRIJS €.	OPMERKINGEN
	JUNI			
Event	21 juni 2003	0.00 uur		
R.E.M. & support act Pete Yorn	22 juni 2003	20.00 uur	€ 40	UITVERKOCHT
Event	24 juni 2003	0.00 uur		
	JULI			
Two 77 Splash	06 juli 2003	13.00 uur	€ 40	Kassa en deuren open: 11.30 uur. Kinderen tot 1.10m gratis toegang.
Hardcore Resurrection	26 juli 2003	22.00 uur	€ 35 / VIP € 45	
	SEPTEMBER			
Event	04 september 2003	0.00 uur		
Event	13 september 2003	0.00 uur		
Event	20 september 2003	0.00 uur		
Event	27 september 2003	0.00 uur		
Event	28 september 2003	0.00 uur		
	OKTOBER			
Kelly Rowland + special guest Solange	02 oktober 2003	20.00 uur	€ 30	Verplaatst concert 10 april 2003
Event	24 oktober 2003	0.00 uur		

Figure 3.6 Program of the events in the [Heineken Music Hall]

The general websites
This part also discusses two websites, the websites [Concert] and [Ticketservice]. The website about concerts has three options. Information can be found about theatre programs, film programs and concert programs. All three options work the same way. At first through the link 'agenda' a search can be made on the following fields:

- Place
- Time (not available for film, but the options are today, this week or the next four weeks)
- Genre

If the search would be done for concert in Amsterdam with the time field on the option 'this week' the results will look like the list in Figure 3.7. This list contains only a part of the complete agenda. The list contains the information about the show, the group, which performs the show, the place where the show will be held and the date and time on which the show will take place. Sometimes a remark is added.

Voorstelling	Gezelschap	Podium	Datum/Tijd	Ticket
Balkankoor Papuscka	Balkankoor Papucska	Vondelpark Openluchttheater	12-06-2003 20:00	
BoomBar	DJ Duo Olwlsey and guests	Leidseplein Theater Boom Chicago	12-06-2003 20:00	
Boeiend	Frank Sanders` Akademie voor Musicaltheater	Fijnhout Trajectum Theater Fijnhout Trajectum Theater	12-06-2003 20:15	
Trein	Fem@il	PleinTheater	12-06-2003 20:15	
Bouncing Souls, T.S.O.L.	Bouncing Souls, T.S.O.L.	Melkweg Oude Zaal	12-06-2003 20:30	
Eric Dillisse Quartet	Eric Dillisse Quartet	De Badcuyp Bovenzaal	12-06-2003 20:30	
Richard Flemming (piano), Marta Dziekanska (cello), Satoko Kato (piano)	Richard Flemming (piano), Marta Dziekanska (cello), Satoko Kato (piano)	Het Concertgebouw Kleine Zaal	12-06-2003 20:30	
The Woods Band	The Woods Band	Paradiso Kleine Zaal	12-06-2003 20:30	Afgelast
Luc Houtkamp`s Pow3	Luc Houtkamp`s Pow3	Bimhuis - Jazz & Improvisatie muziek	12-06-2003 21:00	
Radio 100	Radio 100	Edelsmidschool Theater De Cameleon	12-06-2003 21:00	
Tasha`s World	Tasha`s World	Paradiso Grote Zaal	12-06-2003 21:30	

Figure 3.7 Concerts in Amsterdam with the option 'this week' given by [Concert]

The other website has the possibility to search on the following fields:

- Category
- Organization
- Soon on the market
- New on the market
- Show, which allows us to search on:
  - Title/group
  - Place
  - Date from
  - Date to

If a search is done for a show on the field 'a date from' with the value the 11th of June and the same value in the field 'date to' then the results will be like displayed in Figure 3.8. It contains the information about the show, the date and time and the place.

BOB LOG III

wo 11 jun 2003	23:30	Paradiso kleine zaal	AMSTERDAM	*)

CLASSIC NOODLANDING!

wo 11 jun 2003	23:30	Paradiso grote zaal	AMSTERDAM	*)

PASAR MALAM BESAR / DAGKAART MA-VR

wo 11 jun 2003	12:00	Malieveld	DEN HAAG	*)

Figure 3.8 Shows on the 11th of June 2003 given by [Ticketservice]

3.5 The holidays

When the school vacations and the constructors' vacation will be is known in advance. This information can be found on the Internet on the address [Vakantiekalender]. At first the data of the school vacations is described. After that part the data about the constructors' vacation can be found.

The school vacations
When the school vacation starts not only the children are free, also some parents have to be at home to take care of them. The schoolteachers also have some of these weeks off.
For dividing the start and end of the school vacations over the whole country the Netherlands is build out of three parts: North, Middle and South. The information is put on the website in two parts, the schedule for the summer vacation and one for the other, short vacation. The first schedule contains:

- The year for which the schedule is
- The part of The Netherlands to which this schedule applies
- Whether it is for primary school (denoted by BO/SO) or for secondary school (VO/VSO)
- The date of the vacation period

The other vacations has an other scheme:

- The name of the vacation
- The part of The Netherlands to which this schedule applies
- The number of the week in which the vacation period falls into
- The date of the vacation period

Figure 3.9 shows an example of this information. The first block is the summer vacation and the second block the other vacations. Only the vacation of the school year 2002-2003 is shown.

2003		
Regio Noord	BO/SO	28 juni t/m 10 augustus 2003
	VO/VSO	28 juni t/m 17 augustus 2003
Regio Midden	BO/SO	5 juli t/m 17 augustus 2003
	VO/VSO	5 juli t/m 24 augustus 2003
Regio Zuid	BO/SO	19 juli t/m 31 augustus 2003
	VO/VSO	12 juli t/m 31 augustus 2003

Summer vacation

Schooljaar 2002-2003 Terug

vakantie	*regio*	*week*	*data*
voorjaar	noord	8	16 februari t/m 24 februari 2002
voorjaar	midden + zuid	7	09 februari t/m 17 februari 2002
mei	alle	18	27 april t/m 05 mei 2002
herfst	noord	42	12 oktober t/m 20 oktober 2002
herfst	midden + zuid	43	19 oktober t/m 27 oktober 2002
kerst	alle	52+1	21 december 2002 t/m 5 januari 2003
voorjaar	noord + midden	9	22 februari t/m 2 maart 2003
voorjaar	zuid	10	1 t/m 9 maart 2003
mei	alle	18	26 april t/m 5 mei 2003

Other vacations

Figure 3.9 School vacations of the year 2002-2003 on [Vakantiekalender]

The constructors' vacation

In The Netherlands also a vacation is known under the name constructors' vacation. This vacation is determined by the collective labor agreement of the constructors. Also this vacations is given on this website. This is based on the division in three parts. The following information can be found:

- The part of The Netherlands to which this schedule applies
- The date of the vacation period

In Figure 3.10 an example of this information is given, again the year 2003 is shown.

Regio noord	Regio Midden	Regio Zuid
Maandag 14 Juli t/m Vrijdag 01 augustus	Maandag 21 juli t/m Vrijdag 08 augustus	Maandag 28 juli t/m Vrijdag 15 augustus

Figure 3.10 The constructors' vacations on [Vakantiekalender]

3.6 Traffic measurement along the road

Along the roads around the city of Delft several measurement instruments are installed. Every instrument has its own code. For every instrument it is known on which position along this road it is installed. The information about the number of lanes and the direction of the traffic on its measurement point is also coupled to the code of the instrument.
Every minute of the day the number of cars passing are registered and their average speed is calculated. This information with the date and the time of the measurement is stored in a database and available through [Regiolab Delft]. The information displayed on this website is only accessible for authorized users. Also the reliability of the data is registered. This parameter will indicate if the measurements can be trusted.
The data that the site gives has the following parameters:

- The hexadecimal code of the measurement instrument (HC)
- The place of the instrument (PL) in meters
- The direction (DI) (R for right and L for left)
- The number of road lanes (NL)
- The measurement date (MD) in the format yyyymmdd
- The measurement time (MT) in hh:mm
- Indication of the reliability of the data (IR) (y =yes and n = no)
- The number of cars (NC)
- The average speed (AS) in kilometers/hour

In Table 3.8 an example of the data is given. These measurements are made along the A12 from The Hague to Gouda on the 1st of July 2002 at 7.00 AM. The search function gives two options for the unreliable data. It is possible to receive the data without unreliable data or the values of the number of cars and the average speed can be –1 for the unreliable data. This last option is used to receive the necessary data. Sometimes when it is very quiet on the road no car will come by. The measurements will give a zero for the number of cars, but because no speed is measured also the average speed is zero. Here the data is represented with an average speed of zero, but in the rest of this report the average speed is corrected to the maximum speed allowed. The maximum speed is chosen, because the assumption is made that if there is no car on the road there will also be no traffic jams.
To make the data more visible it is displayed in Figure 3.11. In the figure several little dots can be seen on the A12. The measurements are represented by dots. These dots show two things, the amount of cars and the average speed. The amount of cars determines the size of the dot. The dot will be larger when there are more cars. The average speed is displayed by the color of the dots. The green color is the maximum speed of 120 km per hour. If the color is red the speed is 0. The color legend is displayed in the figure as well.

Table 3.8 The measurements done on the 1st of July 2002 7:00 AM on the A12 from The Hague to Gouda

HC	PL	DI _ NL	MD	MT	IR	NC	AS
03D00C003C50D0074100	1580m	R _ 3	20020701	07:00	Y	31	55
03D00C00405FD0074100	1695m	R _ 2	20020701	07:00	Y	27	58
03D00C00444BD0074100	1775m	R _ 2	20020701	07:00	Y	25	52
03D00C004C37D0074100	1955m	R _ 2	20020701	07:00	Y	14	6
03D00C005C4FD0074100	2379m	R _ 2	20020701	07:00	Y	9	6
03D00C00642ED0074100	2546m	R _ 2	20020701	07:00	Y	13	5
03D00C007020D0074100	2832m	R _ 2	20020701	07:00	Y	16	5
03D00C00744AD0074100	2974m	R _ 3	20020701	07:00	Y	30	8
03D00C007C41D0074100	3165m	R _ 3	20020701	07:00	Y	28	7
03D00C008C37D0074100	3555m	R _ 3	20020701	07:00	Y	20	7
03D00C008C37D0074200	3555m	R _ 1	20020701	07:00	Y	8	11
03D00C00A037D0074100	4055m	R _ 3	20020701	07:00	Y	46	20
03D00C00AC19D0074100	4325m	R _ 3	20020701	07:00	Y	41	19
03D00C00B837D0074100	4655m	R _ 3	20020701	07:00	Y	31	25
03D00C00D455D0074100	5385m	R _ 2	20020701	07:00	Y	0	0
03D00C00D455D0074200	5385m	R _ 2	20020701	07:00	Y	61	62
03D00C00E04BD0074100	5675m	R _ 2	20020701	07:00	Y	0	0
03D00C00E04BD0074200	5675m	R _ 2	20020701	07:00	Y	35	78
03D00C00F437D0074100	6155m	R _ 2	20020701	07:00	Y	23	95
03D00C010805D0074100	6605m	R _ 2	20020701	07:00	Y	30	109
03D00C01184BD0074100	7075m	R _ 2	20020701	07:00	Y	21	106
03D00C01282DD0074100	7445m	R _ 2	20020701	07:00	Y	22	118
03D00C014000D0074100	8000m	R _ 3	20020701	07:00	Y	56	108
03D00C014C5FD0074100	8395m	R _ 3	20020701	07:00	Y	69	113
03D00C014C5FD0074200	8395m	R _ 1	20020701	07:00	Y	0	0
03D00C016450D0074100	8980m	R _ 3	20020701	07:00	Y	72	110
03D00C017800D0074100	9400m	R _ 3	20020701	07:00	Y	70	111
03D00C018855D0074100	9885m	R _ 3	20020701	07:00	y	61	114
03D00C019C37D0074100	10355m	R _ 3	20020701	07:00	y	61	117
03D00C01B037D0074100	10855m	R _ 3	20020701	07:00	y	57	117
03D00C01C405D0074100	11305m	R _ 3	20020701	07:00	y	87	110
03D00C01D419D0074100	11725m	R _ 3	20020701	07:00	y	89	112
03D00C01E84BD0074100	12275m	R _ 3	20020701	07:00	y	70	119
03D00C01F441D0074100	12565m	R _ 3	20020701	07:00	y	61	116
03D00C020046D0074100	12870m	R _ 3	20020701	07:00	y	55	123
03D00C021837D0074100	13455m	R _ 3	20020701	07:00	y	65	121
03D00C023405D0074100	14105m	R _ 3	20020701	07:00	y	62	115
03D00C02502CD0074100	14844m	R _ 3	20020701	07:00	y	56	116
03D00C026042D0074100	15266m	R _ 3	20020701	07:00	y	46	116
03D00C027800D0074100	15800m	R _ 3	20020701	07:00	y	46	118
03D00C029441D0074100	16565m	R _ 2	20020701	07:00	y	58	105
03D00C029441D0074200	16565m	R _ 1	20020701	07:00	y	2	99
03D00C02A837D0074100	17055m	R _ 2	20020701	07:00	y	56	103
03D00C02A837D0074200	17055m	R _ 1	20020701	07:00	y	6	109
03D00C02C037D0074100	17655m	R _ 2	20020701	07:00	y	46	114
03D00C02C037D0074200	17655m	R _ 1	20020701	07:00	y	5	88
03D00C02D405D0074100	18105m	R _ 2	20020701	07:00	y	40	103
03D00C02E41ED0074100	18530m	R _ 2	20020701	07:00	y	44	101
03D00C02FC0AD0074100	19110m	R _ 2	20020701	07:00	y	59	94
03D00C030C55D0074100	19585m	R _ 2	20020701	07:00	y	53	95
03D00C031C0AD0074100	19910m	R _ 2	20020701	07:00	y	47	104
03D00C032805D0074100	20205m	R _ 2	20020701	07:00	y	55	95
03D00C033405D0074100	20505m	R _ 2	20020701	07:00	y	52	99
03D00C03401ED0074100	20830m	R _ 2	20020701	07:00	y	53	96
03D00C034C05D0074100	21105m	R _ 2	20020701	07:00	y	45	97
03D00C036005D0074100	21605m	R _ 2	20020701	07:00	y	48	99
03D00C037405D0074100	22105m	R _ 2	20020701	07:00	y	48	98
03D00C03840FD0074100	22515m	R _ 2	20020701	07:00	y	59	98
03D00C039405D0074100	22905m	R _ 2	20020701	07:00	y	42	95
03D00C03A04BD0074100	23275m	R _ 2	20020701	07:00	y	34	111
03D00C03B00FD0074100	23615m	R _ 2	20020701	07:00	y	35	111
03D00C03B841D0074100	23865m	R _ 2	20020701	07:00	y	42	107
03D00C03C837D0074100	24255m	R _ 2	20020701	07:00	y	36	107
03D00C03D805D0074100	24605m	R _ 2	20020701	07:00	y	39	110
03D00C03E800D0074100	25000m	R _ 2	20020701	07:00	y	45	108
03D00C03F805D0074100	25405m	R _ 2	20020701	07:00	y	40	113
03D00C040805D0074100	25805m	R _ 2	20020701	07:00	y	39	115
03D00C04183CD0074100	26260m	R _ 2	20020701	07:00	y	41	113
03D00C043005D0074100	26805m	R _ 4	20020701	07:00	y	41	112
03D00C043C50D0074100	27180m	R _ 4	20020701	07:00	y	37	120
03D00C04580FD0074100	27815m	R _ 4	20020701	07:00	y	96	107
03D00C046437D0074100	28155m	R _ 4	20020701	07:00	y	89	112
03D00C047454D0074100	28584m	R _ 4	20020701	07:00	y	88	107
03D00C048032D0074100	28850m	R _ 4	20020701	07:00	y	103	102
03D00C049032D0074100	29250m	R _ 3	20020701	07:00	y	103	96
03D00C04A805D0074100	29805m	R _ 3	20020701	07:00	y	106	85

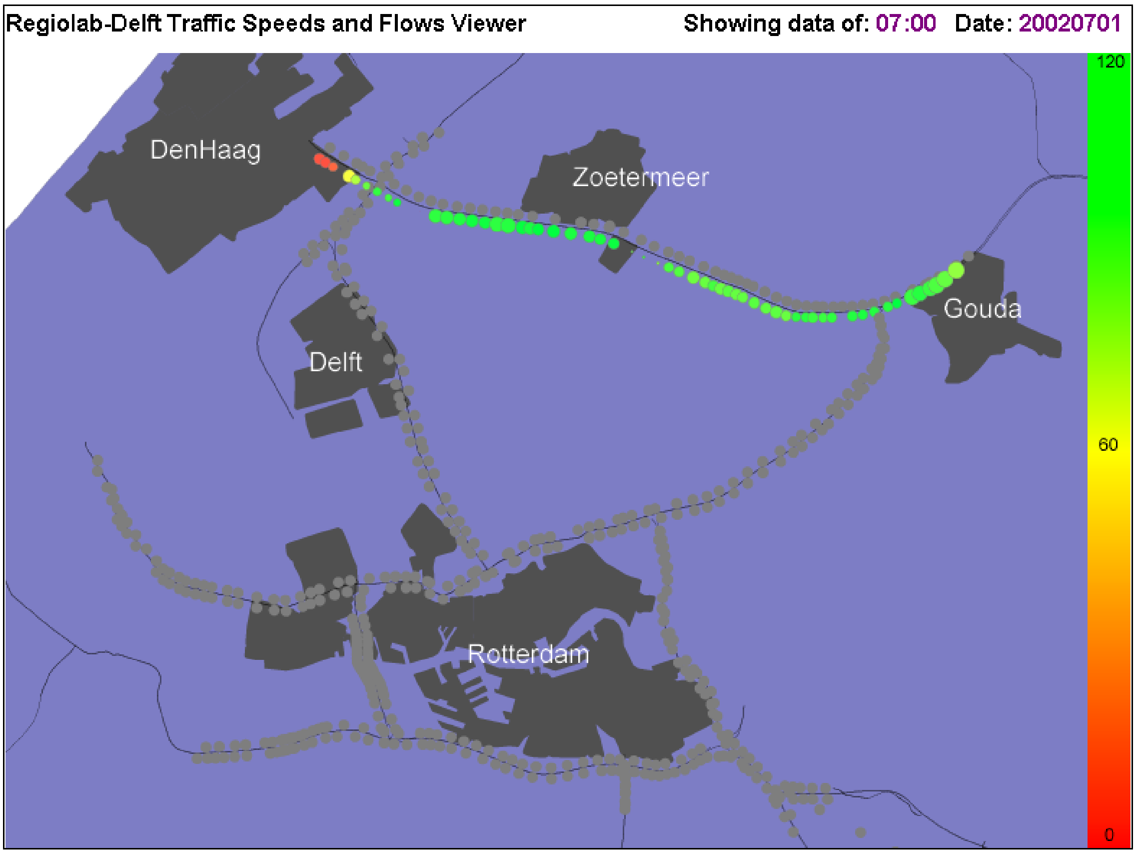


Figure 3.11 The data of the 1st of July 7.00 AM displayed

The red dots in The Hague indicate a traffic jam. The normal color of the dots in The Hague is orange or light green/yellow because the maximum speed is 50-70 km/h. The data in Table 3.8 shows the traffic jam as well, because at the beginning of the list the measured average speed is below the 10 km/h.

Chapter 4

Data analysis

In this chapter it will be demonstrated that the situation on the road depends on different factors. The factors that will be discussed are the days of the week, the month, the weather and the special events like accidents and holidays.
In the figures that display a graph of traffic, for example Figure 4.2, the traffic on the A12 from The Hague to Gouda on the day specified is shown. These graphs give the time versus the average speed. The data is averaged over 5 minutes to remove some of the random fluctuation and noise. To make a more clear representation of the situation on the A12 five points are taken and displayed in the graph by five lines with different colors. Each point has a number. This number represents the position in meters on the road. Point 1580 is a point in The Hague. The maximum speed on at this point is 50 km/h. The second point, 7075, is positioned between The Hague and Zoetermeer just after the viaducts from the junction "prins Clausplein". Between the two exits from Zoetermeer the third point is taken, point 14105. The fourth point, 22905, lies between Zoetermeer and Gouda approximately 4 km before the A20 joins the A12. Point 29805 is positioned nearby Gouda after the merging of the A12 with the A20. The points are displayed in Figure 4.1.

Figure 4.1 Points on the A12 taken to draw the graphs

From Figure 4.2 we can see that the average speed shows a lot of fluctuations. There are several causes for this phenomenon:

- There are different types of cars and drivers on the highway. Even when the maximum speed is 120 km/h and the traffic density is very low some drivers prefer to drive 100 km/h. This can also be the other way around. If the maximum speed is 100 km/h there are always some people driving to fast. Especially when the traffic density is low, the speed of a single driver can have a major impact on the average speed.
- Fluctuations can also be caused by traffic jams. These are the fluctuations that have our special interest. If the speed drops to a low value for more than a few minutes it indicates a traffic jam.
- If the amount of cars is zero the average speed measured is also zero. These values are changed by hand into an average speed that may be expected on that road. It is possible that the maximum speed on that part of the road is 120 km/h, but the average speed there is set on 100 km/h. This results in fluctuations because some of the measurements will have a higher value than 100 km/h.

In the next sections the possible factors that affect the situation on the road will be discussed.

4.1 Time

One cannot deny that the situation on the road is influenced by time. In the morning when most people go to work and in the evening when they come back there is more traffic on the road. It is possible that there is even more traffic than the road can handle. This results into congestion. Figure 4.3 is an example that shows this phenomenon very clearly. In the morning a traffic jam can be seen and in the afternoon another one.

4.2 Days of the week

It can be expected that the situation on the road is influenced by the day of the week. Therefore an analysis is made of the seven different days, Monday to Sunday. After the analysis of the different days a conclusion can be made to which degree the situation on the road is influenced by the days of the week.
For this analysis the month September from the year 2002 is taken. The month September has the most common situation, because it has no holidays or other unusual circumstances. To make a good analysis not only one day of this month is taken, but also all the days from this month are considered. As illustration only one figure is displayed. This is a figure that represents the most common example for the considered situation. All the other figures are represented in Appendix B.

Monday
On a Monday the first traffic congestion can be seen between 7.00 AM and 9.30 AM around Gouda after the merging of the A12 with the A20. Between these times most people are going to work. The congestion arises because there are more cars on the road at the same time than the road can handle. This congestion is shown in Figure 4.2 by the purple line. The average speed around that time goes down to a minimum speed of approximately 10 km/h.

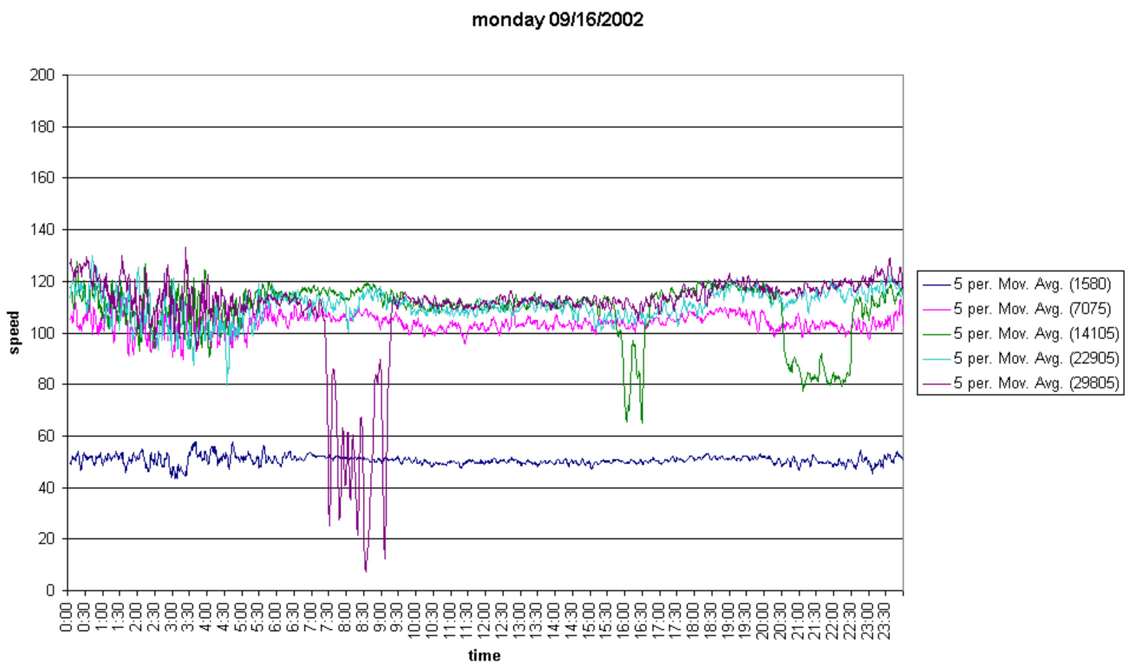


Figure 4.2 Traffic on Monday 16th of September 2002

The situation in the evening is not the same as in the morning. Now there is no congestion visible nearby Gouda. There isn't any real traffic congestion at all. Around Zoetermeer the average speed goes to 65/70 km/h as shown by the green line. This indicates only slow driving traffic. Also slow driving traffic can be seen in the evening between 8.30 PM and 10.30 PM with an average speed of 80/85 km/h, but this is not a common phenomenon. Maybe an accident is the reason for this slow driving traffic. The other figures from Mondays in Appendix B do not show any indications that the slow driving traffic in the evening is a common situation. Mainly it can be said about the Monday that only traffic congestion occurs in the morning around 7.00 AM and 9.30 AM near Gouda and in the afternoon mostly around Zoetermeer, but these are not very serious traffic jams. The rest of the day the traffic on the A12 from The Hague to Gouda does not suffer any delays.

Tuesday
People, who travel on the road on a regular basis, experience Tuesday as one of the busiest days of the week. Whether this is true will be checked here. In the morning between 7.00 AM and 10.30 AM there is traffic congestion nearby Gouda as Figure 4.3 shows. The purple line indicates speeds of less than 10 km/h in this time period. There is also a small traffic jam around 2.30 PM. This is not a usual traffic jam, as the other figures of Tuesdays in Appendix B shall prove. An analysis of the data itself shows that this traffic jam is caused by unreliable data. Unreliable data has the value –1 for the number of cars and the average speed. Therefore the average speed over five minutes is badly influenced by unreliable data. This causes the reaction as if it was a traffic jam.

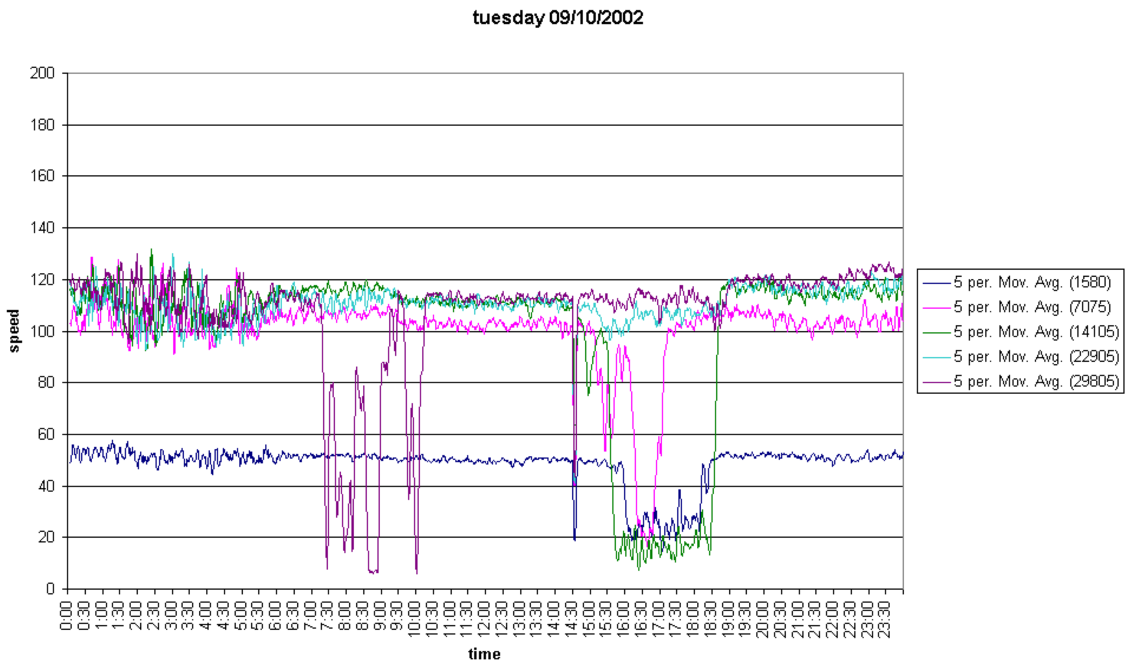


Figure 4.3 Traffic on Tuesday 10th of September 2002

Also traffic congestion can be seen from 3.00 PM to 7.00 PM, but then between The Hague and Zoetermeer. All three points, the point in The Hague, the point between The Hague and Zoetermeer and the point nearby Zoetermeer, are suffering a traffic jam around this time. Usually the congestion is only visible in The Hague and Zoetermeer and not between the two cities, as the figures in Appendix B indicate. The speed drops to 10/15 km/h at that time. The drop of the pink line is not that usual. That the congestion is visible over the whole section can indicate that an accident had happened.
Generally it can be said that on the section of the A12 from The Hague to Gouda traffic jams can be seen in the morning near Gouda and that in the afternoon the same situation occurs in The Hague and nearby Zoetermeer.

Wednesday
In the morning a small traffic jam is visible in Figure 4.4 between 7.30 AM and 10.00 AM near Gouda. The other figures in appendix B show that it is even possible that there is no traffic jam in the morning at all. Typically for Wednesdays is the little bit of slow driving traffic around 12.30 PM. The most probable explanation for this congestion can be found looking at the Dutch school system. In Holland all primary schools close on Wednesday afternoon around 12.30 PM. The children must be picked up at school. Some kids are going to after-school care. All the other kids have to be picked up by their parents or other relatives. Therefore some parents go home on Wednesday afternoon at 12.00/12.30 PM.

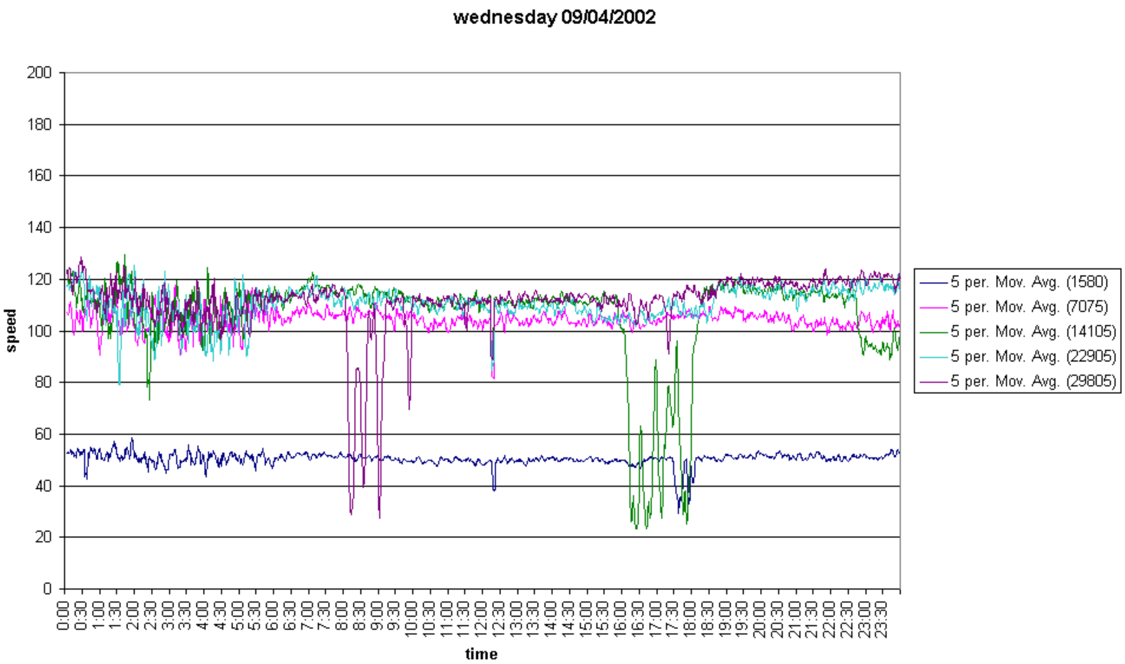


Figure 4.4 Traffic on Wednesday 4th of September 2002

Also can congestion be seen between 4.00 PM and 6.00 PM. This traffic jam is concentrated around Zoetermeer. In The Hague the road is also very crowded and a little slow driving can be observed. The other figures in Appendix B show that it is possible that congestion will not occur near Zoetermeer or The Hague, but between The Hague and Zoetermeer.

On Wednesday it can be said that there are three moments that it is crowded on the road, one in the morning around Gouda, one around lunchtime and one in the afternoon only now between or around The Hague and Zoetermeer. But these crowded moments last not as long as on a Tuesday and the average speed is not as low as on a Tuesday.

Thursday
Figure 4.5 shows that on a Thursday morning there will be a traffic jam around Gouda. The time this traffic jam can be seen is between 8.00 AM and 11.00 AM. Together with the other figures of Thursday in Appendix B the conclusion can be made that normally trouble on the road begins at 7.30 AM and can last until 11.00 AM. These figures also tell that sometimes between Zoetermeer and Gouda congestion occurs.
The figure states as well that it is very crowded on the road in the afternoon. The traffic jam begins at 3.30 PM and is finally gone at 8.00 PM. On this particular Thursday it is very crowded, only between The Hague and Gouda no traffic congestion is visible. Not every Thursday afternoon is as crowded as the one of Figure 4.5. Most likely is a traffic jam around Zoetermeer, but also Gouda and The Hague suffer from congestion most of the time. In this figure also slow driving traffic can be seen around Gouda from 10.00 PM. Analysis of the data gives no further information about the reason for this situation. No unreliable data can be found around that time. Therefore the average speed on the A12 around Gouda was really 80/85 km/h. Maybe this is caused by an accident.

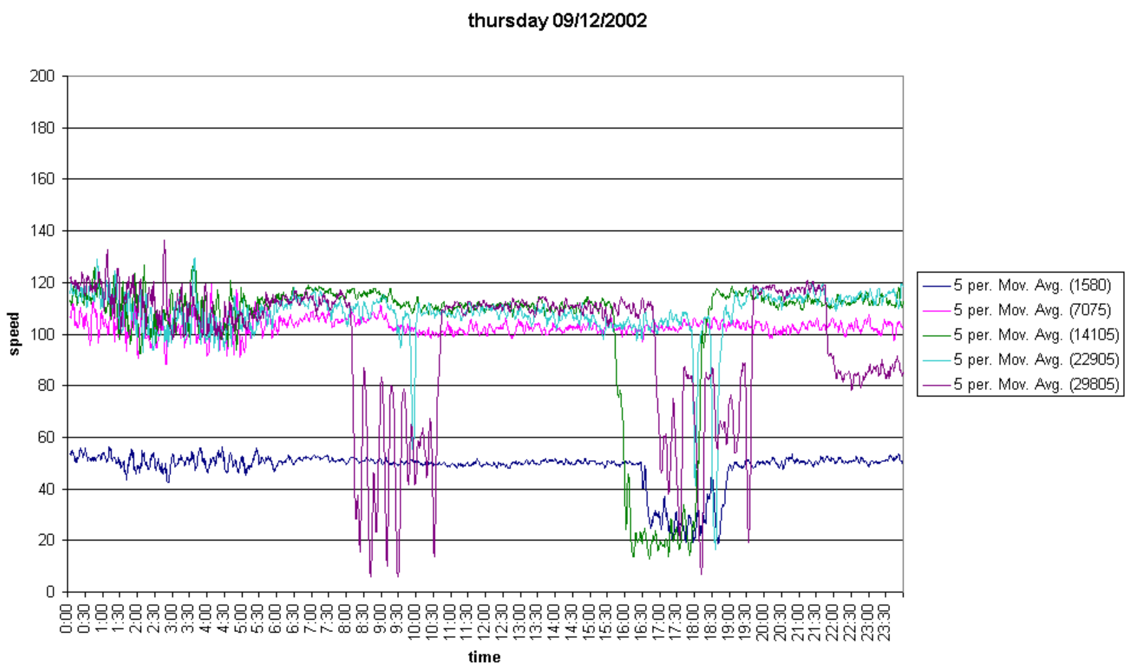


Figure 4.5 Traffic on Thursday 12th of September 2002

Friday
For most people Friday is the last day of the week on which they have to work. From all those people a lot are going home early. This can be seen in the figures in Appendix B. In the morning most of the time there will be no traffic jam. If there occurs congestion, it will be for a short period and most of the time around Gouda. In Figure 4.6 an example is shown of no traffic congestion in the morning.

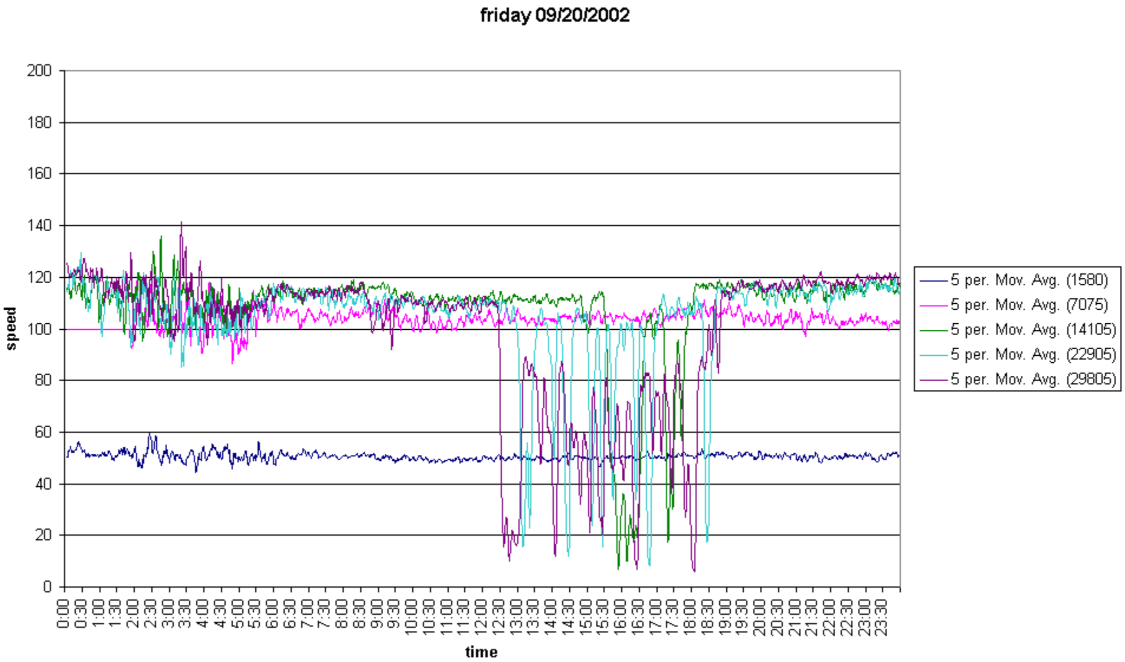


Figure 4.6 Traffic on Friday 20th of September 2002

In the afternoon the traffic jams begin a lot earlier than usual. Figure 4.6 shows a traffic congestion from 0.30 PM until 6.30 PM. This amount of traffic congestion is not usual for this day also the time is very early. On this particular day it was very crowded on the road. Generally it can be said that the time on which this situation starts is around 2.30 PM as shown by the figures in Appendix B. The time that the congestion is over is the same as mentioned before, 6.30 PM. These traffic jams occur almost always around Zoetermeer and Gouda and sometimes between Zoetermeer and Gouda. The Hague does not suffer from congestion in the afternoon. This is remarkable because all the other weekdays traffic congestion can be seen there.

Saturday
On a Saturday not many people have to go to work. Because of that under normal circumstances no congestion will occur. In Figure 4.7 a normal Saturday is shown. Only when there is an unusual situation a traffic jam can be seen. These unusual situations are extreme weather conditions, major events (like football matches) or an accident that had happened.

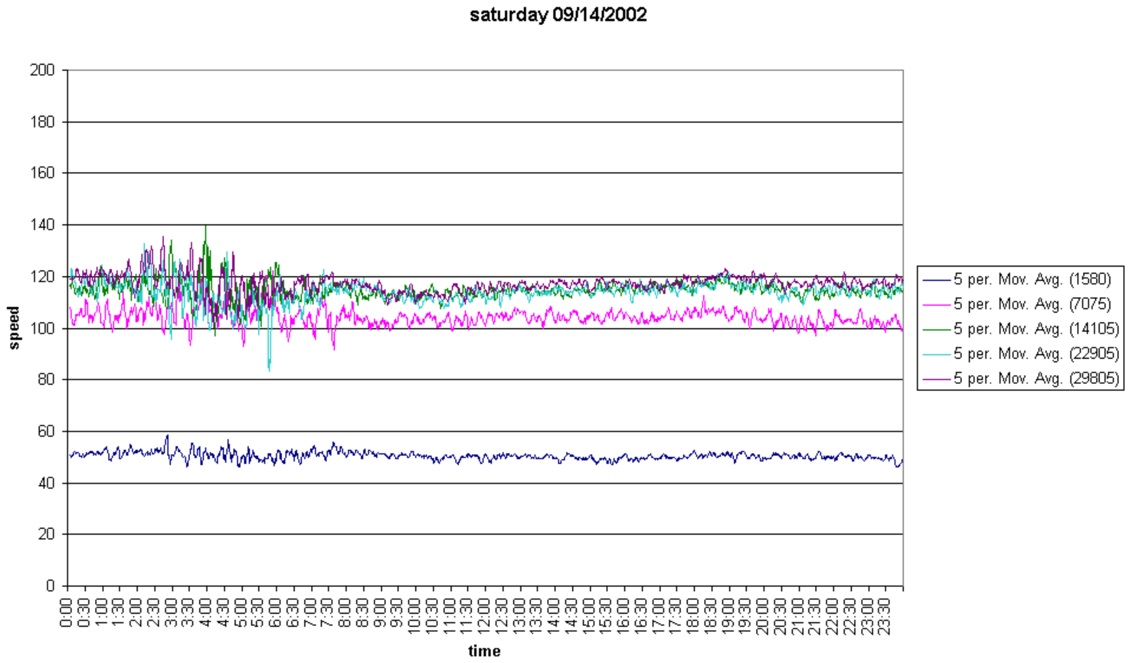


Figure 4.7 Traffic on Saturday 14th of September 2002

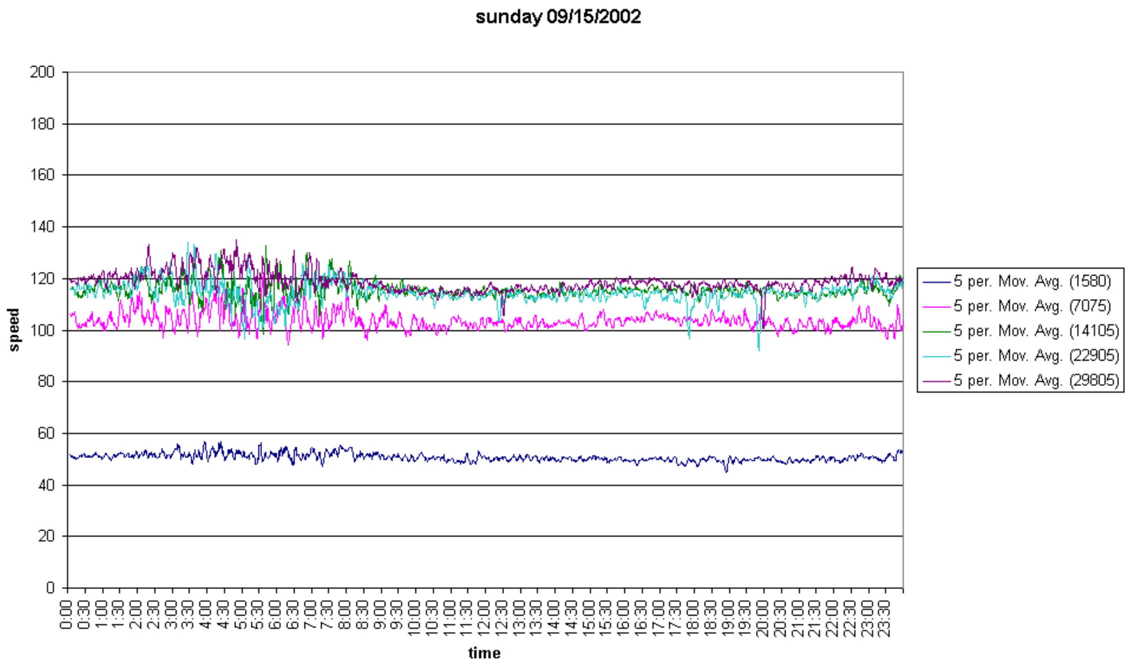


Figure 4.8 Traffic on Sunday 15th of September 2002

Sunday
The Sundays are very similar to Saturdays. Again most people don't have to go to work and also Figure 4.8, a normal Sunday, does not show any signs of congestion. There can be a little bit slow driving traffic around 11.00 AM when some people, Protestants and Catholics, are going to church. Also this is the time that some families are on their way to visit their relatives.
Further only in special situations congestion can occur. These situations are the same as mentioned by the Saturday above, bad weather, major events and accidents.

Conclusion
The most crowded days are the Tuesday and Thursday. Monday, Wednesday and Friday show less signs of congestion, but they can be there as well. Wednesday is the only day on which a traffic jam occurs around 12.30 PM, this is probably due to the Dutch school system. On a Friday the traffic jam arises earlier than usual, because most people like to be home early to enjoy their weekend. In the weekend there will be almost no congestion at all, only on Sunday around 11.00 AM some slow driving traffic can be possible. Most likely the small traffic jams are due to church visits and for example people visiting their grandparents. While discussing the Saturday and the Sunday it is mentioned that always traffic congestion can occur because of the bad weather and accidents. Of course this does not only apply to the weekends. The traffic on the other days can suffer from these circumstances as well.
This analysis shows that the day of the week is a major influencing factor of the situation on the road.

4.3 Date

The days of the week affect the situation on the road. Does the date also influence this situation? Because a situation does not change drastically by the transition of the month, using a combination of the day and the month may give a contribution to the result while predicting average speed. The year is not really a parameter that will contribute to a better prediction, because the only information it contains is the long term fluctuation in traffic load for example, a small possible decrease of maximum speed due to larger traffic density. By training a network on a regular basis this behavior will already be captured. Therefore the year will not be considered. In this section the influence of the date is considered by analyzing the differences between the months. Therefore the first Monday of every month in 2002 will be used and displayed. We realize that the day could be a bad representation of the month, but it is sufficient for the purpose of explorative data analysis.

January
The first Monday in January, shown in Figure 4.9, suffers from traffic congestion in the morning. From 7.00 AM until 9.30 AM traffic congestions occur near Gouda, near Zoetermeer and between those two cities. Probably this is not the normal situation because in the morning most of the time no traffic congestion occurs near Zoetermeer and between Zoetermeer and Gouda. The best explanation for congestion at that place is that several accidents have happened that day. The traffic congestion near Gouda is split in two smaller congestions. The first one starts at 7.00 AM and ends at 9.00 AM. Around 9.30 AM the second congestion takes place. The reason for this split is most likely the traffic congestions on the road a few kilometers before. Because of those congestions the flow of the cars stops for a short time and there will be less traffic near Gouda.

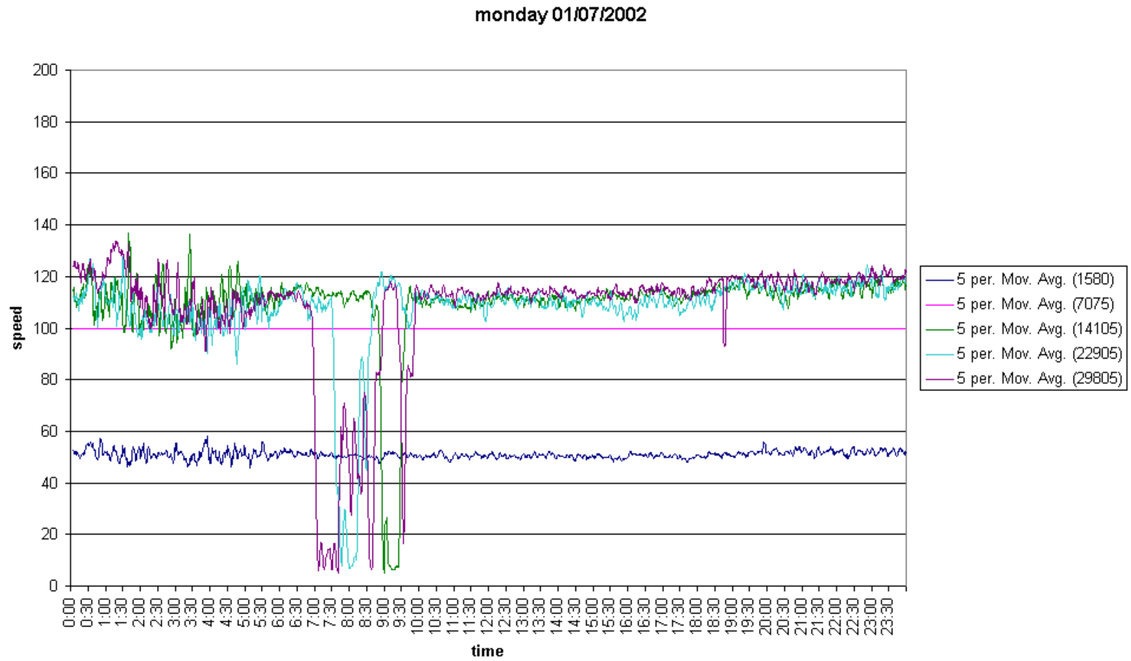


Figure 4.9 Traffic on Monday 7th of January 2002

In the afternoon no traffic congestion can be seen. Normally traffic congestion occurs near Zoetermeer, but this day it stays quiet. This is probably due to the holidays just before that. Also that is the time that a lot of people are away skiing can have influenced the amount of traffic congestion.
The reason that the pink line, the part of the road between The Hague and Zoetermeer, is a straight line on 100 km/h is that no reliable data was available for that place on this day. The value of 100 km/h is given to be able to display it. Therefore no conclusions can be made for that part of the road from this figure.

February

February has a very short period in the morning in which traffic jams occur. From 7.00 AM until 9.00 AM people have suffered from traffic congestion. This is the place traffic congestion occurs normally, but only then it will last longer (see Figure 4.10). The afternoon shows also small signs of congestion. Around 3.00 PM slow driving traffic is almost everywhere on the A12. Maybe a few people suddenly hit the brakes around that time. This results most of the time in a wave of somewhat slower driving traffic. From 6.00 PM until 7.30 PM a traffic jam is visible on the A12 in The Hague. This is a more serious traffic jam than normally on that part of the road and on a Monday. So maybe an accident is the reason for this congestion. Further no real traffic jams can be seen in the afternoon. This can indicate that there are fewer cars on the road and that the capacity of the road is sufficient this day. There is an explanation possible for the smaller number of cars. In February also a lot of people are going on a ski vacation.

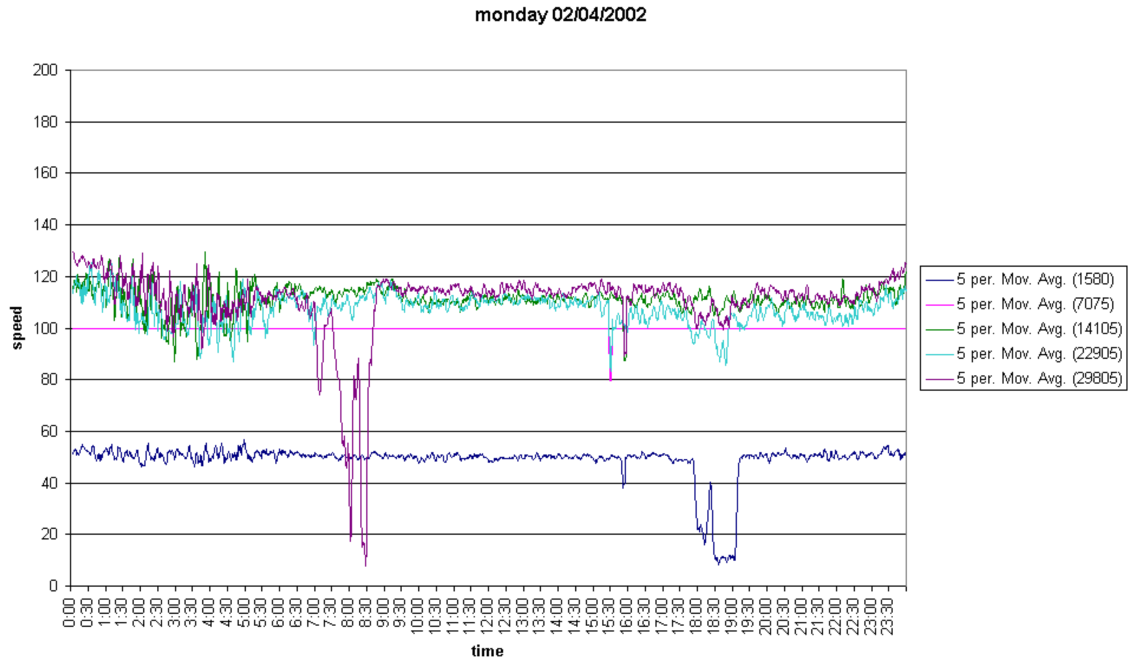


Figure 4.10 Traffic on Monday 4th of February 2002

March
Figure 4.11 is very similar to Figure 4.10. Also a small traffic jam in the morning around Gouda can be seen. This traffic jam last from 7.00 AM until 8.30 AM. The rest of the day no congestions are visible. The congestion of Figure 4.10 in The Hague has disappeared.

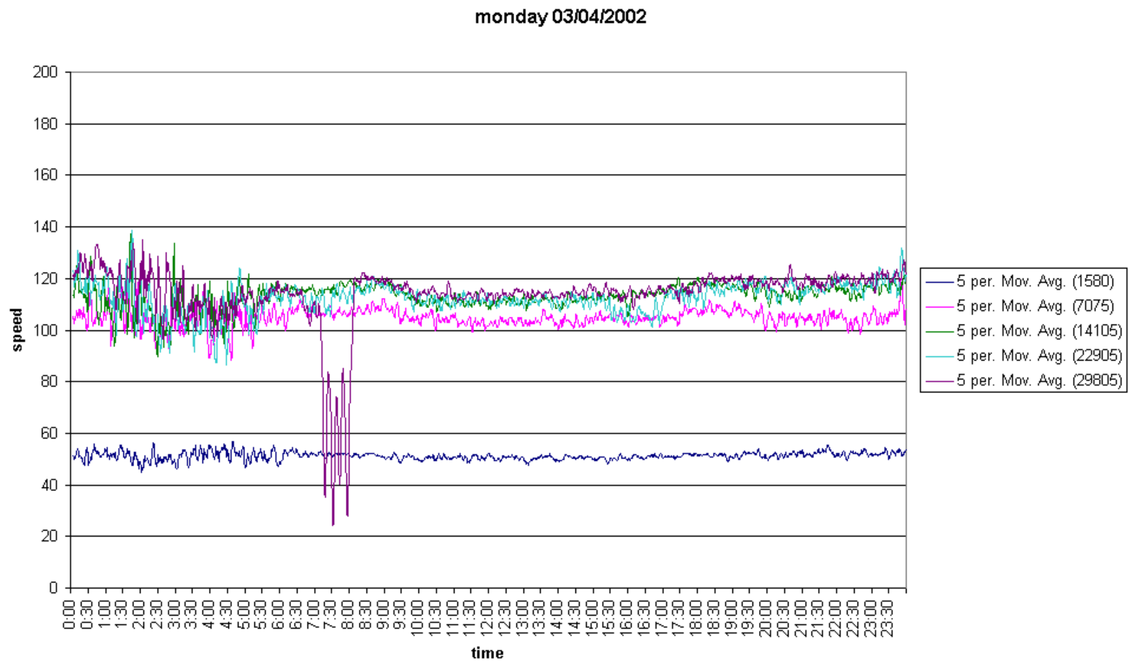


Figure 4.11 Traffic on Monday 4th of March 2002

The capacity of this road isn't sufficient in the morning, but in the afternoon on a Monday in March it is. The reason for the fewer cars is probably the same as in January and February. In March there still can be found a lot of snow in the mountains in Europe. And the people who like to ski and like it most with somewhat higher temperatures go skiing in March.

April

The graph for the first Monday of April showed a remarkable of congestion. Inspection of the calendar revealed, however, that this day was special as it concerned the second Easter Day. This is a clear example of the relation between a special day and the traffic situation. As a consequence another day had to be chosen. Because in the section above it was made clear that the day of the week influenced the situation on the road, it had to be a Monday as well. The second Monday of April is chosen for that reason. This day is displayed in Figure 4.12.
April seems to have a normal pattern of congestions. Traffic jams appear in the morning and in the afternoon. In the morning the usual traffic congestion near Gouda can be seen. It last from 7.00 AM until 10.00 AM. Also a small traffic jam can be seen between Zoetermeer and Gouda. Maybe an accident is the cause of this traffic jam or the road was that busy that day that even the capacity of that part of the road was not sufficient.

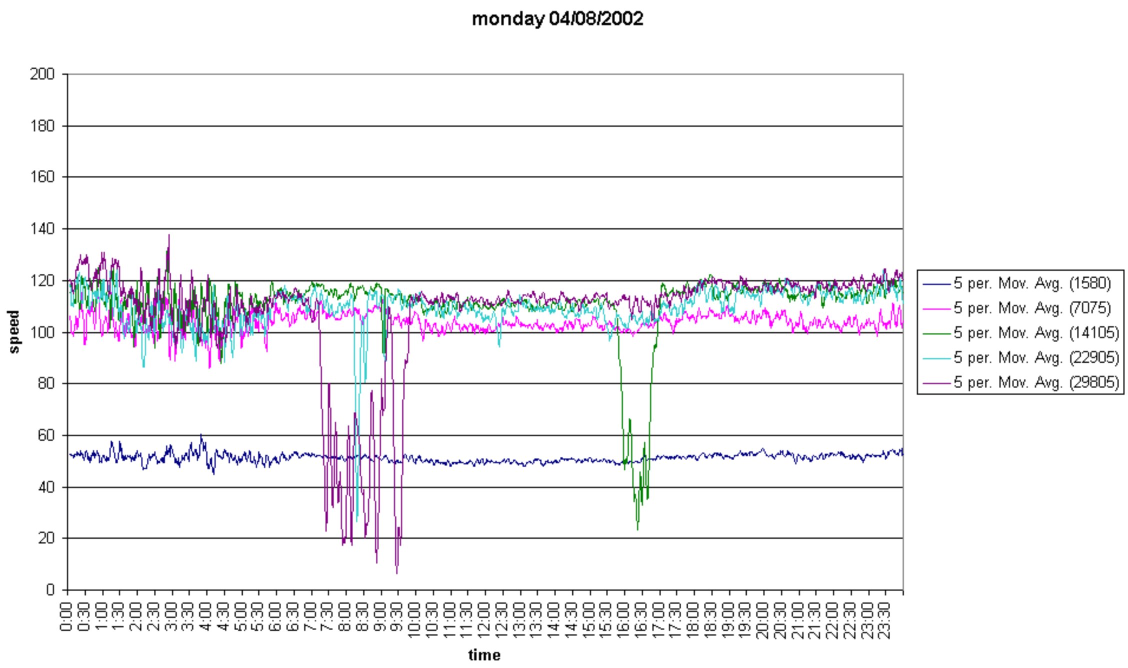


Figure 4.12 Traffic on Monday 8th of April 2002

In the afternoon the traffic congestion arises from 4.00 PM until 6.30 PM near Zoetermeer. This is the normal place for the afternoon traffic jam. Nothing unusual happened on that day. The skiing season is over and therefore the congestions are back.

May

The traffic situation seems also normal in May. Between 7.30 AM and 9.00 AM traffic congestion can be seen in Figure 4.13. This congestion is as usual near Gouda.

Around 11.00 AM some unreliable data was present for all the five points. Because of this no information about the situation around that time is available. The value of the unreliable data was –1 and this results in a disappearing line for the duration of the unreliable data.

Figure 4.13 Traffic on Monday 6th of May 2002

In the evening a little bit slow driving traffic can be detected, but no real congestion occurs. The conclusion for a Monday was that there was a little bit congestion, but mostly just slow driving traffic. Therefore this figure represents a normal situation. There is a sign of a little traffic jam around 1.30 PM near Zoetermeer, the green line, but the reason for that small traffic jam is unknown. Maybe a small accident was the reason or just someone hit the brakes suddenly which results in slow driving traffic for a short time.

June

June has normal morning and afternoon traffic jams. Near Gouda the usual traffic jam occurs at 7.00 AM and lasts until 9.00 AM. Just before 10.00 AM a little bit of slow driving traffic is visible. This is probably due to a person that hit the brakes and caused a row of cars to hit the brakes too. This is all shown in Figure 4.14.
Also in this figure unreliable data is shown, but this time just after 12.00 PM. The data was reliable again just before 1.30 PM.
In the afternoon the first traffic congestion can be seen around 4.00 PM near Zoetermeer. This seems the normal traffic congestion, but it occurs very early. Maybe an accident has caused this congestion. Between Zoetermeer and Gouda a traffic jam occurs from 5.00 PM to 5.30 PM. Around 6.00 PM traffic slows down in The Hague. These are the normal traffic jams and probably due to the capacity of the road, which seems too small.

Figure 4.14 Traffic on Monday 3rd June 2002

July

The traffic of the first of July, the first Monday in July, is shown in Figure 4.15. At the beginning of July the traffic jams are smaller, but are still there. The traffic is able to drive somewhat faster than usual around that time. In the morning a small traffic jam occurs near Gouda. Also in The Hague traffic congestion can be seen. This is not normal, but maybe an accident had happened.

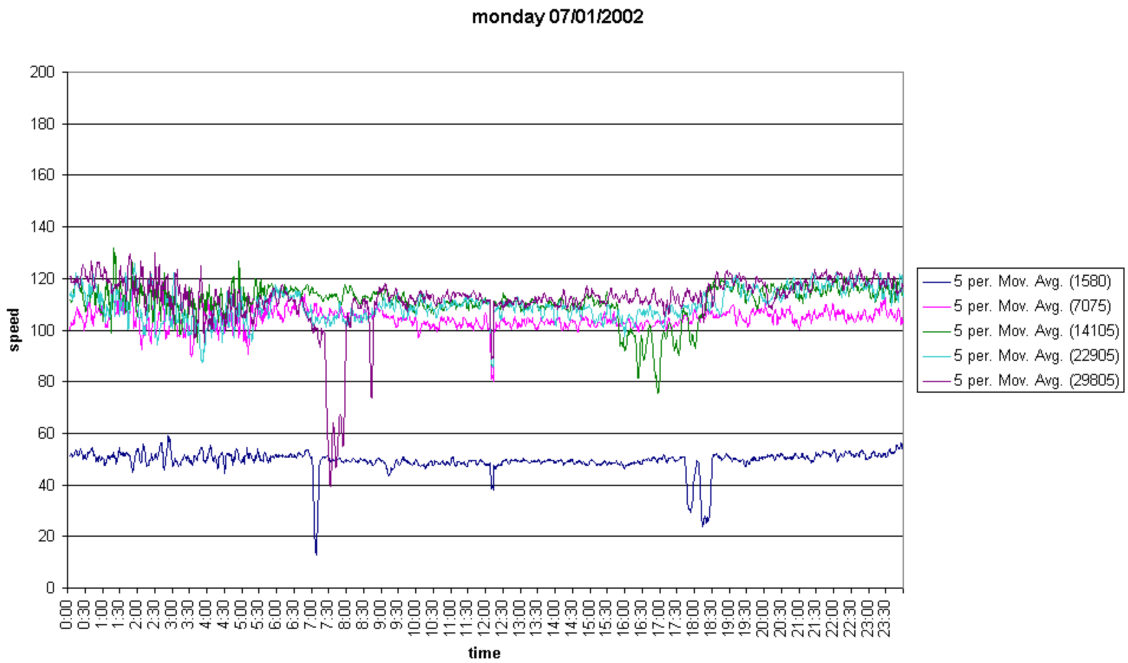


Figure 4.15 Traffic on Monday 1st July 2002

Around 12.00 PM unreliable data caused a small 'traffic jam' in the figure. In the afternoon almost no congestion at all, only slow driving traffic near Zoetermeer and in The Hague is visible. There is one possible reason for the lack of traffic jams. The summer vacation is about to start. Maybe some people are already on a holiday.

August
No information was available from the month August 2002. It only contained unreliable data. Therefore the last Monday from July had to be taken, but also no reliable information about that day could be found. In this case the 22nd of July is considered. There is no congestion on this day. This is caused by the summer vacation. Many people are away or at least free from work. Only if an accident happens or when there is very bad weather a traffic jam can be seen around this time. Figure 4.16 gives the traffic for the 22nd of July 2002.

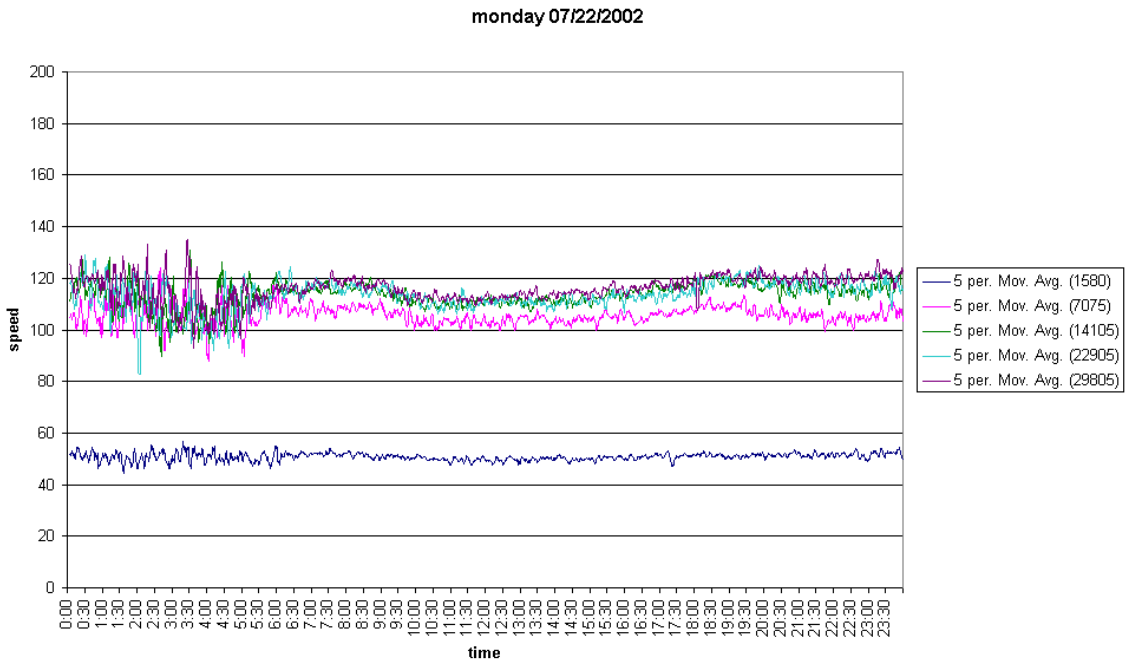


Figure 4.16 Traffic on Monday 22nd of July 2002

September
In the beginning of September the traffic congestions are returning. In the morning the traffic jam is normal again. Traffic congestion occurs near Gouda between 7.30 AM and 9.30 AM. There is also traffic congestion visible between Zoetermeer and Gouda around 8.30 AM. Normally on Monday no traffic congestion is seen there around that time. The most probable explanation is that an accident had happened. In the afternoon, as Figure 4.17 shows, the traffic jams are still gone. They will return in the first week of September, because most people are then back from their vacation and have to go to work again. Figure 4.2 gives a more representative Monday in September.

Figure 4.17 Traffic on Monday 2nd of September 2002

October
The month October brings back all the traffic jams. In the morning the traffic jam near Gouda is occurring again. This congestion lasts from 7.00 AM until 8.30 AM. The traffic jams in the afternoon are somewhat less usual. Normally traffic congestion occurs near Zoetermeer. This afternoon traffic jams are seen in The Hague and between The Hague and Zoetermeer. Maybe one or more accident had happened. This would explain why there was no congestion near Zoetermeer. If some riding tracks are closed and all the traffic have to go through with one ore more tracks less then the cars come more spread in the direction of Zoetermeer. The capacity of the road near Zoetermeer is then sufficient. This is all shown by Figure 4.18.

November
November also has the regular traffic jams. In the morning around 8.00 AM the speed is limited around 20 km/h. In the afternoon the situation is somewhat better, but traffic congestion is still there, but again on unusual places. Near Gouda and between The Hague and Zoetermeer congestion occurs for a very short time as Figure 4.19 shows. Maybe some accidents happened or someone suddenly hit the brakes and caused a little traffic jam. Probably the reason that no congestion occurs near Zoetermeer is the same explained by October, that the traffic jam between The Hague and Zoetermeer spreads the traffic.

Figure 4.18 Traffic on Monday 7th of October 2002

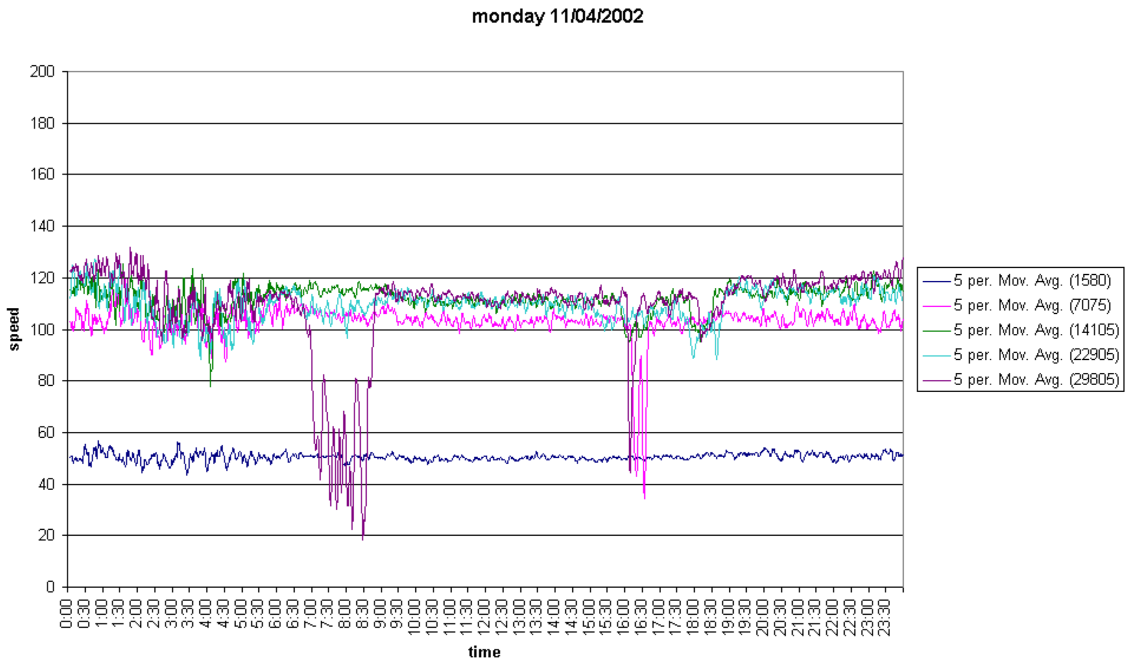


Figure 4.19 Traffic on Monday 4th of November 2002

December
In December the traffic congestions can be seen clearly. In the morning traffic congestion occurs on several points. None of them are the usual. Between The Hague and Zoetermeer, near Zoetermeer and between Zoetermeer and Gouda traffic jams can be seen in Figure 4.20. The most plausible explanation for these events is that accidents had happened. This can also explain why there was no traffic congestion near Gouda. The traffic from the A12 slowly flows through all traffic jams. And the amount of traffic near Gouda at the same time decreases.

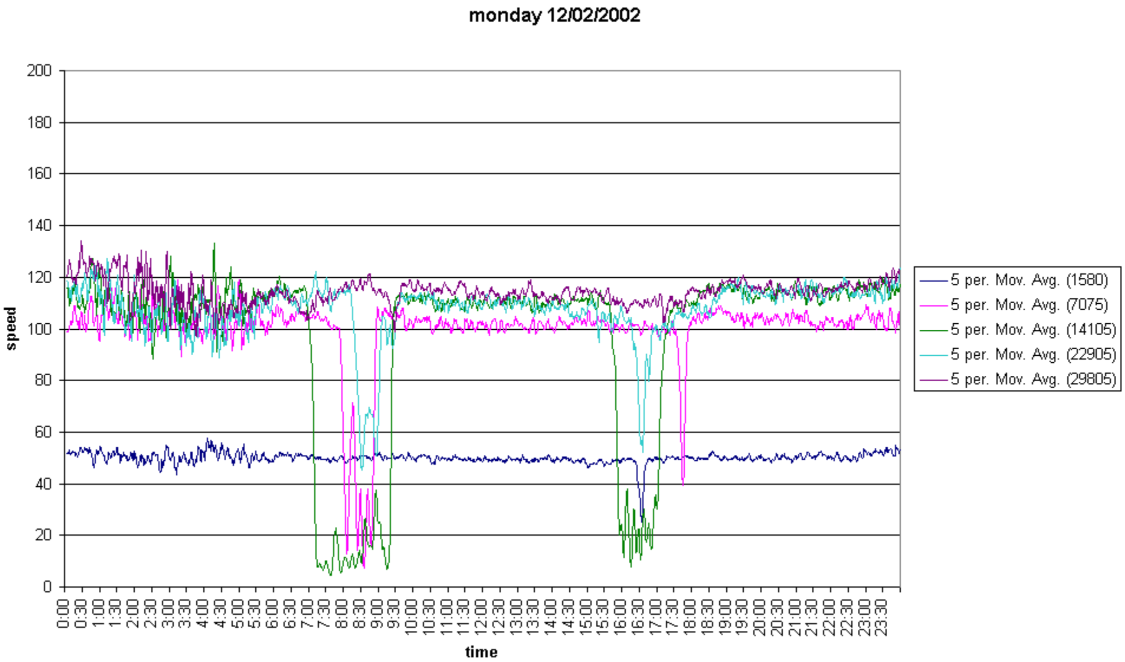


Figure 4.20 Traffic on Monday 2nd of December 2002

The evening shows a normal situation: congestion near Zoetermeer. It lasts from 3.30 PM until 6.30 PM. Except for the point near Gouda, all the other points suffer from little traffic jams. Maybe this is caused by accidents, but because these jams are very small it is more likely that some people hit their brakes and caused there small jams with that action.

Conclusion
Above all the first Mondays of every month in 2002 is described. At the beginning of the year, the first three months, almost no traffic congestion can be seen. Only in the morning sometimes traffic congestion occurs. April, May and June have normal congestion. In the morning and in the afternoon traffic jams occur. In July and August most people are on holiday, therefore almost no congestion can be seen. In the beginning of September the people are returning from their holiday and the congestion starts to appear again. October, November and December show serious traffic jams in the morning and the afternoon.
From this information it can be concluded that it does matter which month it is for the situation on the road. In January, February, March, July and August the traffic jams (almost) disappear.

4.4 Weather

It seems plausible that weather conditions influence the situation on the road. People tend to slow down when it is heavily raining, when it is snowing or when the sight is limited because of fog. But is it possible to make it visible? To prove it one has to compare all the weather information and road situations of all the days in a year. Further a little bit of rain will not cause enormous traffic jams. It is possible that some kinds of weather do have a large impact on the situation of the road. In this section the effects of some extreme kinds of weather on the traffic of the road will be discussed.
The figures with the weather consist of two graphs, one of Rotterdam Airport and one of Valkenburg, near The Hague. Because the information available about the kind of weather is only in text, the kinds of weather are divided in 11 categories. 0 means no clouds and sun. 10 stands for very bad weather, like heavy snow. The distribution of the kinds of weather over the 11 categories is mentioned in Appendix A. The red bars represent the kind of weather in numbers from 0 to 10. The blue line gives the temperature in degrees Celsius.

Fog
It is possible that fog effects on the situation on the road. The visibility on the road is very bad, which can lead to dangerous situations. The risk of collisions is much larger. So probably more traffic jams can be seen. Figure 4.21 displays the weather on 27th of March 2003. On this day the visibility was very bad in the morning on both places caused by fog. The number 9 represents fog. In the evening the station Valkenburg reports fog again.

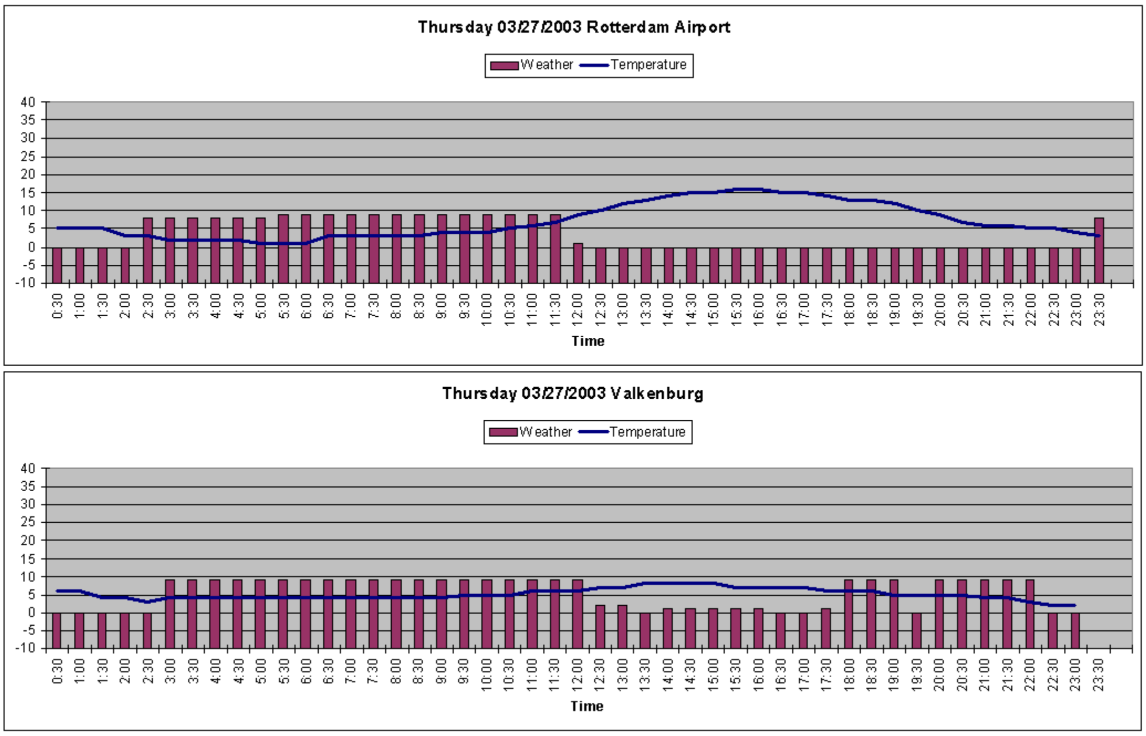


Figure 4.21 Weather on Thursday 27th of March 2003

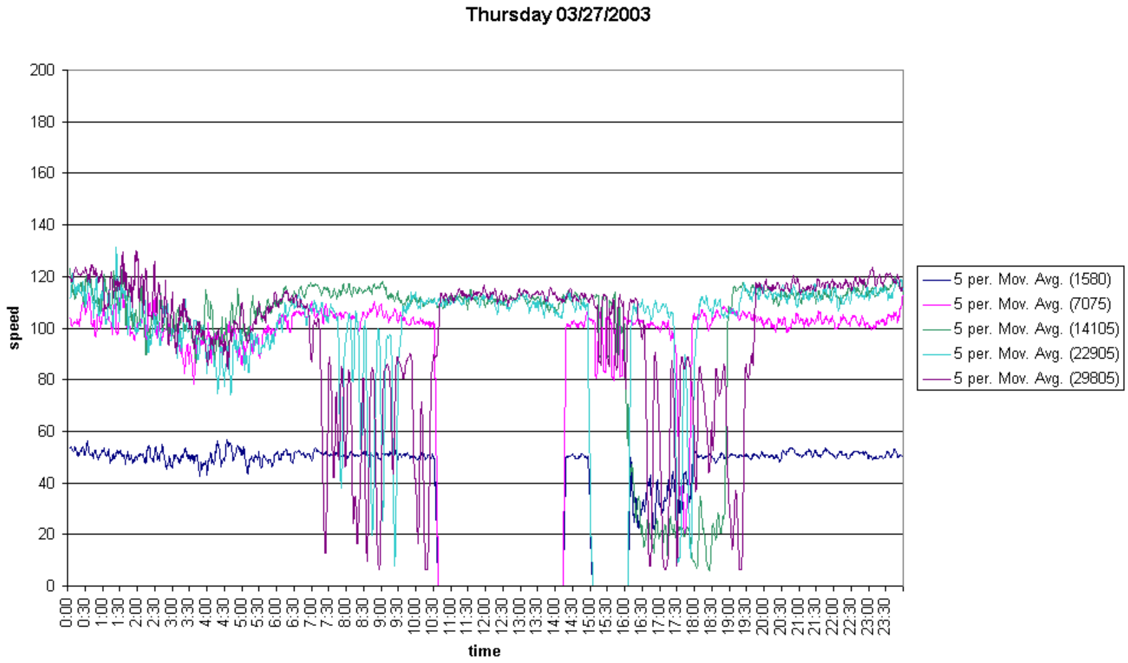


Figure 4.22 Traffic on Thursday 27th of March 2003

The traffic on this day, displayed in Figure 4.22, has unreliable data on two measuring points from 10.30 AM until 2.00 PM. These measuring points are The Hague, the blue line, and the point between The Hague and Zoetermeer represented by the pink line. The values for the average speed in this time interval are –1 and therefore not visible in the figure. So no conclusions can be drawn in that period. Also from 3.00 PM until 4.00 PM two points give unreliable data. This time these points are the point in The Hague with the blue line and the point between Zoetermeer and Gouda with the cyan line. The other lines also have unreliable data around that time, but only a few values. This causes the bumpiness.
From 3.00 AM until 12.00 PM fog was reported by both stations. In the evening from 6.00 PM until 7.00 PM and from 8.00 PM until 10.00 PM again fog was reported by Valkenburg. In the early morning most people are asleep and no traffic jams occur. From 7.00 AM until 10.30 AM the situation on the road is very bad. Only in The Hague the traffic is not suffering from the fog. The most probable reason for this phenomenon is that this measuring point is lying in a city. Fog always seems less thick there. The three measuring points still available after that period do not indicate that the situation is still very bad. In the evening when the fog comes again, the traffic jams are there again. This figure can be compared to Figure 4.5, another Thursday. It shows that the usual traffic jams are there; this are the traffic jams in The Hague and around Zoetermeer and Gouda. But also other traffic jams arise. They other points also report traffic jams, only very short ones. The influence of the fog in the evening isn’t very visible, because traffic congestion is very usual around 6.00 PM on Thursday. But the reason that the speed is that low on this particular Thursday is probably due to the fog.

Snow
On the 1st of February 2003 something unusual happened in Holland, a lot of snow came down from the sky. And because it was very cold at the same time it didn't melt away directly.
It was a Saturday. On a normal Saturday no traffic jams occur on the Dutch highways in the weekend as was stated before. Did this weather have impact on the situation on the road?

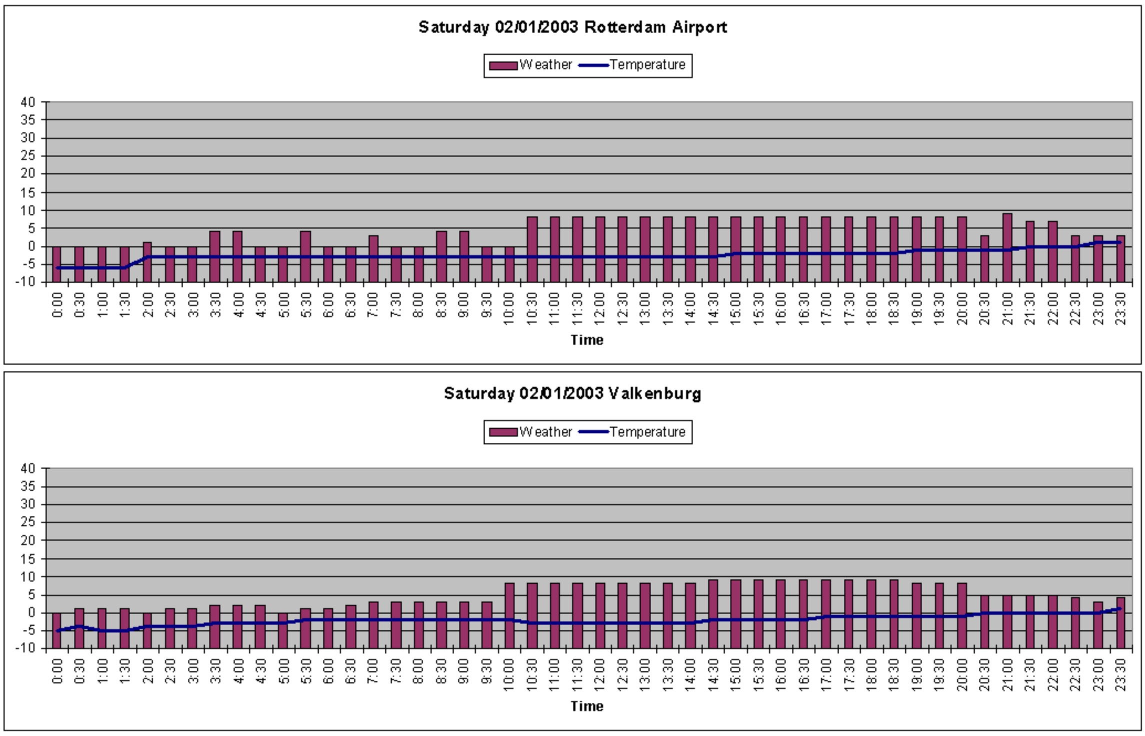


Figure 4.23 Weather on Saturday 1st of February 2003

Figure 4.23 gives the weather of that day. The amount of snow falling down determines the number given to it. The number lies between 8 and 10. The snowfall starts at 10.30 AM and continues until 8.00 PM. Figure 4.24 shows that it is not a normal Saturday, because traffic jams occurred. The traffic begins to drive slower at 10.30 AM. The cars do not reach the maximum speed allowed until 8.00 PM. This is the complete time that it was snowing. So the conclusion can be made that these traffic jams where caused by the snow. People where driving slower because of the slipperiness. This slipperiness also makes the risk of collisions a lot larger.

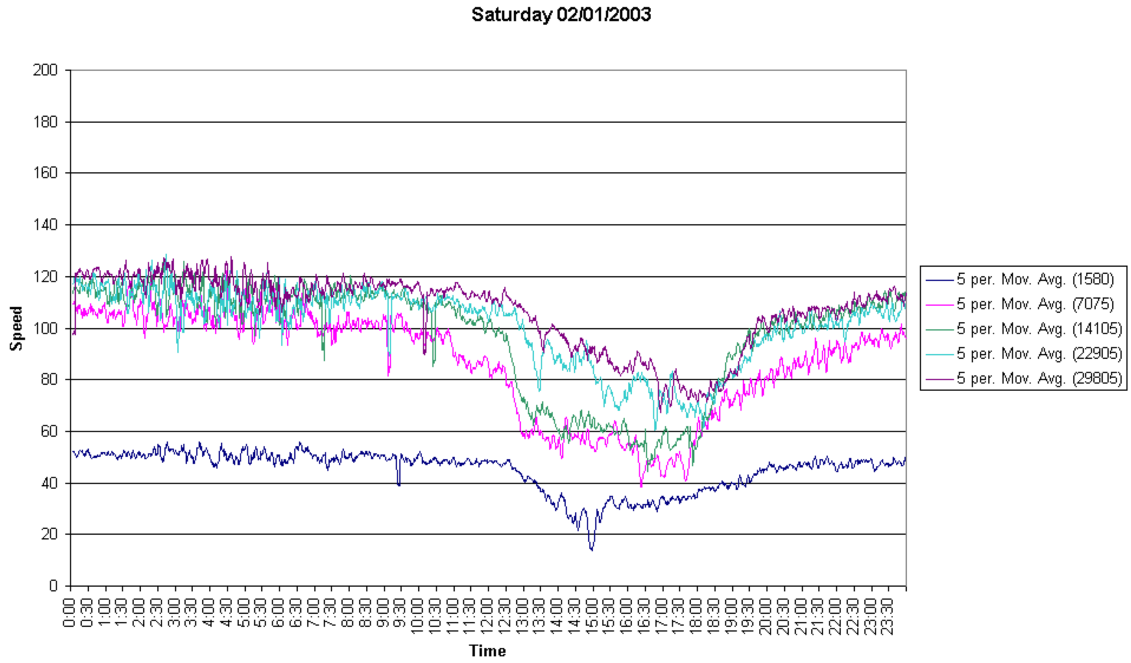


Figure 4.24 Traffic on Saturday 1st of February 2003

Warm and sunny weather

If it is warm and sunny and people are free from work, they go to the beach. A lot of people are on the beach at the same time and have to leave again. The A12 from The Hague to Gouda is one of the routes away from the Scheveningen, one of the seaside resorts of Holland.

Figure 4.25 shows it was nice and warm weather on 28th of July 2002. It was a Sunday. Therefore normally no congestion will occur.

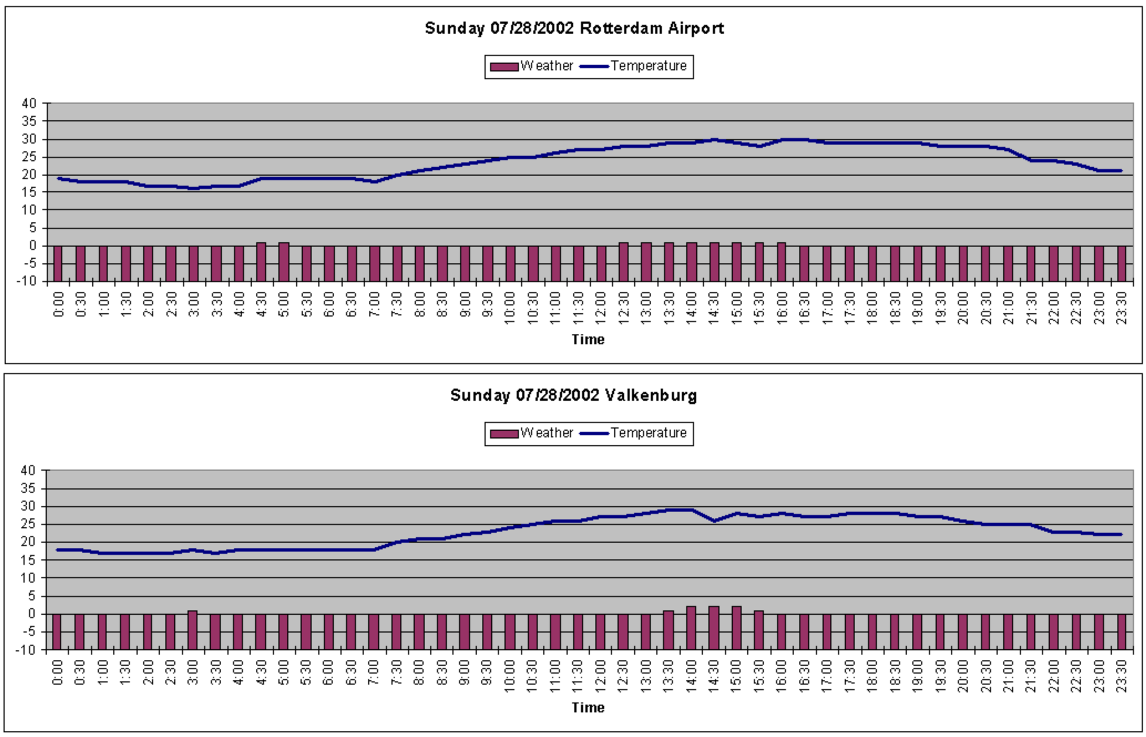


Figure 4.25 Weather on Sunday 28th of July 2002

In Figure 4.26 still no traffic jams can be seen. Maybe this lack of influence is caused by the fact that it is summer vacation or maybe no influence can be expected at all, because these people do not leave the beach at the same time.

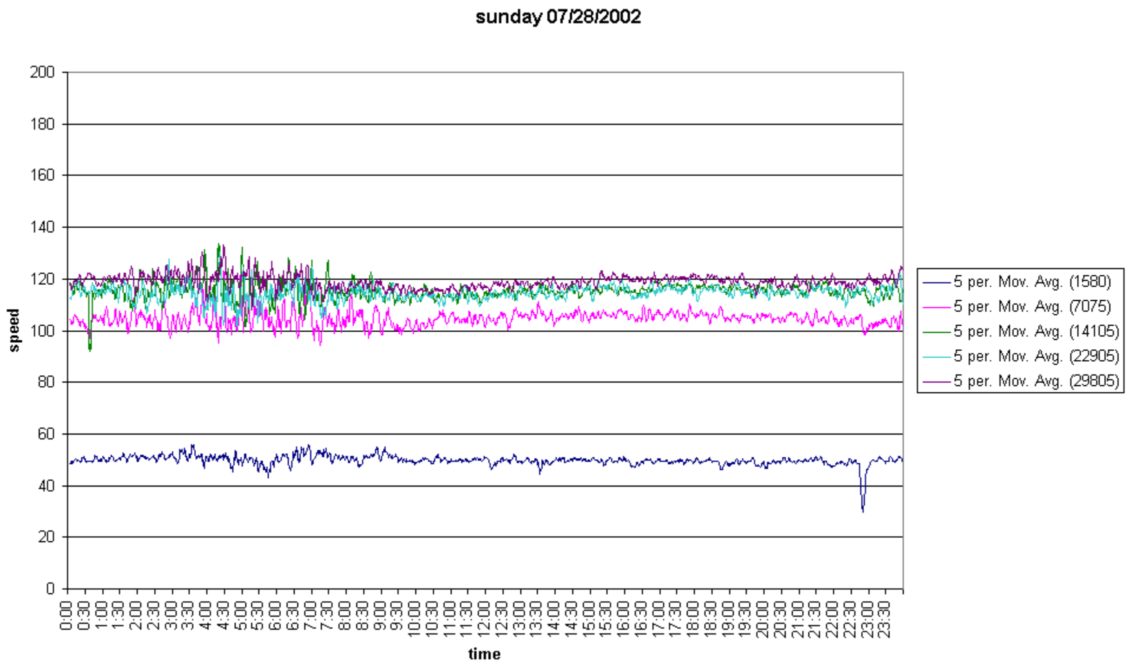


Figure 4.26 Traffic on Sunday 28th of July 2002

Conclusion
Bad weather, like snow and fog, does result in traffic jams. Warm and sunny weather does not, at least not on this road away from the beach. Maybe the same road in the other direction shows traffic congestion with warm and sunny weather. The conclusion can be drawn that weather can have a certain impact on the conditions on the road.

4.5 Special events

There are two kinds of holidays, school vacation and days of celebration, like Easter and Christmas. Some of these days of celebration are in a period of vacation. This section will discuss the possible influence of these periods on the situation on the road. Also the impact of major events and accidents will be discussed here.

Vacation
During school vacations often at least one of the parents will be at home as well. If both parents are working, one of them has to take a certain period off. Besides that, most families are going away to foreign countries for a few weeks. Figure 4.16 shows that there is much less congestion on the road on 22nd of July 2002 than on the 1st of July 2002, displayed in Figure 4.15. The 1st of July the schools are still open. The kids are going to school and most people are still working. The summer vacations starts in July and therefore on the 22nd the schools are closed. The traffic congestions disappear.

Only on the first day of a school vacation the risk of traffic jams rises. Most people leave for their vacation, as soon their children are free from school. A while ago the Dutch government has changed the school vacation system and divided the country in three parts. All the schools in the same area get their vacation in the same period. Between the areas the start dates of the vacation differ one or more weeks. This reduces the risk for traffic jams at the start of the vacation.

Day of celebration

Monday the 1st of April 2002 is second Easter Day. Most of the people don't have to go to work. Figure 4.27 displays the traffic on this day. If this figure is compared with Figure 4.12, a normal Monday a week later, it can be seen that no congestion is present at second Easter Day, but a week later the traffic jams are back again.
Most of the time days of celebration cause a weekend to extend with one or more days. Some people like to go away for a few days in that time. This increases the risk of traffic congestions on the afternoon this extended weekend starts. This does not indicate that traffic congestions have to occur. It also depends on the weather. Experience of a lot of people tells that more people are leaving if the weather is good.

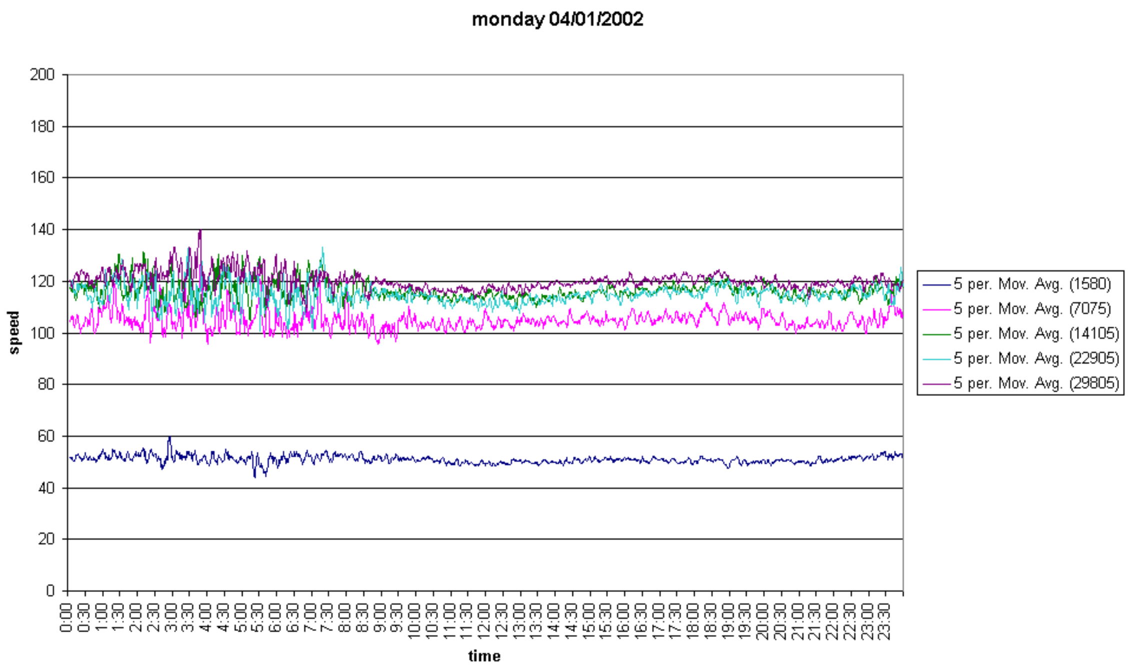


Figure 4.27 Traffic on Monday 1st of April 2002

Events

Before an event has any influence on the average speed, it has to satisfy certain conditions. The most important condition is that it has to be a large event, for example a festival, a large concert or a football match from Ajax or Feyenoord in respectively Amsterdam or Rotterdam. A lot of people have to come to that place at the same time. Another point is that most people must come by car, so people must be able to reach the place of the event very well by car and there has to be a lot of parking space. Now most of those places are situated near a train station or other kinds of public transport to prevent large traffic jams. Sometimes this cannot be prevented. In Figure 4.28 the traffic is displayed on a day it was North Sea Jazz Festival, the 13th of July. As can be seen from 1.00 PM until 6.30 PM it a large traffic jam took place. Now it can be discussed whether the festival or something else caused this congestion. The festival started around 6.00 PM. The most likely reason is that people wanted to be early to have good seats. Another reason can be that an accident had happened, but this is probably not a very good explanation. A traffic jam caused by an accident will not last that long.

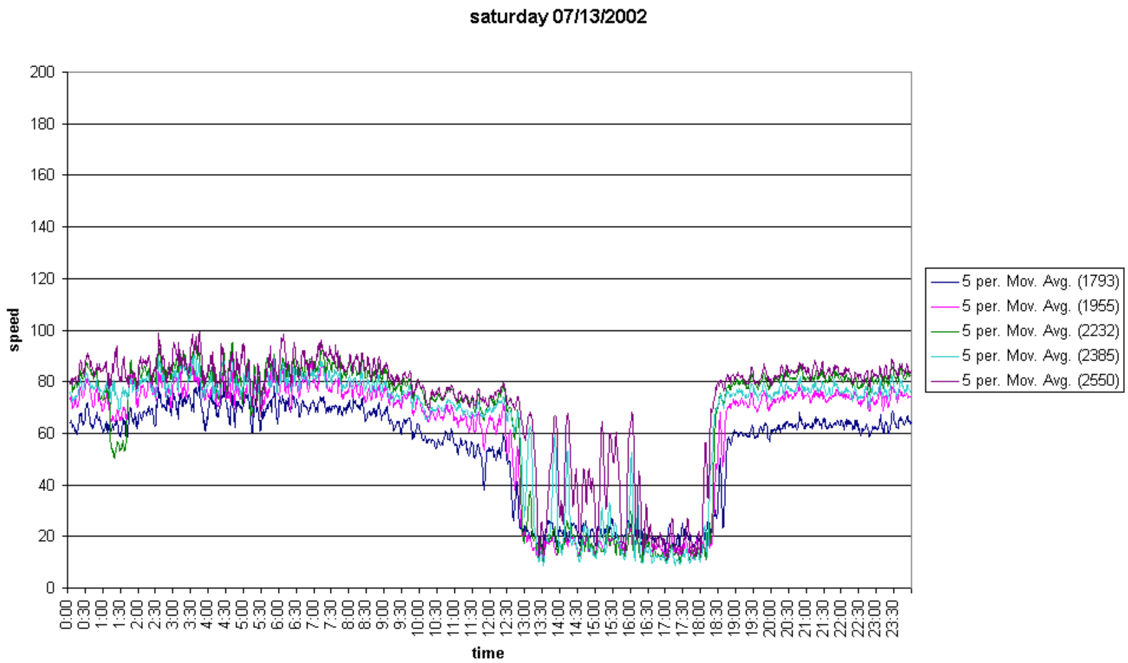


Figure 4.28 Traffic on Saturday 13th of July 2002

Accidents

Accidents are not predictable. It isn't possible to tell when and where an accident will happen. It can only be said that the probability that collisions will occur increases in some situations. When there is more traffic on the road the risk of collisions is much larger than when there is almost no one there. Also weather causing slipperiness increases the probability of accidents. One thing is certain. An accident causes a traffic jam. The size of the accident determines the size of the jam. Most of the time one or more riding tracks have to be closed and all the traffic has to be send through a sort of funnel.

Conclusion
It doesn't matter whether it is vacation or a day of celebration. When most people are free from work or away, the traffic jams will disappear. This information shows that holidays influence the situation on the road.
Accidents aren't predictable. It is only possible to say when the risk of collisions rises.

4.6 Conclusions

After this analysis it is possible to distinguish six possible influences:

- Time
- Day of the week
- Month
- Weather
- Holiday
- Events

In the sections above all these influences are discussed and we have strong indications that these factors have an impact on the average speed. The time has the largest impact on the situation on the road. People going to and leaving from their work cause most traffic jams. Also the days of the week have a major influence. A large difference can be seen between the different days. Tuesdays and Thursdays suffer more from congestion than the other weekdays. The weekend does not suffer from congestion at all except when there is extreme weather or a (special) event takes place. Another major influence is whether or not it is a holiday has a major influence. These are the days that people are free from work and traffic congestions will not occur. Major events have also a large impact on the average speed. If there is a large event to which a lot of people have to go by car then congestion can occur. The influence of the month is somewhat less clear. There are periods that more people are take days off. The situation on the road does not change its behavior drastically when the number of the month changes. Maybe it is a more influencing parameter if it is also known how far the month has passed. This can be done by using the day of the date, like the 24th, or the start middle of end of the month. It wasn't possible to see the influence of the weather very clearly, only extreme weather as showed in the figures did have a major influence on the situation. Just using this factor while predicting can show what the influence on the prediction will be.

Chapter 5

Neural networks

5.1 Introduction to neural networks

The main branch of Artificial Intelligence (AI) research in 1960s-1980s produced Expert Systems [ANN1]. These systems use logic rules for reasoning. The problem with this approach is that one has to define these rules. So all factors that may influence the outcome of a process and the relations between them have to be identified. Such knowledge is often not available or the intuition of the system designer may not be correct. Therefore it would be appealing if a system could learn the complex relations between its input features and its output itself simply from seeing examples of the input factors and corresponding outputs. The human mind is obviously capable of doing so; therefore researchers thought that simulating the structure of the human mind might be the way to build truly intelligent systems. The neural network was born.

The most basic element of the human brain is a specific type of cell, which provides us with the abilities to remember, think, and apply previous experiences to our every action. These cells are known as neurons; each of these neurons can connect with up to 200000 other neurons. The power of the brain comes from the numbers of these basic components and the multiple connections between them.
All natural neurons have four basic components, which are dendrites, soma, axon, and synapses. Basically, a biological neuron receives inputs from other sources, combines them in some way, performs a generally nonlinear operation on the result, and then output the final result. Figure 5.1 below shows a simplified biological neuron and the relationship of its four components [ANN2].
The result is then given as an input to other neurons.

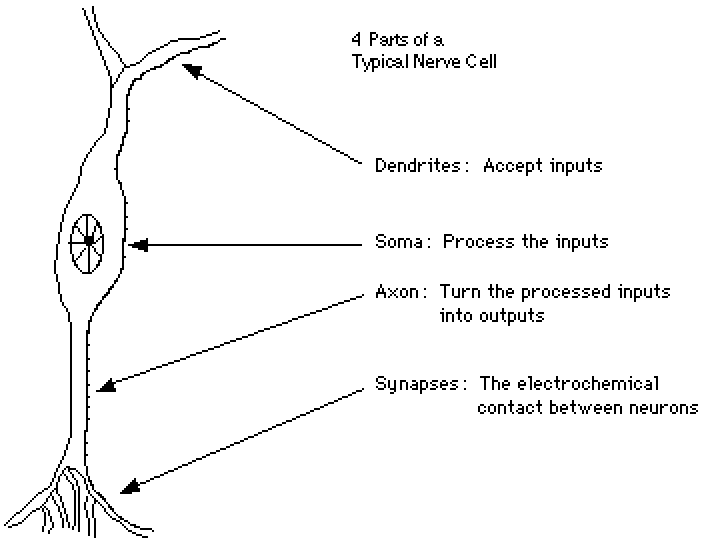


Figure 5.1 A simplified biological neuron

A neural network is also build of several simple elements called neurons. Each of these neurons is linked to some of the other elements. Information is transmitted through these links (connections). A neuron has input parameters and output parameters and an activation function. Each connection has a certain weight. The input parameters are multiplied with the weight of the connection and the parameters are summed. This result is put through the activation function, which yields a certain output. The activation function must represent the nonlinear operation as done in the brain. It transforms the activation level of a unit (neuron) into an output signal. In Figure 5.2 an example of a neuron is displayed.

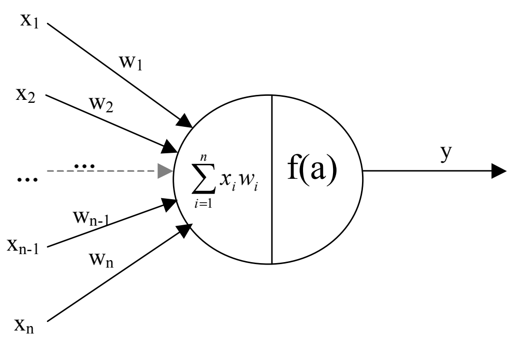


Figure 5.2 A neuron

There are several activation functions. The most common used function is the sigmoid function. This function is shown in Figure 5.3 and has the following formulas stated in (5.1).

$$f(a) = \frac{1}{1+e^{-a}} \tag{5.1}$$

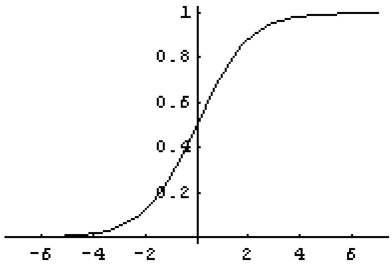


Figure 5.3 The sigmoid function

Some people like to compare a neural network to a coffee filter. Pouring water on the top results in coffee coming out at the bottom [ANN3].
Input is put into a neural network. Several calculations will be done, while the input parameters flow down through all the neurons. Eventually the parameters will reach the output, which will give the result. In Figure 5.4 a simple neural network is given.

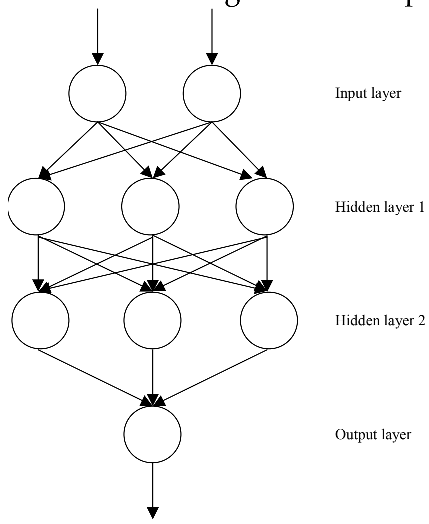


Figure 5.4 A simple neural network

The network of Figure 5.4 is a feedforward neural network. This is a neural network with a distinct layered structure, with all connections feeding forwards from inputs towards outputs.

5.2 Training of neural networks

In the previous section a neural network is described, but how is it possible that for a certain input the neural network gives the right output, how do these connections get the right weights? Training of neural networks takes care of this. To train a neural network a large amount of examples of input and output parameters must be available. These examples are divided in three independent sets, a training set, a validation set and a test set:

- The training set is used to train the network. The error on this dataset is minimized during training.
- The validation set is used to determine the performance of a neural network on patterns that are not trained during learning.
- A test set is used for the for the final check for the overall performance of a neural net.

The learning should be stopped in the minimum of the validation set error, because this is the point that the network generalizes the best. When learning is not stopped, overtraining occurs and the performance of the net on the whole data decreases, despite the fact that the error on the training set still gets smaller. The principle of overtraining can best be explained using Figure 5.5. The green dots represent data from the training set; the orange dots represent the data from the validation set. The function that must be approximated is the blue linear line. This is the line that represents the training data and the validation data best. Overtraining means that the function adapts itself to the training data. The red line represents the function obtained by overtraining, it does not approximate any of the orange dots anymore. After finishing the learning phase, the net should be finally checked with the third dataset, the test set.

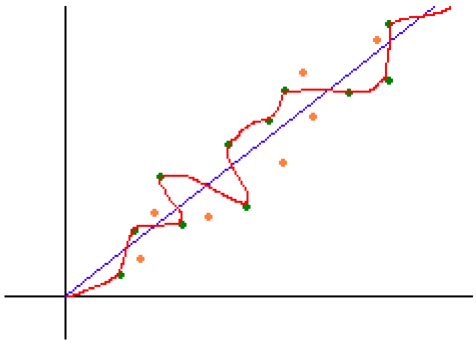


Figure 5.5 Example of overtraining

Using these examples and a trainings algorithm it is possible to adjust the weights of the connections. There are two ways of training, supervised learning and unsupervised learning. Supervised learning is done by training algorithms which adjust the weights in a neural network by executing the network on input cases with known outputs, and using the error between the actual and target outputs to adjust the weights. Training algorithms, which adjust the weights in a neural network by reference to a training data set including input variables only, belong to the group of unsupervised learning. These algorithms try to locate clusters in the input data. Supervised learning is the most common way of training. The backpropagation algorithm is an example of supervised learning and will be described here.

5.2.1 The backpropagation algorithm

At first before using the algorithm it is necessary to give all the weight random generated values. After giving these values it is possible to start with the algorithm. The algorithm has one parameter, a learning rate. This parameter controls the step size when weights are iteratively adjusted.

The first step of the algorithm is to choose an input/output pair from the training set. The input is then propagated forward in the net until activation reaches the output layer. In the second step the output of the output layer is compared to the teaching output. The error (desired output (o_d) – output given by the neural network (o_n)) has to be propagated backwards through the network. For each neuron an error term has to be determined (δ_x). w_{ij} indicates the weight between neuron i and j and a_i the activation of neuron *i*. net_i is defined by the formula in (5.2) where a_i is the activation of *i*. The functions to calculate the error terms for the output and the hidden neurons can be seen in respectively (5.3) and (5.4).

$$net_i = \sum_{i=1}^{n} w_{ij} a_i \tag{5.2}$$

$$\delta_x = f'_x(net_x)(o_d - o_n) \tag{5.3}$$

$$\delta_x = f'_x(net_x) \sum_{i=1}^{n} w_{ix} \delta_i \quad \text{with i the children of the hidden neuron} \tag{5.4}$$

If all the error terms are known the third step is to adjust the weights using the formula in (5.5). x_i represents the activation of the preceding unit neuron *i* and η the learning rate.

$$w_{ij} = w_{ij} + \eta \delta_j x_i \tag{5.5}$$

These steps must be repeated as many times until an acceptable approximation of the function it is being trained on is found.
The test set can be used to test the performance of the neural network.

5.2.2 The parameters of the backpropagation algorithm

The backpropagation algorithm is based on the principle to adjust the weights in such a way that the difference between the output of the network and the desired output is decreased. The learning rate η is used to determine the size of the steps towards the global minimum of the error throughout the training process. If this learning rate has a value that is too low this can result in longer training time. It is also more likely that the network becomes trapped in a local minimum in the error surface. On the other hand, if the learning rate is set too high the weight vector may oscillate widely from one side of the valley to the other while creeping slowly along the length of the valley tot the minimum. This effect is sometimes called cross-stitching [Vancoillie 2003].

5.3 Why using a neural network?

Now the question rises why using a neural network? What are the advantages and disadvantages of neural networks? They will be discussed here.

Advantages of neural networks
A neural network is capable of representing almost every function that maps an input vector *(x1,...,xn)* to an output vector *(y1,...,ym)*. This means that even the difficult non-linear problems can be solved using a neural network.

Building a system that represents a certain domain requires a lot of knowledge or deep understanding about this domain. Gathering this knowledge costs a lot of time. It may even be impossible to collect all the knowledge. If a neural network is used this information does not have to be available. It is trained on a set of samples, so only a good set of training samples is needed.

For neural networks there is no need to assume an underlying data distribution as is usually the case for statistical models, one also does not have to make any assumptions concerning the dependence or independence of input variables.

Training samples often contain noise. Also data can be missing in the training samples. This is not a problem while using a neural network. A neural network is capable of generalizing. This means that the network is able to give an output corresponding to input it has never seen before.

Disadvantages of neural networks
Most of the time, if the function is complex, hundreds or thousands of cycles are needed to train the network until it converges. A cycle means that all the training data is put through the network once. There can be a lot of training samples and one cycle can take a while. So training a neural network is very time consuming.

Probably the most important drawback of a neural network is that it is a black box. The network is capable of relating the input to the output, but it cannot give an explanation of the relation. The knowledge is distributed over the weights of the connections and these values are just number that have no semantic interpretation.

Generalization cannot be guaranteed because of overtraining. If the network is trained to 'good' the network adapts itself to the samples in the training set. The noise and missing data will make the network a bad representation of the model and input that the network has never seen before will result in a wrong output. Undertraining can cause similar problems. This situation can occur when the dataset is to small or contains insufficient representative examples of the problem. Then also the wrong function is learned.

New learning will overwrite the old representations that the neural network has learned. In order to keep the network up to date representing the old data as the new data as well the data sets have the be a combination of old data and new data. Otherwise the network will adapt to the new data only.

Another problem is that it is not guaranteed to find the optimal settings of the weights. This is because the training algorithm can get trapped in local peaks, and may never reach the largest peak if the multidimensional function they try to maximize (or minimize) has many local peaks. And even if the best possible weight settings are found, they may still be nowhere near what is needed for a truly useful solution to the problem. The architecture of the network may not be the right one, or the problem may simply not be solvable with neural networks.

5.4 The Architecture of a Feedforward Neural Network

A feed forward neural network is build out of layers. Each of these layers contains a number of neurons. In [Vancoillie 2003] is stated that the generalizing capabilities of a network are influenced by the following factors:

- The size of the training set and how representative it is of the problem at hand
- The architecture of the neural network
- The complexity of the problem

This problem can be seen from perspectives: the size of the training set is fixed or the architecture of the network is fixed. In this case only the first perspective is considered. In this chapter will be discussed what the architecture of a feed forward neural network should look like using a fixed size of the training set.
A feed forward neural network consists of one or more layers. Each layer contains a number of neurons. At first it has to be determined how much layers are needed to solve the problem. Secondly the question about the number of neurons in each layer must be answered. These answers can be found in the following sections.

5.4.1 How many layers are necessary?

In Section 5.3 the advantages of neural networks were mentioned. One of these advantages was that almost every input-output mapping could be performed using a neural network. The Universal Approximation Theorem, Theorem 5.1, includes this [Vancoillie 2003] and can be stated as follows:

Let $\varphi(\cdot)$ be a non-constant, bounded, monotone-increasing continuous function. Let I_{m_0} denote the m_0-dimensional unit hypercube $[0,1]^{m_0}$. The space of continuous functions on I_{m_0} is denoted by $C(I_{m_0})$. Then, given any function $f \in C(I_{m_0})$ and $\varepsilon > 0$, there exist an integer m_1 and sets of real constants α_i, b_i, and w_{ij}, where $i = 1,\ldots,m_1$ and $j = 1,\ldots,m_0$ such that we may define

$$F(x_1,\ldots,x_{m_0}) = \sum_{i=1}^{m_1} \alpha_i \varphi\left(\sum_{j=1}^{m_0} w_{ij} x_j + b_i\right)$$

as an approximate realization of the function $f(\cdot)$; that is,

$$\left| F(x_1,\ldots,x_{m_0}) - f(x_1,\ldots,x_{m_0}) \right| < \varepsilon$$

for all $x_1, x_2, \ldots, x_{m_0}$ that lie in the input space.

Theorem 5.1 Universal Approximation Theorem

This theorem can directly be used for the neural networks, because the sigmoid activation function $\frac{1}{1+e^{-a}}$ is a non-constant, bounded and monotone-increasing function. Therefore it satisfies the conditions of $\varphi(\cdot)$. Theorem 5.1 leads to the following conclusions for the architecture of a neural network:

- A network has m_0 and a single hidden layer consisting of m_1 neurons. The inputs are denoted by $x_1,...,x_{m_0}$.
- Hidden neuron *i* has synaptic weights $w_{i1},...,w_{im_0}$, and bias b_i.
- The network output is a linear combination of the outputs of the hidden neurons, with $\alpha_1,...,\alpha_{m_1}$, defining the synaptic weights of the output layer.

So it is stated that only one hidden layer is needed to compute an approximation with an error ε given a training set with the inputs $x_1,...,x_{m_0}$ and an output $f(x_1,...,x_{m_0})$.

This theorem gives no guarantees that the learning time, the ease of implementation and the generalization are optimal with a single, sufficiently large hidden layer. But if one layer is capable to approximate a function to a certain accuracy, is there a reason to use more? This reason exists because the layer has to be sufficiently large. Using more hidden layers can lead to approximations with the same accuracy, but using less weights and biases. Not much is known about this issue. According to [Vancoillie 2003] if the number of nodes in the hidden layer becomes large then two hidden layers with fewer nodes could be more efficient. Also when the weight values become very large using one hidden layer and these values are bounded, because of for example hardware limitations, it is possible that more hidden layers give a more efficient neural network.

5.4.2 How many neurons must a layer contain?

It was stated in the section above that one hidden layer with sufficient large number of neurons is able to approximate any continuous function, but does not guarantee that the generalization is optimal. A very large network will fit the training data, but not the function that produces the data. So it can be said that the best size of a feed forward neural network depends on the training data. In [Vancoillie 2003] seven factors are mentioned that influence the best number of hidden neurons.

1. The number of training patterns
2. The number of input and output units
3. The amount of noise in the data
4. The complexity of the function to be learned
5. The type of hidden activation function
6. The training algorithm
7. The regularization method used

If there are too few hidden units, large training and generalization errors will occur. This is to blame on under-fitting. On the other side if too many units are used, the result will be low training errors, but high generalization errors due to over-fitting. The number of hidden nodes should be somewhere in between, but how many? No definitive answer to this question has been found yet, but several suggestions have been made about this problem. These suggestions are stated below. N_i will represent the number of input nodes, N_o the number of output nodes.

1. The suggestion of Hecht-Nielsen is based on Kolmogorov's theorem of Theorem 5.2.

    > Given any continuous function $f : [0,1]^n \rightarrow R^m$, $f(x) = y$, f can be implemented exactly by a three-layer feedforward network having n fanout processing elements in the first (x - input) layer, $(2n + 1)$ processing elements in the middle layer, and m processing elements in the top (y - output) layer. [Hecht-Nielsen 1990]

    Theorem 5.2 Kolmogorov's Mapping Neural Network Existence Theorem

    According to Hecht-Nielsen any continuous function can be implemented using a single layer neural network having $2N_i + 1$ hidden nodes.
2. 'The hidden unit should be *10%* of the sum of the input and output unit count.' (Shavlik)
3. Hush says that a good indication of the number of hidden units is $3N_i$.
4. Rippley thinks that the number should be $(N_i + N_o)/2$.
5. $2N_i/3$ is the best number of hidden nodes according to Wang.
6. Blum has no concrete solution for the problem. According to Blum the rule of thumb is for the size of the hidden layer to be somewhere between the input layer size and the output layer size.
7. The number of nodes in the hidden layer has to lie somewhere between $2N_i + 1$ and $2\sqrt{N_i} + N_o$ according to Fletcher and Goss.
8. Berry and Linoff mention that a rule of thumb is that the number of nodes in the hidden layer should never be more than twice as large as the input layer.
9. Sigillito and Hutton do not share this opinion. They say that the number of hidden nodes required usually is less than half the number of input nodes.
10. The use of $2N_i$ *or* $3N_i$ hidden units is proposed by Kanellopoulos and Wilkinson.
11. Paola gives another suggestion:

$$\frac{2 + N_o \cdot N_i + \frac{1}{2} N_o (N_i^2 + N_i) - 3}{N_i + N_o}$$

These rules of thumb are not always reliable because they ignore the amount of training data, noise in this data and the complexity of the classification. N will represent the number of training samples and the amount of noise is indicated by the symbol r.

12. Garson came with a rule that does include the amount of training samples and the noise. The noise usually lies between 5, for clean data, and 10, for somewhat noisier data. But it is possible that very clean data gives r a value of 2 and very noise data 100. The rule looks as follows:

$$\frac{N}{r(N_i + N_o)}$$

5.4.3 Conclusion

It is not possible to say which amount of neurons in a hidden layer is the best for the problem at hand. Trial and error will give this solution. However there is one conclusion that can be made: a single hidden layer should be enough to solve any problem, even though it does not always give the most optimal solution in terms of ease of implementation, generalization and learning time.

Chapter 6

Stuttgart Neural Network Simulator

6.1 Introduction

Java Neural Network Simulator (JavaNNS) is a simulator for neural networks developed at the Wilhelm-Schickard-Institute for Computer Science (WSI) in Tübingen, Germany. It is based on the Stuttgart Neural Network Simulator (SNNS) 4.2 kernel, with a new graphical user interface written in Java set on top of it. As a consequence, the capabilities of JavaNNS are mostly equal to the capabilities of the SNNS. Some complex, but not very often used features of the SNNS (e.g. three-dimensional display of neural networks) have been left out. Some new features, like the log panel, have been introduced in JavaNNS.
Neural networks can be built using this program. The functionalities are discussed in more detail in the next section.

6.2 The most important functions of JavaNNS

JavaNNS makes it possible to simulate different kinds of neural networks, like the basic feedforward and the recurrent neural networks. With this program predefined networks can be used or one can be created. The networks can be trained using the different implemented training algorithms. The program also provides tools for testing the network. In this section the most important functions of JavaNNS are described shortly.
The data is fed to the network in pattern files. These pattern files are discussed in the next chapter. In that chapter also the result files are discussed. These files can be created by JavaNNS and will contain the results of the network on the data of the pattern file.

6.2.1 View and creating a network

Viewing a network

When the program is started one view of the network is open. If the file xor_untrained.net is opened there will be a network visible consisting of 4 units (neurons) and links between them, in its main part. Another view of the network can be opened through View/Network. Neurons and links have different colors, representing different values of unit activations and link weights. The colored bar on the left edge of the window shows which color corresponds to which value and can be used as reminder. The colors - and appearance in general - can be adjusted through View/Display Settings. In Figure 6.1 the network is displayed.

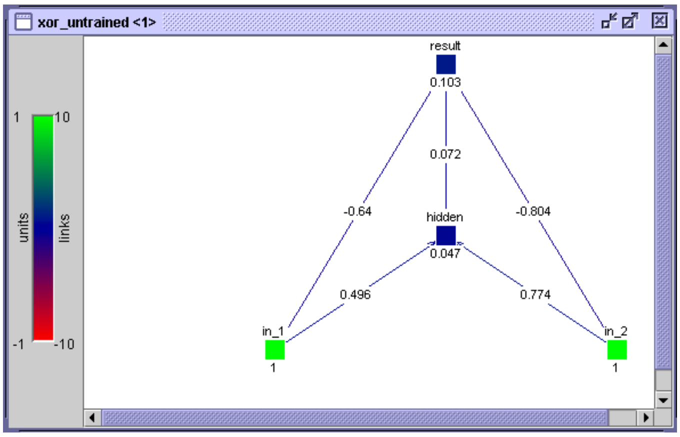


Figure 6.1 A network in JavaNNS

Creating a network

It is also possible to create a new network. With File/New loaded networks can be removed and a blank window appears. In the example directory some predefined networks can be found. Create/Layers from the Tools menu makes it able to create the layers of a neural network. The window shown in Figure 6.2 will appear. The preferred width, height and unit type can be chosen. There are several activation functions implemented in JavaNNS. In this window also the activation function of the neurons in the layer can be chosen. Pushing the button “Create” will create this new layer. This can be repeated for all the other layers that must be created. Using special unit types for example a recurrent neural network can be created. A lot of special networks can be found among the examples included with JavaNNS.
To connect the created units, use Create/Connections from the Tools menu. The window that appears is shown in Figure 6.3. In this window a few options about connecting the units can be chosen. By clicking "Connect" the chosen option will be executed.
If both steps are performed a network is created.

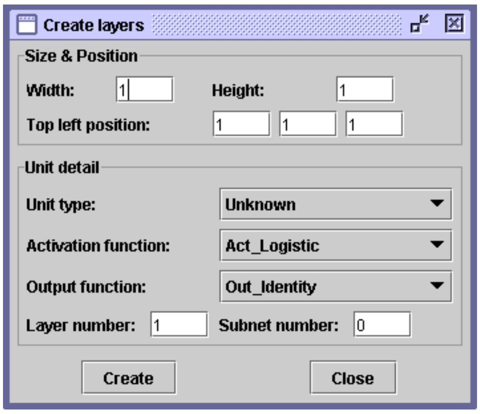


Figure 6.2 Creating layers in JavaNNS

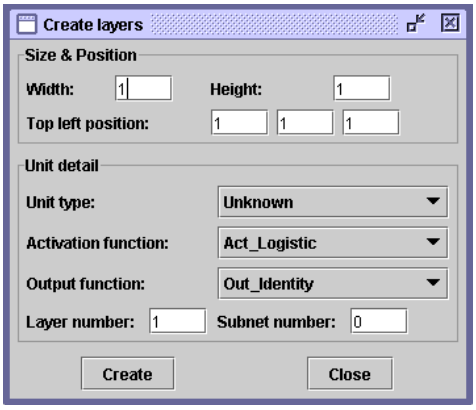


Figure 6.3 Creating connections in JavaNNS

6.2.2 Training a network

If there is a network available the network should be trained. To train a network with JavaNNS the Control Panel is needed. This panel can be found in the Tools menu and is displayed in Figure 6.4. The Control Panel is the most important window in the simulator, because almost all modifications and manipulations of the network are done through it.
Through the Error Graph window the training progress can be watched. And the Log window can give some textual and numerical information. Both windows can be found in the View menu.

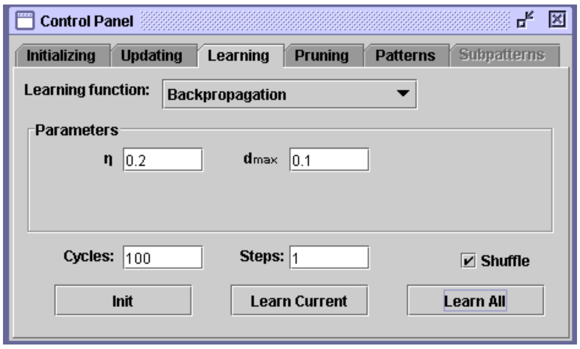


Figure 6.4 The control panel in JavaNNS

The control panel
The Control Panel is divided into six tabs, each containing controls for specific purpose. Here three tabs are discussed, the initialization tab, the learning tab and the pattern tab.
A network can be initialized through the “Initializing” tab. On this tab the initializing function can be selected and the parameters belonging to this function can be filled in. The initialization itself can be executed from two tabs, the initialization tab and the learning tab.
The learning tab makes it possible to choose the learning function, set its parameters, number of learning cycles and update steps and finally perform the learning. The classic Backpropagation is the default learning function. For each learning function default parameters are provided. Learning is performed by pressing one of the buttons: "Learn current", which performs training with the currently selected pattern, and "Learn all", which trains the network with all patterns from the pattern set.

On the pattern tab a remapping function can be chosen, but also the loaded pattern files, the training set and validation set, can be selected.

Error graph and log
During learning, the error graph displays the error curve - the type of error to be drawn is set on the left edge of the window. The error is also written into the log window. These two windows are shown in Figure 6.5.

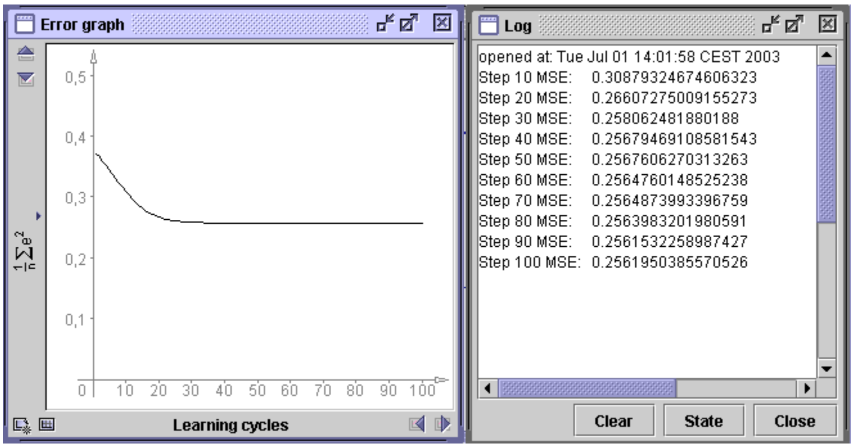


Figure 6.5 The error graph and the log in JavaNNS

6.2.3 Analyzing a network

Weights and projection panels, accessible through the "View menu", and the Analyzer can be used for getting insight into a network. These options are discussed in this section.

The projection panel
The projection panel shows activation of a hidden or output unit as a function of activations of two input units. The panel can only be opened when the three units are selected in a network view. The activations of the input units are drawn on the x- and y-axis, while corresponding activations of the output unit are represented by different pixel colors. Arrows at the panel edges are used for moving the projection window in the input space. The two buttons on in the top left corner are used for zooming, and the buttons in the bottom left corner for adjusting the view resolution. Zooming can also be performed manually, by dragging a rectangle in the projection area. An example of the panel is given in Figure 6.6.

The weights panel
The weights panel presents link weights as colored rectangles. The x-axis is used for source units and the y-axis for the target units of the links. The two buttons at the panel bottom are used for toggling grid and for auto zoom for optimal display. As in the projection panel, zooming can be performed manually. The weights panel is showed in Figure 6.7.

Analyzing tool
For analyzing the network and its performance the analyzer tool (in the Tools menu) can be used. The analyzer is used to show output or activation of a unit as a function of other unit's activation or of the input pattern. An example of the analyzer window is shown in Figure 6.8.

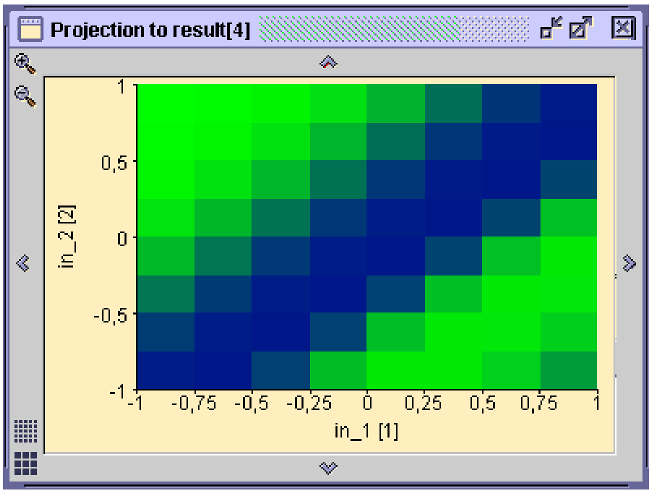


Figure 6.6 The projection panel in JavaNNS

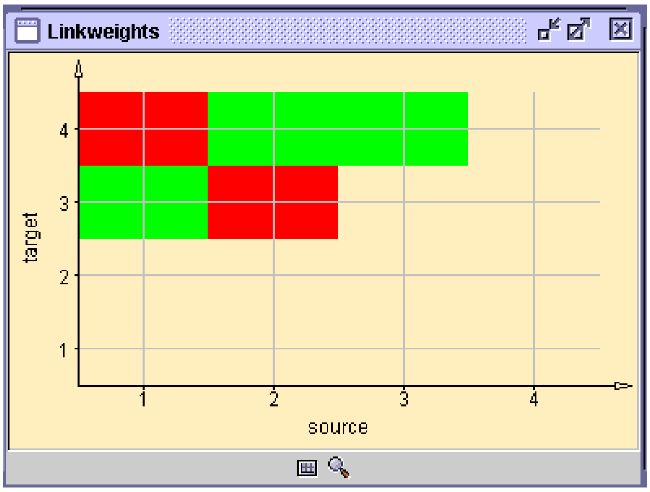


Figure 6.7 The weights panel in JavaNNS

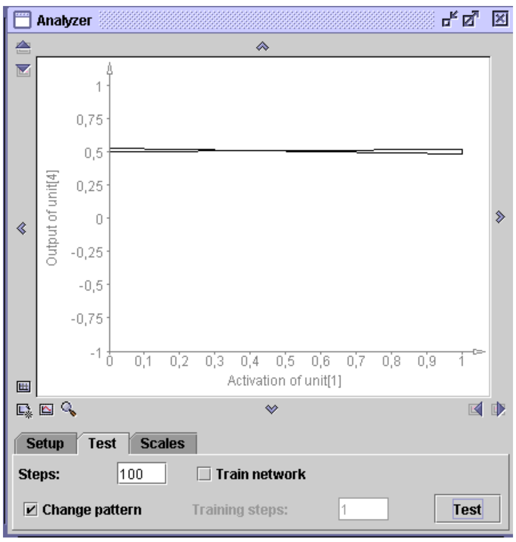


Figure 6.8 The analyzer in JavaNNS

Chapter 7

Data preparation

In Chapter 4 the influencing factors are discussed. From these factors information is known in the form as discussed in Chapter 3. But this information must be fed to the system and therefore it must be put in a format that the system understands. From the available websites the information must be extracted that will be useful for the prediction.
To train a neural network data samples are needed. The data samples consist of input and output parameters. At first it must be determined what these parameters should look like. A neural network prefers the parameters to have values between 0 and 1. To achieve this requirement the parameters have to be normalized. The program JavaNNS only accepts data samples in a pattern file. This pattern file contains the normalized parameters.

7.1 The parameters

In Chapter 4 the influence of the following parameters on the average speed is discussed.

- Time
- Date
- The day of the week
- The weather
- Events
- Holidays

These parameters must be fed to the system. The data of these parameters is available through the websites discussed in Chapter 3. The information useful for the prediction should be extracted from these sites at first. A program must be written to get the data in the appropriate format.

The system only accepts numbers, so the parameters in text must be converted to a number. This is already done for the weather (see Chapter 4). This list of weather types with their corresponding numbers can be found in Appendix A.
Time can be put into a number by seeing it in minutes. One day consists of 24 hours, each having 60 minutes. There will be 1440 minutes in one day. For example 60 minutes will represent 1.00 AM.
The date consists of three parts, the day, the month and the year. In Chapter 4 is already explained why the year is not considered. So only the day and month must be used. In one year there are 365 to 366 days, depending whether it is a leap year. So this information can be encoded in one single parameter starting with 1 for the 1st of January and so on.
The day of the week is a value in text, but how to make a number out of, for example, Monday? There are 7 different days. To put it in a number make an agreement about what the first day of the week will be and this day will get the number 1, the day next to it the number 2 and so on. In this report Monday will be considered as the first day of the week en will therefore get the number 1.
The events on the day for which the prediction must be made are a more serious problem to put in numbers. There are many different events on different places. A way to solve this problem is to divide an event in four parts, the place, the start time, the end time and the size. The place can be numbered. All places, where events can be held, must be categorized and than given a number starting again with the number 1. The start and end time can be converted in to a number in the same way the time will be. A possible solution to put the time into a number format is to categorize an event in categories from small to very large with 1 for small and 10 very large. If all these parameters have the value 0 than no event will take place on that day.
To put holidays into a number format can be very easy. It is possible to say that if there is a holiday this parameter should be 1 and 0 otherwise.

7.1.1 Normalization

The normalization must bring the original values of the parameters to values between 0 and 1. This goal can be achieved by dividing the parameter by the largest possible value for the parameter. For example normalizing the month:
The month march is the 3rd month. There are twelve months, so the maximum value for a month is 12. To bring the months to a value between 0 and 1 the number of the month must be divided by 12. In this case it will make the smallest number 0.083333 for the first month and 1 for the last one.

7.2 Pattern file

A pattern file contains a header and the data itself. The header is build of two parts, the definition of the file and the format of the file. In the definition of the file is indicated that it is a pattern file with its version number and the date on which this pattern file is generated. The format of the file consists of three layers. The first layer contains the number of patterns. The second one defines the number of input units and the third the number of output units.
The data is inserted below the header. A space, a tab or a paragraph can divide the parameters of a data sample. A data sample must contain first the input parameters followed by the output parameters. All the data samples can be put underneath each other.

The pattern file can also contain comments. To indicate that a line consists of commentary the line starts with a '#'. In Figure 7.1 an example of a pattern file is given.

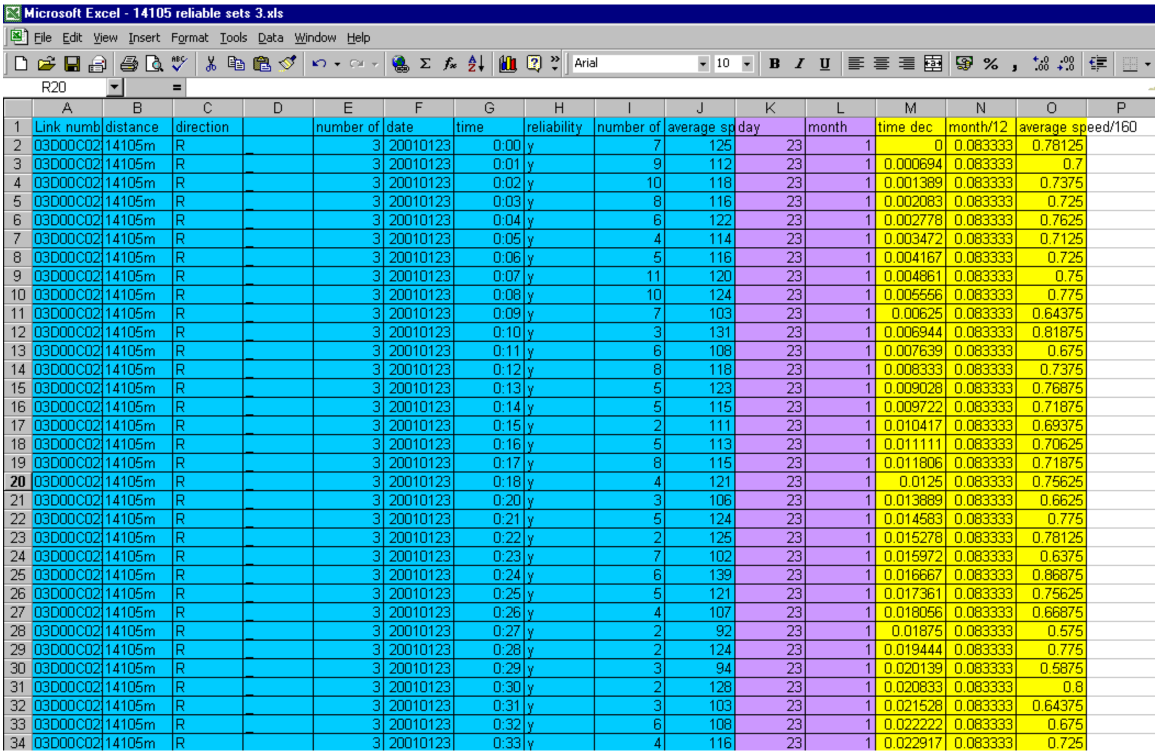
```
SNNS pattern definition file V3.2
generated at Mon Apr 25 15:58:23 1994


No. of patterns : 4
No. of input units : 2
No. of output units : 1

# Input pattern 1:
0 0
# Output pattern 1:
0
# Input pattern 2:
0 1
# Output pattern 2:
1
# Input pattern 3:
1 0
# Output pattern 3:
1
# Input pattern 4:
1 1
# Output pattern 4:
0
```

Figure 7.1 A pattern file

7.2.1 Constructing the pattern file

Before data can be put into a pattern file it has to be normalized and put in a format that can easily be put into a text file. To perform these operations on the data it is imported in Microsoft Excel. In this program copying some columns and converting those columns to the right format and applying formulas on some columns to normalize the data makes it possible to construct the columns necessary for the pattern file. These columns can than be copied and put into a pattern file with the right header. In Figure 7.2 an example is given of an excel document.

- The blue columns contains the imported data from [Regiolab Delft].
- The purple columns are copies from the 'date' column where all other information than respectively the day and the month are filtered using find and replace statements.
- The yellow columns indicate the columns that are copied into the result file. The 'time dec' column is a copy from the 'time' column and the format of that column is set to 'general' which results in numbers from 0 to 1. The month and the average speed have to be normalized by hand. For normalizing the month the value from the cell containing the number of the month is taken and this value is divided by 12. The average speed is normalized in the same way, only now the value has to be divided by a much larger number. The number chosen here is 160, because rarely the average speed is larger than 160 km/h.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Link numb	distance	direction		number of	date	time	reliability	number of	average sp	day	month	time dec	month/12	average speed/160	
2	03D00C02	14105m	R	_	3	20010123	0:00	y	7	125	23	1	0	0.083333	0.78125	
3	03D00C02	14105m	R	_	3	20010123	0:01	y	9	112	23	1	0.000694	0.083333	0.7	
4	03D00C02	14105m	R	_	3	20010123	0:02	y	10	118	23	1	0.001389	0.083333	0.7375	
5	03D00C02	14105m	R	_	3	20010123	0:03	y	8	116	23	1	0.002083	0.083333	0.725	
6	03D00C02	14105m	R	_	3	20010123	0:04	y	6	122	23	1	0.002778	0.083333	0.7625	
7	03D00C02	14105m	R	_	3	20010123	0:05	y	4	114	23	1	0.003472	0.083333	0.7125	
8	03D00C02	14105m	R	_	3	20010123	0:06	y	5	116	23	1	0.004167	0.083333	0.725	
9	03D00C02	14105m	R	_	3	20010123	0:07	y	11	120	23	1	0.004861	0.083333	0.75	
10	03D00C02	14105m	R	_	3	20010123	0:08	y	10	124	23	1	0.005556	0.083333	0.775	
11	03D00C02	14105m	R	_	3	20010123	0:09	y	7	103	23	1	0.00625	0.083333	0.64375	
12	03D00C02	14105m	R	_	3	20010123	0:10	y	3	131	23	1	0.006944	0.083333	0.81875	
13	03D00C02	14105m	R	_	3	20010123	0:11	y	6	108	23	1	0.007639	0.083333	0.675	
14	03D00C02	14105m	R	_	3	20010123	0:12	y	8	118	23	1	0.008333	0.083333	0.7375	
15	03D00C02	14105m	R	_	3	20010123	0:13	y	5	123	23	1	0.009028	0.083333	0.76875	
16	03D00C02	14105m	R	_	3	20010123	0:14	y	5	115	23	1	0.009722	0.083333	0.71875	
17	03D00C02	14105m	R	_	3	20010123	0:15	y	2	111	23	1	0.010417	0.083333	0.69375	
18	03D00C02	14105m	R	_	3	20010123	0:16	y	5	113	23	1	0.011111	0.083333	0.70625	
19	03D00C02	14105m	R	_	3	20010123	0:17	y	8	115	23	1	0.011806	0.083333	0.71875	
20	03D00C02	14105m	R	_	3	20010123	0:18	y	4	121	23	1	0.0125	0.083333	0.75625	
21	03D00C02	14105m	R	_	3	20010123	0:20	y	3	106	23	1	0.013889	0.083333	0.6625	
22	03D00C02	14105m	R	_	3	20010123	0:21	y	5	124	23	1	0.014583	0.083333	0.775	
23	03D00C02	14105m	R	_	3	20010123	0:22	y	2	125	23	1	0.015278	0.083333	0.78125	
24	03D00C02	14105m	R	_	3	20010123	0:23	y	7	102	23	1	0.015972	0.083333	0.6375	
25	03D00C02	14105m	R	_	3	20010123	0:24	y	6	139	23	1	0.016667	0.083333	0.86875	
26	03D00C02	14105m	R	_	3	20010123	0:25	y	5	121	23	1	0.017361	0.083333	0.75625	
27	03D00C02	14105m	R	_	3	20010123	0:26	y	4	107	23	1	0.018056	0.083333	0.66875	
28	03D00C02	14105m	R	_	3	20010123	0:27	y	2	92	23	1	0.01875	0.083333	0.575	
29	03D00C02	14105m	R	_	3	20010123	0:28	y	2	124	23	1	0.019444	0.083333	0.775	
30	03D00C02	14105m	R	_	3	20010123	0:29	y	3	94	23	1	0.020139	0.083333	0.5875	
31	03D00C02	14105m	R	_	3	20010123	0:30	y	2	128	23	1	0.020833	0.083333	0.8	
32	03D00C02	14105m	R	_	3	20010123	0:31	y	3	103	23	1	0.021528	0.083333	0.64375	
33	03D00C02	14105m	R	_	3	20010123	0:32	y	6	108	23	1	0.022222	0.083333	0.675	
34	03D00C02	14105m	R	_	3	20010123	0:33	y	4	116	23	1	0.022917	0.083333	0.725	

Figure 7.2 An Excel document for the pattern file

7.3 The result file

The program JavaNNS can give a result file. This result file contains the values of the pattern file and the values calculated by the neural network. An example of a result file is displayed on the left side of Figure 7.3. A result file contains a header and the data itself, just like the pattern file. Also in this case the header is build of the two parts discussed in the previous section.
To be able to import the data into Microsoft Excel the data has to be put into a different format. To change the file a small program is written. Feeding the result file to this program results in a new result file like the one on the right side of Figure 7.3. While importing a file into Excel it is possible to ignore the first lines of that file. So it is possible to ignore the header and to import only the data. Excel offers the functionality to make graphs out of the data.

```
SNNS result file V1.4-3D
generated at Fri Jun 06 14:48:33 2003

No. of patterns     : 3
No. of input units  : 2
No. of output units : 1
startpattern        : 1
endpattern          : 3
input patterns included
teaching output included
#1.1
0 0.08333
0.78125
0.73203
#2.1
0.00069 0.08333
0.7
0.73131
#3.1
0.00139 0.08333
0.7375
0.7306
```

```
SNNS result file V1.4-3D
generated at Fri Jun 06 14:48:33 2003

No. of patterns     : 3
No. of input units  : 2
No. of output units : 1
startpattern        : 1
endpattern          : 3
input patterns included
teaching output included

#1.1 0 0.08333 0.78125 0.73203
#2.1 0.00069 0.08333 0.7 0.73131
#3.1 0.00139 0.08333 0.7375 0.7306
```

Figure 7.3 A result file generated by JavaNNS and a modified result file

Chapter 8

Model

In this chapter a model for predicting the average speed based on the factors discussed in Chapter 3 and Chapter 4 is presented.

8.1 Introduction

Traffic congestion is among others depended on the traffic load and the capacity of the road. If the traffic load is larger than the capacity a traffic jam will occur. In theory the traffic load at a given place x can be determined by the traffic on all roads leading to x. Unfortunately the problem is not as easy as it seems. As an example take a look at Figure 8.1. In this figure the Prins Clausplein is shown. This is the intersection of the A12 with the A4.
A lot of traffic goes from one road to another on this point. On the A12 there are, besides this intersection with the A4, also a lot of other on- and off-ramps. It has to be determined how much traffic will leave the road or enter the road.
The literature of Chapter 2 is about travel time prediction based on models mostly using the situation on the road. In these papers the travel time is predicted using neural networks. Using a neural network is a black box approach; therefore no complex model, like described above, has to be constructed. However these forecasts only look at the situation on the road, like the number of cars and average speed.
In these researches no factors influencing the travel times are taken into account. By using the factor time for example the trends of traffic congestion during peak times are incorporated in the system. This can make it possible to predict travel times over a long time period.
In Chapter 4 some factors are described that influence the average speed on the road. Using all these factors as parameters while predicting the average speed may lead to a more accurate and reliable prediction.

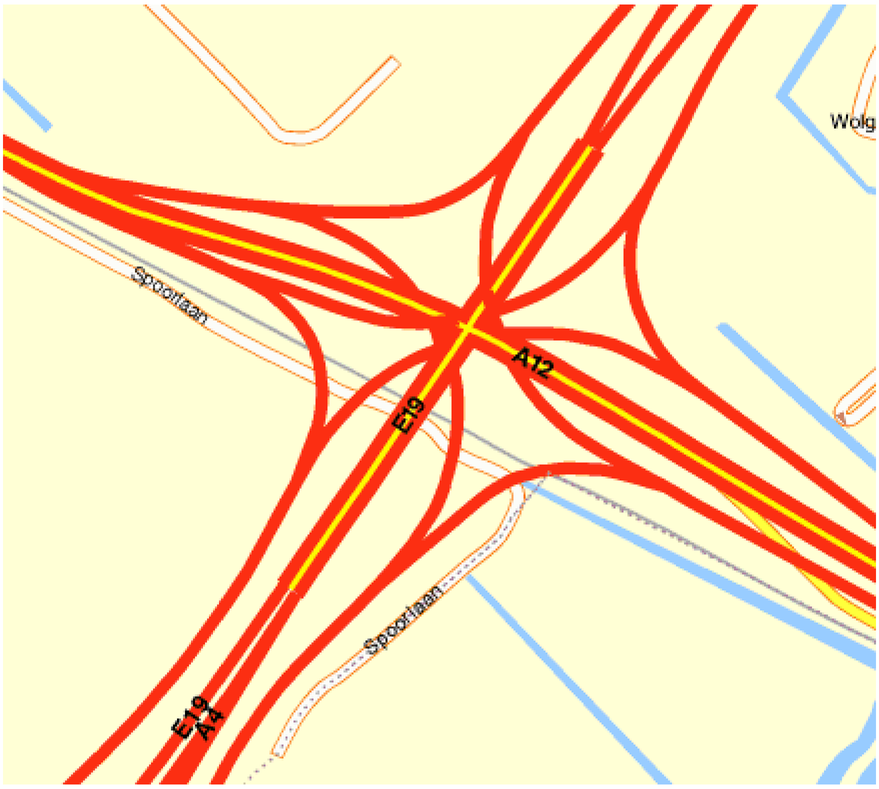


Figure 8.1 The Prins Clausplein, intersection of the A4 and the A12

There are three possible ways to predict the average speed on a certain point of the road. An expert system can be used or a mathematical model or, like the papers of Chapter 2, a neural network. The first option will result in a lot of rules, which are very difficult to establish. It is also complicated to define the interaction between the influencing factors. Also a mathematical model is difficult develop because of the complexity of the problem. Since a neural network can be used in a black box approach this seems the best solution for this problem.
The next section will discuss the parameters that can be used for the prediction. After that three possible models will be introduced.

8.2 Parameters

The goal of the system is to predict the average speed on a specific point of the road. The average speed on the road is influenced by several factors. In Chapter 7 is discussed how to put these influences into parameters in number format. The parameters from these influences will be:

- Time
- Date
- The day of the week
- The weather
- Place of the event
- Start time of the event
- End time of the event
- Size of the event
- Holidays

Using these parameters may result in a reliable prediction. However there may be some problems with these parameters, because they have varying measurement levels ranging from metric, ordinal to normal data. For example the day of the week is represented by 1 to 7. A 1 represents the Monday and a 7 the Sunday. The number 6.5 has no meaning at all. This kind of data is called nominal data. To avoid problems with artificial neural networks using this kind of data, a new neural network can be trained for each possibility. So a network can be made for each day of the week. Also event is a nominal value, because it is possible that no event at all will take place. Therefore two neural networks can be trained to represent the days with or without an event. In case of an event the model used will not be the same as for the other networks. The parameters for the event will be added to the model. In this chapter the models contain also the nominal values, but these are shaded.

8.3 Architecture

8.3.1 The first model

The first model is the most basic model. All the parameters of the possible influences, which are stated in the previous section, are fed to the predictor. So the predictor tries to forecast the speed on a certain time t and place given this time, the day of the week ($d(t)$), the date ($da(t)$), the weather ($w_x(t)$), the place of the event ($p_e(t)$), start time of the event ($s_e(t)$), the end time of the event ($e_e(t)$), the size of the event ($si_e(t)$) and holiday $h(t)$. In Figure 8.2 the model is displayed. The gray boxes indicate parameters that are not taken into account while testing the model.

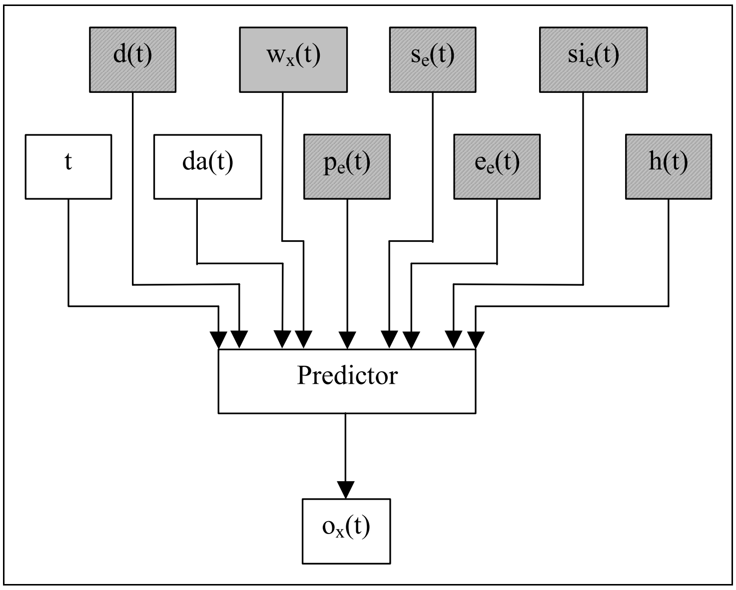


Figure 8.2 The first model

8.3.2 The second model

The model presented in the section above does not look at the previous situation on the place on the road for which the prediction is made. It only uses the data of time t. But because the data shows certain trends it might be helpful to use a time series of average speed. A time series is a sequence of measurements, typically taken at successive points in time. Figure 8.3 shows the model as it was presented in this section.

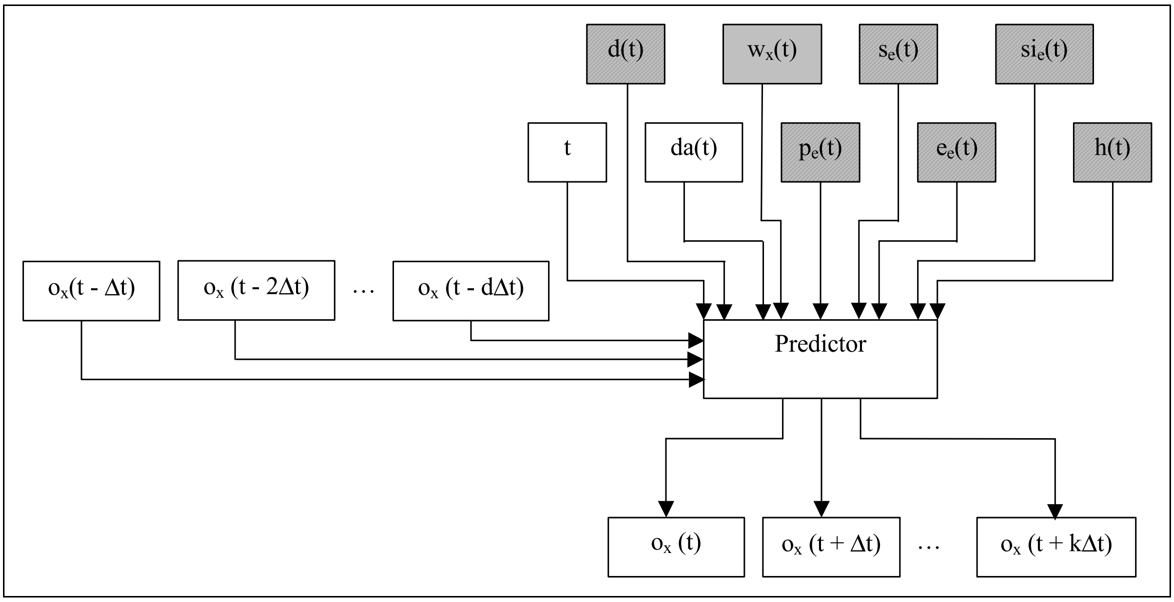


Figure 8.3 The second model using a time series

The difference between the first and the second model is that in the second model besides the parameters used in the first model also *d* values of the average speed in the past are given as input to predict *k + 1* values of the average speed. The advantage of using this model is that the first model may result in a system that will give an average speed depended on the average situation and will not be able to distinguish abnormal situations. By using the average speed known on the moment the prediction will be made it is possible that these situations are recognized resulting in a more accurate average speed.

8.3.3 The third model

In the introduction was stated that the traffic load at a given place x can be determined by the traffic on all roads that lead to x. In the literature models were introduced using the measurements of one minute before on the road to predict the travel time. However these systems are not capable of predicting the average speed into the future, because the measurements are not known yet. If these measurements are predicted over a large time interval t ahead it may be possible to predict the travel times over t + 1 minute into the future.

However it may be possible that using the time series of the average speed on previous points on the road results in a more accurate prediction than the previous two. The third model is based on this assumption, because also trends may be visible on these points and these trends may be related to the average speed on the place for which the prediction must be made. For example if there is congestion on the previous points and no congestion is visible on the place for which the prediction must be made the reason for the lack of congestion can be that the traffic density is low because of the congestion on the previous part, maybe an accident had happened and one ore more lanes are closed. In Figure 8.4 the model is shown.

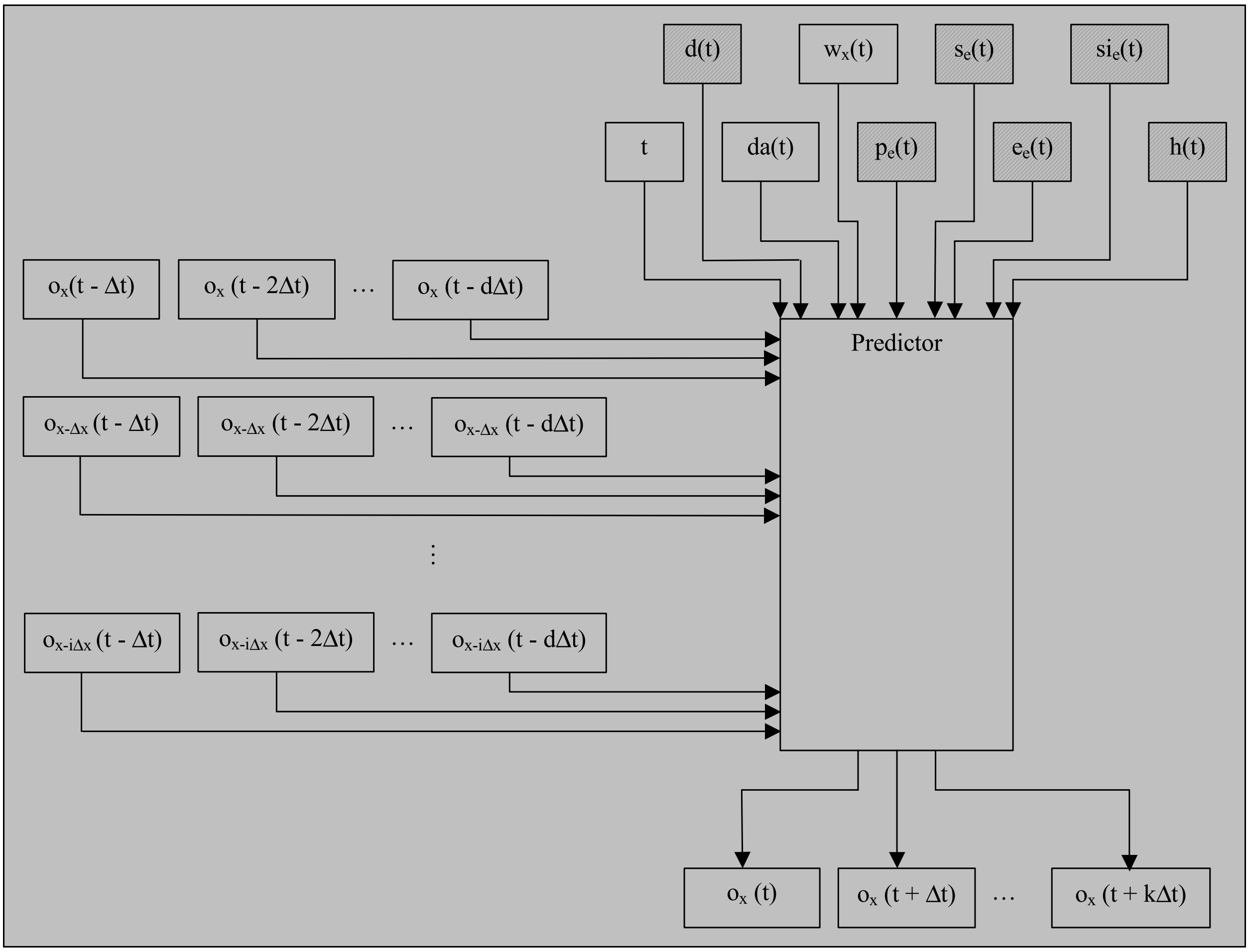


Figure 8.4 The third model using time series from more than one place

Chapter 9

Results

9.1 Introduction

In Chapter 8 three different models were introduced. One of the models, the simplest one, only uses the data of time t. The second one contains time series. In this chapter these two models will be tested. The third model includes time series of other places on the road and might improve the prediction but will not be tested yet. To perform the tests two data sets are needed, one with time series and one without.
Predicting average speed on a highway is a very complex problem. Speed is influenced by several factors. As proof of concept only the average speed on a Tuesday is forecasted. Also the factors weather, events and holidays are ignored. The reason not to take these factors in to account is that collecting this data is a very time-consuming process.
Another problem was using the date for prediction. Indicating a date by a single value between one and 365 or 366 resulted in too much data fragmentation as the number of examples per value is limited. Using two separate parameters, one for the month and one for the day resulted in similar problems, especially since only Tuesdays are used for training, which reduces the number of examples per day considerably. Therefore only the month is used as a parameter, giving a value between one and twelve.
The datasets used for the tests contain only reliable data. All unreliable is filtered out of these sets. The data where the average speed is zero because no car passed in that minute are corrected to the maximum speed as was also mentioned in Chapter 4.
For testing two places on the A12 are taken, one near Zoetermeer and one near Gouda. Near Zoetermeer congestion can be seen in the afternoon. The place near Gouda has low speeds in the morning. In Figure 9.1 the places are marked on the map.

Figure 9.1 The places '14105' near Zoetermeer and '29805' near Gouda on the A12

In this case a feed forward neural network is chosen. This network is trained using the back propagation algorithm.

The architectures of 5.4 will be followed during testing. At first the model without the time series is tested. The results of the predictions for Zoetermeer and Gouda will be discussed in Section 9.2. In Section 9.3 the results of the forecasting for Zoetermeer and Gouda with time series is reported. In Section 9.4 it will be tried to make the best prediction from Section 9.2 and 9.3 even better by tuning the parameters of the model. Section 9.5 will contain the conclusions.

For each network the summed squared error, the mean squared error, variance and standard deviation are calculated to show the accuracy of the network. The formulas are listed below in respectively (9.1), (9.2), (9.3) and (9.4). N is the number of test cases, o_t is the target output and o_c is the calculated output.

Summed squared error (SSE) $$\sum_{i=1}^{N}(o_{ti}-o_{ci})^2 \qquad (9.1)$$

Mean squared error (MSE) $$\frac{1}{N}\sum_{i=1}^{N}(o_{ti}-o_{ci})^2 \qquad (9.2)$$

Variance (VAR) of $(o_t - o_c)^2$ $$\frac{n\sum_{i=1}^{N}(o_{ti}-o_{ci})^2-\left(\sum_{j=1}^{N}(o_{tj}-o_{cj})\right)^2}{n(n-1)} \qquad (9.3)$$

Standard Deviation (StDev) of $(o_t - o_c)^2$ $$\sqrt{\frac{n\sum_{i=1}^{N}\left(o_{ti}-o_{ci}\right)^2-\left(\sum_{j=1}^{N}\left(o_{tj}-o_{cj}\right)\right)^2}{n(n-1)}} \qquad (9.4)$$

9.2 Zoetermeer and Gouda without time series

In Section 5.4 a few formulas were introduced. Each of them estimates the number of neurons that should be used in the hidden layer. It is not proved that one of these functions is the best for all problems. This means that all the topologies resulting from the formulas should be tested in order to determine which one is the best for the problem discussed in this report. In Table 9.1 the formulas are stated and the number of hidden neurons for this problem. For both problems the number of input nodes (N_i) is 2, the time and the month, and the number of outputs (N_o) is 1, the average speed. For the point near Zoetermeer (14105) the number of training samples is 127450 and for Gouda (29850) this number is 127314. Because all unreliable data is filtered out of the dataset the data is rather clean. Therefore the unreliable factor (r) in the formula of Garson is set on *5*.

Table 9.1 Formulas to determine the number of hidden nodes and their results (without time series)

Author	Formula	Number of nodes	
Hecht-Nielsen	$2N_i + 1$	5	
Shavlik	$0.1(N_i + N_o)$	0.3 → 1	
Hush	$3N_i$	6	
Rippley	$(N_i + N_o) / 2$	1.5 → 2	
Wang	$2N_i / 3$	1.33 → 2	
Blum	between N_i and N_o	between 2 and 1	
Fletcher and Goss	from $2N_i + 1$ to $2\sqrt{N_i} + N_o$	from 5 to ≈ 4	
Berry and Linoff	$< 2N_i$	< 4	
Sigillito and Hutton	$< N_i / 2$	< 1	
Kanellopoulos and Wilkinson	$2N_i$ or $3N_i$	4 or 6	
Paola	$\dfrac{2 + N_o \cdot N_i + \frac{1}{2}N_o(N_i^2 + N_i) - 3}{N_i + N_o}$	1.33 → 2	
Garson	$\dfrac{N}{r(N_i + N_o)}$	**Zoetermeer (14105)**	**Gouda (29805)**
		8497	8488

The number of nodes necessary for this problem according to Garson is very large and it is not possible to create such a large network with JavaNNS. The topologies resulting from this formula are not taken into account while testing.

The topologies that result from the formulas above and that will be tested in this section are:

- 2-1-1
- 2-2-1
- 2-3-1
- 2-4-1
- 2-5-1
- 2-6-1

All the networks are trained on 50000 cycles of the same training data. The parameters of the back propagation algorithm are equal (learning rate η = 0.2 and maximum error d_{max} = 0.1). For each network the summed squared error, the mean squared error, variance and standard deviation are calculated. Also a figure containing the error graphs of the network is displayed for each topology. In this figure two graphs can be seen, the one on the left is the graph of the prediction near Zoetermeer and on the right one near Gouda. The black line is the error of the training set and the pink line shows the error value of the validation set. Unfortunately the values given by the error graphs cannot be used. In JavaNNS the mean squared error seems dependent on the size of the dataset, the larger the dataset, the larger the values. However the mean squared error is supposed to be size-independent. In this case this is not a problem. The error graphs can be used to show the improvements made by training. The results of the test set will be calculated using Microsoft excel.
The figures displayed are the predictions for the 12th of November 2002 for respectively Zoetermeer and Gouda. This day is chosen because it represents the normal situation for both places, near Zoetermeer traffic congestion can be seen in the evening and for the place near Gouda there is traffic congestion in the morning.

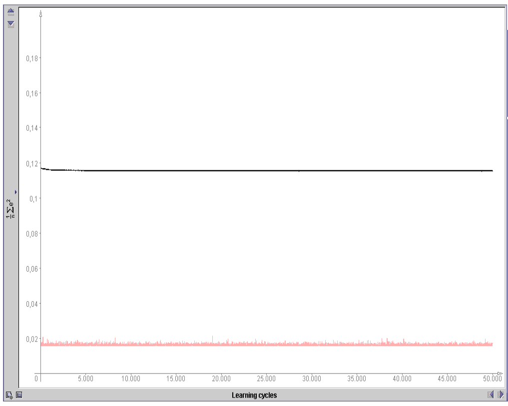
9.2.1 The 2-1-1-topology

Figure 9.2 shows the error graphs of the 2-1-1-topology. In these graphs it can be seen that at first the error drops (because of the number of training cycles is large the drop is visible as a very small slope) and after that error stays almost the same for both sets. The networks will not converge to an error of 0. In that case the black line will go to 0 and the line of the validation set will move up. Because the error drops only a little bit it can take millions of training cycles before it will converge to 0 if it goes to zero at all. In this case the training is stopped after 50000 training cycles. The difference in height of the lines in the figure can be explained by the difference in scaling. The first graph has a range up to 0.18 on the error-axis and the second one up to 0.09.
In Table 9.2 the results of the test set are listed of the 2-1-1-topology. The SSE is a lot smaller on location 29805 than with 14105. This can be explained because it is the sum of the squared errors and there are fewer examples in the test set of 29805. So the number of errors summed is smaller. Remarkable is that the MSE, the SSE divided by the number of examples, and the variance and the standard deviation are also lower. This means that the overall error is smaller for 29805.

Table 9.2 Results of the 2-1-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	389.2156	280.082	216.507	197.8741
MSE	0.023305	0.016832	0.012964	0.011891
VAR	0.02313	0.016492	0.012709	0.011818
STDev	0.152085	0.128422	0.112733	0.108712

The lower error can be explained by Figure 9.3 and Figure 9.4. Because the prediction results (the pink line) is an almost straight line and because the traffic jams are not always serious near Gouda, the predicted speed is much closer to the real speed near Gouda than near Zoetermeer. The average speed over all Tuesdays has a smaller error. Therefore it can be said that the average speed over all Tuesdays is a better prediction than the prediction made by the neural network.



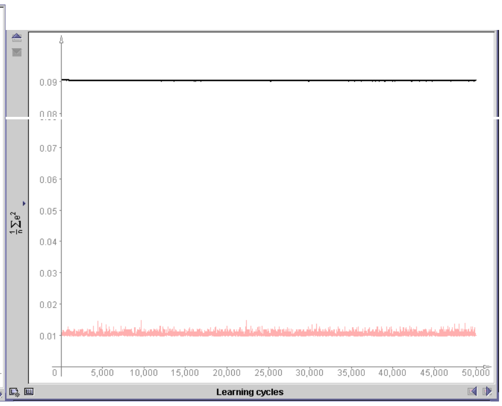


Figure 9.2 The error graphs of the 2-1-1-topology

These figures can also tell that the prediction is not accurate. If all the graphs resulting from this topology are compared, a few of them are displayed in Appendix C, then the conclusion can be made that the prediction is the same for all figures. So this topology is not capable of predicting the average speed. Probably this topology is too small to approach the complex function of this problem.
It can be said that the results listed in Table 9.2 are not very accurate, because most of the errors are made during the busiest hour of the day. Near Zoetermeer the busiest hours are from 4.00 PM to 7.00 PM. For Gouda these times are usually from 7.00 AM to 10.00 AM. The errors for these times are displayed in Table 9.3.

Table 9.3 Results of the 2-1-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	240.24672	162.92637	141.2843	116.1987
MSE	0.11138	0.0751505	0.0655	0.053597
VAR	0.0674597	0.0535165	0.054492	0.051933
STDev	0.2597301	0.2313363	0.233436	0.227887

As can be concluded from the value of the SSE most of the error is made in these times, so the assumption from above is correct. Because of this the MSE is about five times higher. Also all the other calculations are a lot higher than the values in Table 9.2. To be able to determine which network is the most accurate, important conclusions can be drawn from Table 9.3. If a network is accurate in the busiest times then the network will be able to predict the average speed of traffic jams.

From the straight line displayed in figures can also be concluded that the parameter 'time' does have an impact on the prediction. Otherwise this line will not always be the same over time.

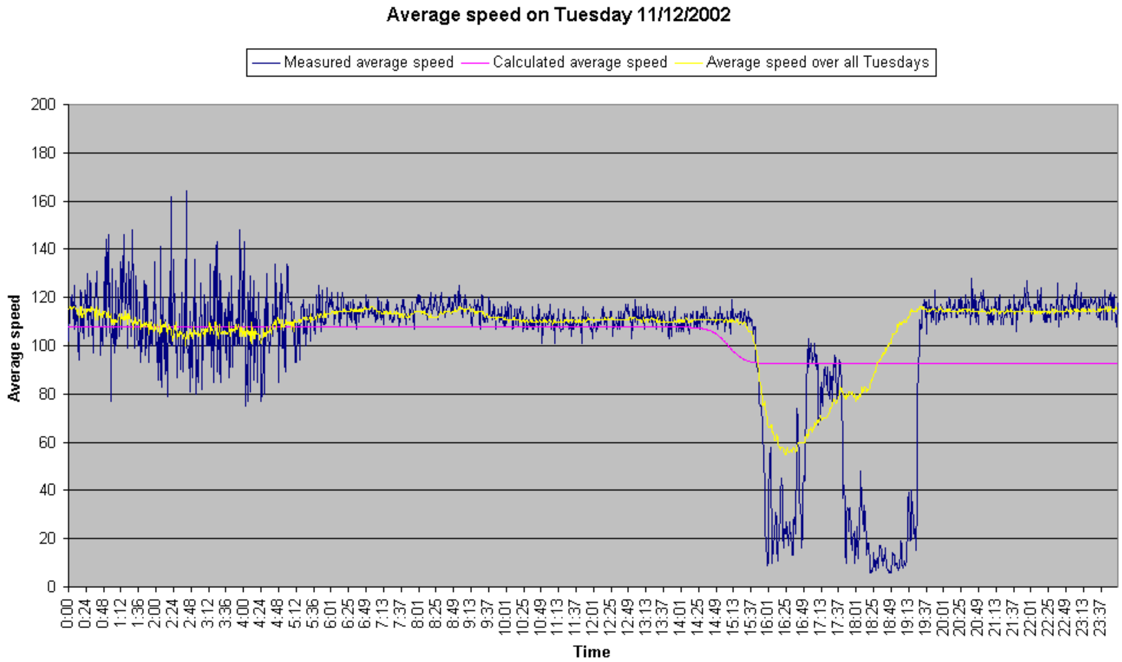


Figure 9.3 The average speed on 12th of November 2002 near Zoetermeer with the 2-1-1-topology

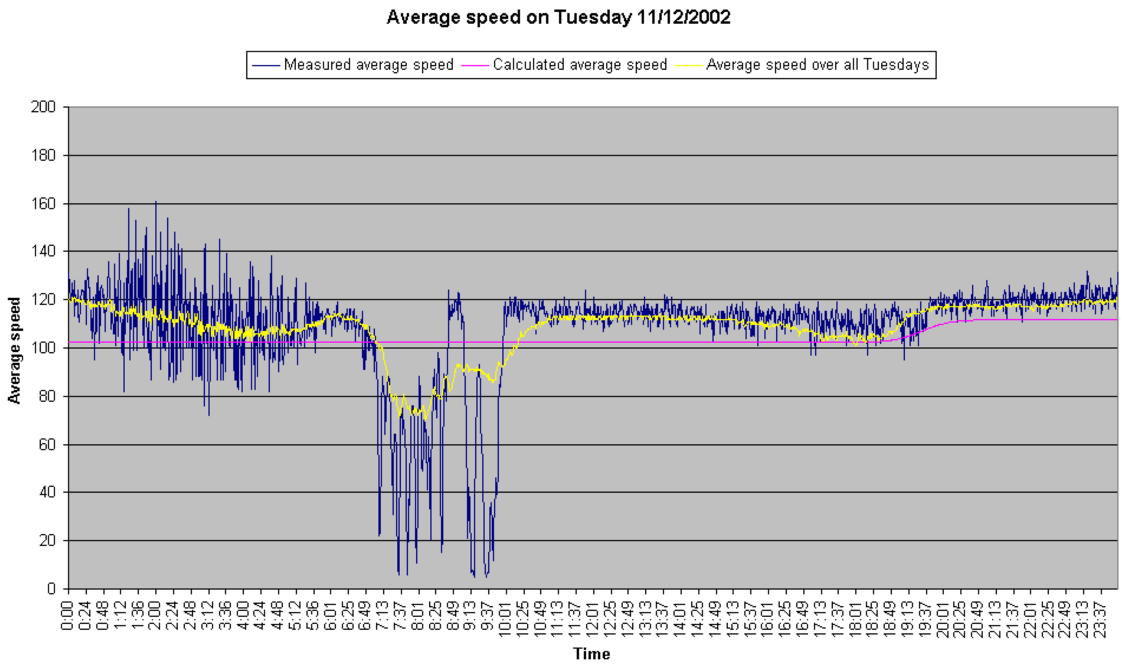


Figure 9.4 The average speed on 12th of November 2002 near Gouda with the 2-1-1-topology

9.2.2 The 2-2-1-topology

The results of the 2-2-1-network are improved for both of the networks. The results for the network are displayed in Table 9.4. The best prediction however is still the average speed over all Tuesdays.
The error graphs of these networks are shown in Figure 9.5. About these graphs can be said the same as in the previous section. The networks did again not converge to an error of 0.

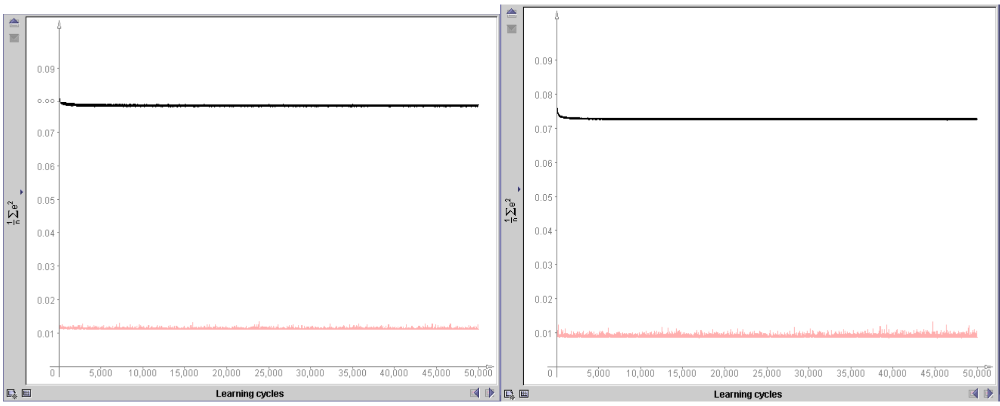


Figure 9.5 The error graphs of the 2-2-1-topology

Table 9.4 Results of the 2-2-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	241.2076	232.3884	216.507	197.8741
MSE	0.014443	0.013966	0.012964	0.011891
VAR	0.014407	0.013887	0.012709	0.011818
STDev	0.120028	0.117844	0.112733	0.108712

Table 9.5 Results of the 2-2-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	157.76194	127.01441	141.2843	116.1987
MSE	0.0731395	0.058586	0.0655	0.053597
VAR	0.0593918	0.0547429	0.054492	0.051933
STDev	0.2437044	0.2339719	0.233436	0.227887

In Figure 9.6 and Figure 9.7 one graphs for each point resulting from the test on the 2-2-1-network is shown. It can be concluded from these figures and the other figures in Appendix C that this network fits to the average speed over all Tuesdays. The yellow line displays this average speed.
Table 9.5 contains the results for the busiest times. These times are near Zoetermeer from 4.00 PM to 7.00 PM. Near Gouda it is most of the time busy in the morning from 7.00 AM to 10.00 AM. Again it can be seen that most of the inaccurateness is in the times. Because the predicted lines, displayed in the figures by the pink line, are not always the same for each day the conclusion can be made that the parameter 'month' does have a little impact on the prediction.

It can also be concluded that this network follows the line of the average speed over all Tuesdays. This can be expected because the network is trained only on time and month. If the network was trained only with the parameter 'time' an approximation of the average speed over all Tuesdays would be the result of the prediction. Time is the most influencing factor, as could be expected, while predicting the average speed. The month cannot make a real difference at least not in this topology.

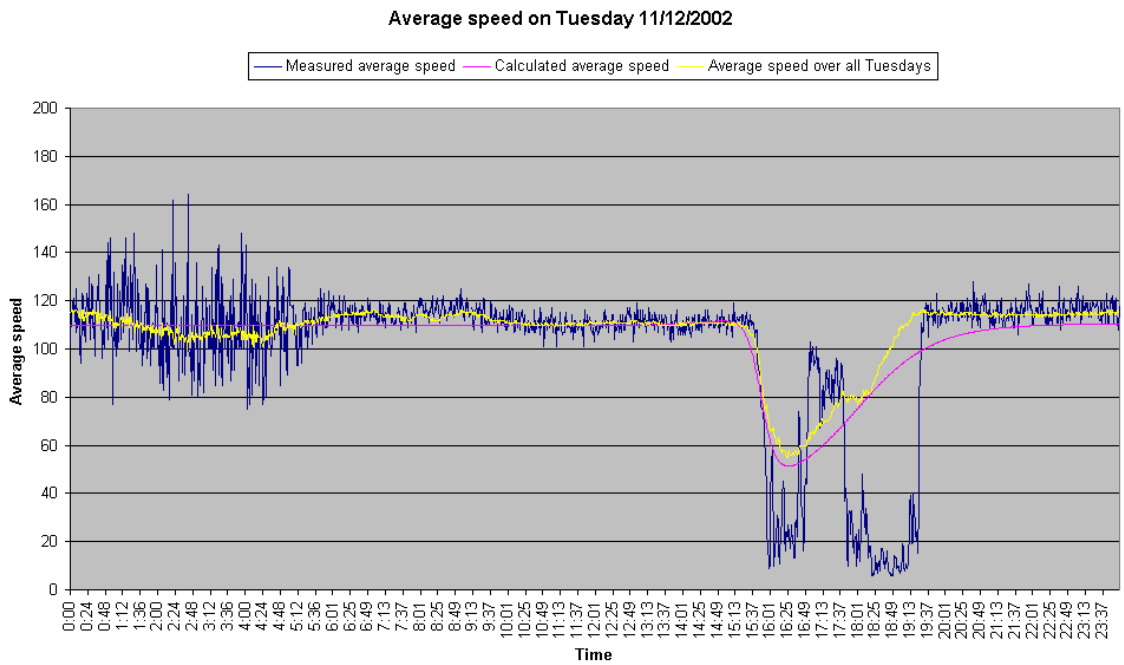


Figure 9.6 The average speed on 12th of November 2002 near Zoetermeer with the 2-2-1-topology

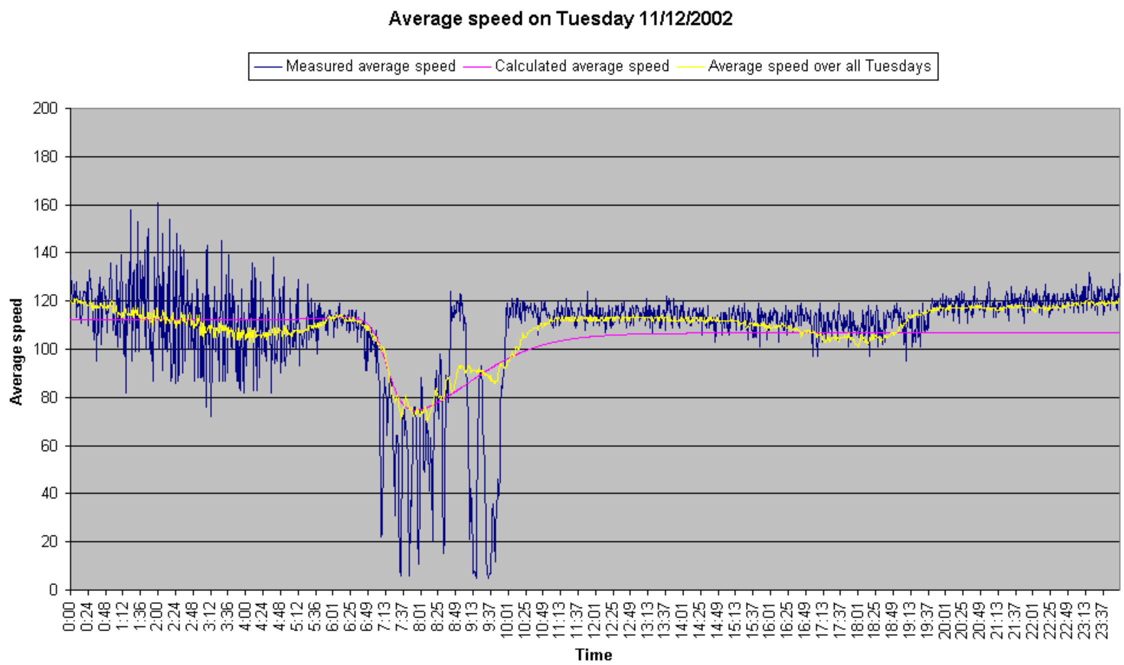


Figure 9.7 The average speed on 12th of November 2002 near Gouda with the 2-2-1-topology

9.2.3 The 2-3-1-topology

This topology has errors in Table 9.6 that are worse than those of the previous topology, especially the results for the place near Gouda. The variance and standard deviation are lower and therefore the spread of the errors is smaller. Also this topology did not give networks that will converge within 50000 training steps. Figure 9.8, the error graphs of this topology, shows again a drop at the beginning and an almost straight line during the rest of the training.

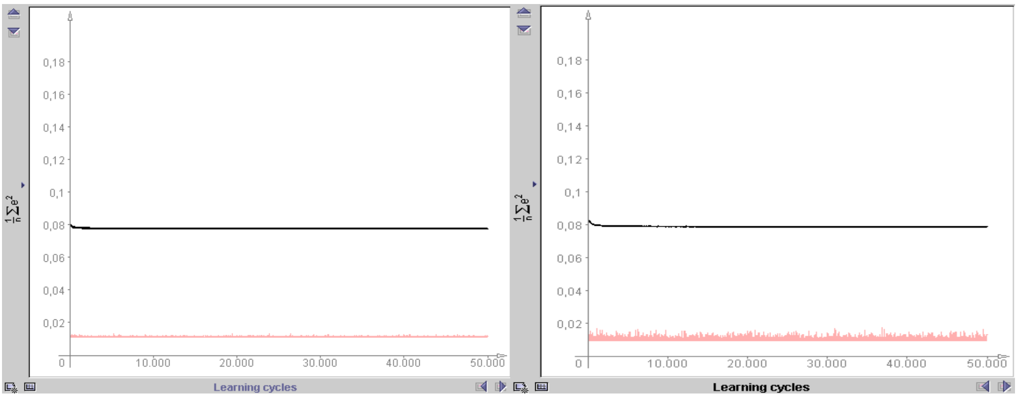


Figure 9.8 The error graphs of the 2-3-1-topology

This network compared to the 2-2-1-network gives a better prediction during the busiest hours of days. These results can be found in Table 9.7. The mean squared error is lower and the variance and standard deviation do not differ very much of those of the previous network. These results for Zoetermeer show that a better prediction is made for this place. For Gouda the error is smaller, but the variance and standard deviation are somewhat higher; so the spread of the error is a little bit wider. But the conclusion can be made that the busiest times are better predicted with this network.

Table 9.6 Results of the 2-3-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	247.7492	259.3073	216.507	197.8741
MSE	0.014834	0.015583	0.012964	0.011891
VAR	0.014106	0.013682	0.012709	0.011818
STDev	0.118769	0.116968	0.112733	0.108712

Table 9.7 Results of the 2-3-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	139.18687	119.9653	141.2843	116.1987
MSE	0.064528	0.0553345	0.0655	0.053597
VAR	0.0578023	0.0551908	0.054492	0.051933
STDev	0.2404212	0.2349273	0.233436	0.227887

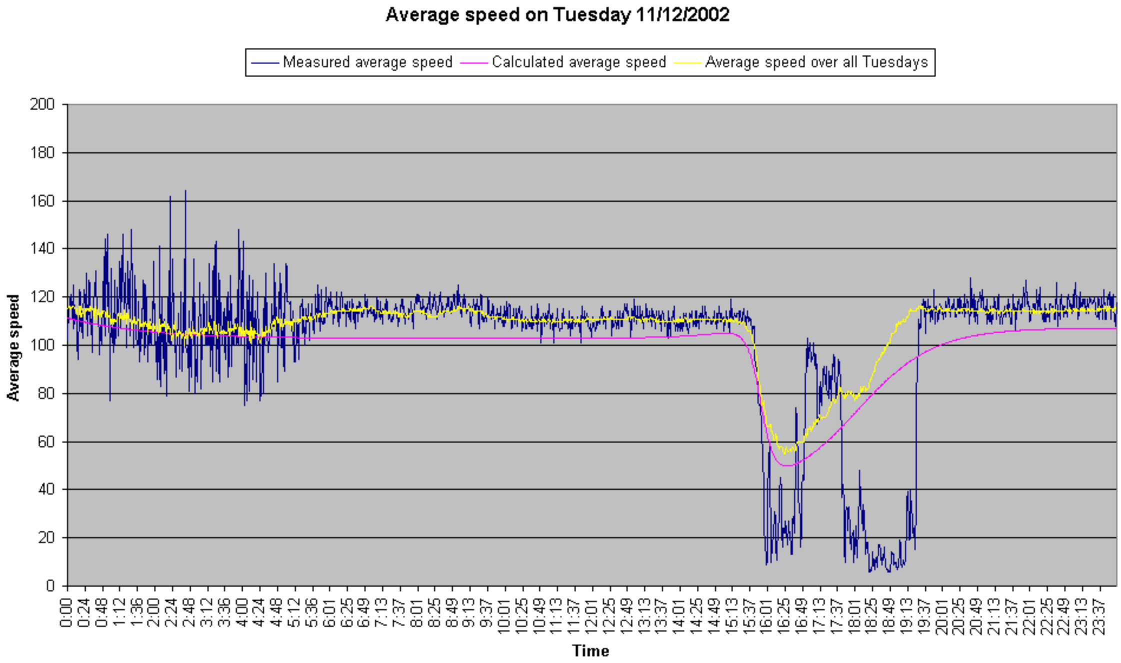


Figure 9.9 The average speed on 12th of November 2002 near Zoetermeer with the 2-3-1-topology

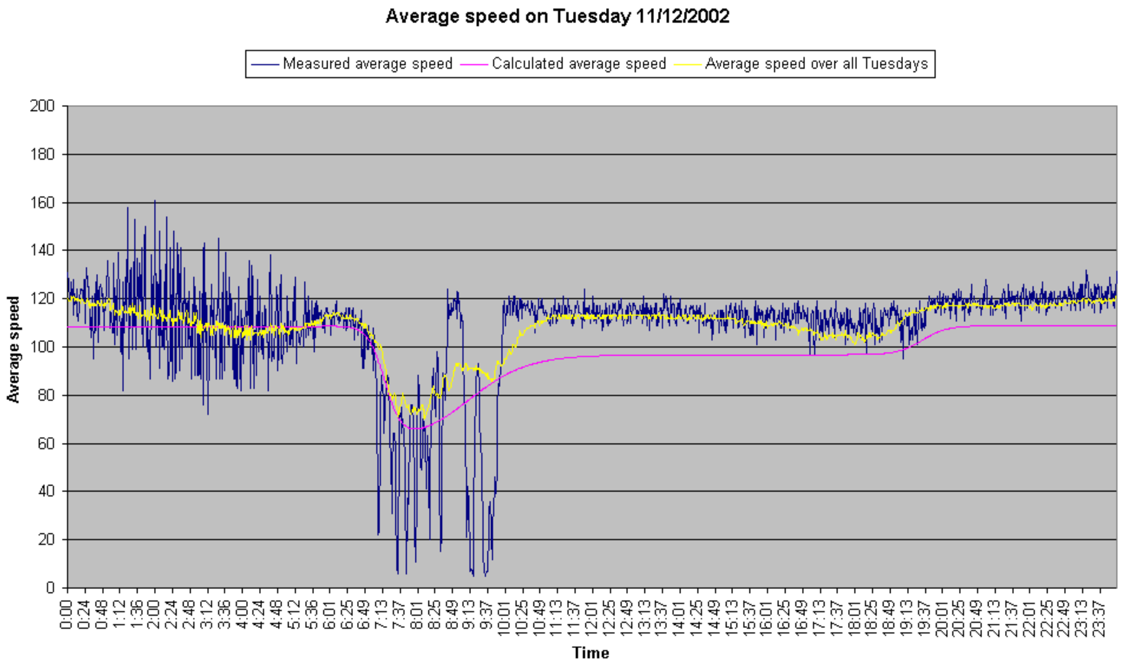


Figure 9.10 The average speed on 12th of November 2002 near Gouda with the 2-3-1-topology

The explanation for the smaller error for the busiest times for Gouda compared to Zoetermeer and the higher error for the whole prediction can be given by looking to the Figure 9.9 and Figure 9.10. If these figures are compared with respectively Figure 9.6 and Figure 9.7 it can be seen that the reason is that the pink line, the prediction, has somewhat lower values than the pink line in the figures above. This leads to a worse prediction for the quiet times, because the prediction for these times was almost accurate. However the prediction for the busiest times will improve because of the downtrend of the line. The prediction for these times was far from accurate. The predicted speed was far too high. This will make the prediction for the busiest times somewhat better. This is also the reason that the prediction for the busiest times near Zoetermeer is better than the average speed over all Tuesdays. However all the rest of the results show worse predictions than this average.

9.2.4 The 2-4-1-topology

As Table 9.8 shows the errors for Zoetermeer and Gouda are almost the same. These results show a lower error than the 2-3-1-topology, but the prediction for the situation around Zoetermeer is has a higher variance and standard deviation for both of the networks. In all cases the errors of the average speed over all Tuesdays is smaller.

Table 9.8 Results of the 2-4-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	242.5549	242.9572	216.507	197.8741
MSE	0.014523	0.014601	0.012964	0.011891
VAR	0.01434	0.014331	0.012709	0.011818
STDev	0.119749	0.119711	0.112733	0.108712

The error graphs of the 2-4-1-topology, Figure 9.11, give the same picture as error graphs in the previous sections. Also this time the difference in height of the lines is cause by different scaling.

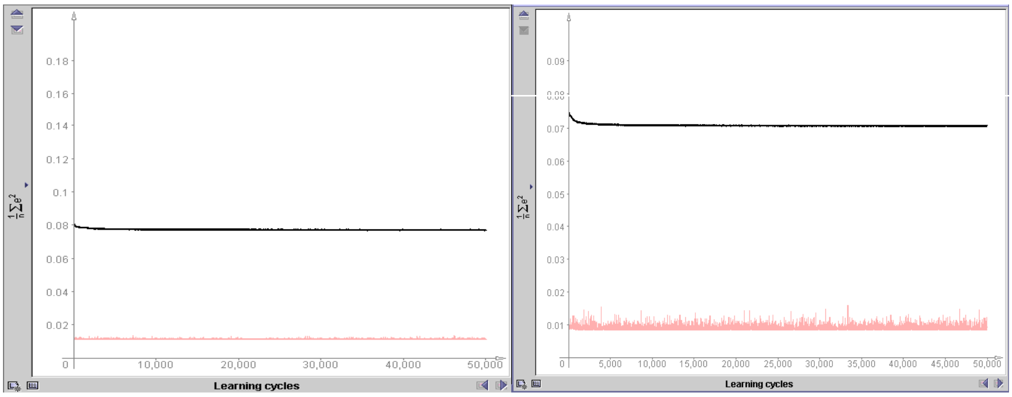


Figure 9.11 The error graphs of the 2-4-1-topology

Table 9.9 shows that the prediction for Zoetermeer is worse for the busiest times. Two third of the error is made in this small period of three hours. This also indicates a good prediction for the quiet times. For Gouda only half of the total error lies in the three-hour period of busyness; so this prediction is good for the busiest times, but does not give a good forecast for the quiet times.

The error of these networks is higher than the error of the 2-2-1-network, but the differences are small. The variance and standard deviation is lower for the prediction near Zoetermeer. If the results of the prediction near Gouda is compared to the results of the 2-2-1-network near Gouda, the variance and the standard deviation is higher for the prediction with the 2-4-1-topology.

Table 9.9 Results of the 2-4-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	161.42195	129.63354	141.2843	116.1987
MSE	0.0748363	0.0597941	0.0655	0.053597
VAR	0.0579763	0.0580849	0.054492	0.051933
STDev	0.2407827	0.2410082	0.233436	0.227887

In the Figure 9.12 and Figure 9.13 it can be seen that the prediction of this topology will follow the average speed over all Tuesdays again. Only in the case of the prediction for Gouda the prediction gives a value to low for the period after the busy times.
Maybe this difference has something to do with training. It is possible that using other weight values by the start of the training will result in other graphs. Perhaps somewhat more like the ones resulting from the 2-2-1-topology.

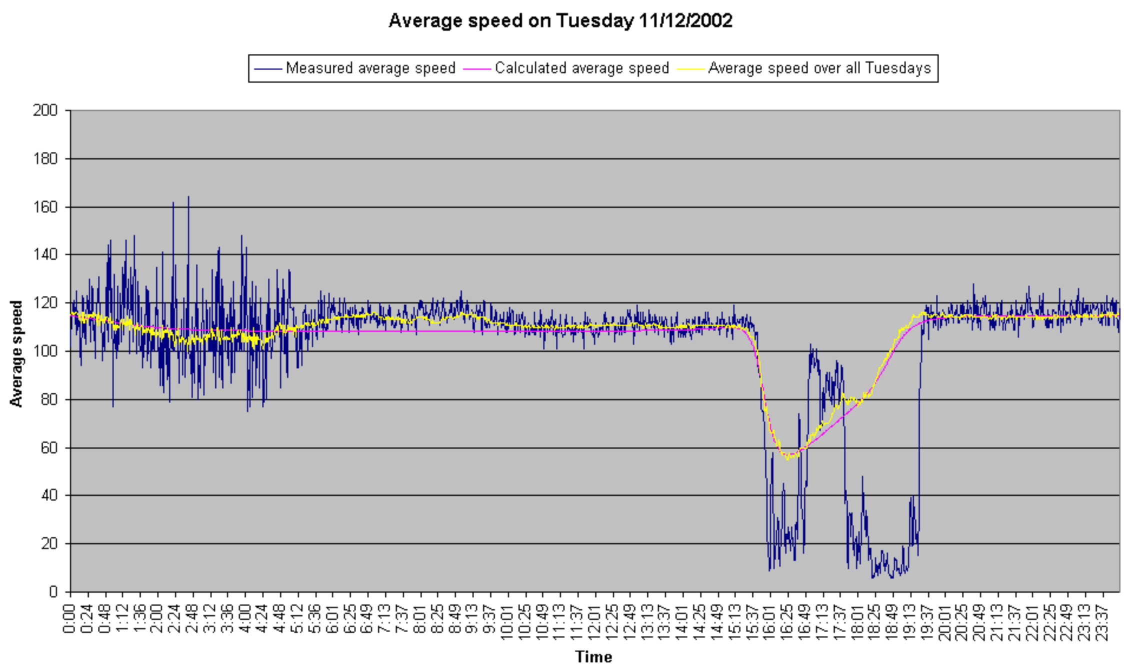


Figure 9.12 The average speed on 12th of November 2002 near Zoetermeer with the 2-4-1-topology

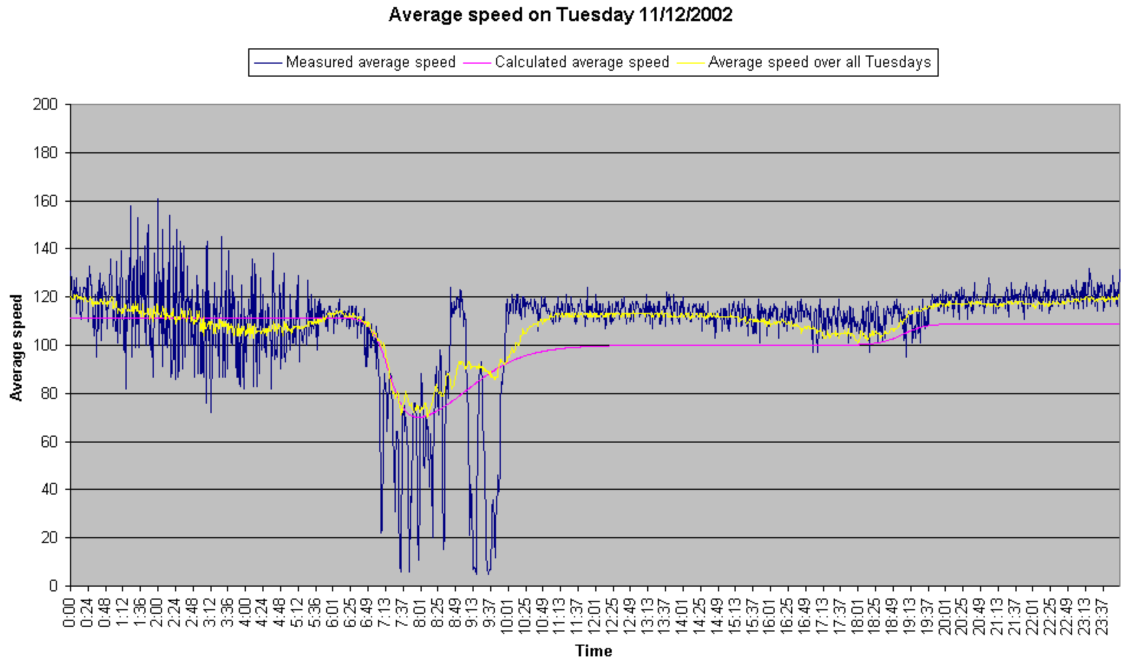


Figure 9.13 The average speed on 12th of November 2002 near Gouda with the 2-4-1-topology

9.2.5 The 2-5-1-topology

The error graphs of this topology, displayed in Figure 9.14, are the same as the error graphs above. A drop can be seen at the beginning of the training and the networks again do not converge to an error of 0. Because the scales are the same in the graphs, this figure indicated that it is possible that the prediction for Gouda will be better than for Zoetermeer. If this information is right can be seen in the results on the test set, Table 9.10.

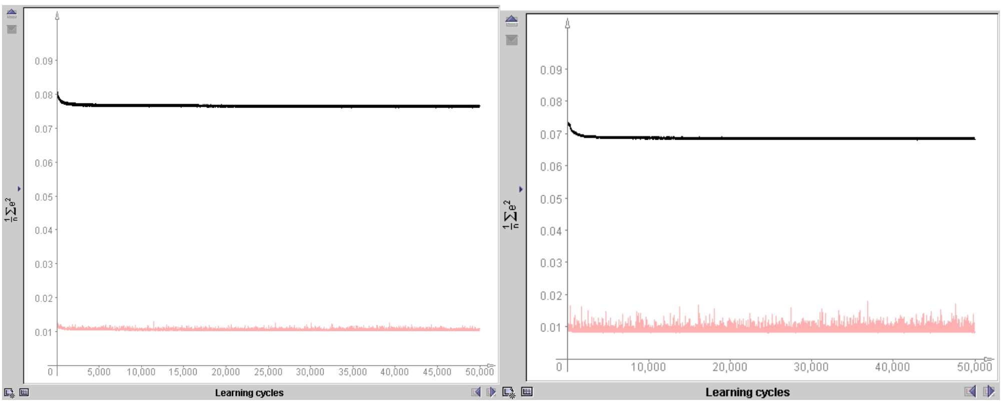


Figure 9.14 The error graphs of the 2-5-1-topology

This network shows even in the table a better prediction for the place near Gouda than the 2-2-1-network. The error is smaller and even the variance and the standard deviation are higher for the 2-2-1-network. This cannot be said for the prediction of the average speed near Zoetermeer. The errors, the variance and the standard deviation are larger for the 2-5-1-network than for the 2-2-1-network.

Table 9.10 Results of the 2-5-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	245.5919	229.5208	216.507	197.8741
MSE	0.014705	0.013793	0.012964	0.011891
VAR	0.014563	0.013741	0.012709	0.011818
STDev	0.120678	0.117222	0.112733	0.108712

Table 9.11 shows the results of the 2-5-1-network for the busiest times. This time the prediction for Zoetermeer seems better than the prediction of the 2-2-1-network and the 2-4-1-network. However the variance of the error and the standard deviation is larger. For Gouda the best prediction for the busiest times, besides the one made by the 2-3-1-network, is still the prediction of 2-2-1-network for Gouda. The difference in results between the prediction and the average speed over all Tuesdays is not very large for the busiest times, but still the average speed over all Tuesdays is the best approximation.

Table 9.11 Results of the 2-5-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	149.46009	132.94418	141.2843	116.1987
MSE	0.0692907	0.0613211	0.0655	0.053597
VAR	0.062335	0.0575234	0.054492	0.051933
STDev	0.2496698	0.2398403	0.233436	0.227887

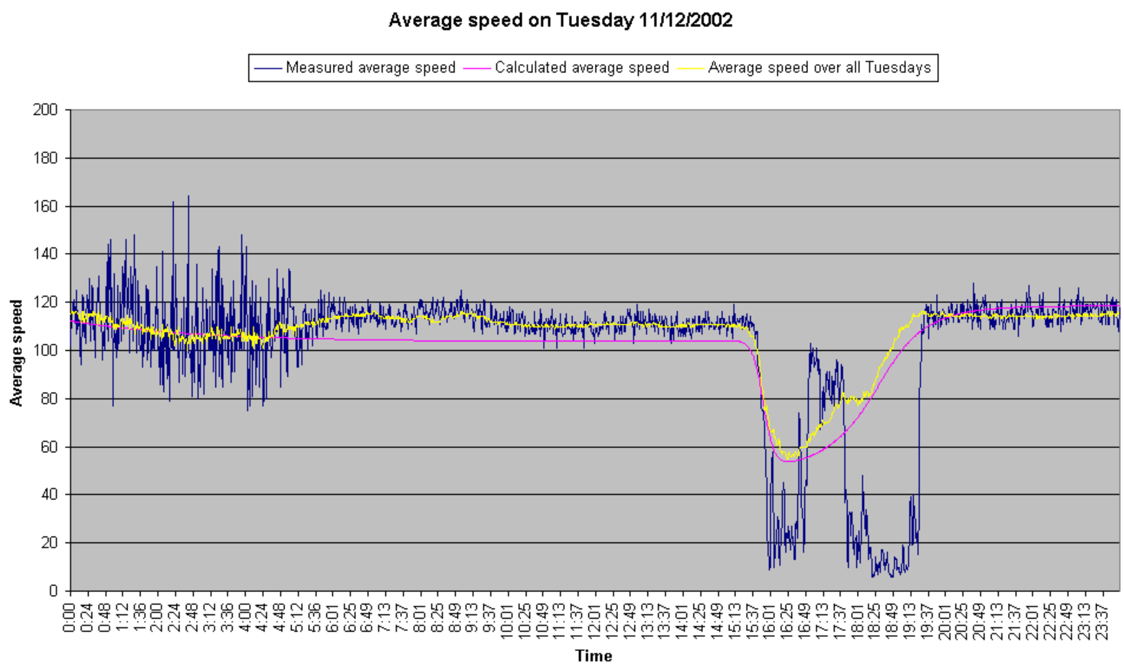


Figure 9.15 The average speed on 12th of November 2002 near Zoetermeer with the 2-5-1-topology

In Figure 9.15 and Figure 9.16 it can be seen that this network also approximate the average speed over all Tuesdays. Especially the prediction near Gouda follows this average speed very accurate. The forecast near Zoetermeer gives a lower speed in the quiet times before the busiest times than necessary.

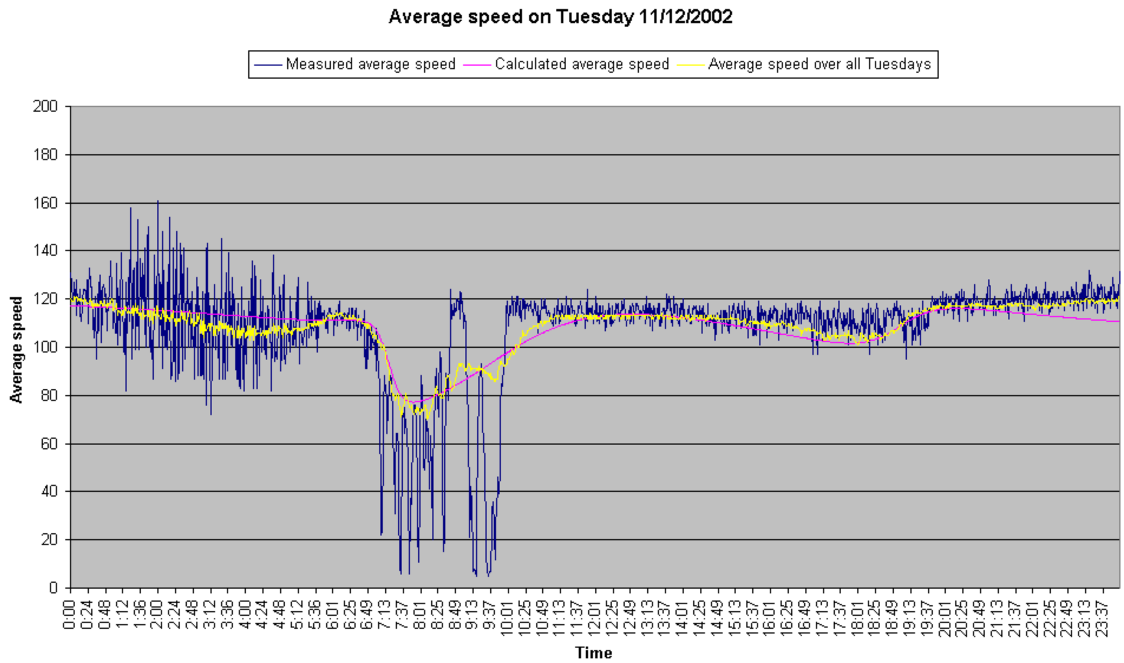


Figure 9.16 The average speed on 12th of November 2002 near Gouda with the 2-5-1-topology

9.2.6 The 2-6-1-topology

If the results from this network, Table 9.12, are compared to those from the 2-5-1-network, the 2-5-1-network gives a better prediction than this network. Not only the error is smaller with the 2-5-1-network, but also the variance and the standard deviation.

Table 9.12 Results of the 2-6-1-topology

	Zoetermeer	Gouda	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	247.6772	232.8028	216.507	197.8741
MSE	0.01483	0.013991	0.012964	0.011891
VAR	0.014816	0.013824	0.012709	0.011818
STDev	0.121722	0.117573	0.112733	0.108712

Table 9.13 Results of the 2-6-1-topology for the busiest times

	Zoetermeer (4.00 PM – 7.00 PM)	Gouda (7.00 AM – 10.00 AM)	All Tuesdays (Zoetermeer)	All Tuesdays (Gouda)
SSE	156.69903	127.18987	141.2843	116.1987
MSE	0.0726467	0.0586669	0.0655	0.053597
VAR	0.0621775	0.0553229	0.054492	0.051933
STDev	0.2493542	0.2352082	0.233436	0.227887

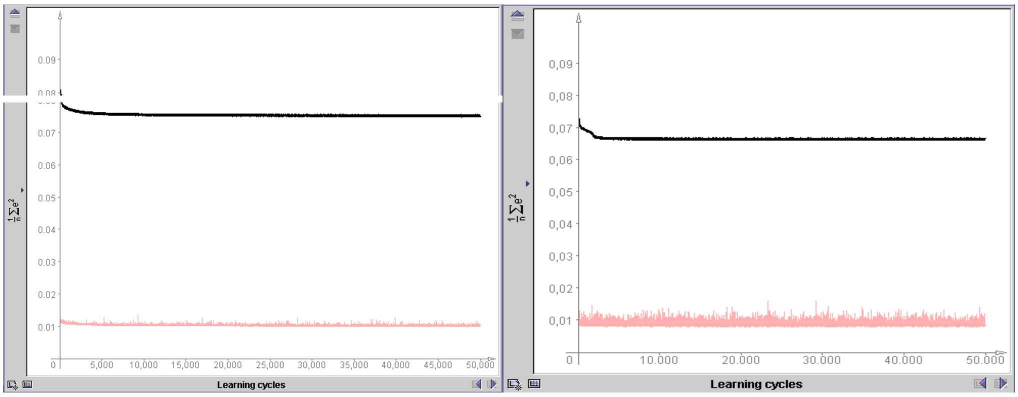


Figure 9.17 The error graphs of the 2-6-1-topology

Figure 9.17 gives the same information as the error-graphs of the 2-5-1-toplogy. The results of Table 9.12 did already reveal this. Further these error graphs are just like all the other ones. Again the networks did not converge to an error of 0 and the training was stopped by 50000 training cycles.
Looking at the results of the busiest times and comparing these values to the ones of the 2-5-1-network shows that for the prediction near Zoetermeer the error is lower again for the 2-5-1-prediction. The variance and the standard deviation however are higher for this prediction, but these differences are very small. The results for this topology during the busiest times are showed in Table 9.13. The results of the prediction of the busiest times near Gouda are better than the results of the 2-5-1-network. Again this time the error of the average speed over all Tuesdays is smaller than the error of the predicted speed.
Figure 9.18 and Figure 9.19 show no real differences with the previous figures. Again the pink line representing the predicted average speed is following the yellow line that gives the average speed over all Tuesdays.

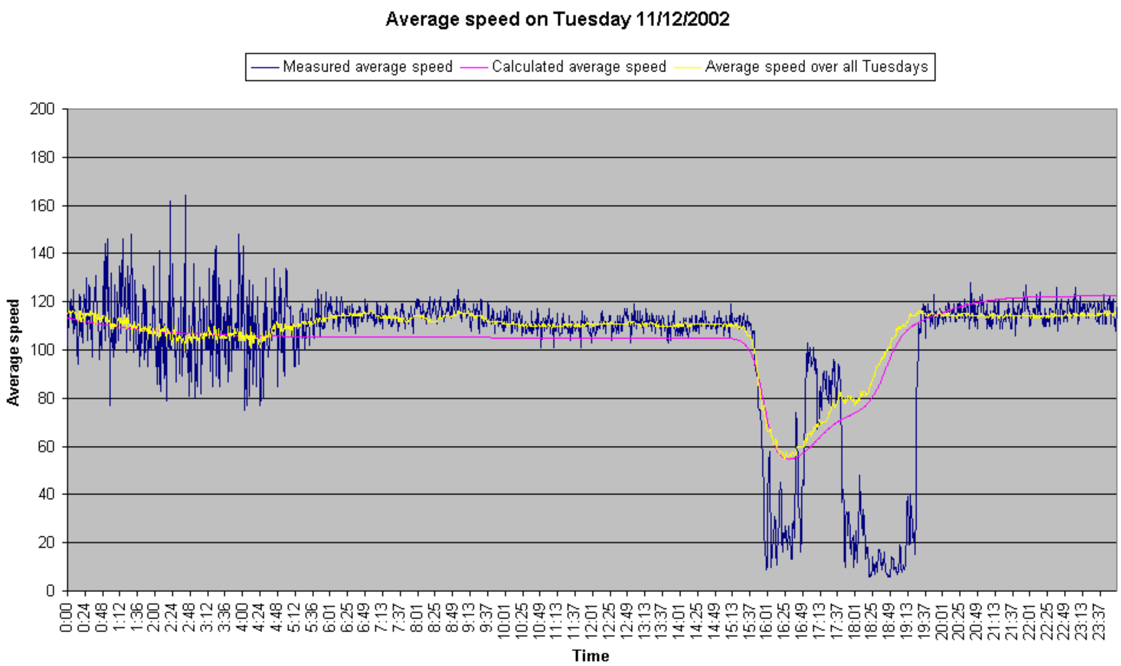


Figure 9.18 The average speed on 12th of November 2002 near Zoetermeer with the 2-6-1-topology

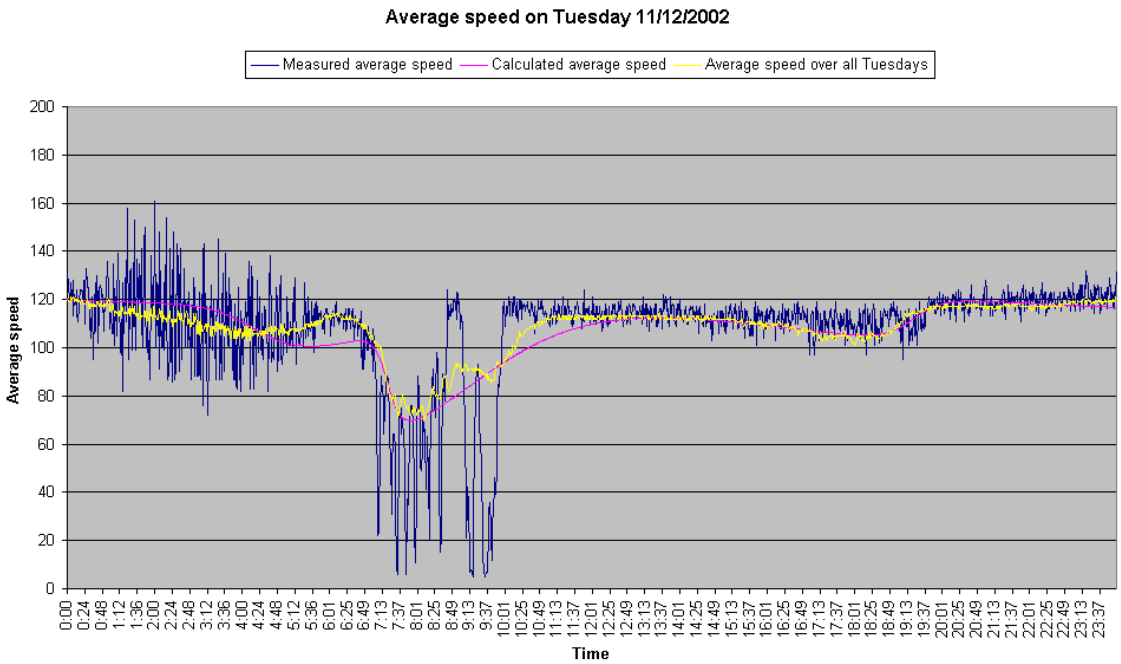


Figure 9.19 The average speed on 12th of November 2002 near Gouda with the 2-6-1-topology

9.2.7 Conclusions

In Section 5.4 twelve formulas where introduced. Each of the formulas estimates the number of neurons in the hidden layer necessary to obtain the best result. However none of the formulas is proved to be giving the number of hidden neurons that is best for all problems. Therefore in this section all the topologies resulting from these formulas for the model without time series are tested as stated above. The results of these tests are summarized here.

In Table 9.14 to Table 9.17 the best results for the column are marked. One of the previous sections showed that the prediction of the 2-1-1-topology resulted in a 'straight' line. It was not able to predict any traffic jams at all.

On the other hand the 2-3-1-topology predicted traffic jams pretty well, but had difficulties recognizing the normal situation. Of this network the results for the busiest times looked promising, however the results of the error over the complete dataset did not show this extreme improvement. The figures gave the reason for these good results. The prediction resulted in a lower speed than the speed actually driven for the quiet times. Also the prediction of the speed driven at the busiest times was lower, but because the prediction was to high in these times for all the other topologies the traffic jams were better predicted. Maybe training the network again and with different initial values will lead to a complete other prediction. However in this case the network is not capable of predicting the normal, quiet situation and this results in a bad network.

The networks described above will not be used to predict the average speed because of the reasons described. But what will be the best network to predict the average speed near Zoetermeer. If the results of only Table 9.14 will be taken into account the best network will be the 2-2-1-network. Table 9.15 gives a different prediction. According to this table the 2-5-1-network will be the best. Because the difference between these two networks for the busiest times is very large, while it is only 1/8 of the complete dataset, and the differences over the complete dataset are a lot smaller, it seems fair to consider the 2-5-1-network as the best network for the problem of predicting the average speed near Zoetermeer without time series.

Table 9.14 The results of all the networks without time series for Zoetermeer

Zoetermeer	SSE	MSE	VAR	StDev
2-1-1	389.2156	0.023305	0.02313	0.152085
2-2-1	**241.2076**	**0.014443**	0.014407	0.120028
2-3-1	247.7492	0.014834	**0.014106**	**0.118769**
2-4-1	242.5549	0.014523	0.01434	0.119749
2-5-1	245.5919	0.014705	0.014563	0.120678
2-6-1	247.6772	0.01483	0.014816	0.121722
All Tuesdays	216.507	0.012964	0.012709	0.112733

Table 9.15 The results of all the networks without time series for Zoetermeer at the busiest times

Zoetermeer	SSE	MSE	VAR	StDev
2-1-1	240.24672	0.11138	0.0674597	0.2597301
2-2-1	157.76194	0.0731395	0.0593918	0.2437044
2-3-1	**139.18687**	**0.064528**	**0.0578023**	**0.2404212**
2-4-1	161.42195	0.0748363	0.0579763	0.2407827
2-5-1	149.46009	0.0692907	0.062335	0.2496698
2-6-1	156.69903	0.0726467	0.0621775	0.2493542
All Tuesdays	141.2843	0.0655	0.054492	0.233436

It must also be decided which network is able to predict the average speed near Gouda. Will this also be a 2-5-1-network? At first the results in Table 9.16 must be taken into account. This table shows that the best network can be the 2-5-1-network. Besides the 2-3-1-network this network has the lowest error, variance and standard deviation. In the busiest times, Table 9.17, the 2-3-1-network will give the lowest error and the 2-1-1-topology the smallest variance and standard deviation. These networks will both not be used to predict speeds because of the reasons mentioned above. The second best option will be the 2-2-1-network. Now it has to be decided which network will be the best for the prediction of the average speed near Gouda without time series. This is not a very easy problem. The results of the 2-2-1-network and the 2-5-1-network do not differ very much.

This means that it is possible to consider both problems, the prediction of the average speed near Zoetermeer and near Gouda, as the same. In that case it is not necessary to try and decide what network is the best for all places for which a prediction has to be made. If this information is considered the best decision will be to use the 2-5-1-network for the prediction of the average speed without time series.

According to the formulas of Section 5.4 it can be said that using the formula of Hecht-Nielsen ($2N_i +1$) leads to the number of hidden neurons resulting in a network giving the best prediction. Because the first part of the formula of Fletcher and Goss also contains this formula it can be said that the network-topology resulting in the best prediction is based on the formulas of Hecht-Nielsen and Fletcher and Goss.

Table 9.16 The results of all the networks without time series for Gouda

Gouda	SSE	MSE	VAR	StDev
2-1-1	280.082	0.016832	0.016492	0.128422
2-2-1	232.3884	0.013966	0.013887	0.117844
2-3-1	259.3073	0.015583	**0.013682**	**0.116968**
2-4-1	242.9572	0.014601	0.014331	0.119711
2-5-1	**229.208**	**0.013793**	0.013741	0.117222
2-6-1	232.8028	0.013991	0.013824	0.117573
All Tuesdays	197.8741	0.011891	0.011818	0.108712

Table 9.17 The results of all the networks without time series for Gouda at the busiest times

Gouda	SSE	MSE	VAR	StDev
2-1-1	162.92637	0.0751505	**0.0535165**	**0.2313363**
2-2-1	127.01441	0.058586	0.0547429	0.2339719
2-3-1	**119.9653**	**0.0553345**	0.0551908	0.2349273
2-4-1	129.63354	0.0597941	0.0580849	0.2410082
2-5-1	132.94418	0.0613211	0.0575234	0.2398403
2-6-1	127.18987	0.0586669	0.0553229	0.2352082
All Tuesdays	116.1987	0.053597	0.051933	0.227887

Because the results are very close to each other, except for the topology with one hidden neuron, it is possible that the conclusion that best network will not be the 2-5-1-network but a network with approximately 5 hidden neurons. To make a real good decision the networks should be trained several times and with different initial values. After that an average must be taken from the results. Comparing these results will lead to the network that is really the best. However to show if the model is capable of predicting the average speed, like in this report, only having the results of one training will be sufficient.
Overall it can be said that the prediction that will be made with the networks without time series will follow the average speed over all Tuesdays. This can be expected if the parameter 'time' is taken more into account then the parameter 'month'. In the analysis, Chapter 4, it was already clear that the average speed depends very strong on the time of the day. However the prediction made by the neural networks gave higher errors than the average speed over all Tuesdays. In this case it is probably better to use the average speed over all Tuesdays than the speed predicted by the neural network. If all the parameters are used as mentioned in Chapter 8 then it is possible that the prediction of the neural network improves and will be better than the average over all Tuesdays.

9.3 Zoetermeer and Gouda with time series

Section 9.2 contains the results from the model without time series. This model resulted in a topology of the neural network based on the formulas introduced by Hecht-Nielsen and Fletcher and Goss. It cannot be assured that this formula will give the number of neurons resulting in the best topology for the model with time series.

To determine the formula that gives the best topology for this model, in this section the formulas are again used to make an educated guess of the number of hidden neurons. The topologies will not be the same as in the section before. This is due to a different number of input and output parameters. Therefore the formulas and their results are stated again in Table 9.18. In this case the number of input nodes (N_i) is 5. These parameters corresponding to these nodes are time, month and a time series of the average speed. This time series consists of the average speed at time t – 30 minutes, at time t – 20 minutes and at time t – 10 minutes. The output nodes (N_o) are also a time series. This time series consists of 4 items, the average speed at time t, at time t +10, at time t +20 and at time t +30. With these network it is tried to predict the average speed 40 minutes in advance.
The number of trainings examples for the point near Zoetermeer (14105) and the point near Gouda (29805) are respectively 118584 and 119311. Again the unreliable data is filtered out. The unreliable factor in the formula of Garson can be set to 5.

Table 9.18 Formulas to determine the number of hidden nodes and their results (with time series)

Author	**Formula**	**Number of nodes**	
Hecht-Nielsen	$2N_i + 1$	11	
Shavlik	$0.1(N_i + N_o)$	0.8 → 1	
Hush	$3N_i$	15	
Rippley	$(N_i + N_o) / 2$	4	
Wang	$2N_i / 3$	3.33 → 4	
Blum	between N_i and N_o	between 5 and 4	
Fletcher and Goss	from $2N_i + 1$ to $2\sqrt{N_i} + N_o$	from 11 to ≈ 8	
Berry and Linoff	$< 2N_i$	< 10	
Sigillito and Hutton	$< N_i / 2$	< 2.5	
Kanellopoulos and Wilkinson	$2N_i$ *or* $3N_i$	10 or 15	
Paola	$\dfrac{2 + N_o \cdot N_i + \frac{1}{2} N_o (N_i^2 + N_i) - 3}{N_i + N_o}$	7.3 → 8	
Garson	$\dfrac{N}{r(N_i + N_o)}$	**Zoetermeer (14105)**	**Gouda (29805)**
		2636	2652

The number of nodes necessary for this problem according to Garson is again very large and it is not possible to create such a large network with JavaNNS. So these topologies are again not taken into account.

The following topologies result from the formulas above:

- 5-1-4
- 5-2-4
- 5-4-4
- 5-5-4
- 5-8-4
- 5-9-4
- 5-10-4
- 5-11-4
- 5-15-4

These topologies are tested in this section. All of them are trained in 50000 cycles of the training data and the parameters of the backpropagation algorithm are set to η = 0.2 (learning rate) and d_{max} = 0.1 (maximum error) in all cases. For each of the networks the summed squared error, the mean squared error, variance and standard deviation are computed.
In the subsections the different topologies are discussed. In each section the figures of the predicted average speed 10 minutes and 40 minutes in advance on the 11th November 2002 will be shown. The other two figures of this day, the prediction 20 and 30 minutes ahead can be found in appendix C. In each section also a figure can be found showing the error graph of the network as displayed by JavaNNS. Again the remark should be made that the left one is the error graph of the network for Zoetermeer and the right one for Gouda. The function calculating the mean squared error is dependent on the size of the dataset as is explained in the previous section, so the values resulting from the graphs cannot be used. The results are calculated by Microsoft Excel.

9.3.1 The 5-1-4-topology

The results of Table 9.19 and Table 9.20 show for ten minutes ahead a fairly good prediction. The Mean Squared error of the prediction 10 minutes ahead is much lower than in the previous section, but in that section a model was tested that was able to predict the average speed over a long time. This is possible, because the network does not use any information of the road itself or other information that is not known in advance. In this case information about the average speed on the road is used. This information is not known in advance and therefore the time period over which the prediction is made is limited. In this case the prediction is made 10, 20, 30 or 40 minutes in advance. Because the goal was to predict the average speed over 25 minutes only the results of the prediction over 30 and 40 minutes is discussed. If these results are considered the prediction 30 minutes ahead is better than the best prediction of the previous section, at least this can be said of the prediction for the point near Zoetermeer.

Table 9.19 Results of the 5-1-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	102.866	172.324	204.839	265.031
MSE	0.00647	0.01084	0.01289	0.01668
VAR	0.00608	0.00967	0.01262	0.01547
STDev	0.07798	0.09835	0.11236	0.12438

Table 9.20 Results of the 5-1-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	169.237	167.677	255.556	245.468
MSE	0.01082	0.01072	0.01633	0.01569
VAR	0.00841	0.01016	0.01173	0.01294
STDev	0.09172	0.10079	0.1083	0.11374

The prediction of the point near Gouda is better 40 minutes ahead than 30 minutes ahead. This is remarkable, because if Figure 9.21 and Figure 9.23 are considered the prediction is the average speed given at the input of which the very low speeds are replaced by somewhat higher values. If all the figures are considered, the figures displayed in this section and those of appendix C, a trend can be seen. From the prediction 10 minutes ahead to the prediction 40 minutes ahead the low values become higher and higher. This means that the error at the busiest times becomes higher the further ahead the prediction will be made; on the other hand just after these times the error will be lower. If the error in the busiest times will not increase as much as the error decreases of the prediction of the times after these busy times the prediction can be better for 40 minutes ahead. However the prediction should be worse for the busiest times. The same trend can be seen in the Figure 9.22 and Figure 9.24, but here apparently the error will increase more on the busiest times than the error decreases on the quiet times after the busy times.
In Figure 9.20 the error graphs are shown. These graphs have the same scaling and can be compared. The error on the training set, the black line, is lower for the graph of Zoetermeer. However the pink lines representing the error on the validation set do not differ very much. In this case again the networks will not converge to an error of 0. If the network would converge if the black line goes to 0 and the pink line of the validation set will move up. Because the error still drops a little bit after 50000 training cycles it can take millions of training cycles before it will converge to 0 if it goes to zero at all.

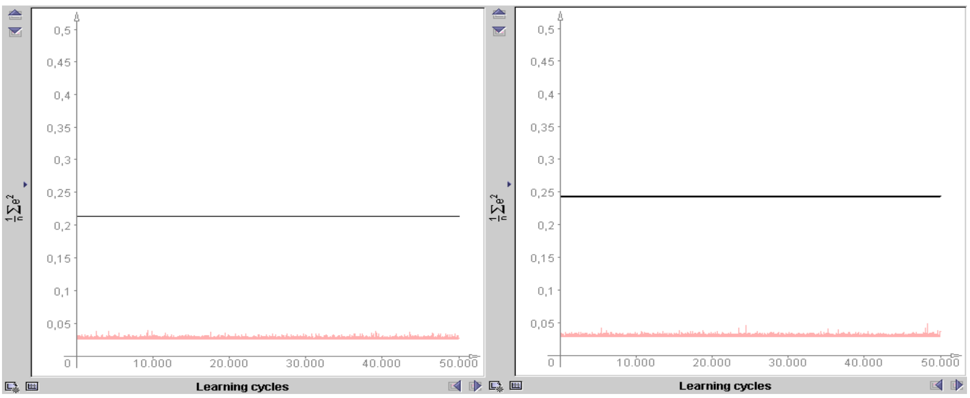


Figure 9.20 The error graphs of the 5-1-4-topology

In Table 9.21 and Table 9.22 the results for the busiest times are given. The error does increase the further ahead the prediction will be made as mentioned before. However the previous model always showed a lower error of the prediction for the point near Gouda than for the point near Zoetermeer. With this model this prediction for the point near Zoetermeer is better.

Table 9.21 Results of the 5-1-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	36.089635	68.394601	107.7134	137.7193
MSE	0.0173926	0.0331048	0.0520606	0.0664348
VAR	0.0170442	0.0324026	0.0481736	0.0603701
STDev	0.1305535	0.1800072	0.2194849	0.2457033

Table 9.22 Results of the 5-1-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	64.991528	80.309479	87.141264	97.947079
MSE	0.0302991	0.0376157	0.0405875	0.0458339
VAR	0.0302958	0.0359345	0.040294	0.0430839
STDev	0.1740569	0.1895639	0.2007335	0.2075667

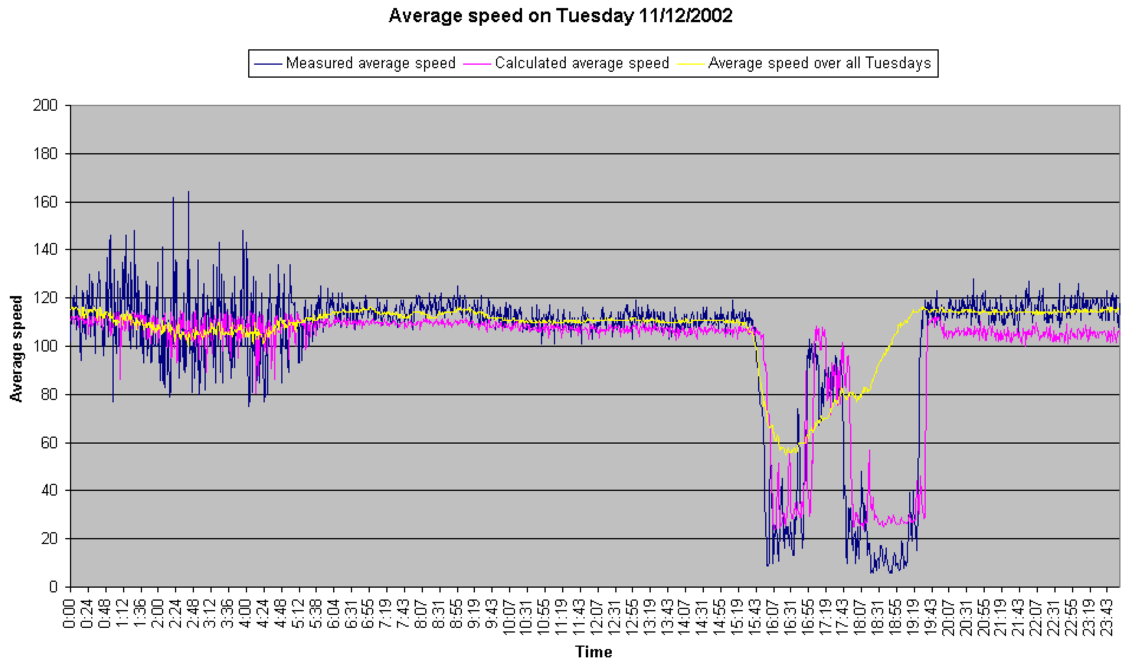


Figure 9.21 The average speed on 12th of November 2002 near Zoetermeer with the 5-1-4-topology 10 minutes ahead

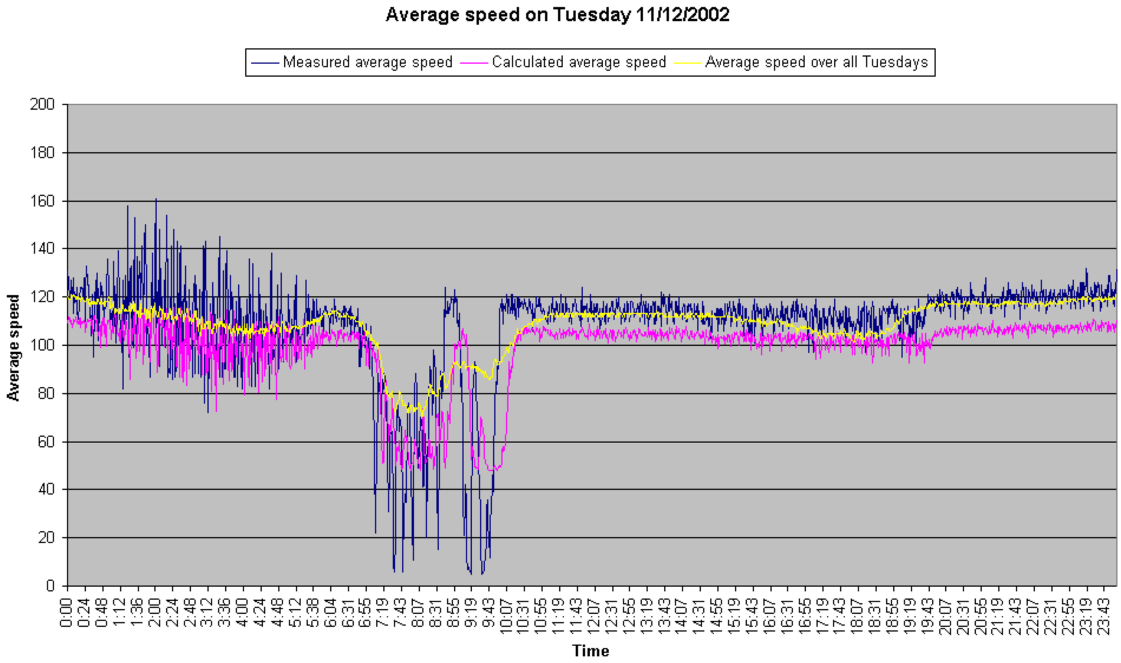


Figure 9.22 The average speed on 12th of November 2002 near Gouda with the 5-1-4-topology 10 minutes ahead

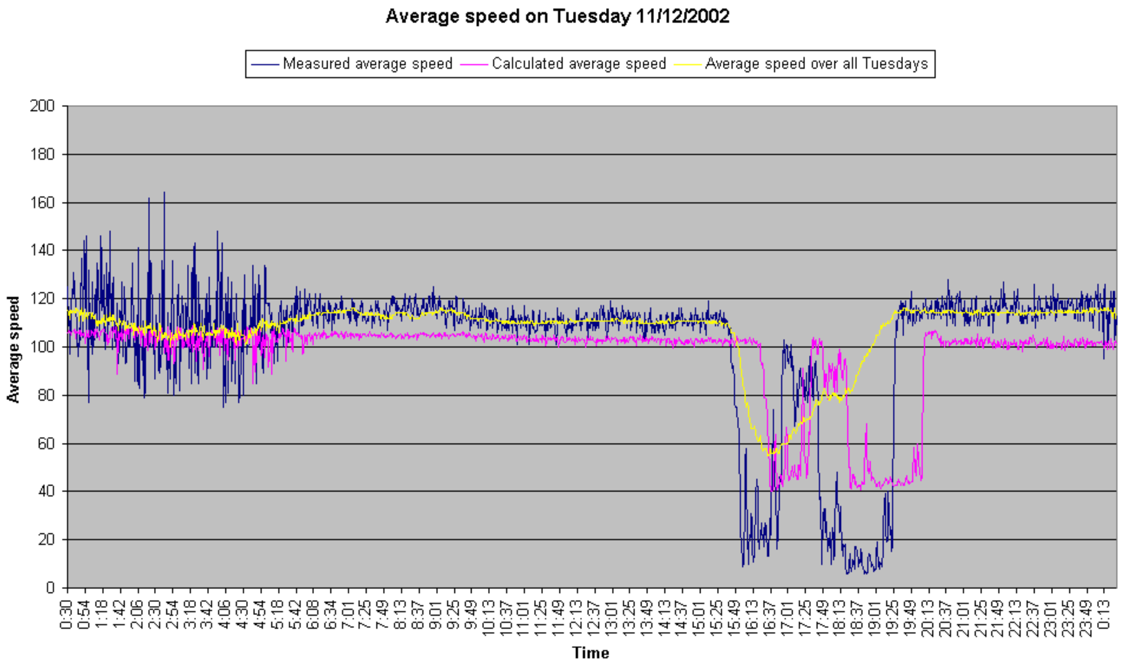


Figure 9.23 The average speed on 12th of November 2002 near Zoetermeer with the 5-1-4-topology 40 minutes ahead

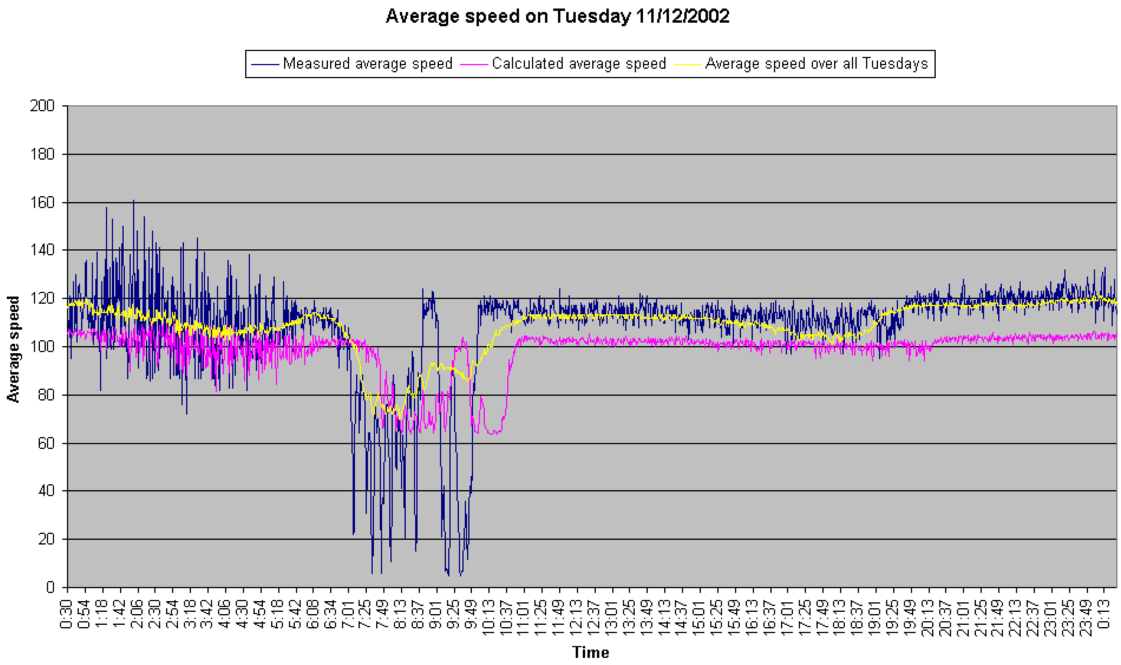


Figure 9.24 The average speed on 12th of November 2002 near Gouda with the 5-1-4-topology 40 minutes ahead

9.3.2 The 5-2-4-topology

The 5-2-4-topology has better results than best topology representing the model in the previous section for the prediction 30 minutes ahead near Zoetermeer, Table 9.23. If the prediction 40 minutes in advance is considered the other model is able to give a better prediction. For Gouda this is not true as Table 9.24 shows. The prediction given by this network is far better for all times ahead than the prediction given by the previous section.

Table 9.23 Results of the 5-2-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	96.1862	165.008	204.203	256.394
MSE	0.00605	0.01038	0.01285	0.01613
VAR	0.00589	0.00955	0.01257	0.01537
STDev	0.07674	0.09773	0.1121	0.12398

Table 9.24 Results of the 5-2-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	146.557	158.011	192.199	200.267
MSE	0.00937	0.0101	0.01228	0.0128
VAR	0.00871	0.01006	0.01141	0.01262
STDev	0.0933	0.1003	0.10682	0.11235

In this case the error graphs of Figure 9.25 cannot be compared on the eye, because the scaling on the y-axis is not the same. The information that is given by these graphs is that the prediction for the point near Zoetermeer is better for the training set. However it is difficult to determine from this figure what the values of the pink lines are.

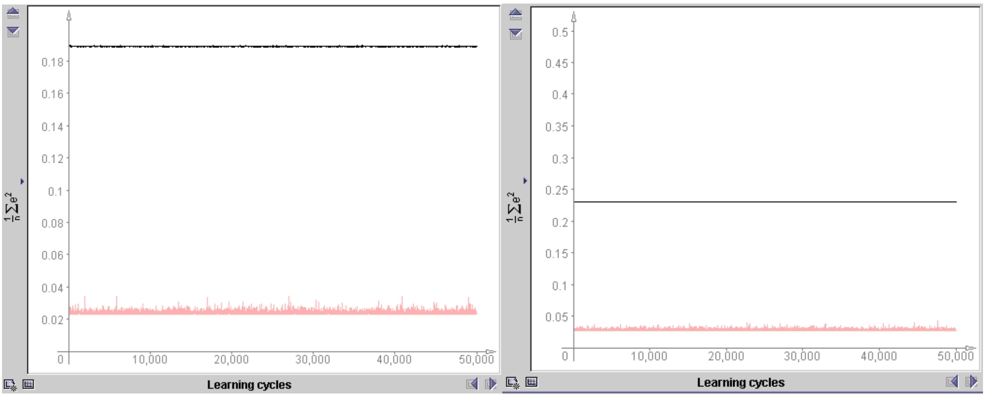


Figure 9.25 The error graphs of the 5-2-4-topology

The first two predictions of the average speed near Zoetermeer are pretty correct. The errors are relatively low, even lower than those of the prediction near Gouda. But the last two, 30 and 40 minutes ahead, are not as good as the prediction of the point near Gouda. Also the busiest times in Table 9.25 and Table 9.26 show these effects, even as the predictions with the other model. It was showed that these lower errors were caused by the fact that the busiest times near Zoetermeer had a greater impact on the average speed than the busiest times near Gouda. Because the trend was that the further ahead the prediction was made the higher the values of the lowest speeds became, it is not a very remarkable effect. The trends can also be seen in Figure 9.26 to Figure 9.29.

Table 9.25 Results of the 5-2-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	34.975341	67.607243	106.19899	138.2161
MSE	0.0168556	0.0327237	0.0513287	0.0666744
VAR	0.0166077	0.0319687	0.0475185	0.0599992
STDev	0.1288707	0.178798	0.2179873	0.2449473

Table 9.26 Results of the 5-2-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	67.130856	80.928722	90.277952	106.41207
MSE	0.0312964	0.0379057	0.0420484	0.0497951
VAR	0.0311851	0.0361079	0.0403849	0.043283
STDev	0.176593	0.1900207	0.20096	0.2080456

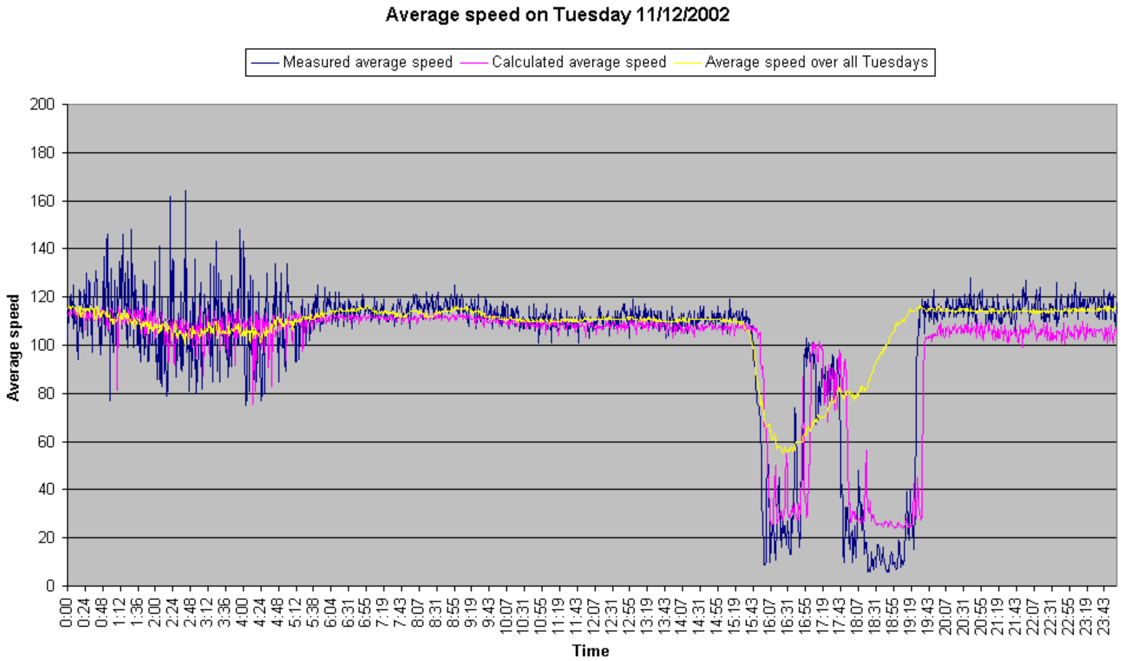


Figure 9.26 The average speed on 12th of November 2002 near Zoetermeer with the 5-2-4-topology 10 minutes ahead

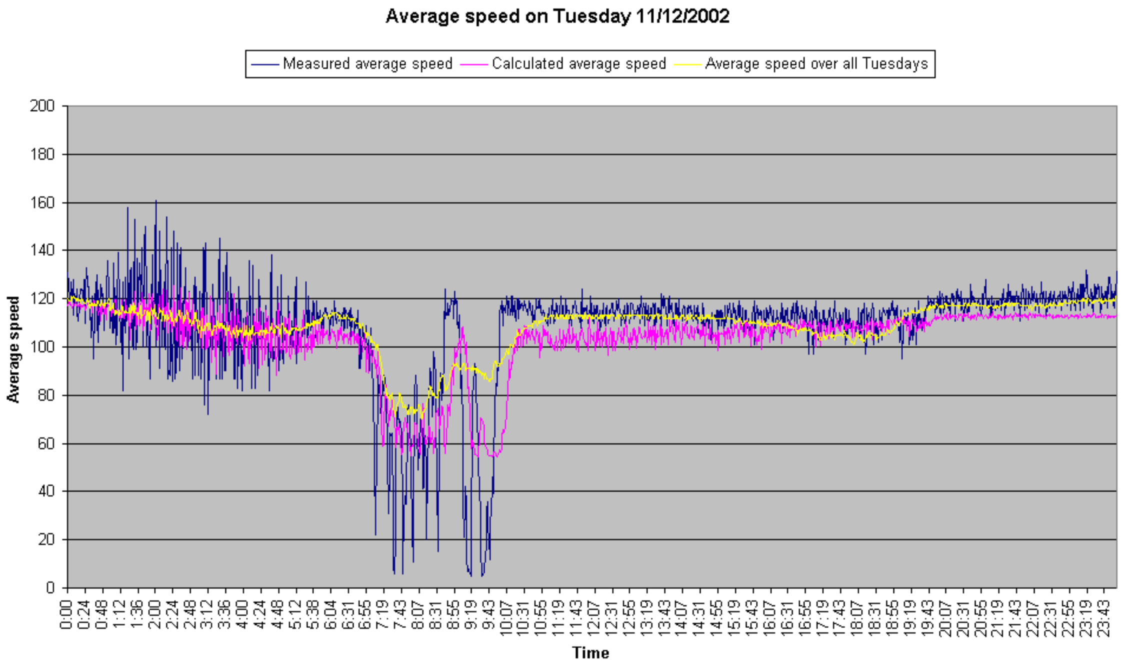


Figure 9.27 The average speed on 12th of November 2002 near Gouda with the 5-2-4-topology 10 minutes ahead

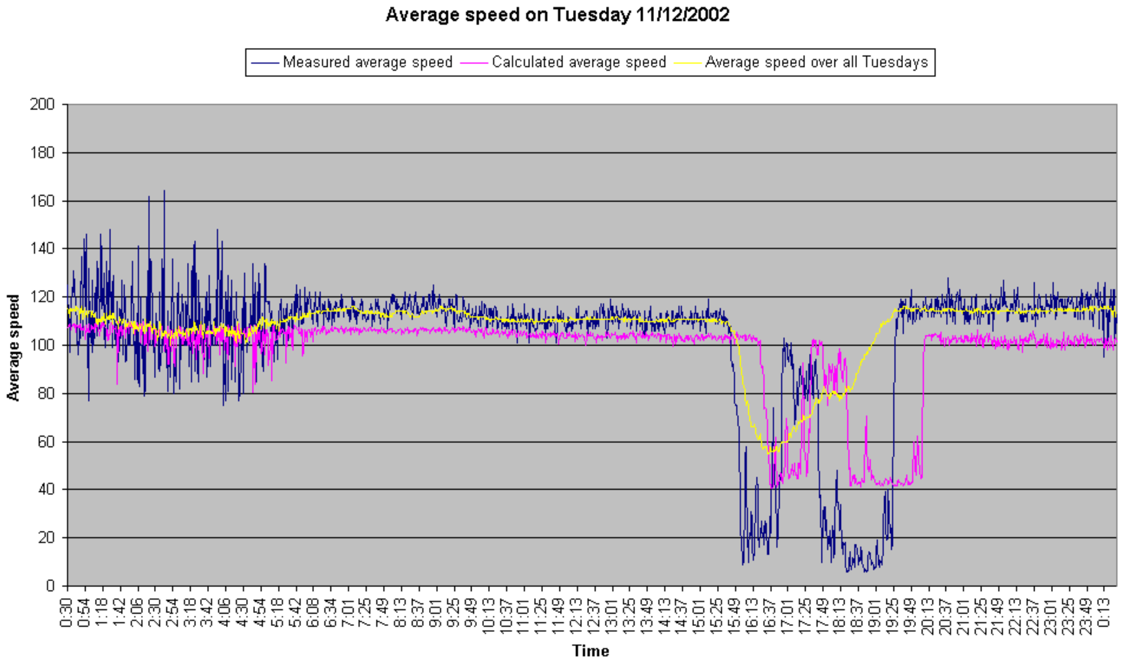


Figure 9.28 The average speed on 12th of November 2002 near Zoetermeer with the 5-2-4-topology 40 minutes ahead

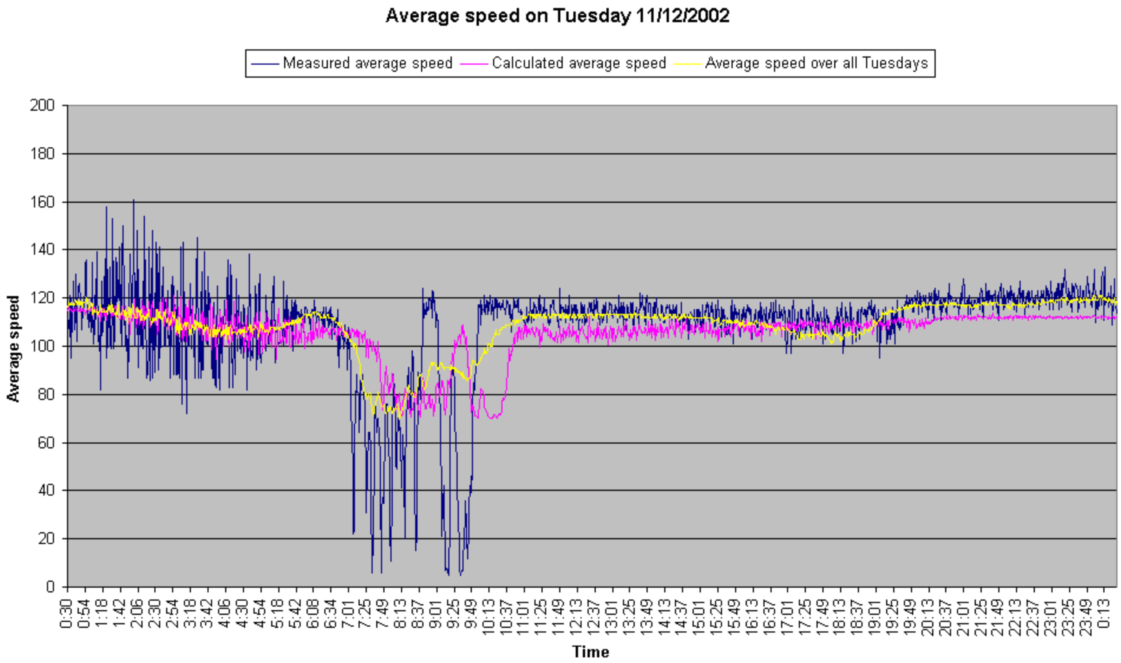


Figure 9.29 The average speed on 12th of November 2002 near Gouda with the 5-2-4-topology 40 minutes ahead

9.3.3 The 5-4-4-topology

Using four hidden neurons has a great impact on the predictions. Considering the point near Zoetermeer the prediction is improving a lot for the prediction 30 and 40 minutes ahead. However the prediction of the point near Gouda is worse than with two hidden neurons. These results can be confirmed by taking a look at Figure 9.31 to Figure 9.34. The prediction for the place near Zoetermeer is improved compared to the figures of the previous discussed topologies. The figures of the point near Gouda seems to show a prediction that is going to fit to the average speed over all Tuesdays the further ahead the prediction is made. This will be correct if the prediction is made over a very long period, but not for 40 minutes in advance. The results are displayed in the Table 9.27 and Table 9.28.

Table 9.27 Results of the 5-4-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	94.3644	127.705	144.126	164.652
MSE	0.00594	0.00804	0.00907	0.01036
VAR	0.00529	0.00769	0.00907	0.01036
STDev	0.0727	0.08767	0.09523	0.10178

Table 9.28 Results of the 5-4-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	145.244	176.101	209.334	219.195
MSE	0.00928	0.01125	0.01338	0.01401
VAR	0.00893	0.01117	0.0129	0.01401
STDev	0.09452	0.10568	0.11357	0.11835

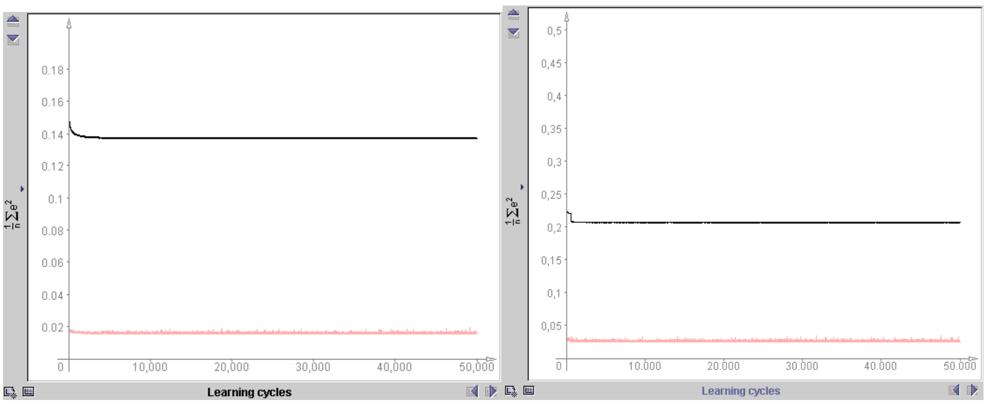


Figure 9.30 The error graphs of the 5-4-4-topology

The error graphs, Figure 9.30, do show the better result of the topology for the point near Zoetermeer. Again the figurers cannot be compared because of the scaling, but the value of the black line, the training set, is lower in the left graph. The value of the pink line, the validation set, is difficult to determine in the right graph, so no conclusions can be made about these lines. Also the network did again not converge to an error of 0.

Table 9.29 Results of the 5-4-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	32.451757	56.900817	79.560132	96.245351
MSE	0.0156394	0.0275415	0.0384534	0.0464281
VAR	0.0155319	0.0273643	0.0376802	0.0451985
STDev	0.1246272	0.1654215	0.1941139	0.2125994

Like the overall error also the error for the busiest times is lower for the point near Zoetermeer than the point near Gouda (Table 9.29 and Table 9.30). A real improvement of the prediction can be seen for Zoetermeer compared to the previous discussed model. But if the results of the prediction for the busiest times near Gouda are compared to those of the previous model, the results of the previous model were better.

Table 9.30 Results of the 5-4-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	72.845705	96.154807	110.68599	125.53434
MSE	0.0339607	0.0450374	0.0515538	0.0587433
VAR	0.0338982	0.0439812	0.0508409	0.0563063
STDev	0.1841147	0.2097171	0.2254793	0.2372895

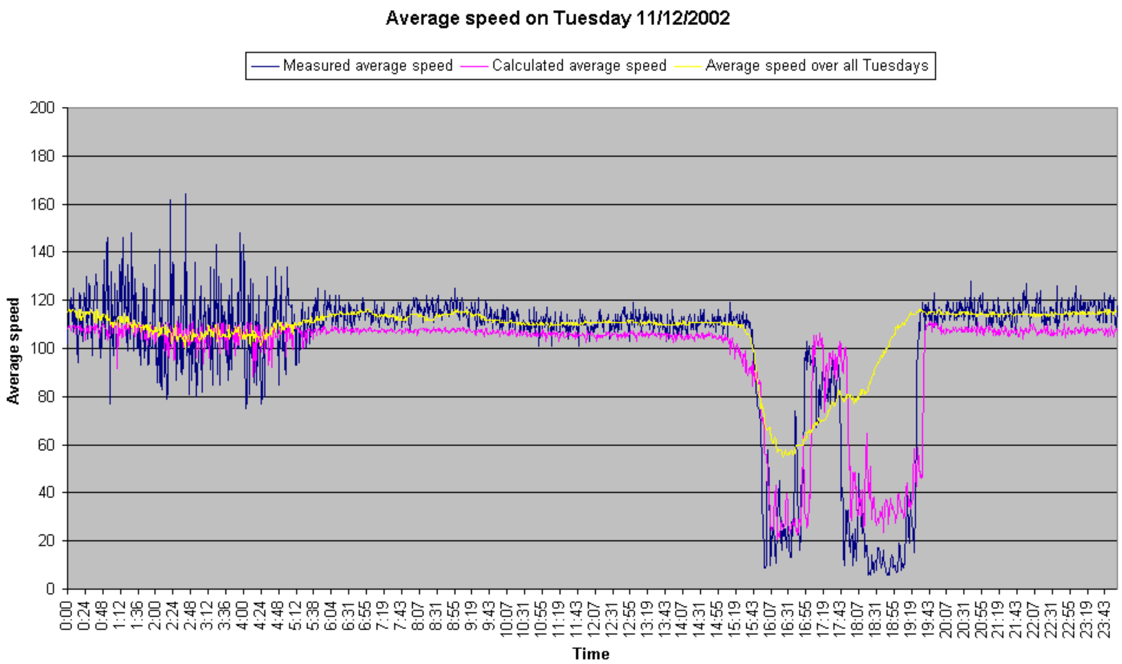


Figure 9.31 The average speed on 12th of November 2002 near Zoetermeer with the 5-4-4-topology 10 minutes ahead

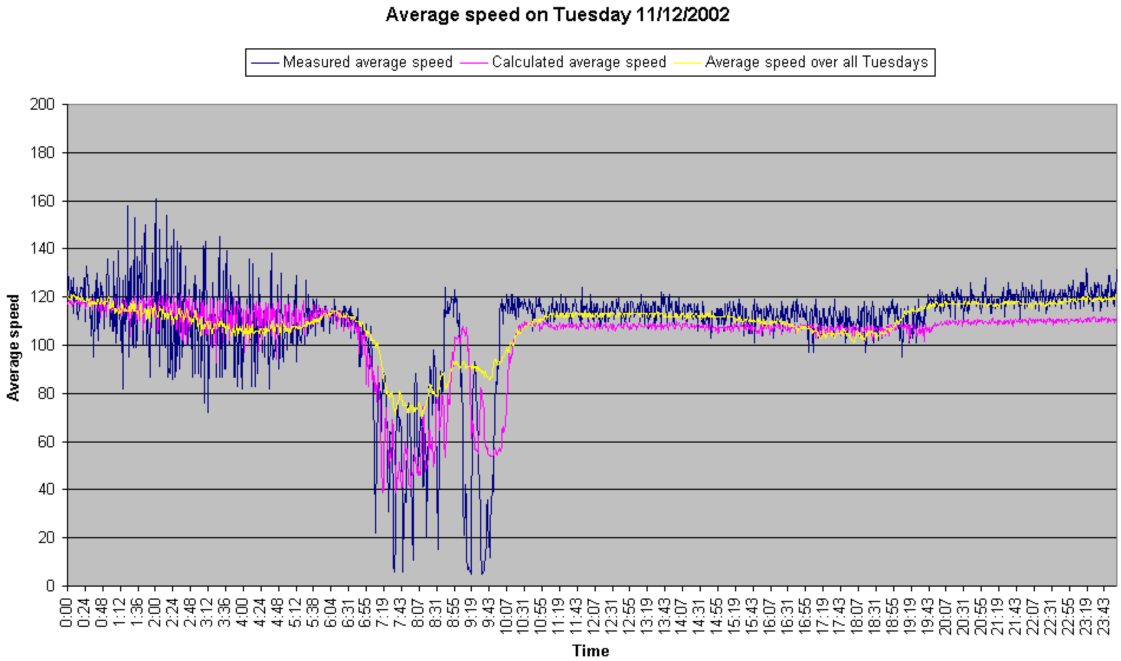


Figure 9.32 The average speed on 12th of November 2002 near Gouda with the 5-4-4-topology 10 minutes ahead

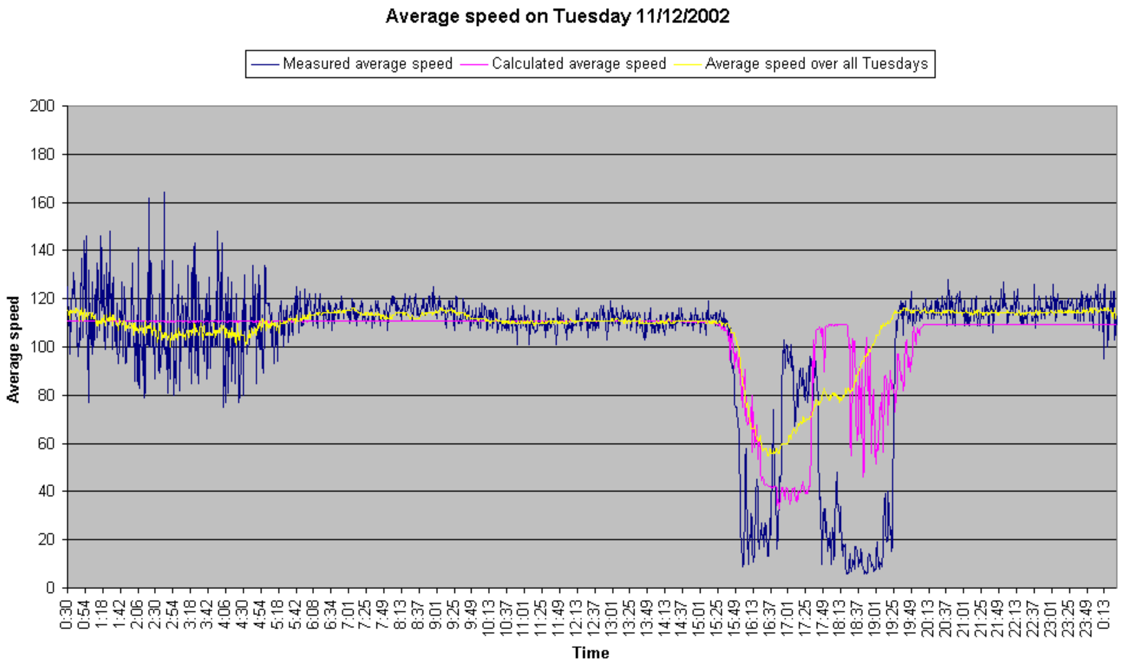


Figure 9.33 The average speed on 12th of November 2002 near Zoetermeer with the 5-4-4-topology 40 minutes ahead

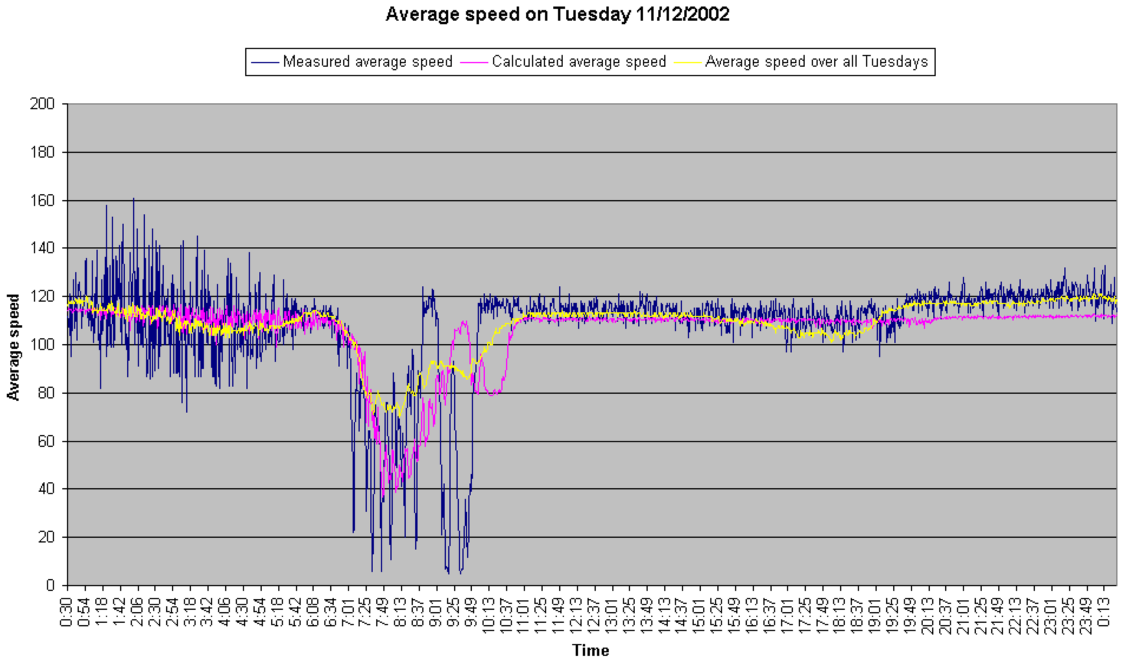


Figure 9.34 The average speed on 12th of November 2002 near Gouda with the 5-4-4-topology 40 minutes ahead

9.3.4 The 5-5-4-topology

In Table 9.31 and Table 9.32 the results for the 5-5-4-topology are displayed. These results show that the prediction for the place near Zoetermeer is better with this topology than the best prediction of the previous discussed model; however the prediction is worse than the one the 5-4-4-toplogy gives. The results of the forecast near Gouda seem to get worse when the size of the network becomes larger.

Table 9.31 Results of the 5-5-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	85.8817	136.465	154.491	171.02
MSE	0.0054	0.00859	0.00972	0.01076
VAR	0.00516	0.00778	0.00935	0.01046
STDev	0.07183	0.08822	0.09668	0.1023

Table 9.32 Results of the 5-5-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	143.576	177.092	207.076	239.423
MSE	0.00918	0.01132	0.01323	0.0153
VAR	0.00883	0.01102	0.0129	0.0146
STDev	0.09399	0.10499	0.11358	0.12081

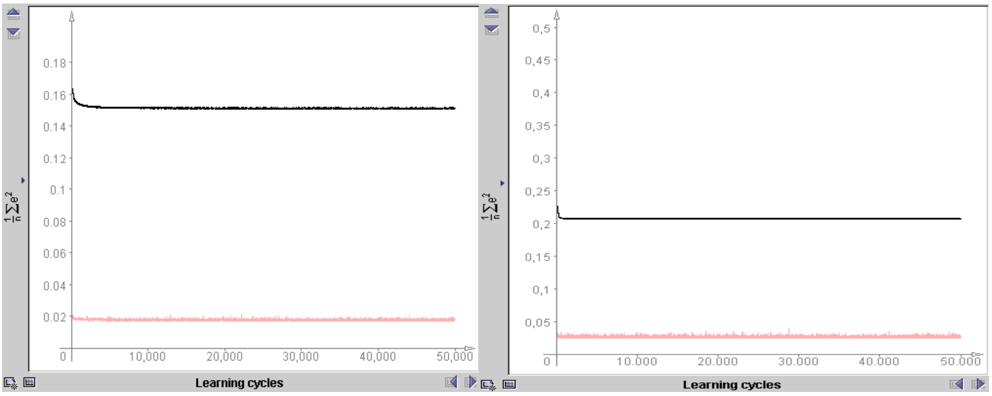


Figure 9.35 The error graphs of the 5-5-4-topology

The results in Table 9.31 and Table 9.32 do show again that the prediction near Zoetermeer is better than the prediction near Gouda. Also the error graphs of Figure 9.35 give this information. Again the scales are not the same, but the value of the black line in the left graph is much lower than the value of the one in the right graph. The value of pink line on the right side is again not clear, so that line is not able to give information.

Table 9.33 Results of the 5-5-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	31.728506	54.448013	75.253464	93.602954
MSE	0.0152908	0.0263543	0.0363719	0.0451534
VAR	0.0150822	0.0263108	0.036041	0.0445523
STDev	0.1228094	0.162206	0.1898448	0.2110742

Table 9.34 Results of the 5-5-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	73.828843	91.921011	109.40986	122.48596
MSE	0.034419	0.0430543	0.0509594	0.0573168
VAR	0.0344054	0.0427769	0.0501957	0.0559645
STDev	0.1854869	0.2068257	0.224044	0.2365681

Table 9.33 and Table 9.34 give the results for the busiest times for both places. For the busiest times the prediction near Zoetermeer and near Gouda are better with this topology than with the 5-4-4-topology. Because the error over the complete test set is larger than the error of the 5-4-4-topology, the conclusion can be made that this network is better predicting the low speeds. However it makes more mistakes while predicting the normal situation. In Figure 9.36 to Figure 9.39 can be seen that the prediction of average speed in the quiet times after the busiest times have passed, is lower than the real speed on that moment.

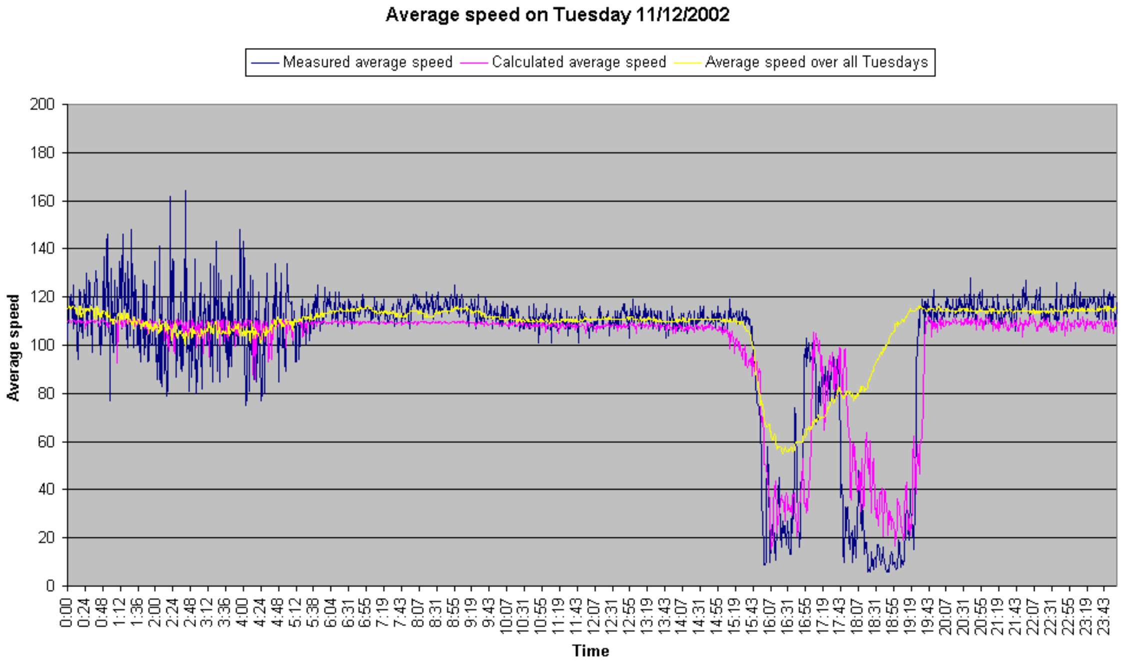


Figure 9.36 The average speed on 12th of November 2002 near Zoetermeer with the 5-5-4-topology 10 minutes ahead

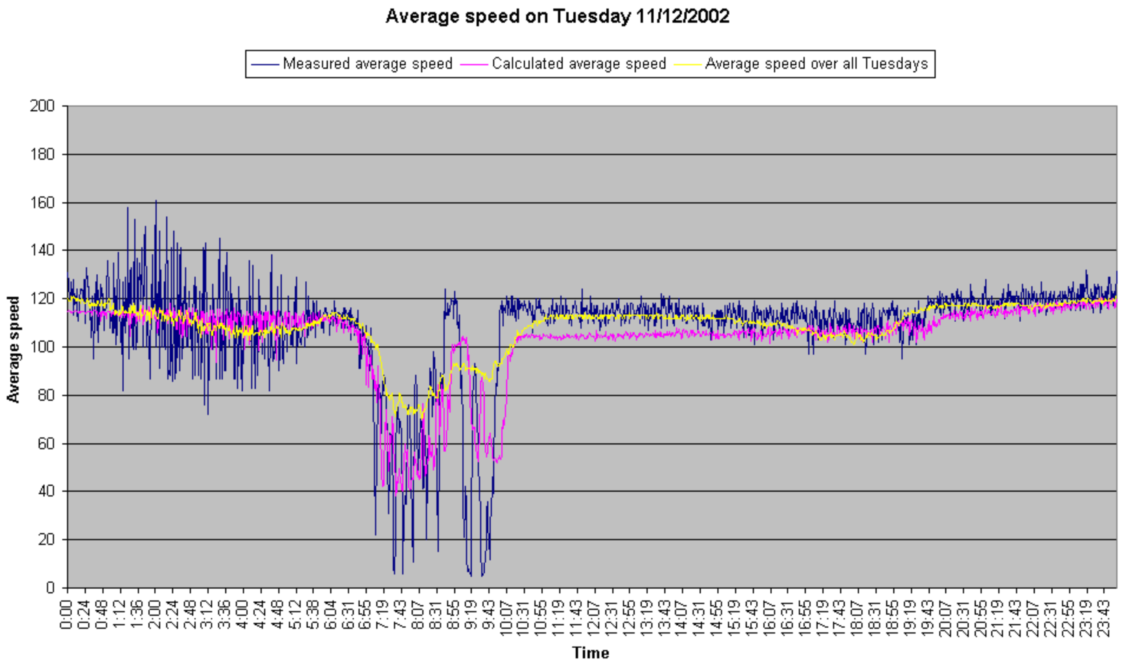


Figure 9.37 The average speed on 12th of November 2002 near Gouda with the 5-5-4-topology 10 minutes ahead

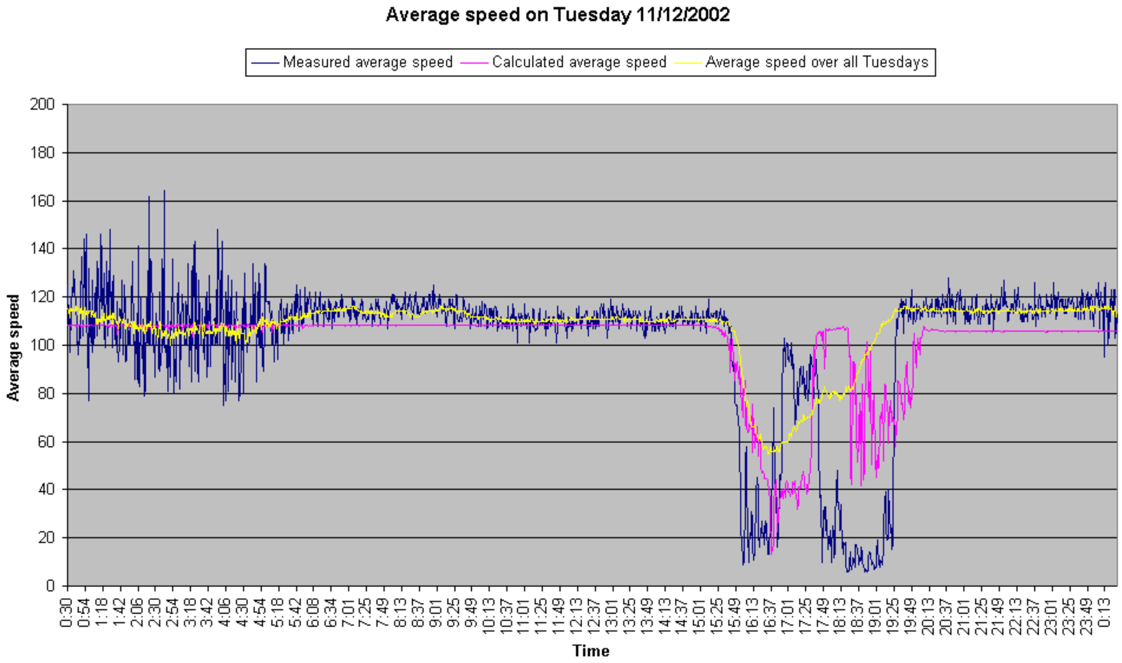


Figure 9.38 The average speed on 12th of November 2002 near Zoetermeer with the 5-5-4-topology 40 minutes ahead

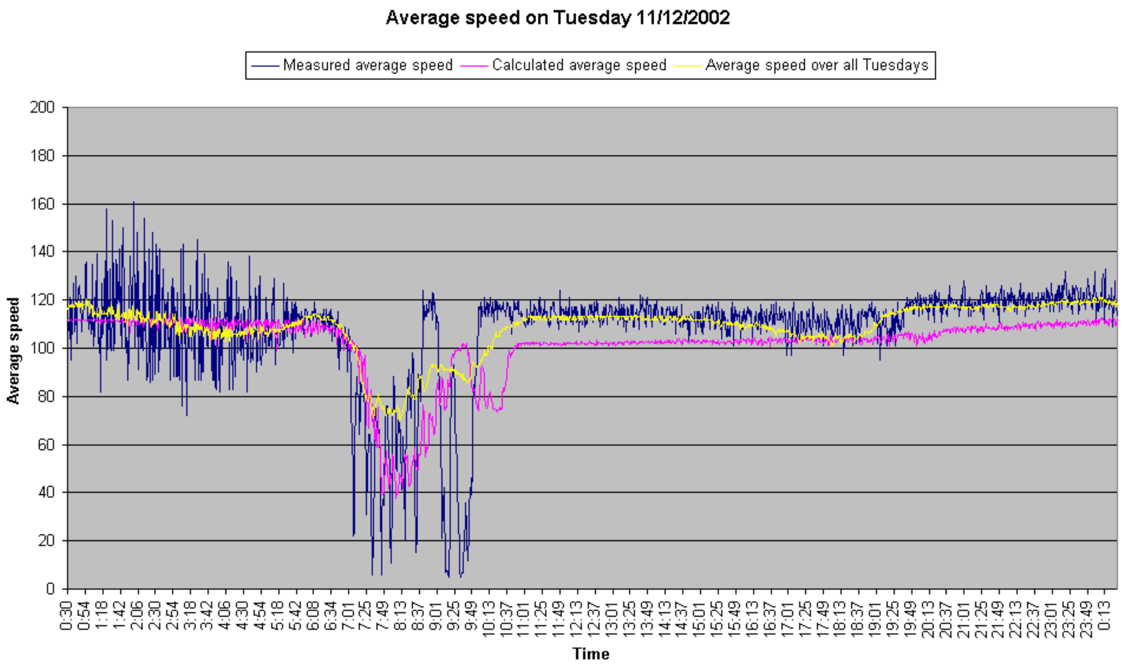


Figure 9.39 The average speed on 12th of November 2002 near Gouda with the 5-5-4-topology 40 minutes ahead

9.3.5 The 5-8-4-topology

The results for Gouda are a lot better with this topology than with the 5-5-4-topology as Table 9.36 shows. The results of the 5-5-4-topology for Zoetermeer compared to those in Table 9.35 do not differ very much.
Still the prediction for Gouda is not as good as it is expected to be with this model. The prediction is only slightly better than the best prediction with the model discussed in Section 9.2.

Table 9.35 Results of the 5-8-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	86.0391	148.449	151.904	171.469
MSE	0.00541	0.00934	0.00956	0.01079
VAR	0.00517	0.0082	0.00912	0.01049
STDev	0.07191	0.09058	0.09549	0.10244

Table 9.36 Results of the 5-8-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	140.761	176.22	207.305	217.085
MSE	0.009	0.01126	0.01325	0.01387
VAR	0.00893	0.01116	0.01317	0.01368
STDev	0.09452	0.10565	0.11477	0.11696

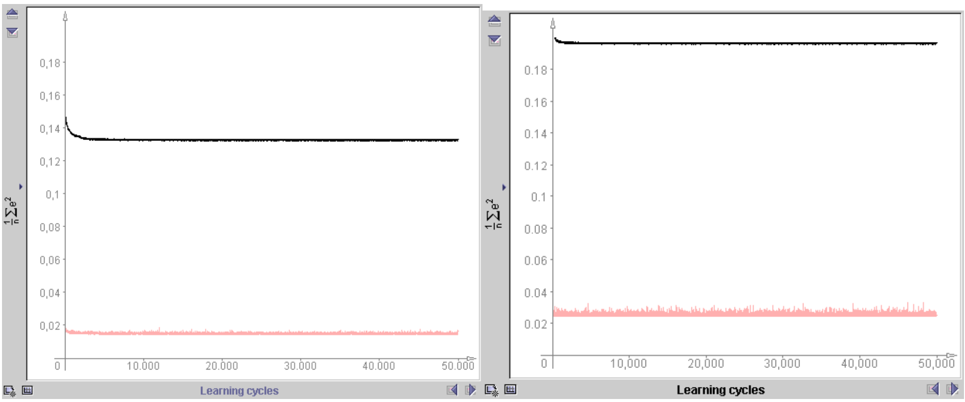


Figure 9.40 The error graphs of the 5-8-4-topology

The error graphs in Figure 9.40 have the same scale. So it is possible to compare these graphs. It can be seen that the error values are lower for the prediction near Zoetermeer, but this information is also given by the results in the tables.

Table 9.37 Results of the 5-8-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	30.031329	52.560051	74.340278	92.507643
MSE	0.0144729	0.0254405	0.0359305	0.044625
VAR	0.0143069	0.0250282	0.0350526	0.0428421
STDev	0.1196114	0.1582031	0.1872234	0.2069833

In Table 9.37 and Table 9.38 the results are displayed of the busiest times. For both places the error is lower than the error with the model without time series. However the difference between the error 40 minutes ahead for Gouda and the error of the best network with the other model is again very small. By comparing Figure 9.42 with Figure 9.44 this can be explained. In Figure 9.42 it can be seen that the prediction will still follow the measured average speed. In the other figure the prediction is more approximating the average speed over all Tuesdays than the real measured speed. This effect can also be seen in Figure 9.41 and Figure 9.43, but the prediction is still very close to the measured speed in these figures.

Table 9.38 Results of the 5-8-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	75.800837	97.43387	118.92678	115.66509
MSE	0.0353384	0.0456365	0.0553921	0.054125
VAR	0.0349123	0.0449004	0.0523049	0.0535409
STDev	0.1868483	0.211897	0.2287026	0.2313891

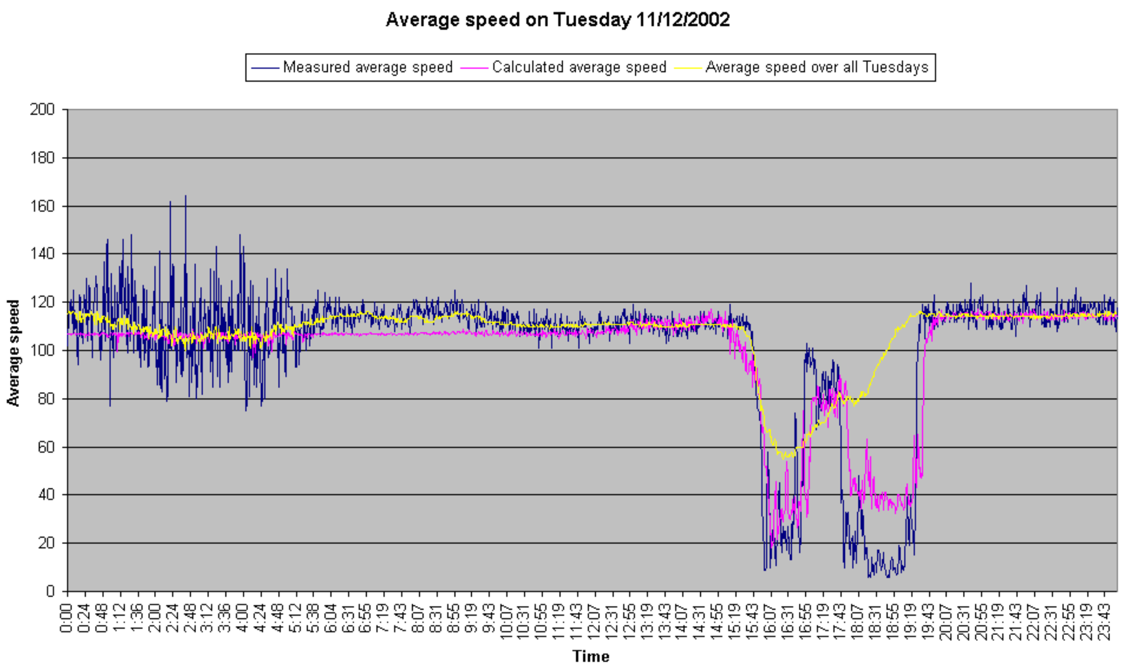


Figure 9.41 The average speed on 12th of November 2002 near Zoetermeer with the 5-8-4-topology 10 minutes ahead

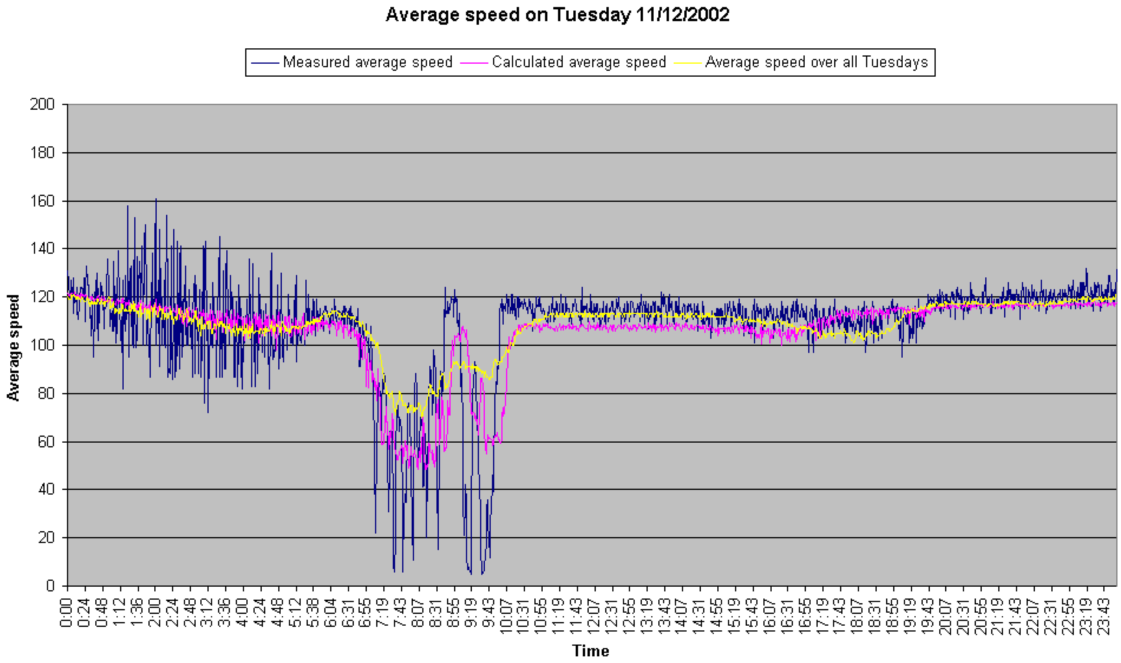


Figure 9.42 The average speed on 12th of November 2002 near Gouda with the 5-8-4-topology 10 minutes ahead

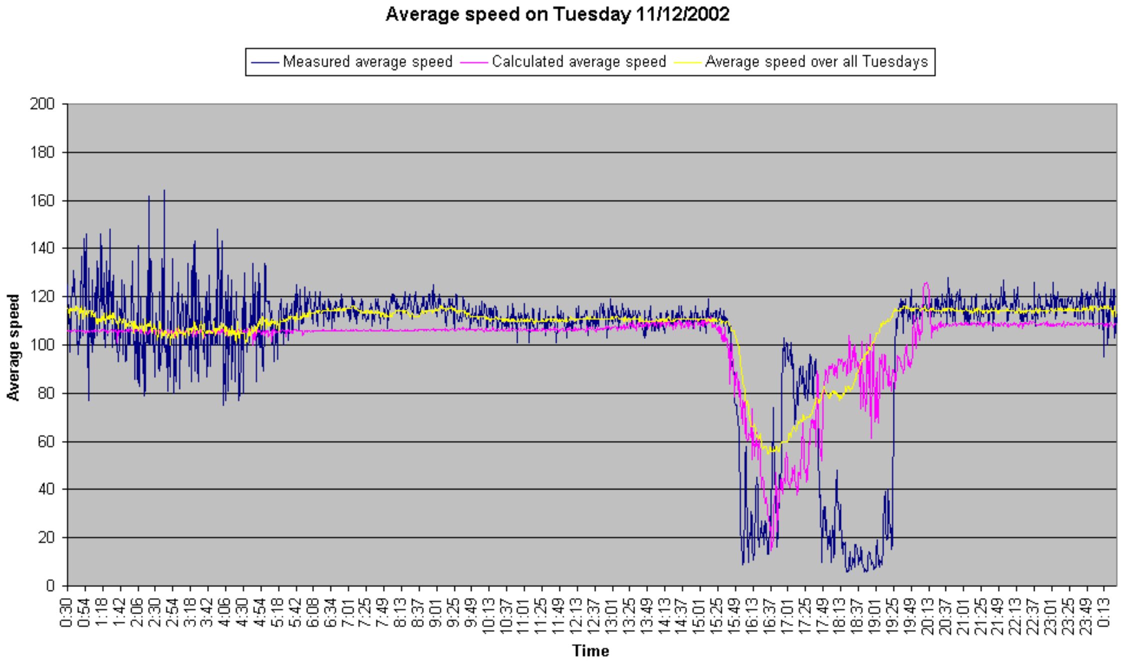


Figure 9.43 The average speed on 12th of November 2002 near Zoetermeer with the 5-8-4-topology 40 minutes ahead

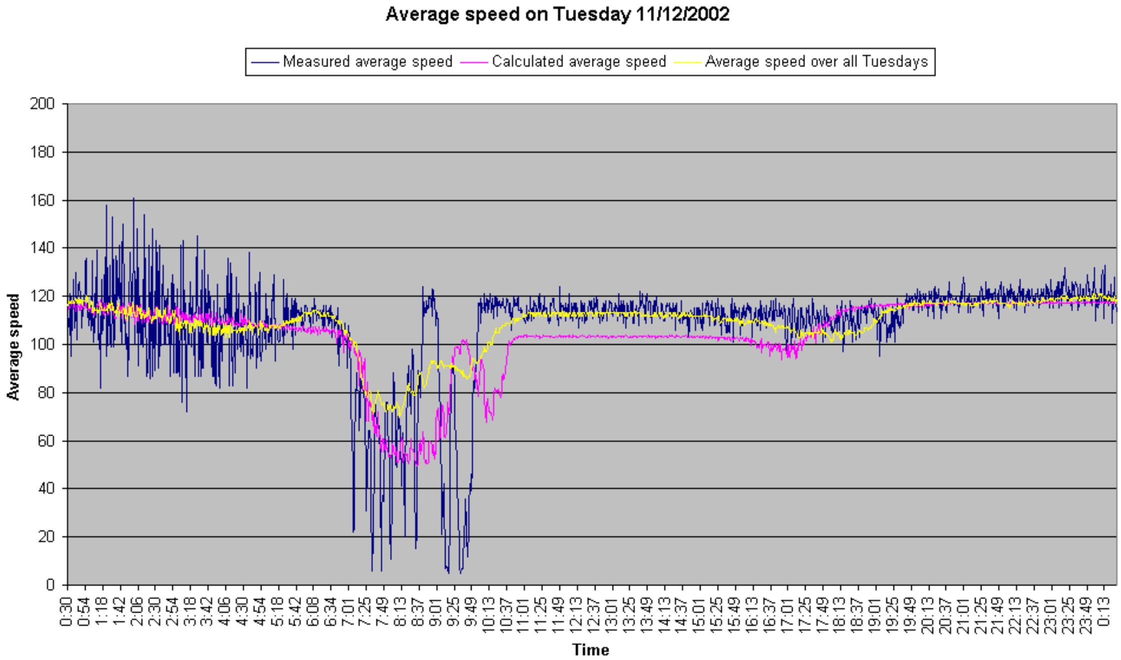


Figure 9.44 The average speed on 12th of November 2002 near Gouda with the 5-8-4-topology 40 minutes ahead

9.3.6 The 5-9-4-topology

For the point near Zoetermeer the results are getting better using a larger network. For Gouda this is not the case. The prediction is somewhat worse than the previous one. The results can be found in Table 9.39 and Table 9.40.

Table 9.39 Results of the 5-9-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	79.9337	116.506	137.843	158.693
MSE	0.00503	0.00733	0.00867	0.00999
VAR	0.00492	0.00722	0.00865	0.00998
STDev	0.07012	0.08497	0.09301	0.09992

Table 9.40 Results of the 5-9-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	151.869	176.967	203.873	217.401
MSE	0.00971	0.01131	0.01303	0.01389
VAR	0.00909	0.01125	0.01303	0.01375
STDev	0.09534	0.10606	0.11414	0.11724

The error graphs in Figure 9.45 have the same scale on the axes and therefore they can be compared. The information given by these graphs is that the prediction for the point near Zoetermeer is more accurate than the prediction for Gouda. The networks will not converge to an error of 0 and so the networks are trained over 50000 cycles and after that the training was stopped.

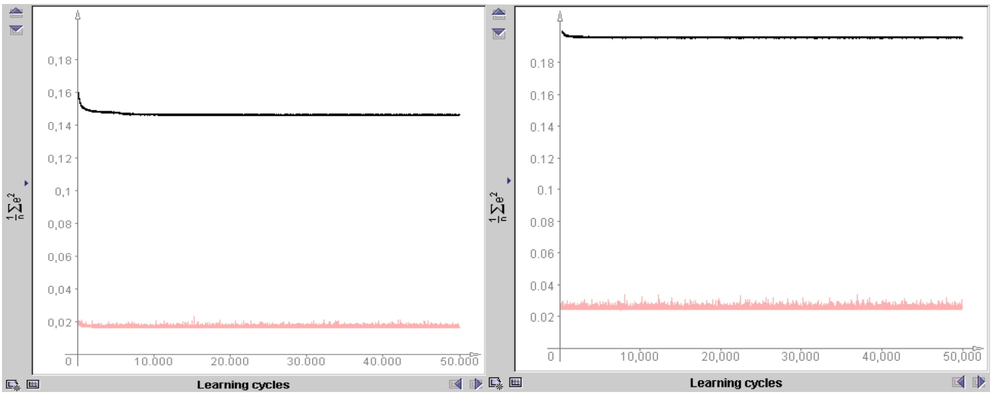


Figure 9.45 The error graphs of the 5-9-4-topology

Table 9.41 Results of the 5-9-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	29.267848	49.913198	70.871822	89.538103
MSE	0.014105	0.0241593	0.0342541	0.0431925
VAR	0.0136996	0.0241622	0.0340554	0.0422227
STDev	0.1170452	0.1554419	0.1845409	0.2054817

Table 9.42 Results of the 5-9-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	74.401027	95.743342	114.36228	117.24065
MSE	0.0346858	0.0448447	0.0532661	0.0548623
VAR	0.0344573	0.0442927	0.051011	0.0535254
STDev	0.1856268	0.2104583	0.2258561	0.2313555

The predictions for both networks in the busiest times, Table 9.41 Table 9.42, are better than the prediction of the 5-8-4-network. However still a remarkable difference can be seen if the errors of the prediction near Zoetermeer are compared to those near Gouda. In the Figure 9.46 to Figure 9.49 the predictions of both networks are shown. In the last figure the predicted speed is adapted to the average speed over all Tuesdays. In Figure 9.48, which is the prediction 40 minutes ahead of the network for the point near Zoetermeer, the predicted line does not seem to approximate the average speed over all Tuesdays.

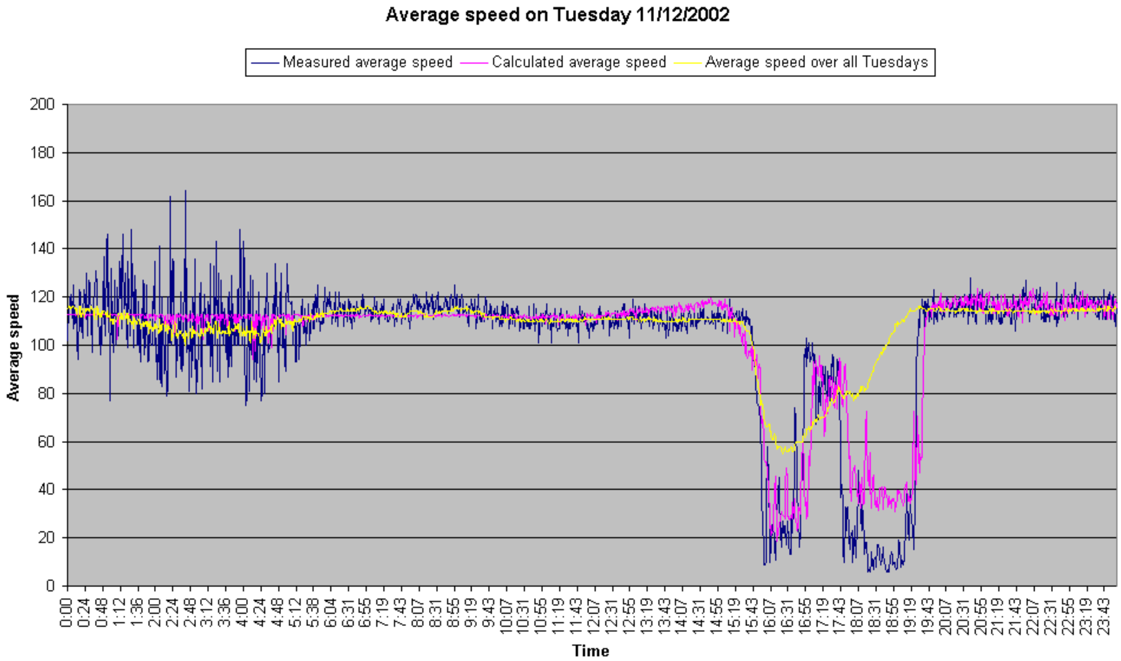


Figure 9.46 The average speed on 12th of November 2002 near Zoetermeer with the 5-9-4-topology 10 minutes ahead

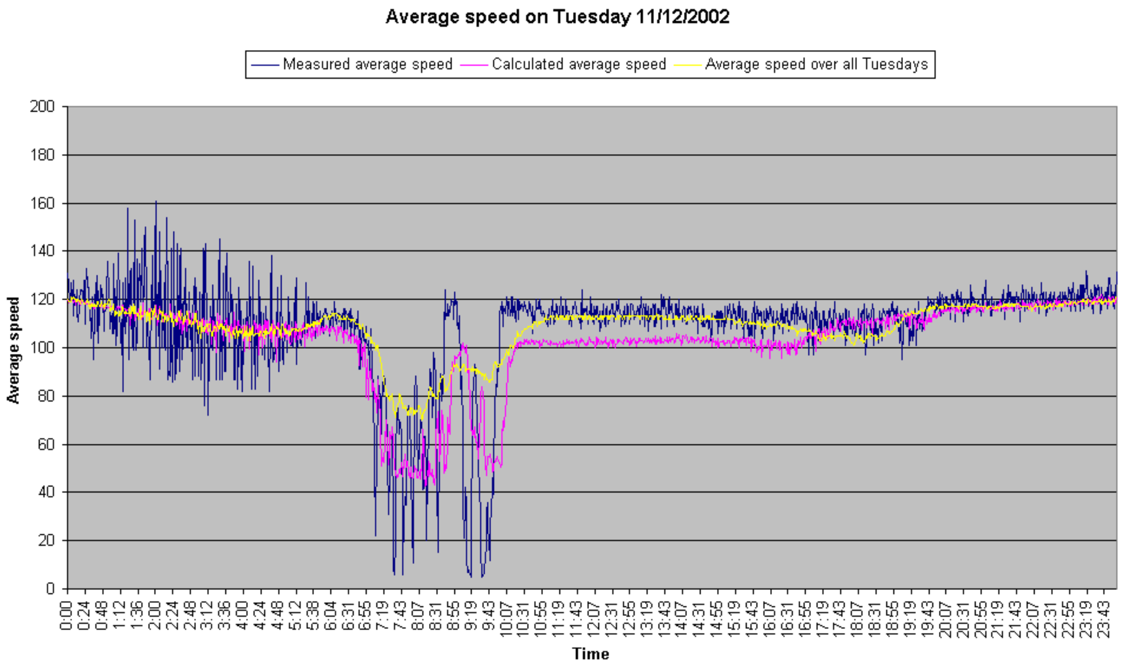


Figure 9.47 The average speed on 12th of November 2002 near Gouda with the 5-9-4-topology 10 minutes ahead

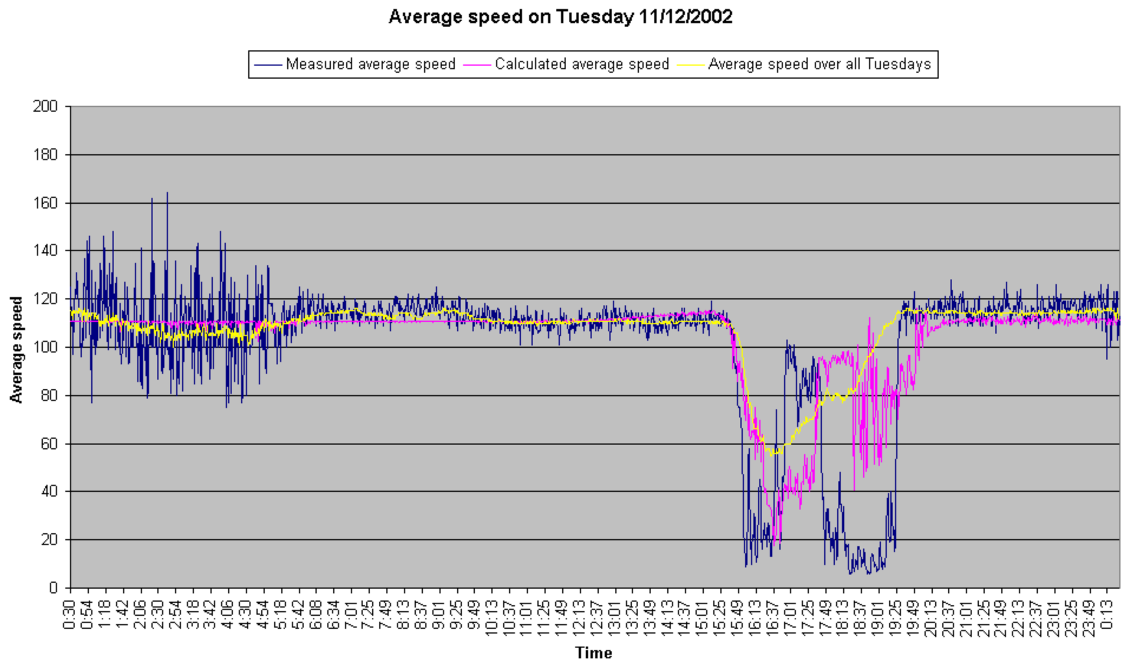


Figure 9.48 The average speed on 12th of November 2002 near Zoetermeer with the 5-9-4-topology 40 minutes ahead

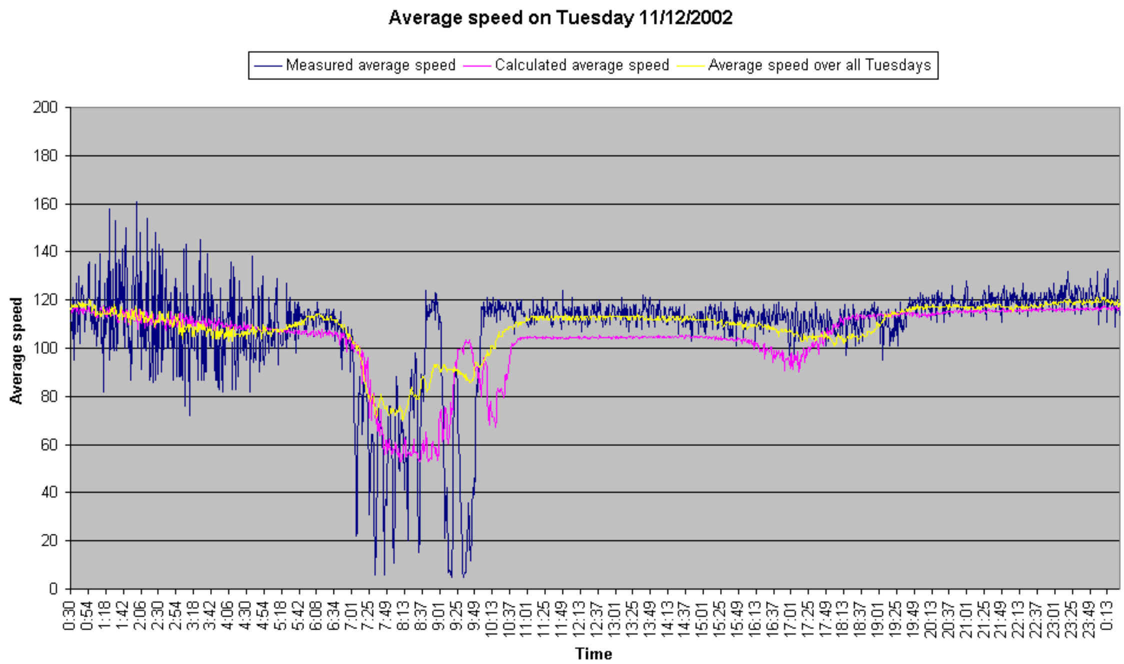


Figure 9.49 The average speed on 12th of November 2002 near Gouda with the 5-9-4-topology 40 minutes ahead

9.3.7 The 5-10-4-topology

Using a 5-10-4-topology will not result in a better prediction. The results of this network are displayed in Table 9.43 and Table 9.44. Most of the values in this table are higher than the ones in the tables of the previous topology, however there are a few values that are lower if this topology is considered.

Table 9.43 Results of the 5-10-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	84.9875	135.834	139.516	156.786
MSE	0.00535	0.00855	0.00878	0.00987
VAR	0.005	0.00761	0.00874	0.00984
STDev	0.07074	0.08724	0.0935	0.09918

Table 9.44 Results of the 5-10-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	146.762	180.312	209.572	230.336
MSE	0.00938	0.01152	0.01339	0.01472
VAR	0.00933	0.01152	0.01335	0.01461
STDev	0.09662	0.10735	0.11556	0.12085

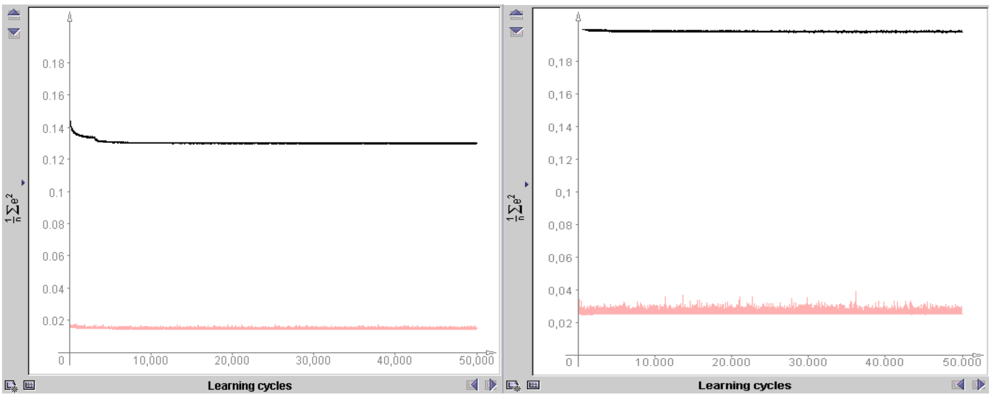


Figure 9.50 The error graphs of the 5-10-4-topology

The errors in the error graphs, Figure 9.50, do not converge to an error of 0. This has resulted in straight lines for the error on the training set, the black line, and the error on the validation set, the pink line. The scaling is equal in both graphs, so they can be compared. The information from these graphs is that the prediction of the point near Zoetermeer is better than the prediction for Gouda. This information can also be seen in the tables of this section.

The prediction for the average speed in the busiest times near Zoetermeer is almost the same as the prediction with the 5-9-4-topology. Some of the differences are positive for this network and others a little bit negative. The results can be found in Table 9.45.

For Gouda the prediction with this network for the busiest times, Table 9.46, is worse on all predictions.

Table 9.45 Results of the 5-10-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	28.421177	50.376321	72.922296	88.103362
MSE	0.013697	0.0243835	0.0352452	0.0425004
VAR	0.0136077	0.0237123	0.0331084	0.0400263
STDev	0.1166522	0.1539878	0.1819573	0.2000659

Table 9.46 Results of the 5-10-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	79.409582	102.35157	116.51867	126.40139
MSE	0.0370208	0.0479398	0.0542705	0.059149
VAR	0.0363374	0.045937	0.0522167	0.0566879
STDev	0.1906236	0.2143292	0.2285098	0.2380921

Figure 9.51 and Figure 9.53 show a fairly good prediction for the place near Zoetermeer. The prediction for the point near Gouda can be seen in the Figure 9.52 and Figure 9.54. Analyzing these figures leads to the conclusion that the prediction is following the average speed over all Tuesdays more than the measured average speed.

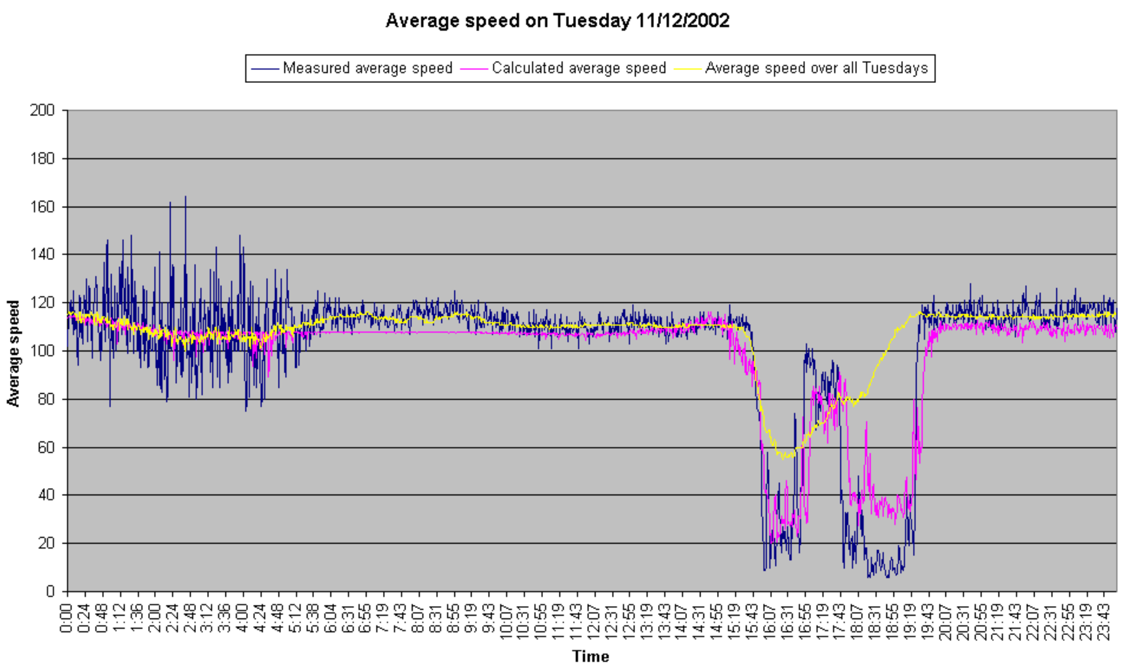


Figure 9.51 The average speed on 12th of November 2002 near Zoetermeer with the 5-10-4-topology 10 minutes ahead

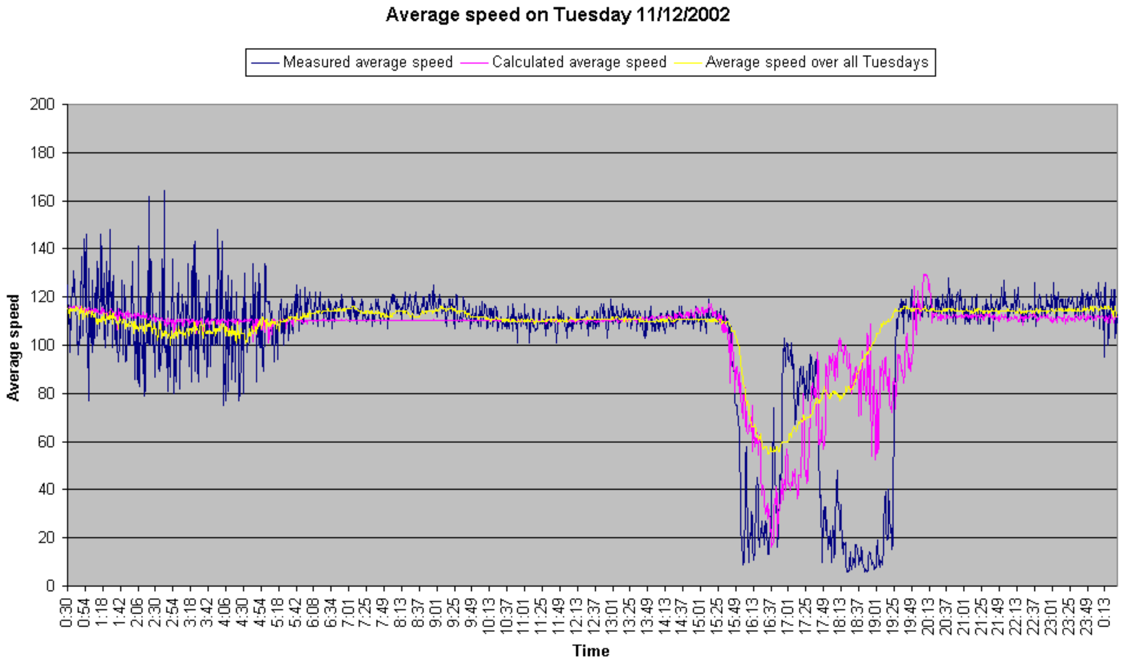


Figure 9.52 The average speed on 12th of November 2002 near Gouda with the 5-10-4-topology 10 minutes ahead

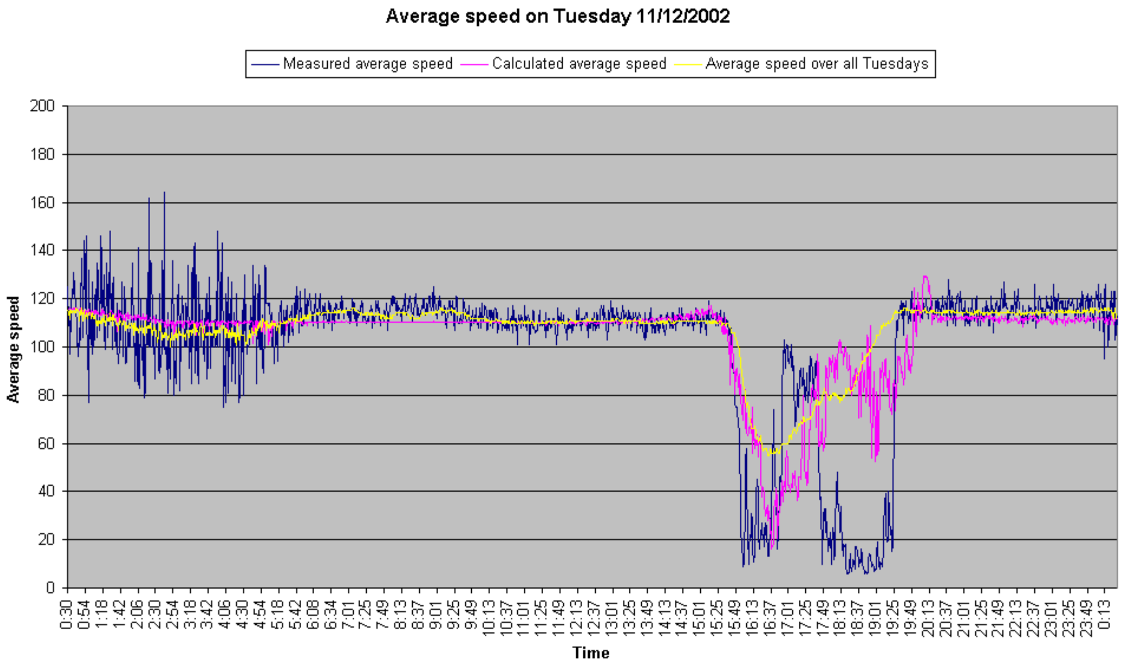


Figure 9.53 The average speed on 12th of November 2002 near Zoetermeer with the 5-10-4-topology 40 minutes ahead

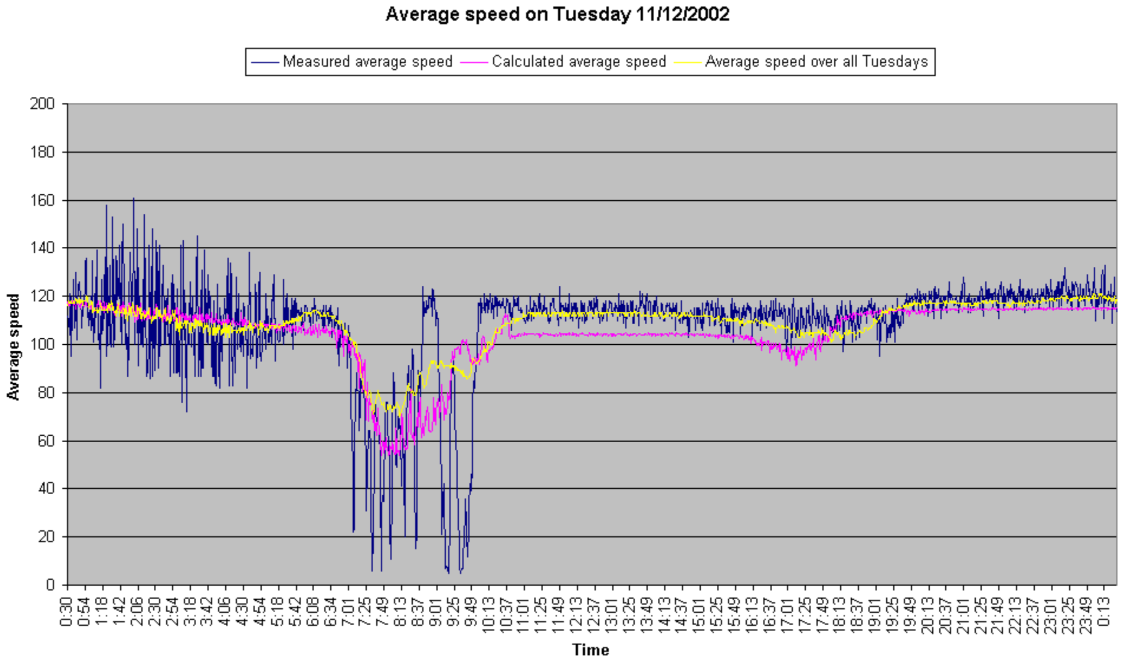


Figure 9.54 The average speed on 12th of November 2002 near Gouda with the 5-10-4-topology 40 minutes ahead

9.3.8 The 5-11-4-topology

The predictions 10 and 20 minutes ahead are better for the place near Zoetermeer compared to those of the 5-10-4-network. The other two predictions, those of 30 and 40 minutes ahead, have larger errors. For Gouda it is the other way around. The predictions for the 10, 20 and 30 minutes ahead are worse than those of the 5-10-4-topology. However the prediction 40 minutes in advance has a much smaller error. Table 9.47 and Table 9.48 show the results for this topology.

Table 9.47 Results of the 5-11-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	83.1708	118.438	154.609	177.002
MSE	0.00523	0.00745	0.00973	0.01114
VAR	0.00523	0.00744	0.00896	0.01041
STDev	0.07231	0.08625	0.09467	0.10205

Table 9.48 Results of the 5-11-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	149.828	185.861	216.947	216.09
MSE	0.00957	0.01188	0.01386	0.01381
VAR	0.00951	0.01177	0.01376	0.01369
STDev	0.09754	0.10847	0.1173	0.11701

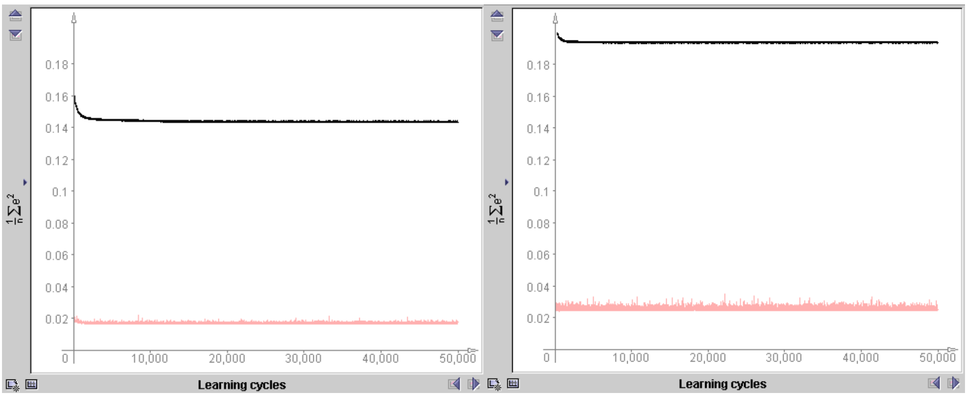


Figure 9.55 The error graphs of the 5-11-4-topology

Like the results in the tables show the prediction for the place near Zoetermeer is still better compared to the prediction on the point near Gouda. This information can also be given by comparing the two error graphs of Figure 9.55.

Table 9.49 Results of the 5-11-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	31.347704	53.727166	81.856648	102.64156
MSE	0.0151073	0.0260054	0.0395634	0.0495135
VAR	0.0140559	0.0248253	0.035296	0.0432135
STDev	0.1185574	0.1575604	0.1878722	0.2078785

Table 9.49 and Table 9.50 give the results of the prediction in the busiest times. The prediction 40 minutes ahead for the place near Gouda is only a little bit better than the prediction without time series, but it is better than the prediction with the 5-10-4-topology. The other three predictions for this place are worse with this topology. The predictions of the network representing the place near Zoetermeer are also higher than those of the previous network, but at least they are much smaller than those of the best network without time series.

Table 9.50 Results of the 5-11-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	83.094721	103.28508	119.02648	117.7906
MSE	0.0387388	0.0483771	0.0554385	0.0551196
VAR	0.0373218	0.0465389	0.0531004	0.0538942
STDev	0.1931884	0.2157288	0.2304352	0.2321512

The figures of this topology, Figure 9.56 to Figure 9.59, show again that the prediction for the point near Zoetermeer approximates the measured speed. The prediction for the point near Gouda is more an estimation of the average speed over all Tuesdays, like the topologies without the time series.

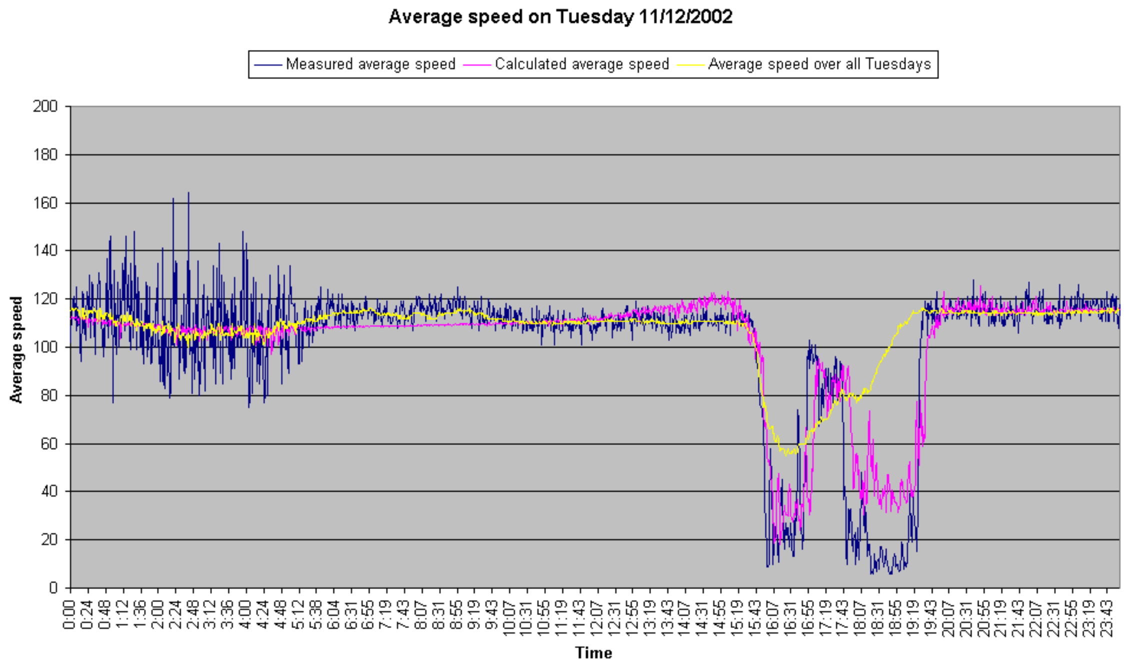


Figure 9.56 The average speed on 12th of November 2002 near Zoetermeer with the 5-11-4-topology 10 minutes ahead

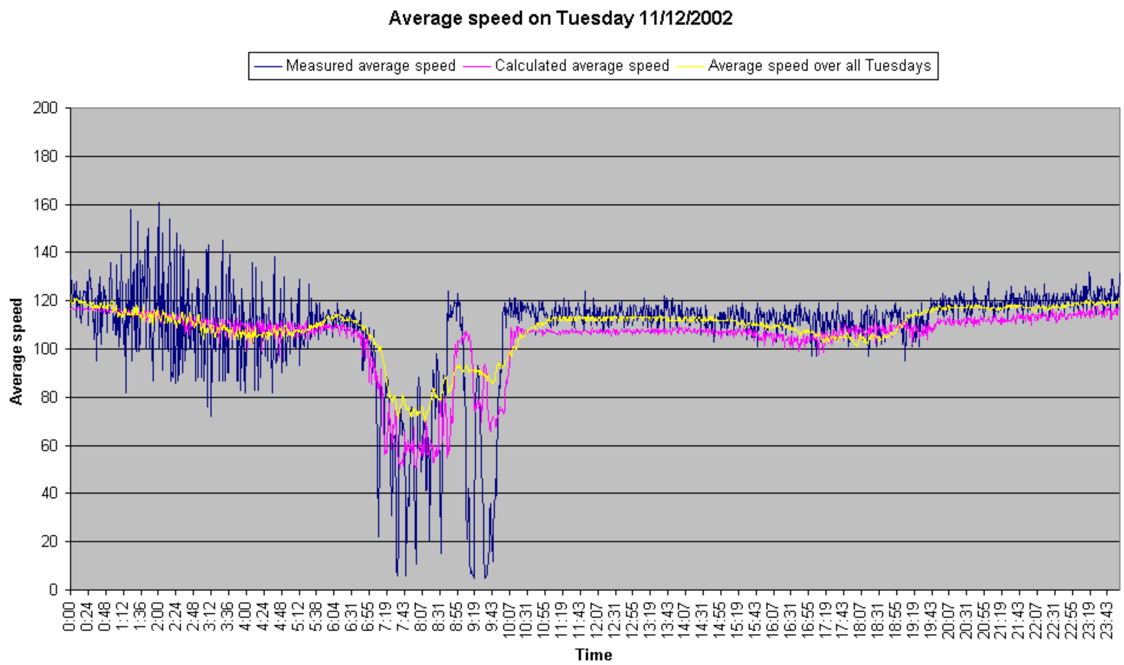


Figure 9.57 The average speed on 12th of November 2002 near Gouda with the 5-11-4-topology 10 minutes ahead

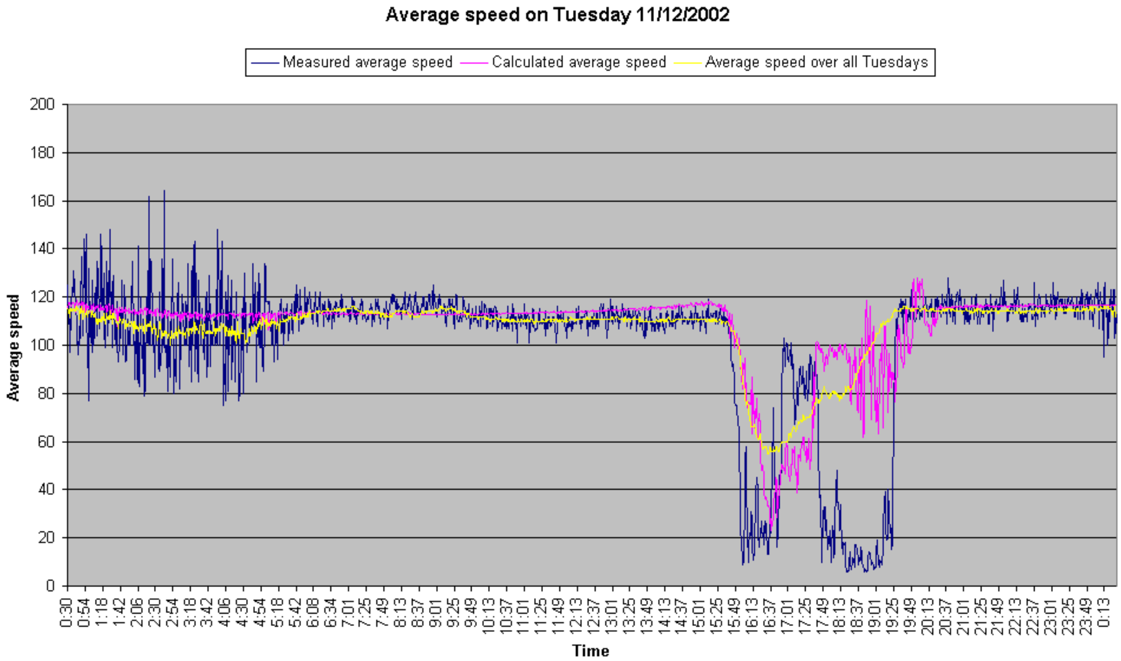


Figure 9.58 The average speed on 12th of November 2002 near Zoetermeer with the 5-11-4-topology 40 minutes ahead

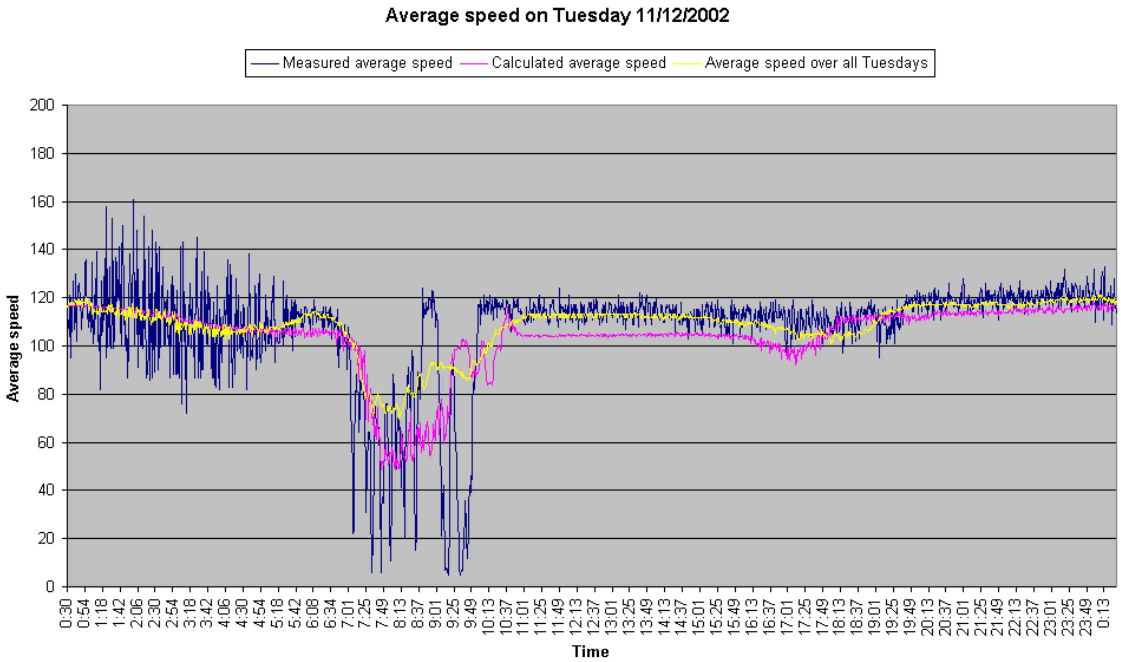


Figure 9.59 The average speed on 12th of November 2002 near Gouda with the 5-11-4-topology 40 minutes ahead

9.3.9 The 5-15-4-topology

The topology with 15 hidden neurons does not contribute anything to the results of the 5-11-4-topologies for the place near Zoetermeer. For the place near Gouda some of the results are better than those of a few topologies before, also some errors have higher values. In Table 9.51 and Table 9.52 these results can be found.

Table 9.51 Results of the 5-15-4-topology near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	84.9339	135.233	170.014	188.823
MSE	0.00534	0.00851	0.0107	0.01188
VAR	0.0052	0.00806	0.00975	0.01186
STDev	0.07212	0.08976	0.09872	0.10891

Table 9.52 Results of the 5-15-4-topology near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	148.514	183.837	203.71	229.825
MSE	0.00949	0.01175	0.01302	0.01469
VAR	0.00941	0.01111	0.01271	0.01406
STDev	0.09699	0.10539	0.11273	0.11859

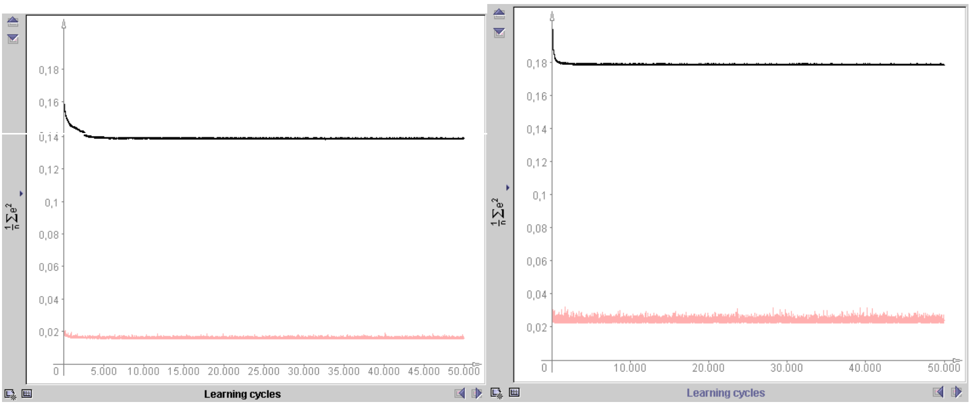


Figure 9.60 The error graphs of the 5-15-4-topology

The networks have the same error graphs like all the other topologies. The networks will not converge to an error of 0, so there are straight lines in the graphs. The graphs can be found in Figure 9.60.

Table 9.53 Results of the 5-15-4-topology for the busiest times (4.00 PM – 7.00 PM) near Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	32.723851	65.871805	93.050333	110.25872
MSE	0.0157705	0.0318837	0.0449736	0.053188
VAR	0.0150692	0.0287572	0.0404715	0.0499521
STDev	0.1227565	0.1695794	0.2011754	0.2234996

Table 9.54 Results of the 5-15-4-topology for the busiest times (7.00 AM – 10.00 AM) near Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	82.343493	95.085737	108.94538	116.84075
MSE	0.0383886	0.0445366	0.0507431	0.0546751
VAR	0.0378396	0.0444959	0.05027	0.0543104
STDev	0.1945241	0.2109406	0.2242096	0.233046

The results for the busiest times near Zoetermeer are worse with this network than with the 5-11-4-network. For Gouda the 5-15-4-topology gives always a better prediction of the measured speed in the busiest times. These results are displayed in Table 9.53 and Table 9.54.
In Figure 9.61 to Figure 9.64 the predicted and the measured average speed is showed. Also the average speed over all Tuesdays can be seen. The prediction for Gouda is approximating the average speed over all Tuesdays. The place near Zoetermeer has a prediction based on the measured speed for 10 minutes ahead. For 40 minutes ahead it is finding a solution to fit to the measured speed as well as to the average speed over all Tuesdays.

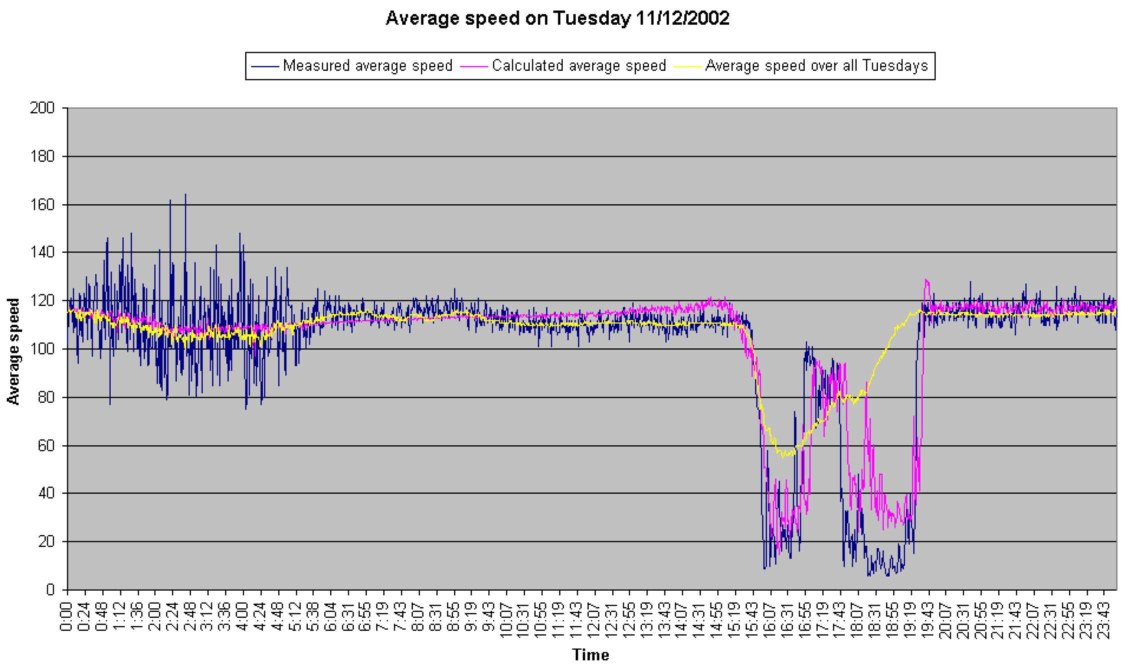


Figure 9.61 The average speed on 12th of November 2002 near Zoetermeer with the 5-15-4-topology 10 minutes ahead

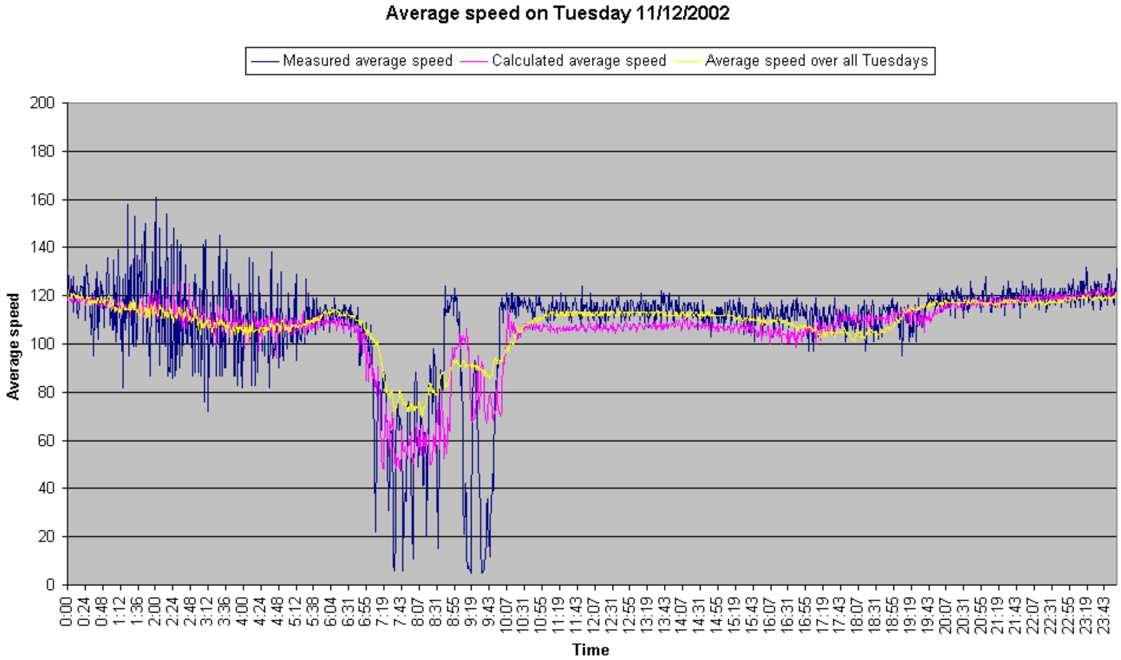


Figure 9.62 The average speed on 12th of November 2002 near Gouda with the 5-15-4-topology 10 minutes ahead

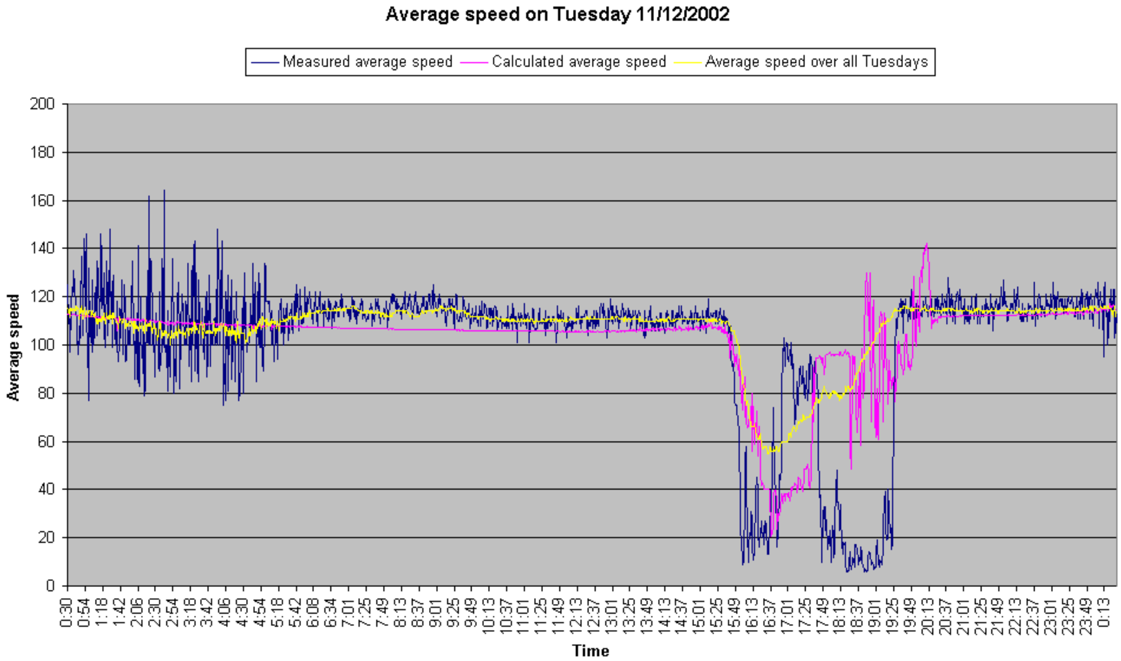


Figure 9.63 The average speed on 12th of November 2002 near Zoetermeer with the 5-15-4-topology 40 minutes ahead

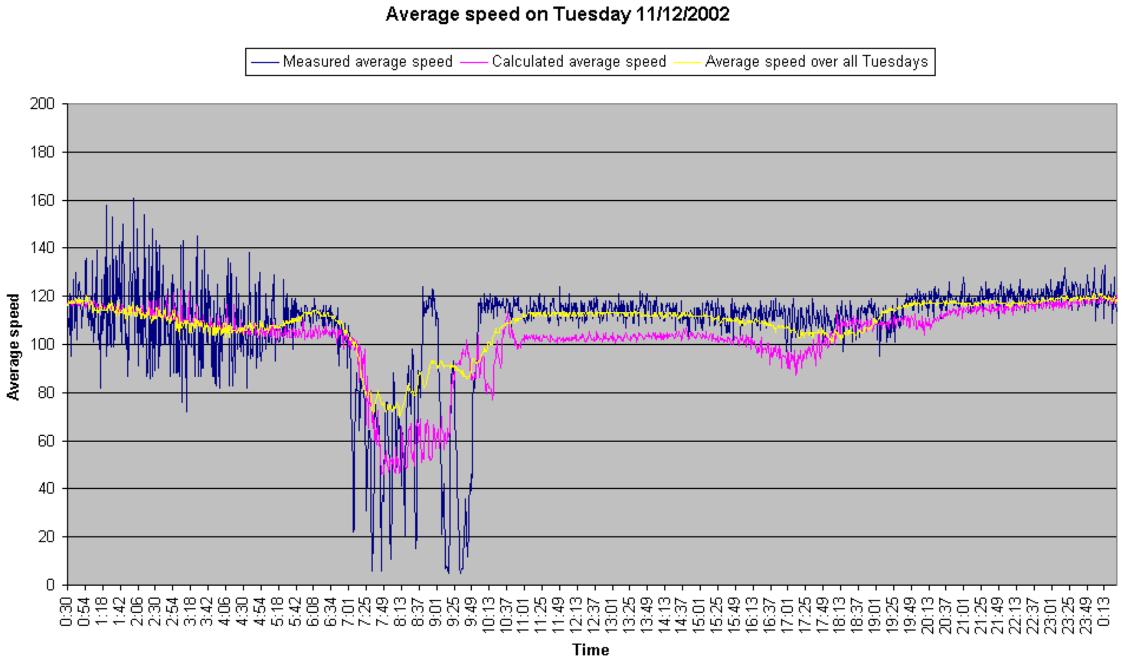


Figure 9.64 The average speed on 12th of November 2002 near Gouda with the 5-15-4-topology 40 minutes ahead

9.3.10 Conclusions

Above the different topologies where discussed resulting from the formulas of Section 5.4. In this part the results of the topologies are listed in tables for each prediction. Table 9.55 to Table 9.58 give the predictions for the place near Zoetermeer. The bold numbers are the lowest value of a column. The predictions 10 20 and 30 minutes ahead are best predicted by the network with the 5-9-4-topology. The last one, the prediction 40 minutes ahead, has the best results with the 5-10-4-network. However in this table, Table 9.58, it can be seen that this result does not differ very much from those of the 5-9-4-network. In general it can be said that the network with the 5-9-4-topology gives the best result for Zoetermeer.

Table 9.55 Results of all the networks with time series near Zoetermeer 10 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	102.866	0.00647	0.00647	0.07798
5-2-4	96.1862	0.00605	0.00589	0.07674
5-4-4	94.3644	0.00594	0.00529	0.0727
5-5-4	85.8817	0.0054	0.00516	0.07183
5-8-4	86.0391	0.00541	0.00517	0.07191
5-9-4	**79.9337**	**0.00503**	**0.00492**	**0.07012**
5-10-4	84.9875	0.00535	0.005	0.07074
5-11-4	83.1708	0.00523	0.00523	0.07231
5-15-4	84.9339	0.00534	0.0052	0.07212

Table 9.56 Results of all the networks with time series near Zoetermeer 20 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	172.324	0.01084	0.00967	0.09835
5-2-4	165.008	0.01038	0.00955	0.09773
5-4-4	127.705	0.00804	0.00769	0.08767
5-5-4	136.465	0.00859	0.00778	0.08822
5-8-4	148.449	0.00934	0.0082	0.09058
5-9-4	**116.506**	**0.00733**	**0.00722**	**0.08497**
5-10-4	135.834	0.00855	0.00761	0.08724
5-11-4	118.438	0.00745	0.00744	0.08625
5-15-4	135.233	0.00851	0.00806	0.08976

Table 9.57 Results of all the networks with time series near Zoetermeer 30 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	204.839	0.01289	0.01262	0.11236
5-2-4	204.203	0.01285	0.01257	0.1121
5-4-4	144.126	0.00907	0.00907	0.09523
5-5-4	154.491	0.00972	0.00935	0.09668
5-8-4	151.904	0.00956	0.00912	0.09549
5-9-4	**137.843**	**0.00867**	**0.00865**	**0.09301**
5-10-4	139.516	0.00878	0.00874	0.0935
5-11-4	154.609	0.00973	0.00896	0.09467
5-15-4	170.014	0.0107	0.00975	0.09872

Table 9.58 Results of all the networks with time series near Zoetermeer 40 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	265.031	0.01668	0.01547	0.12438
5-2-4	265.394	0.01613	0.01537	0.12398
5-4-4	164.652	0.01036	0.01036	0.10178
5-5-4	171.02	0.01076	0.01046	0.1023
5-8-4	171.469	0.01079	0.01049	0.10244
5-9-4	158.693	0.00999	0.00998	0.09992
5-10-4	**156.786**	**0.00987**	**0.00984**	**0.09918**
5-11-4	177.002	0.01114	0.01041	0.10205
5-15-4	188.823	0.01188	0.01186	0.10891

If the busiest times are considered for Zoetermeer 10 and 40 minutes ahead are best predicted by the 5-10-4-network. The 5-9-4-topology resulted in a network giving lower errors for the 20 and 30 minutes ahead. The results can be found in Table 9.59 to Table 9.62. The differences between the results of the 5-10-4-network and the 5-9-4-network are very small. Because it was clear that for the prediction of the average speed on the place near Zoetermeer of all the times the 5-9-4-network was the best it can be said that it gives generally the best results over all the different predictions.

Table 9.59 Results of all the networks with time series for the busiest times near Zoetermeer 10 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	36.089635	0.0173926	0.0170442	0.1305535
5-2-4	34.9753	0.0168556	0.0166077	0.1288707
5-4-4	32.451757	0.0156394	0.0155319	0.1246272
5-5-4	31.728506	0.0152908	0.0150822	1.1228094
5-8-4	30.031329	0.0144729	0.0143069	0.1196114
5-9-4	29.267848	0.014105	0.0136996	0.1170452
5-10-4	**28.421177**	**0.013697**	**0.0136077**	**0.1166522**
5-11-4	31.347704	0.0151073	0.0140559	0.1185574
5-15-4	32.723851	0.0157705	0.015069	0.1227565

Table 9.60 Results of all the networks with time series for the busiest times near Zoetermeer 20 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	68.394601	0.0331048	0.0324026	0.1800072
5-2-4	67.607243	0.0327237	0.0319687	0.178798
5-4-4	56.900817	0.0275415	0.0273643	0.1654215
5-5-4	54.448013	0.0263543	0.0263108	0.162206
5-8-4	52.560051	0.0254405	0.0250282	0.1582031
5-9-4	**49.913198**	**0.0241593**	0.0241622	0.1554419
5-10-4	50.376321	0.0243835	**0.0237123**	**0.1539878**
5-11-4	53.727166	0.0260054	0.0248253	0.1575604
5-15-4	65.871805	0.0318837	0.0287572	0.1695794

Table 9.61 Results of all the networks with time series for the busiest times near Zoetermeer 30 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	107.7134	0.0520606	0.0481736	0.2194849
5-2-4	106.19899	0.0513287	0.0475185	0.2179873
5-4-4	76.560132	0.03845534	0.0376802	0.1941139
5-5-4	75.253464	0.0363719	0.036041	0.0445523
5-8-4	74.340278	0.0359305	0.0350526	0.1872234
5-9-4	**70.871822**	**0.0342541**	0.0340554	0.1845409
5-10-4	72.922296	0.0352452	**0.0331084**	**0.1819573**
5-11-4	81.856648	0.0395634	0.035296	0.1878722
5-15-4	93.050333	0.0449736	0.0404715	0.2011754

Table 9.62 Results of all the networks with time series for the busiest times near Zoetermeer 40 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	137.7193	0.0664348	0.0603701	0.2457033
5-2-4	138.2161	0.0666744	0.0599992	0.2449473
5-4-4	96.245351	0.0464281	0.0451985	0.2125994
5-5-4	93.602954	0.0451534	0.0445523	0.2110742
5-8-4	92.507643	0.044625	0.0428421	0.2069833
5-9-4	89.538103	0.0431925	0.0422227	0.2054817
5-10-4	**88.103362**	**0.0425004**	**0.0400263**	**0.2000659**
5-11-4	102.64156	0.0495135	0.0432135	0.2078785
5-15-4	110.25872	0.053188	0.0499521	0.2234996

For the predictions near Gouda at first it has to be explained why the errors are so much higher than those for the prediction near Zoetermeer. Only the first two networks were able to give a fairly good prediction by giving the input time series as the output with a small modification. This can be seen in the Table 9.63 to Table 9.70. A possible explanation is that the test set consists of data capturing a lot of special events like accidents. If these events are available with low frequency in the training set the network is not trained on these events and is therefore not able to recognize them. It can be checked if this theory is correct. For that reason the network is tested again, but this time with the validation set, on the 5-1-4-network and the 5-9-4-network. These results are the red colored values that are displayed in the tables. The red colored values show a much better prediction than the prediction on the test set. It can also be seen that the 5-1-4-topology gives worse results than the 5-9-4-topology when the validation set is used. Therefore the 5-9-4-topology is taken as the best topology to predict the average speed on both places using the model with time series.

The 5-9-4-topology resulted from the formula of Fletcher and Goss. According to them the number of hidden neurons should be lying between $2N_i+1$ and $2\sqrt{N_i}+N_o$, 11 to 8 hidden neurons in this case. One of the formulas determining the best topology for the neural networks based on the model without time series is also the formula of Fletcher and Goss.

Table 9.63 Results of all the networks with time series near Gouda 10 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	169.237	0.01082	**0.00841**	**0.09172**
5-1-4	146.497	0.00954	0.00663	0.08145
5-2-4	146.557	0.00937	0.00871	0.0933
5-4-4	145.244	0.00928	0.00893	0.09452
5-5-4	143.576	0.00981	0.00883	0.09399
5-8-4	**140.761**	**0.009**	0.00893	0.09452
5-9-4	151.869	0.00971	0.00909	0.09534
5-9-4	122.333	0.00796	0.00683	0.08266
5-10-4	146.762	0.00938	0.00933	0.09662
5-11-4	149.828	0.00957	0.00951	0.09754
5-15-4	148.514	0.00949	0.00941	0.09699

Table 9.64 Results of all the networks with time series near Gouda 20 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	167.667	0.01072	0.01016	0.10076
5-1-4	135.711	0.00883	0.00796	0.08919
5-2-4	**158.011**	**0.0101**	**0.01006**	**0.1003**
5-4-4	176.101	0.01125	0.01117	0.10568
5-5-4	177.092	0.01132	0.01102	0.10499
5-8-4	176.22	0.01126	0.01116	0.10565
5-9-4	176.967	0.01131	0.01125	0.10606
5-9-4	120.665	0.00785	0.00746	0.08639
5-10-4	180.312	0.01152	0.01152	0.10735
5-11-4	185.861	0.01188	0.01177	0.10847
5-15-4	183.837	0.01175	0.01111	0.10539

Table 9.65 Results of all the networks with time series near Gouda 30 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	255.556	0.01633	0.01173	0.1083
5-1-4	224.307	0.0146	0.00895	0.09461
5-2-4	**192.199**	**0.01228**	**0.01141**	**0.10682**
5-4-4	209.334	0.01338	0.0129	0.11357
5-5-4	207.076	0.01323	0.0129	0.11358
5-8-4	207.305	0.01325	0.01317	0.11477
5-9-4	203.873	0.01303	0.01303	0.11414
5-9-4	127.299	0.00829	0.00805	0.08972
5-10-4	209.572	0.01339	0.01335	0.11556
5-11-4	216.947	0.01386	0.01376	0.1173
5-15-4	203.71	0.01302	0.01271	0.11273

Table 9.66 Results of all the networks with time series near Gouda 40 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	245.468	0.1569	0.01294	0.11374
5-1-4	212.694	0.01384	0.01008	0.10042
5-2-4	**200.267**	**0.0128**	**0.01262**	**0.11235**
5-4-4	219.195	0.01401	0.01401	0.11835
5-5-4	239.423	0.0153	0.0146	0.12081
5-8-4	217.085	0.01387	0.01368	0.11696
5-9-4	217.401	0.01389	0.01375	0.11724
5-9-4	145.298	0.00946	0.00867	0.09309
5-10-4	230.336	0.01472	0.01461	0.12085
5-11-4	216.09	0.01381	0.01369	0.11701
5-15-4	229.825	0.01496	0.01406	0.11859

Table 9.67 Results of all the networks with time series for the busiest times near Gouda 10 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	**64.991528**	**0.0302991**	**0.0302958**	**0.1740569**
5-1-4	51.211257	0.0249689	0.0244666	0.156418
5-2-4	67.130856	0.0312964	0.0311851	0.176593
5-4-4	72.845705	0.0339607	0.0338982	0.1841147
5-5-4	73.828843	0.034419	0.0344054	0.1854869
5-8-4	75.800837	0.0353384	0.0349123	0.1868483
5-9-4	74.401027	0.0346858	0.0344573	0.1856268
5-9-4	55.058413	0.0268447	0.0243381	0.1560066
5-10-4	79.409582	0.0370208	0.0363374	0.1906236
5-11-4	83.094721	0.0387388	0.0373218	0.1931884
5-15-4	82.343493	0.0383886	0.0378396	0.1945241

Table 9.68 Results of all the networks with time series for the busiest times near Gouda 20 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	**80.309479**	**0.0376157**	**0.0359345**	**0.1895639**
5-1-4	63.83676	0.0313848	0.0311698	0.1765497
5-2-4	80.928722	0.0379057	0.0361079	0.1900207
5-4-4	96.154807	0.0450374	0.0439812	0.2097171
5-5-4	91.921011	0.0430543	0.0427769	0.2068257
5-8-4	97.43387	0.0456365	0.0449004	0.211897
5-9-4	95.743342	0.0448447	0.0442927	0.2104583
5-9-4	59.319112	0.0291638	0.0286647	0.1693067
5-10-4	102.35157	0.0479398	0.045937	0.2143292
5-11-4	103.28508	0.0483771	0.0465389	0.2157288
5-15-4	95.085737	0.0445366	0.0444959	0.2109406

Table 9.69 Results of all the networks with time series for the busiest times near Gouda 30 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	**87.141264**	**0.0405875**	**0.040294**	**0.2007335**
5-1-4	73.193604	0.0360915	0.0358226	0.1892686
5-2-4	90.277952	0.0420484	0.0403849	0.20096
5-4-4	110.68599	0.0515538	0.0508409	0.2254793
5-5-4	109.40986	0.0509594	0.0501957	0.224044
5-8-4	118.92678	0.0553921	0.0523049	0.2287026
5-9-4	114.36228	0.0532661	0.051011	0.2258561
5-9-4	65.822813	0.032457	0.0323847	0.1799576
5-10-4	116.51867	0.0542705	0.0522167	0.2285098
5-11-4	119.02648	0.0554385	0.0531004	0.2304352
5-15-4	108.94538	0.0507431	0.05027	0.2242096

Table 9.70 Results of all the networks with time series for the busiest times near Gouda 40 minutes ahead

	SSE	MSE	VAR	StDev
5-1-4	97.947079	0.0458339	0.0430839	0.2075667
5-1-4	84.615499	0.0418061	0.0417659	0.2043672
5-2-4	106.41207	0.0497951	0.043283	0.2080456
5-4-4	125.53434	0.0587433	0.0563063	0.2372895
5-5-4	122.48596	0.0573168	0.0559645	0.2365681
5-8-4	115.66509	0.054125	0.0535409	0.2313891
5-9-4	117.24065	0.0548623	0.0535254	0.2313555
5-9-4	73.524657	0.0363264	0.0357396	0.1890492
5-10-4	126.40139	0.059149	0.0566879	0.2380921
5-11-4	117.7906	0.0551196	0.0538942	0.2321512
5-15-4	116.84075	0.0546751	0.0543104	0.233046

9.4 Parameter tuning

From each model a network with the best results is given in the two previous sections. Now it has to be determined which model gives the best results. Therefore the MSE and the variance of the two networks are given in Table 9.71. In this table it can be seen that with the second model a better prediction is made even 40 minutes ahead. In this section the parameters of the backpropagation algorithm will be adjusted for the best network of both models to improve the prediction if possible. Because the second model produced a better prediction the best network is the 5-9-4-network.

Table 9.71 Results of the two models compared

	MSE	VAR
2-5-1 Zoetermeer	0.014705	0.014563
5-9-4 Zoetermeer 10	0.00503	0.00492
5-9-4 Zoetermeer 20	0.00733	0.00722
5-9-4 Zoetermeer 30	0.00867	0.00865
5-9-4 Zoetermeer 40	0.00999	0.00998
2-5-1 Gouda	0.013793	0.013741
5-9-4 Gouda 10	0.00796	0.00683
5-9-4 Gouda 20	0.00785	0.00746
5-9-4 Gouda 30	0.00829	0.00805
5-9-4 Gouda 40	0.00946	0.00867
Busiest times	MSE	VAR
2-5-1 Zoetermeer	0.0692907	0.062335
5-9-4 Zoetermeer 10	0.014105	0.0136996
5-9-4 Zoetermeer 20	0.0241593	0.0241622
5-9-4 Zoetermeer 30	0.0342541	0.0340554
5-9-4 Zoetermeer 40	0.0431925	0.0422227
2-5-1 Gouda	0.0613211	0.0575234
5-9-4 Gouda 10	0.0268447	0.0243381
5-9-4 Gouda 20	0.0291638	0.0286647
5-9-4 Gouda 30	0.032457	0.0323847
5-9-4 Gouda 40	0.0363264	0.0357396

The standard parameters are a learning rate of 0.2 and a maximum error of 0.1. At first it is tried to improve the prediction with adjusting only the learning rate. The learning rate resulting in the best prediction will then be used with an adjusted error.

9.4.1 Learning rate

During the test for determination of the best learning rate, the maximum error for these network remains constant with the same value as used in the previous sections (d_{max} = 0.1).

Table 9.72 Results of the learning rates near Zoetermeer 10 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	**79.9337**	**0.00503**	**0.00492**	**0.07012**
0.4	82.87992	0.005215	0.005104	0.071439
0.6	100.4816	0.006323	0.005366	0.07325
1	130.1465	0.008189	0.008069	0.089828
10	102.425	0.006445	0.006285	0.079277

Table 9.73 Results of the learning rates near Zoetermeer 20 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	**116.506**	**0.00733**	**0.00722**	**0.08497**
0.4	119.5219	0.00752	0.00747	0.08644
0.6	125.0662	0.00787	0.00771	0.08781
1	168.1055	0.01058	0.00876	0.09361
10	173.3415	0.01091	0.01087	0.10425

Table 9.74 Results of the learning rates near Zoetermeer 30 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	**137.843**	**0.00867**	**0.00865**	**0.09301**
0.4	144.3867	0.00909	0.00904	0.09509
0.6	151.8095	0.00955	0.00916	0.09569
1	166.9897	0.01051	0.00998	0.09988
10	248.9101	0.01566	0.01517	0.12315

Table 9.75 Results of the learning rates near Zoetermeer 40 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	**158.693**	**0.00999**	**0.00998**	**0.09992**
0.4	177.7468	0.01119	0.01051	0.10251
0.6	169.4516	0.01066	0.01057	0.10281
1	182.9072	0.01151	0.01142	0.10685
10	304.7808	0.01918	0.01842	0.13572

Table 9.72 to Table 9.75 show that the results for the place near Zoetermeer are always the best with a learning rate of 0.2. If the tables containing the results for the busiest times, Table 9.76 to Table 9.79, are taken into account a learning rate of 0.2 does not always give the best results anymore. For the prediction 10 minutes ahead the error is smaller with a learning rate of 0.4. Besides that the prediction 40 minutes ahead shows that the variance is lower for the network trained with a learning rate of 0.4. However, because the error is almost always minimum for the learning rate of 0.2 it can be said that this learning rate gives the best network of all the learning rates tested.

Table 9.76 Results of the learning rates for the busiest times near Zoetermeer 10 minutes ahead

H	SSE	MSE	VAR	StDev
0.2	29.267848	0.014105	**0.0136996**	**0.1170452**
0.4	**29.093508**	**0.014021**	0.0138866	0.1178414
0.6	31.812947	0.0153315	0.0153388	0.12385
1	49.454938	0.0238337	0.0165229	0.1285414
10	44.317137	0.0213577	0.021163	0.1454752

Table 9.77 Results of the learning rates for the busiest times near Zoetermeer 20 minutes ahead

H	SSE	MSE	VAR	StDev
0.2	**49.913198**	**0.0241593**	**0.0241622**	**0.1554419**
0.4	52.278697	0.0253043	0.0251663	0.1586389
0.6	54.784410	0.0265171	0.0262847	0.1621255
1	86.508195	0.0418723	0.0296509	0.1721943
10	85.737456	0.0411993	0.0411926	0.2029596

Table 9.78 Results of the learning rates for the busiest times near Zoetermeer 30 minutes ahead

H	SSE	MSE	VAR	StDev
0.2	**70.871822**	**0.0342541**	**0.0340554**	**0.1845409**
0.4	74.086494	0.0358079	0.0344635	0.1856436
0.6	75.320821	0.0364045	0.0356027	0.1886867
1	99.566505	0.0481230	0.0397887	0.1994710
10	140.77910	0.0680421	0.068064	0.2608910

Table 9.79 Results of the learning rates for the busiest times near Zoetermeer 40 minutes ahead

H	SSE	MSE	VAR	StDev
0.2	**89.538103**	**0.0431925**	0.0422227	0.2054817
0.4	90.595861	0.0437028	**0.0416757**	**0.2041463**
0.6	95.686271	0.0461584	0.0431713	0.2077770
1	106.72832	0.0514850	0.0468204	0.2163802
10	181.17193	0.0873960	0.0867693	0.2945663

The prediction near Gouda does not a learning rate that gives clearly better network than the others. In this case sometimes the learning rate of 1 gives a better result, but it can also be the learning rate of 0.4 or even 0.2. Because the network trained with a learning rate of 0.4 gives a very high error for the prediction 40 minutes ahead, this network will not be suited to predict the average speed more than 30 minutes in advance. Therefore this learning rate will not be taken in to account for the rest of this report. By comparing the other two learning rates it can be seen that the learning rate 1 probably is best in this case.

Table 9.80 Results of the learning rates near Gouda 10 minutes ahead

H	SSE	MSE	VAR	StDev
0.2	122.333	0.00796	0.00683	0.08266
0.4	102.342	0.00666	0.0065	0.08062
0.6	146.259	0.00952	0.00668	0.08172
1	**96.3438**	**0.00627**	**0.00624**	**0.07901**
10	138.228	0.009	0.00801	0.08952

Table 9.81 Results of the learning rates near Gouda 20 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	120.665	0.00785	0.00746	0.08639
0.4	**119.899**	**0.00780**	**0.00712**	**0.08440**
0.6	128.958	0.00839	0.00777	0.08814
1	122.000	0.00794	0.00793	0.08906
10	162.990	0.01061	0.00878	0.09368

Table 9.82 Results of the learning rates near Gouda 30 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	**127.299**	**0.00829**	0.00805	0.08972
0.4	141.874	0.00923	**0.00766**	**0.08751**
0.6	135.355	0.00881	0.00782	0.08841
1	145.165	0.00945	0.00817	0.09040
10	245.241	0.01596	0.00966	0.09828

Table 9.83 Results of the learning rates near Gouda 40 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	145.298	0.00946	0.00867	0.09309
0.4	197.879	0.01288	0.00835	0.09138
0.6	140.268	0.00913	0.00894	0.09453
1	**136.614**	**0.00889**	**0.00841**	**0.09170**
10	163.802	0.01066	0.01063	0.10312

However the leaning rate can be the best learning rate that is tested in this report for the complete day, but how well does it predict the busiest times. Maybe gives the network trained with a learning rate of 0.2 better results in these times.
In Table 9.84 to Table 9.87 it can be seen that the learning rate of 1 is slightly better for a prediction 10 or 40 minutes ahead. In the other cases the network trained with a learning rate of 0.2 gives a better prediction than the one with a learning rate of 1. The error of the prediction 30 minutes ahead shows a large difference between the network with a learning rate of 0.2 and 1. With 0.2 better results are produced. What should be the best learning rate to train the network for the prediction for the place near Gouda? In the previous sections it was said that it is preferable for both problems the networks would be the same. However the correctness of the prediction will be more important than that. In this case it is possible to keep the learning rate the same. So the argument to have the same networks if possible gives a decision. The network trained with a learning rate of 0.2 gives a good prediction for both cases and will be used as the network to vary the maximum error.

Table 9.84 Results of the learning rates for the busiest times near Gouda 10 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	55.058413	0.0268447	0.0243381	0.1560066
0.4	49.055599	0.0239179	0.0237714	0.1541796
0.6	54.908201	0.0267714	**0.0236375**	**0.1537449**
1	**50.480339**	**0.0246125**	0.0243958	0.1561916
10	64.204676	0.0313041	0.0299398	0.1730312

Table 9.85 Results of the learning rates for the busiest times near Gouda 20 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	59.319112	0.0291638	0.0286647	0.1693067
0.4	**58.510473**	**0.0287662**	**0.0284064**	**0.1685420**
0.6	58.971022	0.0289926	0.0287610	0.1695908
1	60.588628	0.0297879	0.0294039	0.1714755
10	81.901262	0.0402661	0.0347425	0.1863935

Table 9.86 Results of the learning rates for the busiest times near Gouda 30 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	65.822813	0.032457	0.0323847	0.1799576
0.4	**65.290432**	**0.0321945**	**0.0310907**	**0.1763254**
0.6	65.596956	0.0323456	0.0320521	0.1790311
1	76.863051	0.0379009	0.0334002	0.1827571
10	126.47402	0.0623639	0.0380374	0.1950318

Table 9.87 Results of the learning rates for the busiest times near Gouda 40 minutes ahead

η	SSE	MSE	VAR	StDev
0.2	73.524657	0.0363264	0.0357396	0.1890492
0.4	76.309289	0.0377022	**0.0351757**	**0.1875519**
0.6	72.188796	0.0356664	0.0356283	0.1887546
1	**71.196459**	**0.0351761**	0.0351783	0.1875587
10	98.461185	0.0486468	0.0436209	0.2088563

9.4.2 Maximum error

The maximum error is the parameter, which determines whether a backpropagation should take place. If this parameter is set to 0 always backpropagation will always take place if the measured and calculated error are not equal. Therefore training will last longer. Also it can be expected that eventually the minimum error should be reached if the learning rate is large enough not to get trapped in local minima and not to large to step over the absolute error. However if Table 9.88 to Table 9.91 are taken into account it can be said that the maximum error of 0.1 gives the best result. At least for the prediction 30 and 40 minutes ahead and it was to goal of the research to predict the average speed more than 25 minutes in the future.

Table 9.88 Results of the maximum errors near Zoetermeer 10 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	79.9337	0.00503	0.00492	0.07012
0.01	**76.5701**	**0.00482**	**0.00481**	**0.06936**
0.001	80.6432	0.00507	0.00491	0.07003
0	78.2346	0.00492	0.00489	0.06993

Table 9.89 Results of the maximum errors near Zoetermeer 20 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	116.506	0.00733	**0.00722**	**0.08497**
0.01	**115.400**	**0.00726**	0.00723	0.08502
0.001	116.872	0.00735	0.00733	0.08559
0	116.363	0.00732	0.00732	0.08556

Table 9.90 Results of the maximum errors near Zoetermeer 30 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**137.843**	**0.00867**	**0.00865**	**0.09301**
0.01	140.639	0.00885	0.00885	0.09406
0.001	140.592	0.00885	0.00879	0.09375
0	139.074	0.00875	0.00872	0.09340

Table 9.91 Results of the maximum errors near Zoetermeer 40 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**158.693**	**0.00999**	**0.00998**	**0.09992**
0.01	161.740	0.01018	0.01016	0.10081
0.001	160.584	0.01010	0.01009	0.10047
0	161.472	0.01016	0.01002	0.10010

For the busiest times Table 9.92 to Table 9.95 show that the prediction with a maximum error of 0.1 gives overall the best results. However it can be said that the result for the prediction over the whole day and for the busiest times are very close to each other, so maybe if the networks where trained several times with different initial values the results will be different.

Table 9.92 Results of the maximum errors for the busiest times near Zoetermeer 10 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**29.267848**	**0.014105**	**0.0136996**	**0.1170452**
0.01	29.398881	0.0141681	0.0141534	0.1189682
0.001	31.660498	0.0152581	0.0143852	0.1199385
0	29.550281	0.0142411	0.0142278	0.1192805

Table 9.93 Results of the maximum errors for the busiest times near Zoetermeer 20 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**49.913198**	**0.0241593**	**0.0241622**	**0.1554419**
0.01	53.622216	0.0259546	0.0258348	0.1607319
0.001	55.696675	0.0269587	0.0259938	0.1612260
0	54.576305	0.0264164	0.0258578	0.1608036

Table 9.94 Results of the maximum errors for the busiest times near Zoetermeer 30 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**70.871822**	**0.0342541**	**0.0340554**	**0.1845409**
0.01	77.381076	0.0374002	0.0372670	0.1930465
0.001	77.583936	0.0374983	0.0359001	0.1894733
0	75.025273	0.0362616	0.0357975	0.1892024

Table 9.95 Results of the maximum errors for the busiest times near Zoetermeer 40 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	**89.538103**	**0.0431925**	**0.0422227**	**0.2054817**
0.01	94.426314	0.0455506	0.0447927	0.2116429
0.001	92.903060	0.0448158	0.0432831	0.2080460
0	92.809639	0.0447707	0.0427912	0.2068603

The prediction near Gouda does give the prediction as expected. From almost all results as displayed in Table 9.96 to Table 9.99 the prediction with a maximum error of 0 gives the best results. The error of the prediction with a maximum error of 0.1 is even very much higher than the prediction with a maximum error of 0.
If the results of Table 9.100 to Table 9.103 are used the opinion cannot change the prediction with a maximum error 0 is still the best.

Table 9.96 Results of the maximum errors near Gouda 10 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	122.333	0.00796	0.00683	0.08266
0.01	**94.4032**	**0.00614**	**0.00613**	**0.07829**
0.001	94.9728	0.00618	0.00618	0.0786
0	97.318	0.00633	0.00615	0.07842

Table 9.97 Results of the maximum errors near Gouda 20 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	120.665	0.00785	0.00746	0.08639
0.01	108.969	0.00709	0.00708	0.08414
0.001	110.457	0.00719	0.00713	0.08447
0	**107.014**	**0.00697**	**0.00693**	**0.08324**

Table 9.98 Results of the maximum errors near Gouda 30 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	127.299	0.00829	0.00805	0.08972
0.01	118.434	0.00771	0.00769	0.08769
0.001	119.112	0.00775	0.00771	0.08780
0	**116.876**	**0.00761**	**0.00754**	**0.08683**

Table 9.99 Results of the maximum errors near Gouda 40 minutes ahead

d_{max}	SSE	MSE	VAR	StDev
0.1	145.298	0.00946	0.00867	0.09309
0.01	132.754	0.00864	0.00849	0.09213
0.001	130.327	0.00848	0.00847	0.09201
0	**125.633**	**0.00818**	**0.00817**	**0.09039**

Table 9.100 Results of the maximum errors for the busiest times near Gouda 10 minutes ahead

η	SSE	MSE	VAR	StDev
0.1	55.058413	0.0268447	0.0243381	0.1560066
0.01	49.880469	0.0243201	0.0243249	0.1559643
0.001	50.102813	0.0244285	0.0244398	0.1563323
0	**49.474535**	**0.0241222**	**0.0233609**	**0.1528427**

Table 9.101 Results of the maximum errors for the busiest times near Gouda 20 minutes ahead

η	SSE	MSE	VAR	StDev
0.1	59.319112	0.0291638	0.0286647	0.1693067
0.01	59.214571	0.0291124	0.0291196	0.1706446
0.001	59.676243	0.0293394	0.0292388	0.1709935
0	**57.279238**	**0.0281609**	**0.0278368**	**0.1668436**

Table 9.102 Results of the maximum errors for the busiest times near Gouda 30 minutes ahead

η	SSE	MSE	VAR	StDev
0.1	65.822813	0.032457	0.0323847	0.1799576
0.01	65.317170	0.0322077	0.0321545	0.1793168
0.001	65.260778	0.0321799	0.0318328	0.1784176
0	**63.319125**	**0.0312224**	**0.0308529**	**0.1756498**

Table 9.103 Results of the maximum errors for the busiest times near Gouda 40 minutes ahead

η	SSE	MSE	VAR	StDev
0.1	73.524657	0.0363264	0.0357396	0.1890492
0.01	73.927566	0.0365255	0.0363029	0.1905331
0.001	73.813491	0.0364691	0.0364469	0.1909107
0	**70.391027**	**0.0347782**	**0.0347505**	**0.1864148**

9.4.3 Conclusions

Above the best network is trained with several different parameters. The results of these test show that the best network trained with a learning rate of 0.2 and a maximum error of 0.1 gives the best results for the prediction of the average speed near Zoetermeer. However this time it cannot be said anymore that the network will be the same for each network. The difference is only the maximum error. For the prediction for the place near Gouda this parameter should be set to 0 to get the best results.

9.5 Results and conclusions

In Section 5.4 it was stated that a neural network with one hidden layer should be capable to approximate all functions. During the years a few formulas are introduced. These formulas are also given In Section 5.4. Each of them gives the number of neurons in the hidden layer. However it cannot be said that one of the formulas gives the best results for all problems, so all the topologies resulting from these formulas where tested in this chapter.

In the previous sections of this chapter it is determined what the architecture of the network and what the parameter of the backpropagation algorithm are to get the best prediction of the average speed for both places. At least it is determined which network gives the best prediction of the networks that are tested. To determine which parameters are the best just a few combinations are tried. So there is a small possibility that the results can be better using another architecture or other parameters.

For the model without time series the formulas of Hecht-Nielsen ($2N_i+1_o$) and Fletcher and Goss (between $2N_i+1$ and $2\sqrt{N_i}+N_o$) resulted in the number of hidden neurons capable of predicting the problem. The formula of Fletcher and Goss did also determine the number of neurons for the best network for the model with time series. From these results it can be concluded that the number of hidden neurons should be lying somewhere between $2N_i+1$ and $2\sqrt{N_i}+N_o$ for predicting the average speed. However the formula only set boundaries for the number of neurons. So still a small number of different topologies have to be tested.

Calculating the number of hidden neurons for the model without time series using the formula of Fletcher and Goss gives the range 4 to 5 hidden neurons. For the model with timeseries this resulted in 8 to 11 hidden neurons.
For Zoetermeer the best prediction where made with the 5-9-4-network based on the model with time series. The parameters of the backpropagation algorithm are a learning rate of 0.2 and a maximum error of 0.1. The results of this network for the whole day are shown in Table 9.104.

Table 9.104 Results of the best network for Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
SSE	79.9337	116.506	137.843	158.693
MSE	0.00503	0.00733	0.00867	0.00999
VAR	0.00492	0.00722	0.00865	0.00998
STDev	0.07012	0.08497	0.09301	0.09992

The best network for the place near Gouda produced the results of the whole day as displayed in Table 9.105. This network is also the 5-9-4-network with a learning rate of 0.2, but with a maximum error of 0. As can be seen the topology of the network is the same as for the prediction near Zoetermeer.

Table 9.105 Results of the best network for Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
SSE	97.318	107.014	116.876	125.633
MSE	0.00633	0.00697	0.00761	0.00818
VAR	0.00615	0.00693	0.00754	0.00817
STDev	0.07842	0.08324	0.08683	0.09039

However this result says nothing over the difference relatively to the speed (the difference expressed in percentages). These results are stated in Table 9.106 and Table 9.107. The first table shows that the average difference over the complete dataset is more than 20% for the prediction 40 minutes ahead. Analyzing these results shows that the maximum difference is for 40 minutes ahead more than 2000%. The measured average speed of that item is very low (between 0 and 10). This is an exception and therefore the calculated average speed is a lot higher resulting in a very high difference. These items do have a major influence of the average difference over the complete data set.
Now he complete dataset has 15892 items. From this dataset minimal 12107 items (76%) are predicted with a difference of equally or less than 10%. Between the items with a difference of more then 10% there are minimal 748 items with a difference of 100%. This is 4.7% of the complete dataset. If the items of this 4.7% are considered as exceptions and are not taken into account the average speed is predicted with an average difference of maximal 8.6%.

Table 9.106 Results in percentages of the best network for Zoetermeer

	Zoetermeer 10 minutes ahead	Zoetermeer 20 minutes ahead	Zoetermeer 30 minutes ahead	Zoetermeer 40 minutes ahead
Average difference	14.1342	16.46217	19.51092	22.7088
Maximum difference	1461.92	2034.432	2090.272	2112.608
Number of items of 10% or less	12416	12402	12437	12107
Number of items over 10%	3416	3490	3455	3785
Number of items over 100%	439	480	622	748
Average difference skipping the predictions with difference over 100%	7.898429	8.572624	8.334925	8.184104

The second table, Table 9.107, gives the same kind of results as stated above. The average difference on the complete dataset is a little bit more than 18% for the prediction 40 minutes ahead. The largest difference is a difference of 2866,64%. Again it can be said that if the percentage that produces differences over 100% are seen as exceptions the average difference will maximal be 7.17%. In the complete dataset of 15364 item there are maximal 283 with a value over the 100%. This is less than 2% of the dataset.
From this set minimal 12203 items are predicted with a difference of 10% or less. This is almost 80% of the predictions.

Table 9.107 Results in percentages of the best network for Gouda

	Gouda 10 minutes ahead	Gouda 20 minutes ahead	Gouda 30 minutes ahead	Gouda 40 minutes ahead
Average difference	14.95683	15.87649	17.05146	18.05702
Maximum difference	2007.6	2377.76	2509.44	2866.64
Number of items of 10% or less	12351	12318	12203	12207
Number of items over 10%	3013	3046	3161	3157
Number of items over 100%	238	261	269	283
Average difference skipping the predictions with difference over 100%	7.070497	6.870318	7.172643	7.136715

In Figure 9.65 and Figure 9.66 the graphs of the predicted average speed for respectively Zoetermeer and Gouda can be seen. The prediction for Gouda seems to adjust to the average speed over all Tuesdays. This can be explained by the large fluctuations that are on exactly the same place as the minimum of the average speed over all Tuesdays. The further ahead the prediction is made the prediction will go in the direction of the average speed over all Tuesdays, but will also try to follow the measured data itself. Giving the average speed a certain time before as input causes this last behavior. Because the fluctuations are very large the prediction lies closer to the average speed over all Tuesdays than normally like in Figure 9.65.

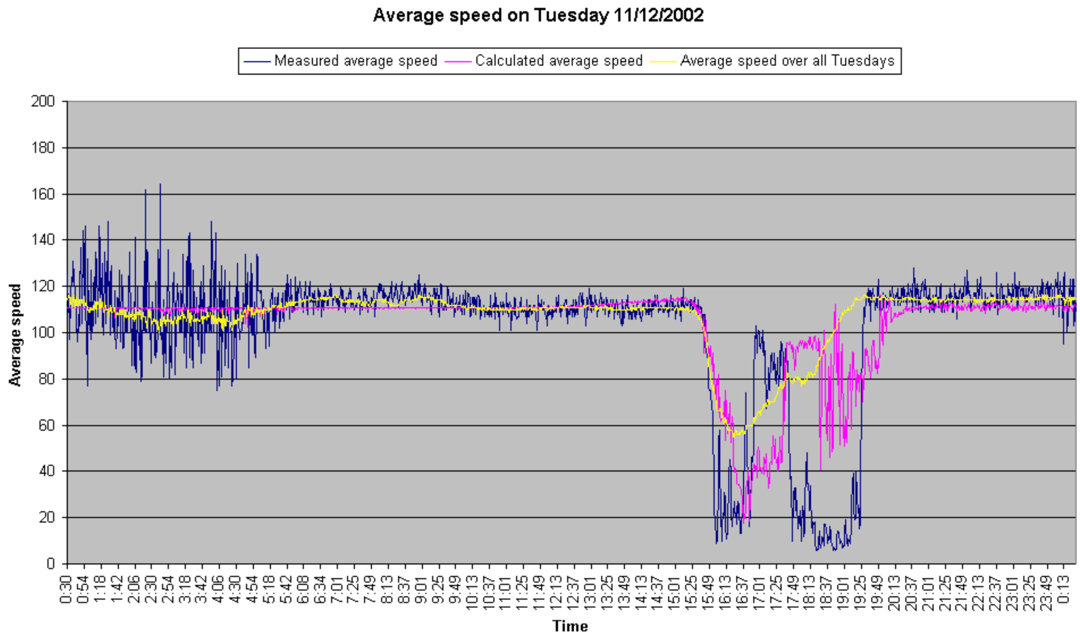


Figure 9.65 The average speed on 12th of November 2002 near Zoetermeer 40 minutes ahead with the 5-9-4-topology, η = 0.2 and d_{max} = 0.1

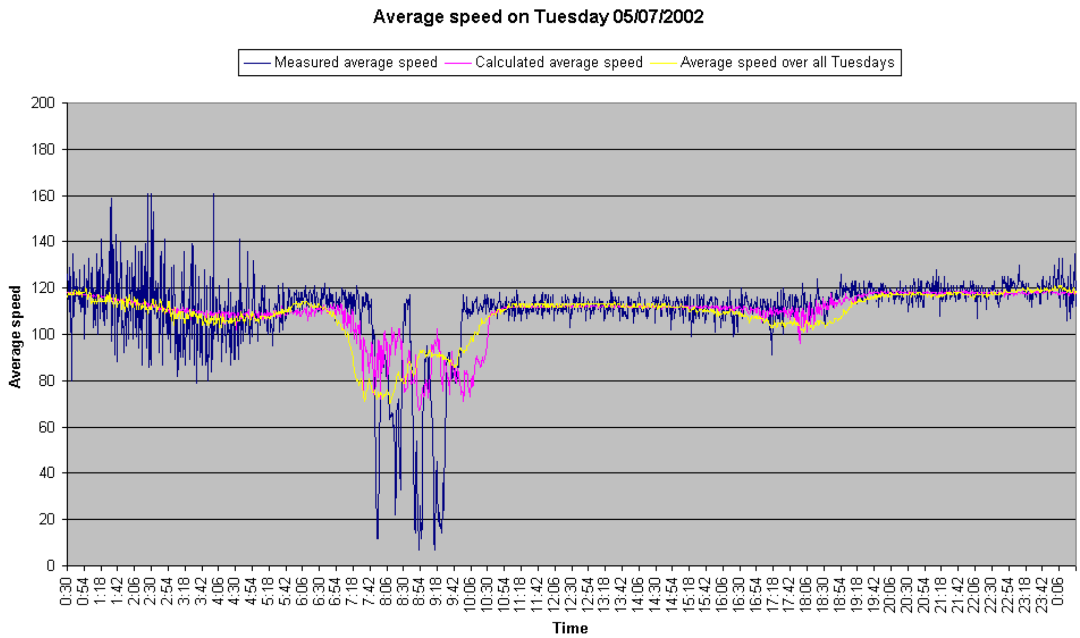


Figure 9.66 The average speed on 7th of May 2002 near Gouda 40 minutes ahead with the 5-9-4-topology, η = 0.2 and d_{max} = 0

During the training of the neural networks the error did not converge to the minimum resulting in figures like Figure 9.17. The error first decreases and the rest of the training it will stay the same. So the error does not converge to 0. It was stated that it takes a lot of training cycles to reach the minimum 0 or it will not go to 0 at all. The last option is the most probable explanation. This can be explained by the values in the training set. There is variation in speed measurements on the same place, time and month. So each input pair can have different output values. For example at 7.00 AM in January on a certain place the speed can be 70, 110 or 115 km/h. In Figure 9.67 it can be seen that there is a normal distribution of values on a certain time. Usually it is assumed that only one output can belong to an input combination instead of a distribution of values. This output can contain noise. Using input and output combinations leads to neural network approximating the function indicated by these combinations. The network is most generalizing after the number of training cycles when the error on the validation set is at its lowest value. At this point the neural network has removed most of the noise that the data contains.
In this case however there is not one output belonging to an input combination, but there can be several. The weights in the network will be adjusted to each of the value every training cycle again. The predicted value will therefore always balance between the possible solutions, not able to reach an error of approximately 0 on the training set, but only an average minimum error not equal to 0.

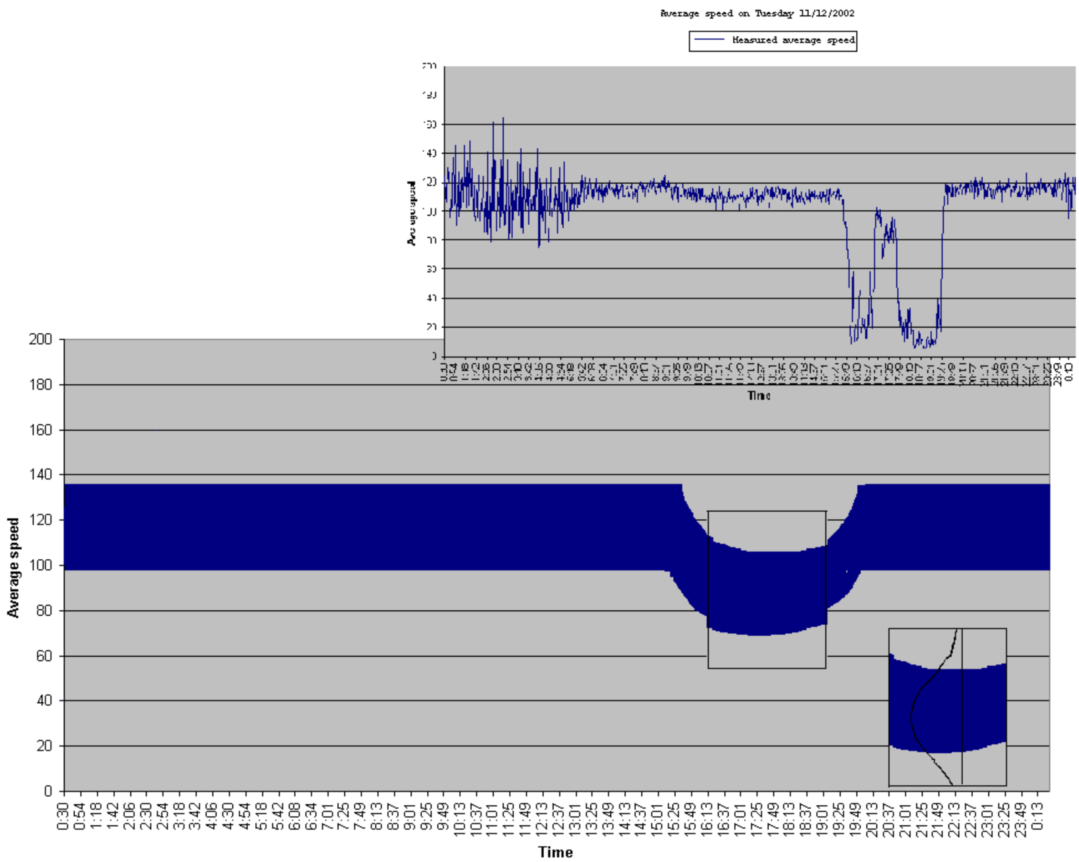


Figure 9.67   Average speed and its variation on a certain time

Chapter 10

Conclusions

In Chapter 1 the goals for this research were stated. The research was divided into three parts:

- Research on traffic
- Research on Artificial Neural Networks
- Research on travel time prediction

These parts and the results of our research are discussed in this chapter.

10.1 Research on traffic

The literature survey
The literature of Chapter 2 describes systems that are able of predicting travel times using measured data of one minute before. These systems are however not capable of predicting the travel times in the future. But if the measured data can be predicted and this predicted data is used to predict the travel times with this system, it is possible to predict the real travel times in the future.

Traffic data
The traffic data can be collected trough the website of [Regiolab Delft]. The information displayed on this website is only accessible for authorized users. This data is collected by measurement instruments along the highways. These instruments count the number of cars and register the average speed of these cars with a time interval of one minute. This data was successfully used to train the Artificial Neural Network.

10.2 Research on Artificial Neural Networks

In the literature about traffic neural networks are used to predict travel times. The reason for the use of neural networks is that the interaction between the influencing factors is very complex and therefore difficult to define. Therefore research is done to determine the architecture of a neural network.

The architecture of a neural network
In literature several ideas can be found about the architecture of a neural network. At first it had to be determined what sort of neural network can be used to solve traffic problems. Because it is known that a feed-forward neural network is a robust network this network is used in this research. Another reason to choose a feed-forward neural network is because only a proof of concept is needed.
A feed forward neural network consists of an input layer, an output layer and one or more hidden layers. Each of these layers contains one or more neurons. The number of neurons in the input and output layer is determined by the number of input and output parameters. It is, however, not known what the number of hidden layers should be and how many neurons they should contain. A lot of research is done to determine these numbers. It is proved that a feed forward neural network with one hidden layer is capable of approximating every function. Also a number of formulas is presented to calculate the number of hidden neurons. However none of these formulas is proved to be best for all problems. Therefore the networks resulting from these formulas are tested to determine the formula that is best for thus problem.

Experiments
After testing all topologies resulting from the formulas mentioned above the conclusion could be made that a feed forward neural network with one hidden layer, from which the number of hidden neurons is determined by the formula of Fletcher and Goss (between $2N_i+1$ and $2\sqrt{N_i} +N_o$), gave the best results while predicting the average speed. However this formula only gives a boundary between which the number of hidden neurons should lie. So still a few topologies should be tried.

10.3 Research on travel time prediction

In Chapter 1 the use of predicted travel times is stated. Route planners and navigation systems can use these travel times to guide traveler along the real fastest route. Another advantage of knowing the real travel times on parts of the road is that the complete real travel time of a certain route can be calculated and can be used by the traveler to determine the departure time.

Research on the influences on the speed
During the research on the influences on the speed it has to be determined which factors influence travel times. This study has determined that the average speed on a road is influenced by:

- Time
- Day of the week
- Month
- Weather
- Holiday
- Events

Developing a model
Using the influences on the road a model has to be developed capable of predicting the average speed on a given place x and time t. Basically the function F(x, t) on every link is considered. This function can be approximated by the average values. But a computing device is wanted. An analytic representation of F is one way to do it. Also an implicit representation of F by Artificial Neural Networks can be used. This is one of the goals of our research. The first model developed is capable of reaching this goal. This model can be found in Section 8.3.1.
To be able to use this average speed for predicting the travel times into the future the following goal should be reached:

Study the possibilities of predicting the average speed on a certain place over 25 minutes in the future with an acceptable accuracy. An acceptable accuracy of the predicted speed is defined as a difference of at most 10 percent of the real speed.

There are two models designed for the prediction of the average speed into the future. These models can be found in Section 8.3.2 and 8.3.3.

Experiments
To test the models the different network architectures are used described by the formulas mentioned in Section 10.2. Only the first two models are tested to make sure whether or not the goal is reached.
The results of these test shows that it is possible for at least 76% of the predictions to make a prediction with a difference smaller or equal to 10% even if the prediction is made 40 minutes ahead.
In Table 10.1 and Table 10.2 these results are stated again.

Table 10.1 Results in percentages of the best network for Zoetermeer

	Zoetermeer 10 minutes ahead	**Zoetermeer 20 minutes ahead**	**Zoetermeer 30 minutes ahead**	**Zoetermeer 40 minutes ahead**
Average difference	14.1342	16.46217	19.51092	22.7088
Maximum difference	1461.92	2034.432	2090.272	2112.608
Number of items of 10% or less	12416	12402	12437	12107
Number of items over 10%	3416	3490	3455	3785
Number of items over 100%	439	480	622	748
Average difference skipping the predictions with difference over 100%	7.898429	8.572624	8.334925	8.184104

Unfortunately it can be seen that for the complete dataset the average difference is larger than 10%. It can even be more than 20%. This is caused by a very small part of the dataset with differences of more than 100%. These large differences are the results of very low values that are measured. If the difference is 10 kilometers while the measured speed is 1 kilometer the difference is already 1000%. However it is not very likely that traffic will move between 1 and 10 kilometer per hour when no accident or when there is no maintenance on or beside the road. So if these items, less 5% of the dataset, are not taken into account the goal is reached. The average speed has then an average difference of less than 9%.

Table 10.2 Results in percentages of the best network for Gouda

	Gouda 10 minutes ahead	**Gouda 20 minutes ahead**	**Gouda 30 minutes ahead**	**Gouda 40 minutes ahead**
Average difference	14.95683	15.87649	17.05146	18.05702
Maximum difference	2007.6	2377.76	2509.44	2866.64
Number of items of 10% or less	12351	12318	12203	12207
Number of items over 10%	3013	3046	3161	3157
Number of items over 100%	238	261	269	283
Average difference skipping the predictions with difference over 100%	7.070497	6.870318	7.172643	7.136715

If these neural networks will be used in real systems they have to be retrained over and over again. The situation on the road itself will also change over the years. A few reasons can be found for these changes:

- The number of cars that will use the road has increased over the years. It is very likely that the number of cars using the road will not be the same over the years. Most probable is that it will increase over the following years. But if the public transport improves maybe this number will decrease again. At least fluctuation can be expected in the number of cars over the years.
- Also new roads will be put down to remove a bottleneck. These roads lead to new possible routes. For these roads a new network has to be developed over the years. But because a part of the traffic will use these new roads, other roads will have to deal with less traffic.
- Besides new roads also an extra lane can be added to a road leading less congestion on that road.
- The last point is that when a large company moves from one place to another, a large company goes away forever or a large company will settle somewhere, the traffic flow of employees will change and because of this traffic jams will move, disappear or be added.

Chapter 11

Recommendations

To improve the prediction of the average speed a few suggestions can be made.

The model

- Most of the results of tested feed-forward neural networks did not differ very much. Only the networks where the number of hidden neurons was not sufficient the error was much higher. To determine the real best network for this problem, the networks must be trained several times with different initial values.
- The parameters of the backpropagation algorithm can be suboptimal. Therefore more combinations of parameters should be investigated.
- The month is used as a parameter to determine a part of a year. It can be useful to divide each month in three parts, the beginning, the middle and the end.
- In Chapter 8 the models are described. The first two models are tested. The third model should be tested as well.
- The models of Chapter 8 are tested using a feed-forward neural network with the backpropagation algorithm. It should be tried to use a different type of networks like a recurrent neural network, using for example the 'back-propagation through time' algorithm.

Accidents and Incidents

- There are a few problems with the models described in this report. These models are not capable to deal with accidents and incidents, like roadblocks. The system can be expanded by an expert system to adjust the predicted speed if accidents or incidents are known. However at first an analysis should be made about the influence of the accidents and incidents on the average speed to determine the impact of the adjustment.

Travel time

- This thesis provides a necessary base to compute the travel time into the future on a link. The next step is to research the possibilities of combining our system described in this thesis and the systems described in TRAIL literature predicting travel times using the measurements of along the road of one minute before in order to predict the travel time into the future.

Collecting data

- The approach in this report is based on the measurements of speed on a well-defined time and place. But it is not necessary to be dependent of the data measured along the road. At this moment there is a project running by CMG. This project is aimed at collecting information about cars along the road using the GSM system. An extension of our system can be to use the GSM-data so it is not necessary to install and maintain expensive measurement instruments along the roads.

Chapter 12

References

[Abdel-Aty 1997] Abdel-Aty, M.A., Kitamura, R., Jovanis, P.P., Using stated preference data for studying the effect of advanced traffic information on drivers' route choice, Transportation Research Part C: Emerging Technologies v 5C n 1 Feb. 1997, p 39-50

[Amsterdam Arena] http://www.amsterdamarena.nl

[ANN1] http://www.statsoftinc.com/textbook/stneunet.html

[ANN2] http://hem.hj.se/~de96klda/NeuralNetworks.htm

[ANN3] http://www.pmsi.fr/neurin2a.htm

[Concert] http://www.concert.nl

[Eggenkamp 2001] Eggenkamp, G., Dynamic multi modal route planning: an artificial intelligence approach, Master Thesis, Delft University of Technology, 2001

[Hansen 2001] Hansen, I.A., Determination and evaluation of traffic congestion costs, EJTIR, 1, no.1 (2001), pp. 61-72

[Hecht-Nielsen 1990] Hecht-Nielsen, R., Neurocomputing, Addison-Wesley Publishing Company Inc., 1990, pp. 122-124

[Heineken Music Hall] http://www.heinekenmusichall.nl

[Hoogendoorn 2001] Hoogendoorn, S.P., Lint, J.W.C. van, 'Short-term Prediction of Traffic Flow Conditions in a Multilane Multiclass Network', Full paper to be presented at the TRISTAN IV conference, june 2001, Ponta del Gado (The Azores, Portugal)

[KNMI] http://www.knmi.nl

[Park 1999] Park, D., Rilett, L.R., Han, G., Spectral Basis Neural Networks For Real-Time Travel Time Forecasting, Journal of Transportation Engineering, 1999/11. 125(6), p. 515-523

[Regiolab Delft] http://www.regiolab-delft.nl

[Ticketservice] http://www.ticketservice.nl

[Ubbels 2001] Ubbels, B., Robdenburg, C., Nijkamp, P., Different Perspectives on the Global Development of Transport, EJTIR, 1, no. 1 (2001), pp. 9 – 28

[Vakantiekalender] http://www.vakantiekalender.nl

[Vancoillie 2003] Vancoillie, F.,Design and application of artificial neural networks for digital image classification of tropical savannah vegetation, PhD thesis, Ghent University, 2003

[Van Lint 2000] Lint, J.W.C. van, Hoogendoorn, S.P., Zuylen, H.J. van, Robust and adaptive Travel Time prediction with Neural Networks, proceedings of the 6th annual TRAIL Congress (part 2), December 12th 2000

[Van Lint 2002 BARI] Lint, J.W.C. van, Hoogendoorn, S.P., Zuylen, H.J. van, Robust Freeway Travel time Prediction with State-Space Neural Networks, Proceedings of the 13th Mini-EURO Conference "Handling Uncertainty in the Analysis of Traffic and Transportation Systems" June 10 to 13, 2002, BARI, Italy

[Van Lint 2002 Madrid] Lint, J.W.C. van, Hoogendoorn, S.P., Zuylen, H.J. van, State Space Neural Networks for Freeway Travel Time Prediction, Proceedings of the INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, ICANN 2002, August 27-30, 2002, Madrid, Spain

[Van Lint 2002 TRB] Lint, J.W.C. van, Hoogendoorn, S.P., Zuylen, H.J. van, Freeway Travel time Prediction with State-Space Neural Networks, Transportation Planning and Traffic Engineering Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Accepted for presentation & publication at TRB 2002

[Van Zuylen] Zuylen, H.J. van, Dynamisch verkeersmanagement, Pao Cursus Model- en Evaluatiestudies verkeersbeheersingsmaatregelen, Delft University of Technology, February 2000

[Weeronline] http://www.weeronline.nl

Appendix A

Haze and smoke are not very influencing for the situation on the road and therefore they will not make a difference while categorizing the weather.
The scaling is from 0 to 10.

0 = -
0 = -, rook
0 = -, nevel
0 = Mooi
0 = Onbewolkt
0 = Onbewolkt, rook
0 = Onbewolkt, nevel
0 = Weinig / geen bewolking
0 = Vrijwel onbewolkt
0 = Vrijwel onbewolkt, rook
0 = Vrijwel onbewolkt, nevel
1 = Licht bewolkt
1 = Licht bewolkt, rook
1 = Licht bewolkt, nevel
2 = Half bewolkt
2 = Half bewolkt, rook
2 = Half bewolkt, nevel
3 = Bewolkt
3 = Bewolkt, rook
3 = Bewolkt, nevel
4 = Geheel bewolkt
4 = Geheel bewolkt, rook
4 = Geheel bewolkt, nevel
4 = Droog
5 = Motregen
5 = Lichte buien
5 = Lichte regen
6 = Bui
6 = Regen
7 = Onweer
7 = Bui met korrelhagel
7 = Vriespunt motregen
8 = Mistbanken
8 = Lichte sneeuw
8 = Lichte sneeuwbui
9 = Mist
9 = Korrelsneeuw
9 = Sneeuwval
10 = Krachtige sneeuw
10 = IJsregen

Appendix B

B.1 Monday

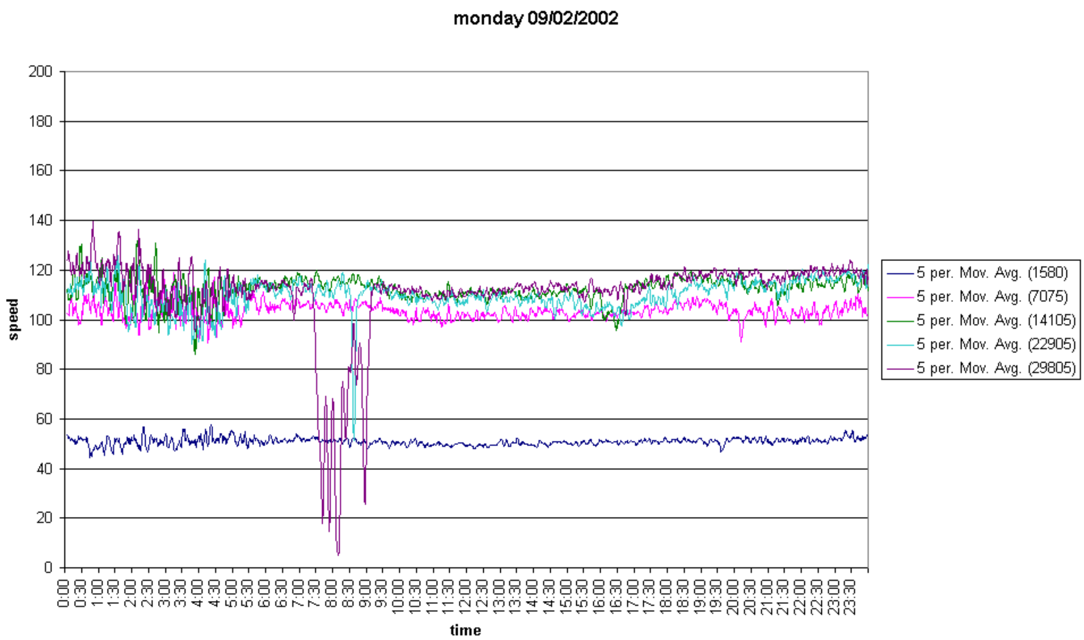


Figure B.1 Traffic on Monday 2nd of September 2002

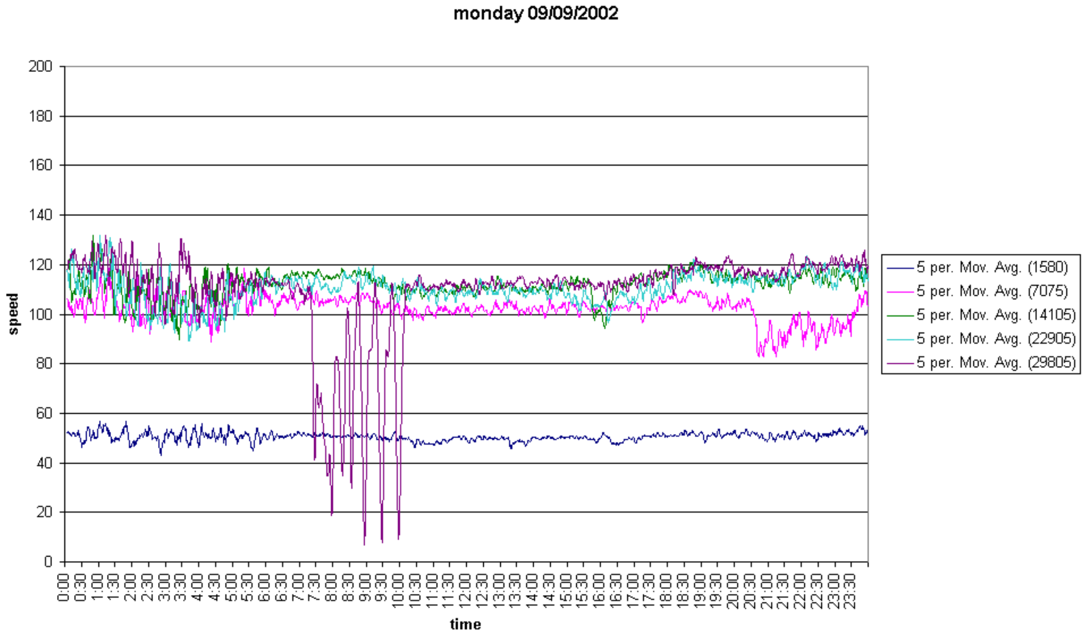


Figure B.2 Traffic on Monday 9th of September 2002

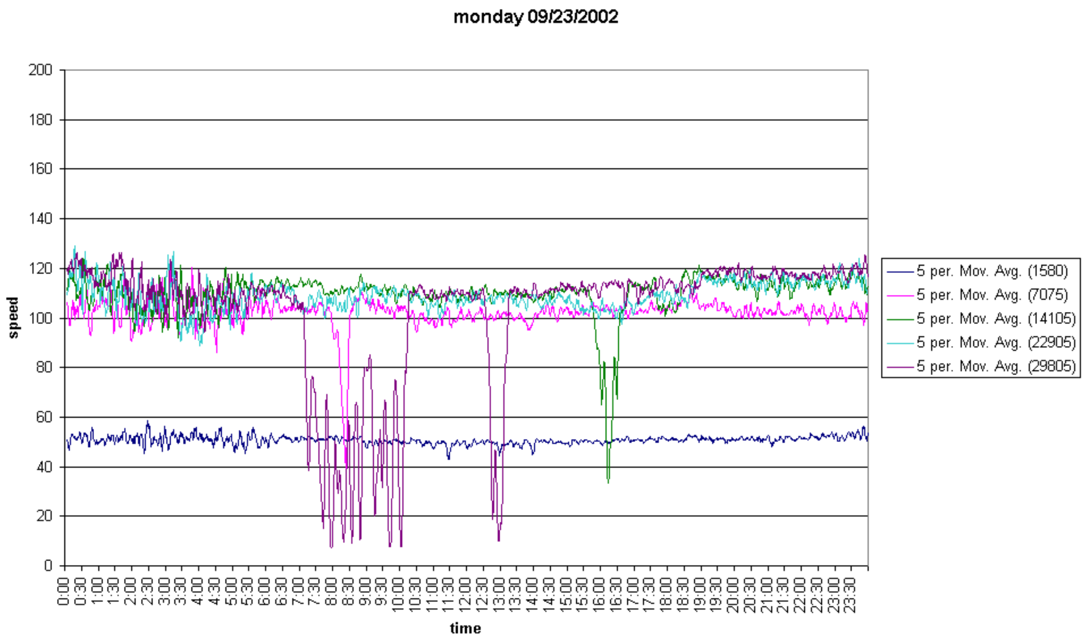


Figure B.3 Traffic on Monday 23rd of September 2002

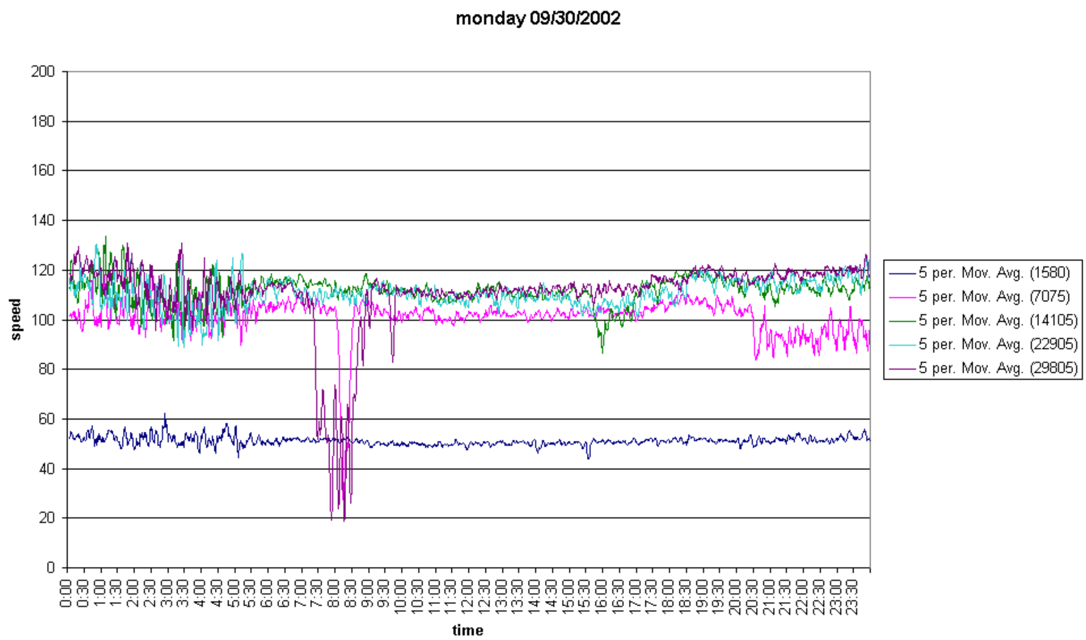


Figure B.4 Traffic on Monday 30th of September 2002

B.2 Tuesday

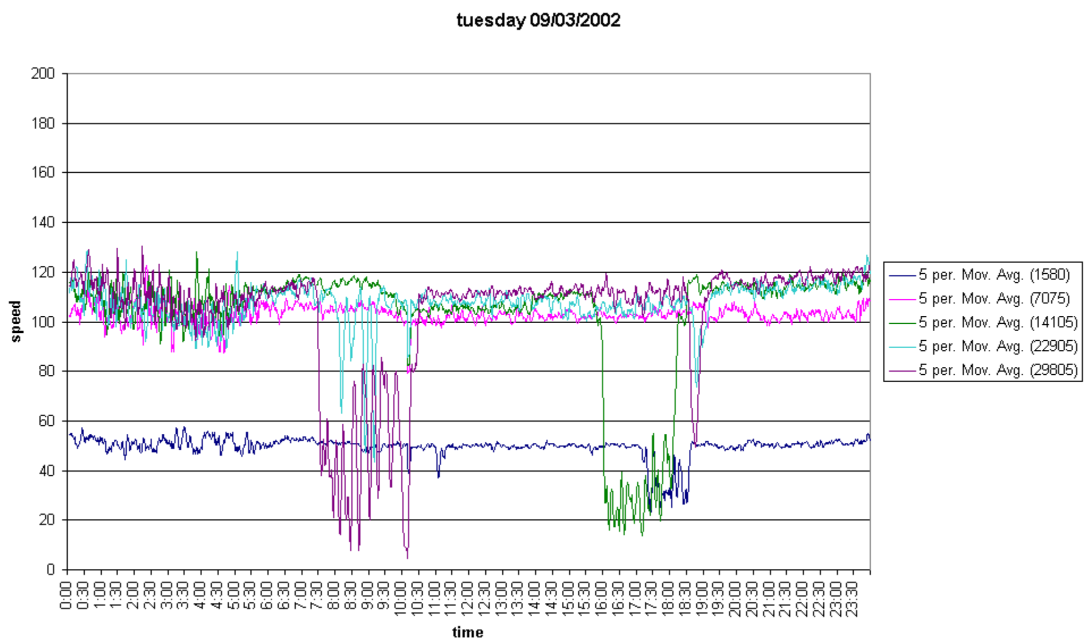


Figure B.5 Traffic on Tuesday 3rd of September 2002

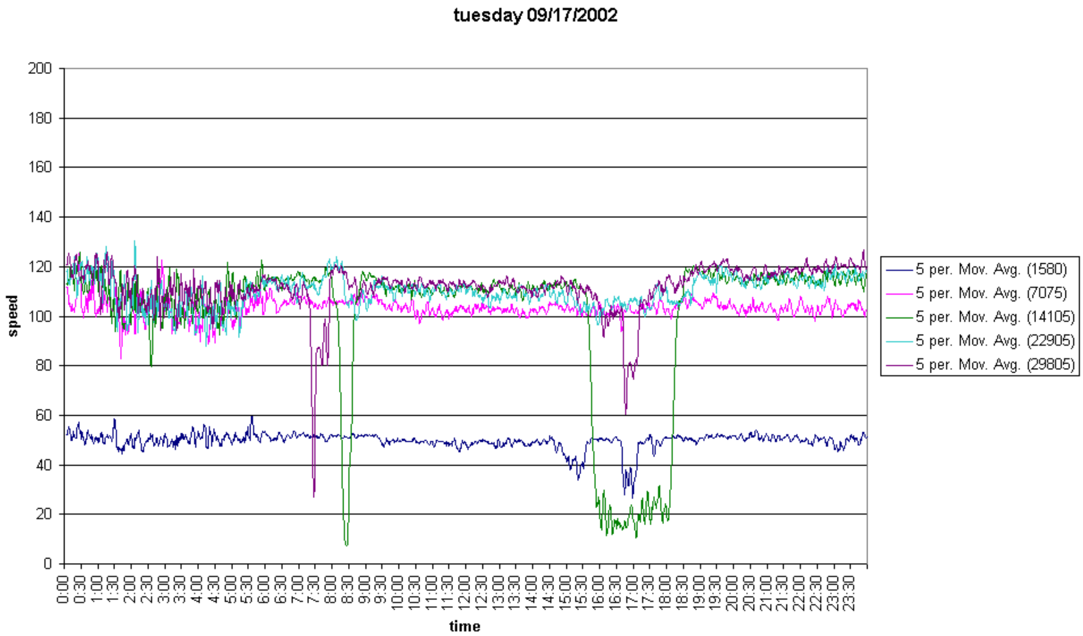


Figure B.6 Traffic on Tuesday 17th of September 2002

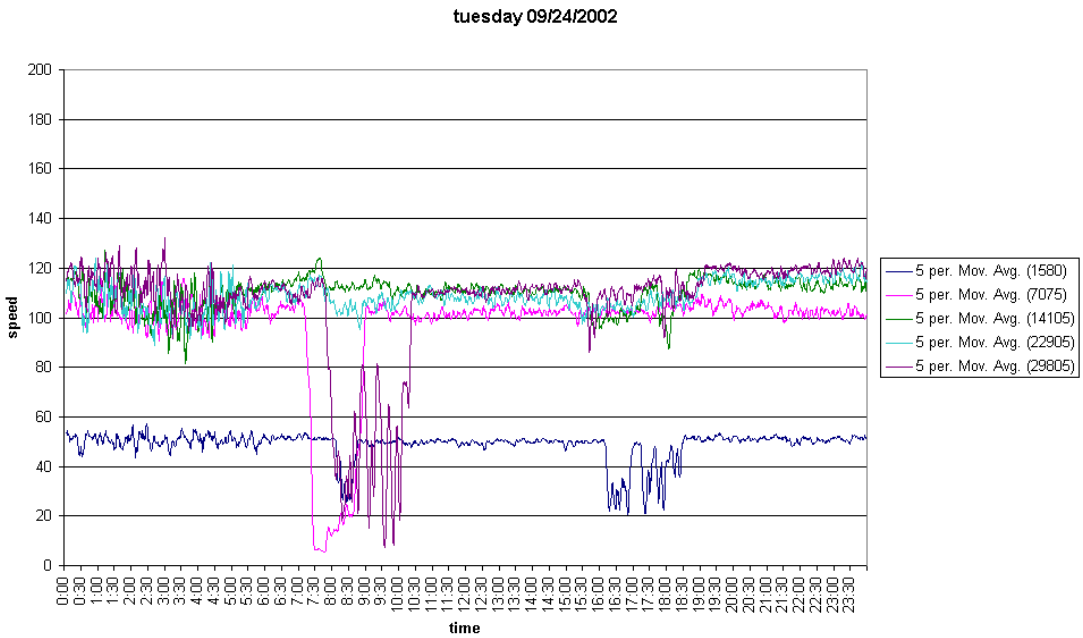


Figure B.7 Traffic on Tuesday 24th of September 2002

B.3 Wednesday

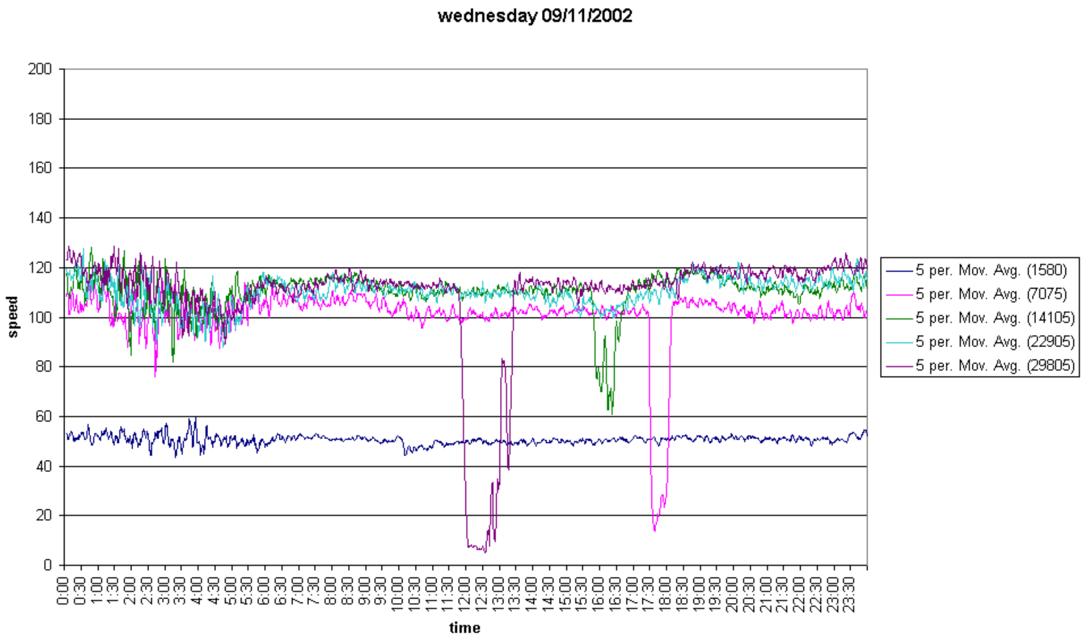


Figure B.8 Traffic on Wednesday 11th of September 2002

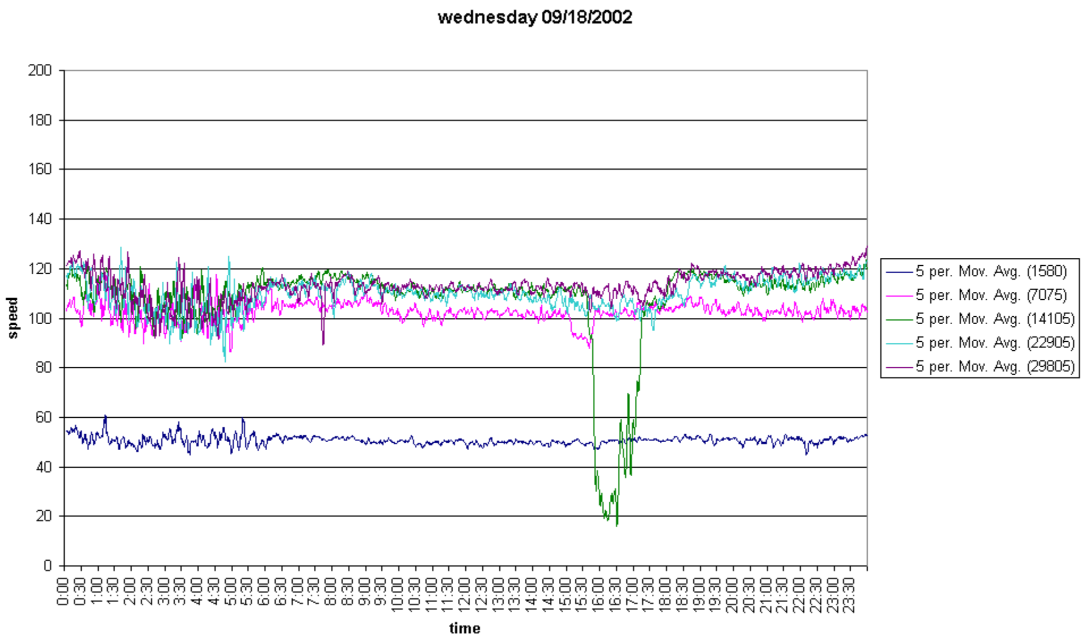


Figure B.9 Traffic on Wednesday 18th of September 2002

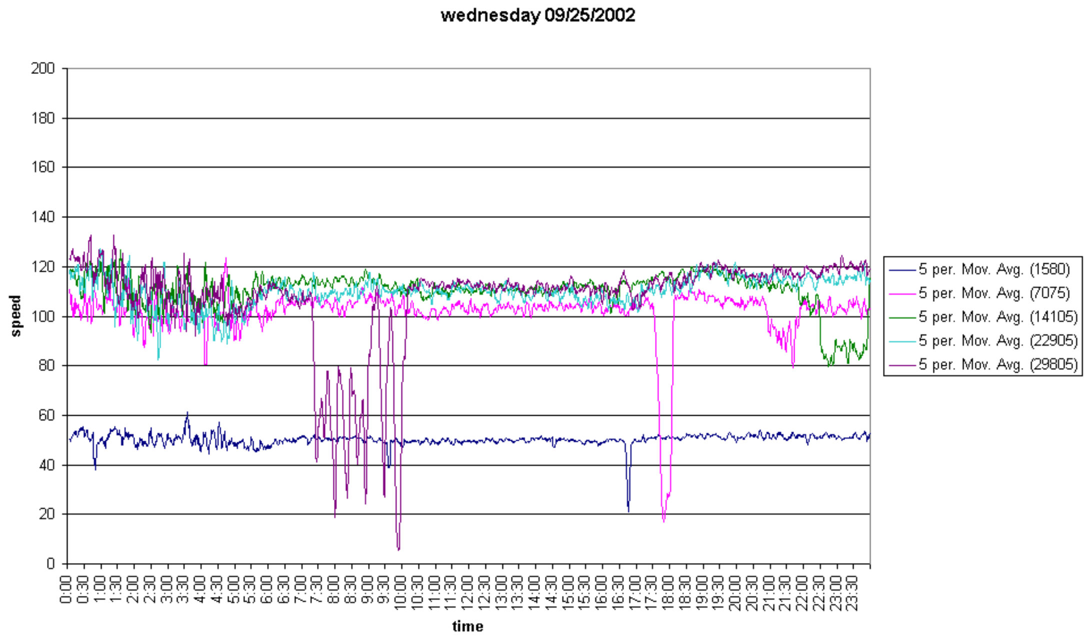


Figure B.10 Traffic on Wednesday 25th of September 2002

B.4 Thursday

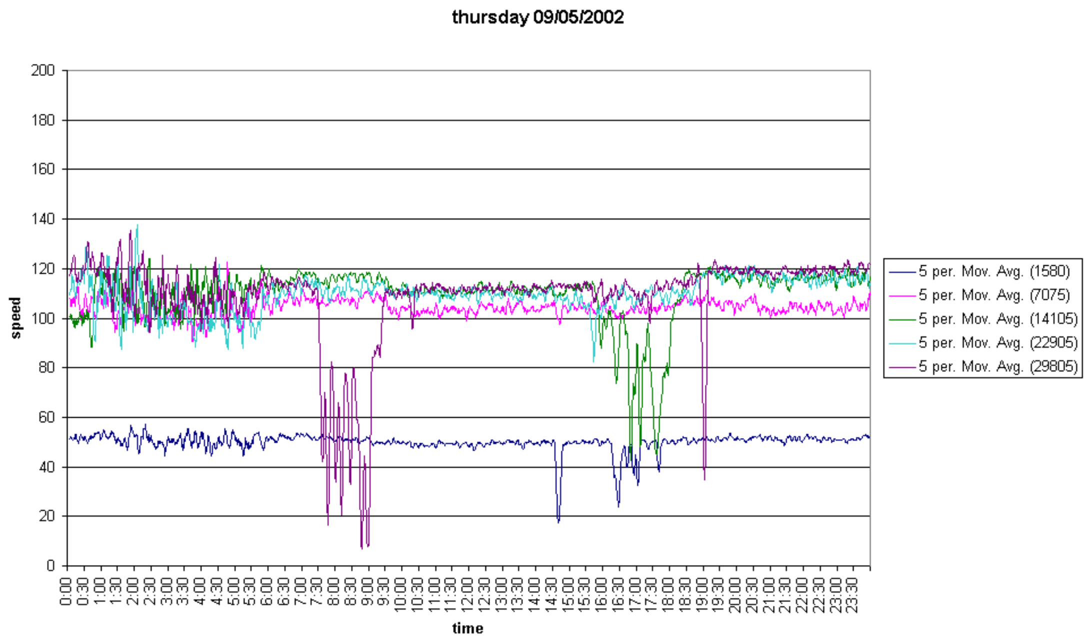


Figure B.11 Traffic on Thursday 5th of September 2002

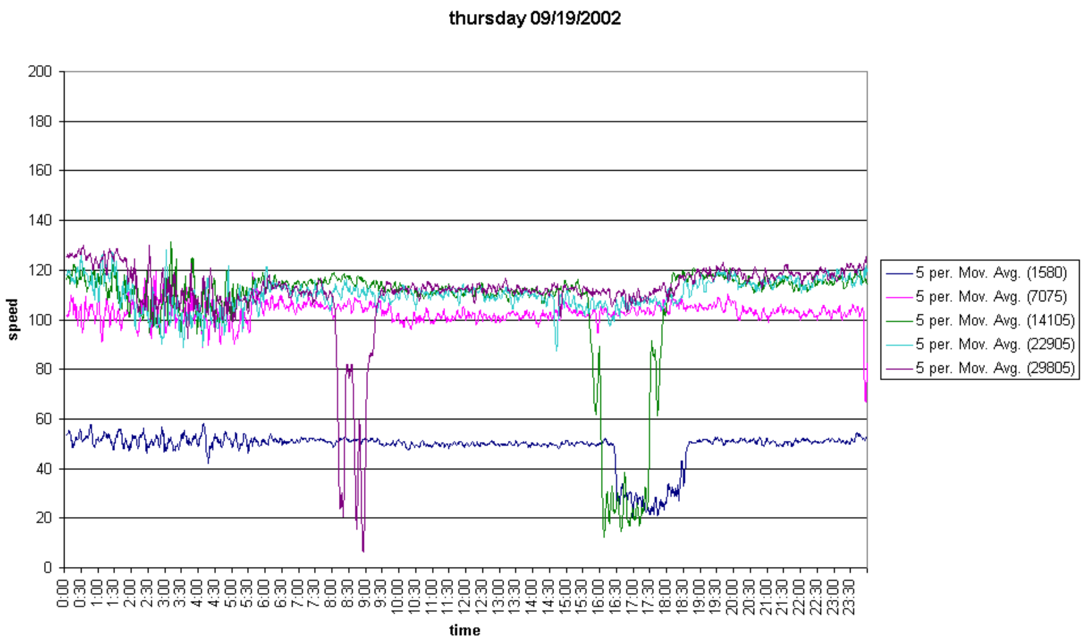


Figure B.12 Traffic on Thursday 19th of September 2002

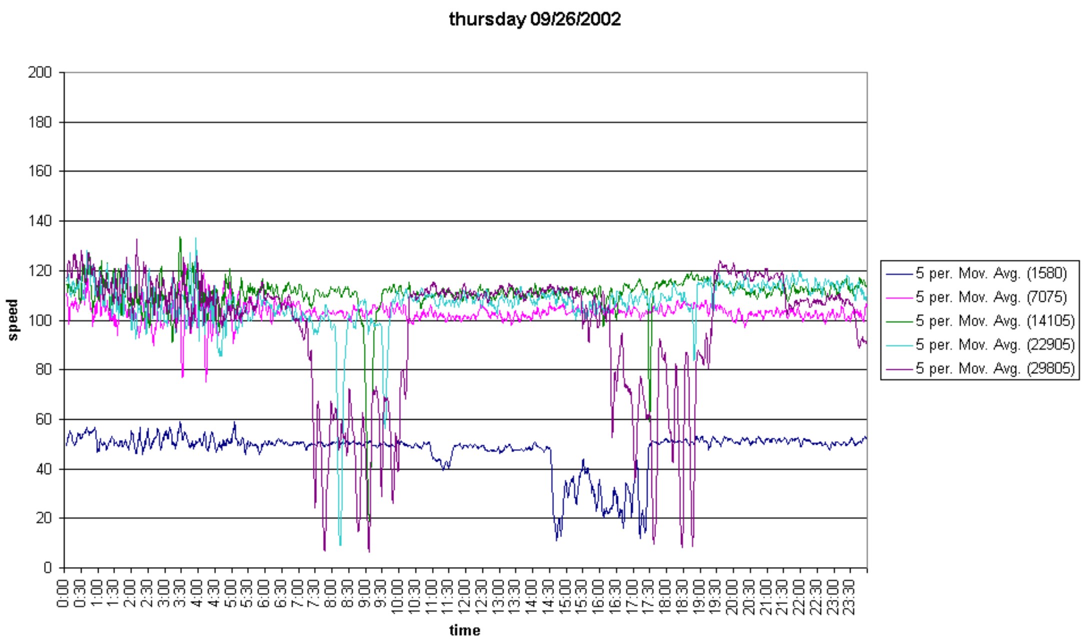


Figure B.13 Traffic on Thursday 26th of September 2002

B.5 Friday

Figure B.14 Traffic on Friday 6th of September 2002

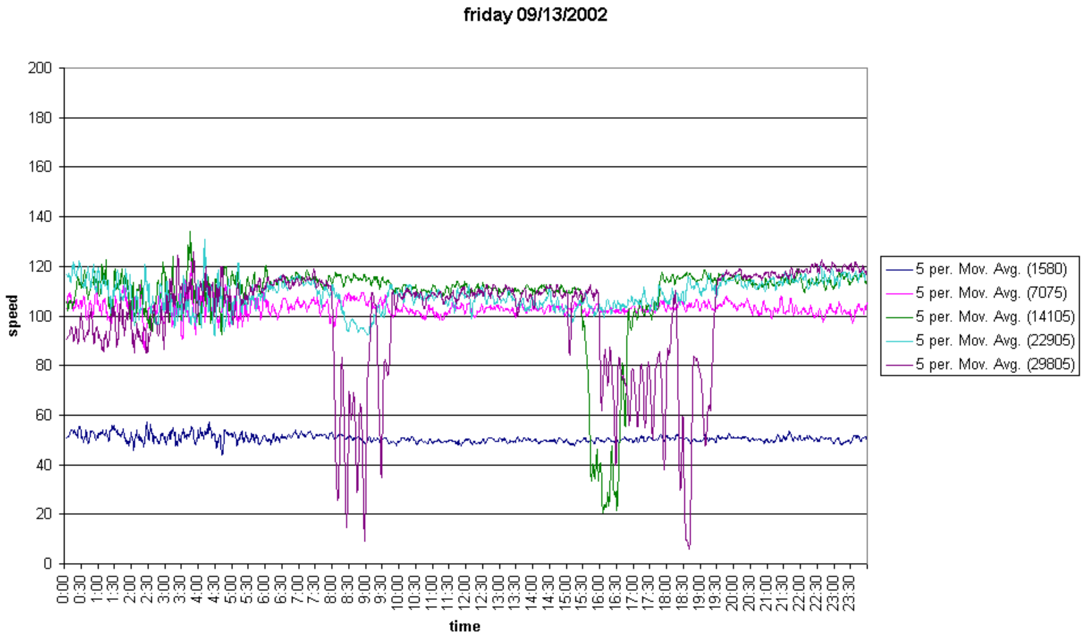


Figure B.15 Traffic on Friday 13th of September 2002

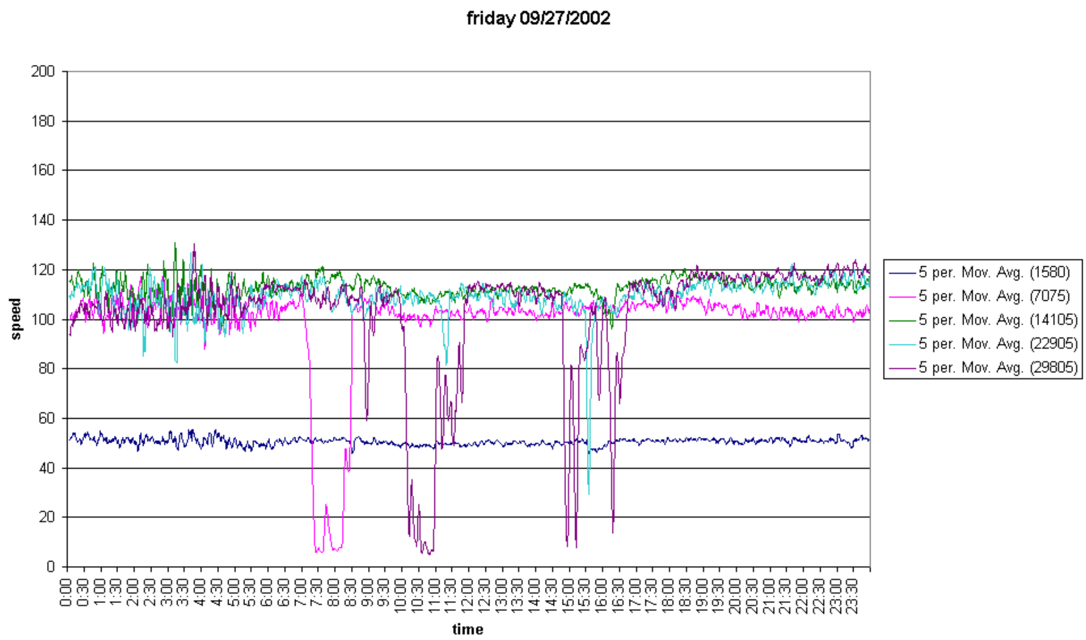


Figure B.16 Traffic on Friday 27th of September 2002

B.6 Saturday

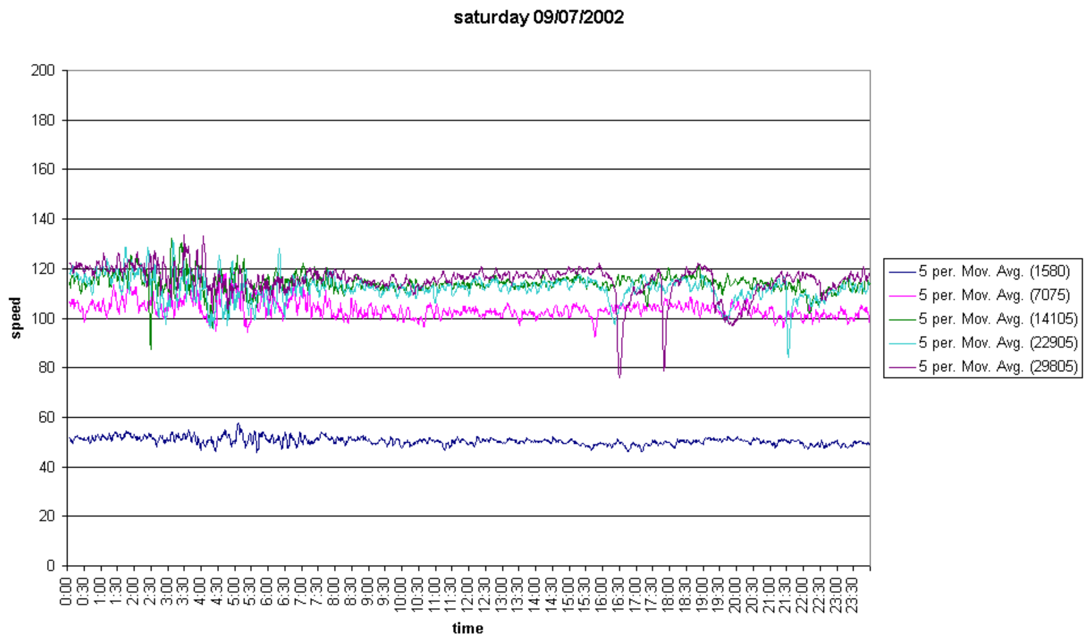


Figure B.17 Traffic on Saturday 7th of September 2002

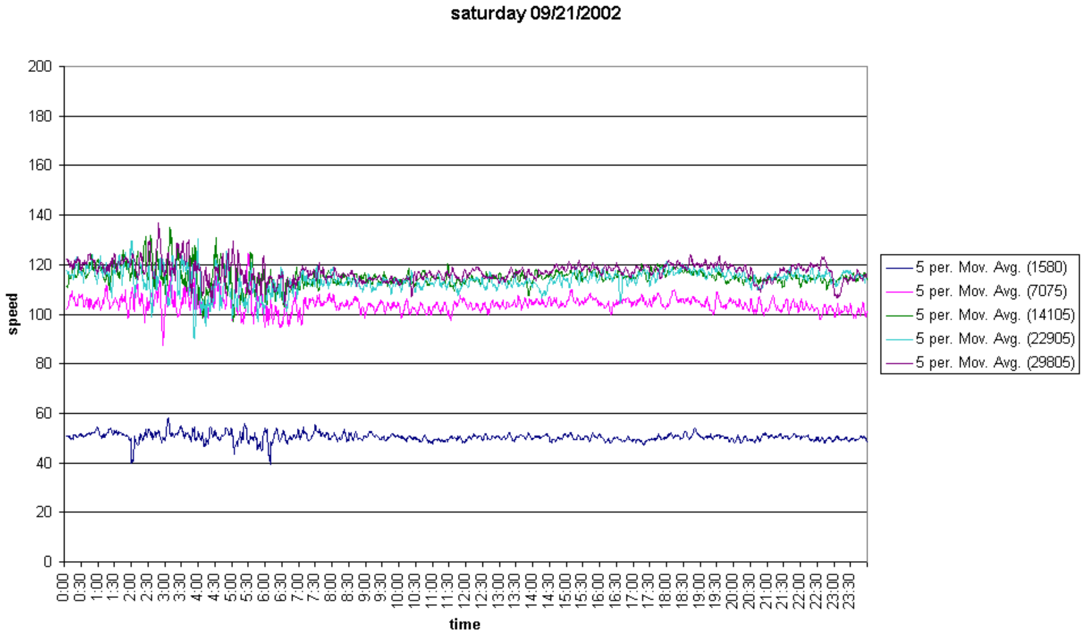


Figure B.18 Traffic on Saturday 21th of September 2002

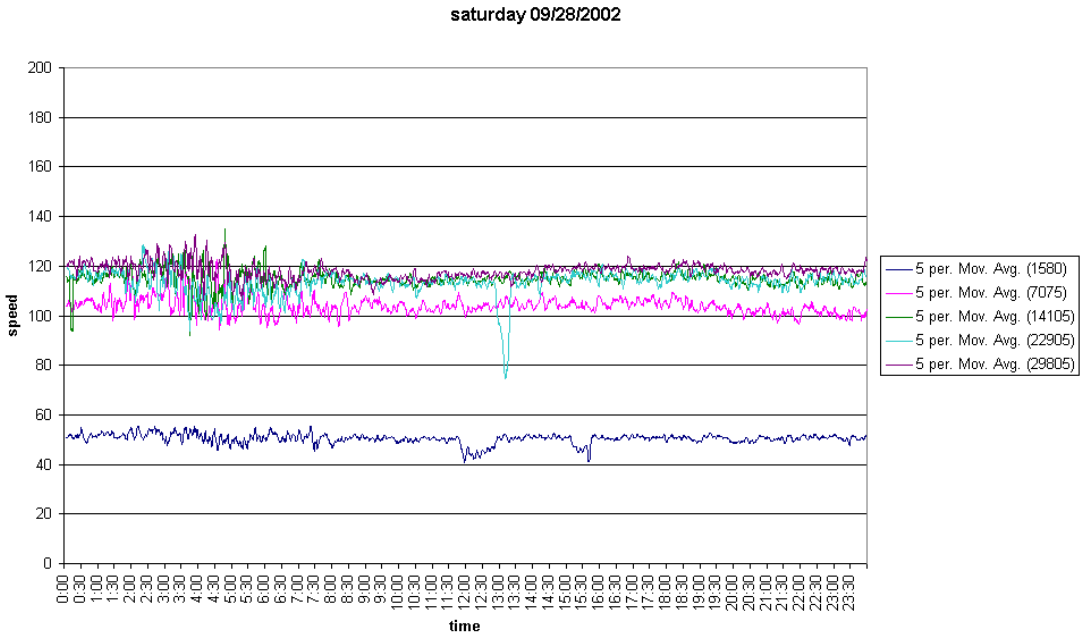


Figure B.19 Traffic on Saturday 28th of September 2002

B.7 Sunday

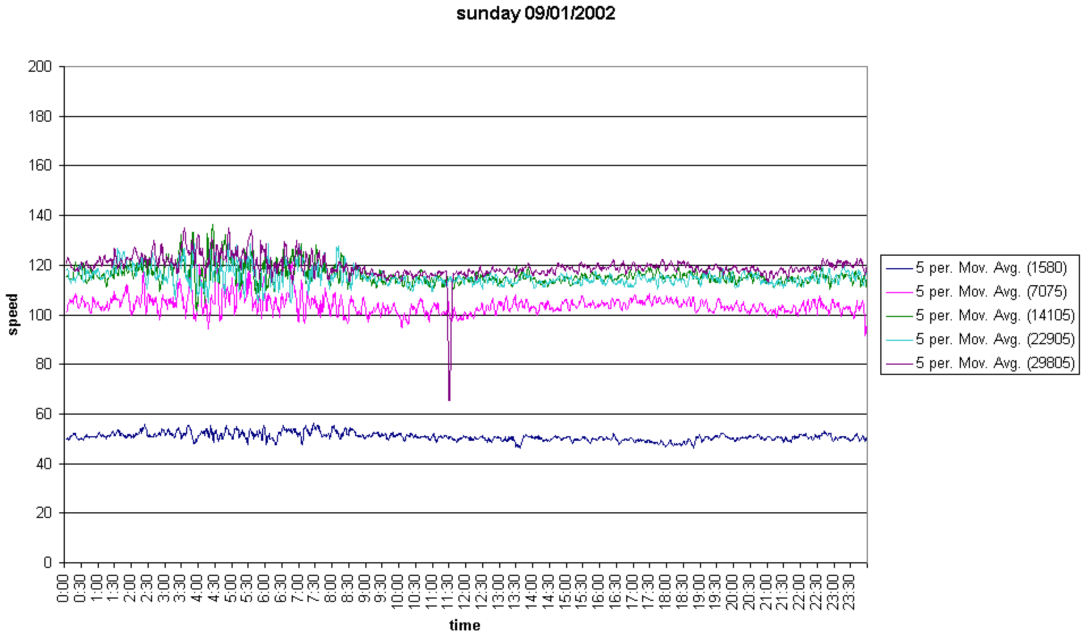


Figure B.20 Traffic on Sunday 1st of September 2002

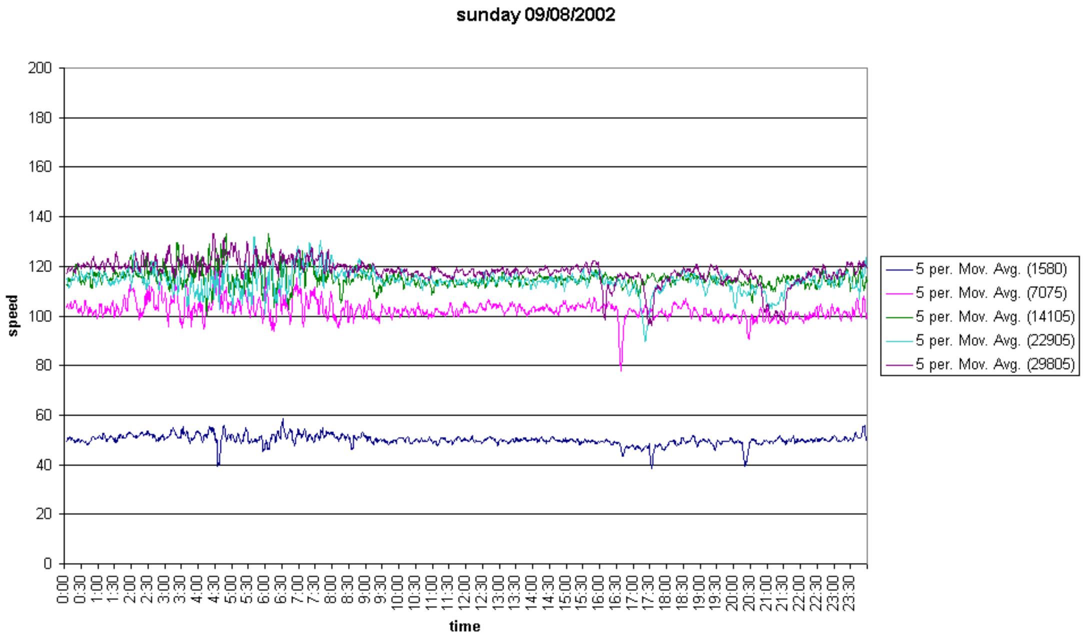


Figure B.21 Traffic on Sunday 8th of September 2002

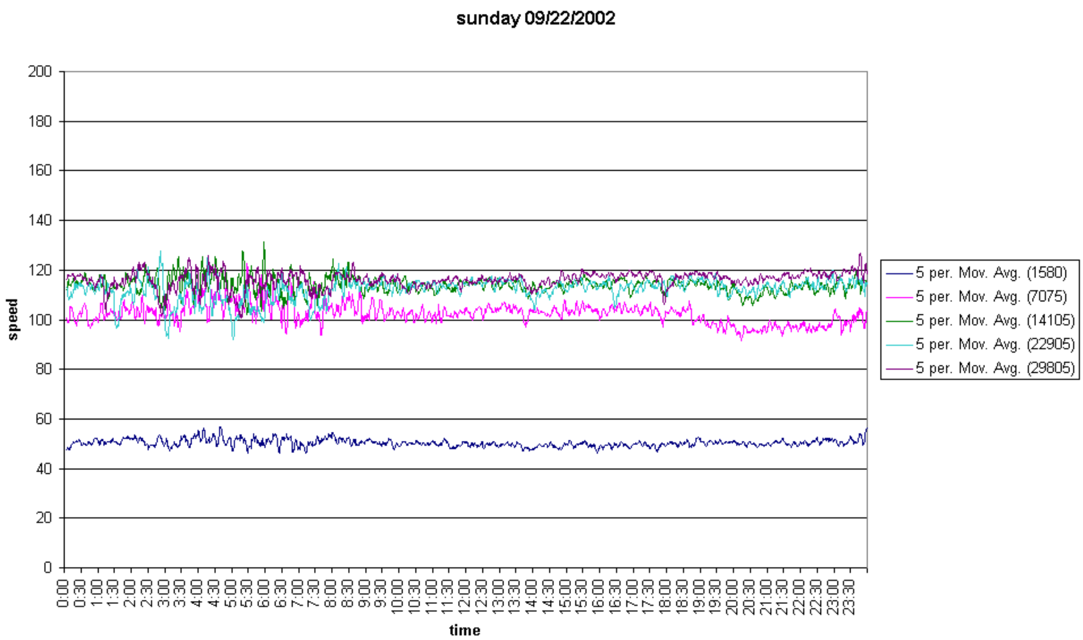


Figure B.22 Traffic on Sunday 22nd of September 2002

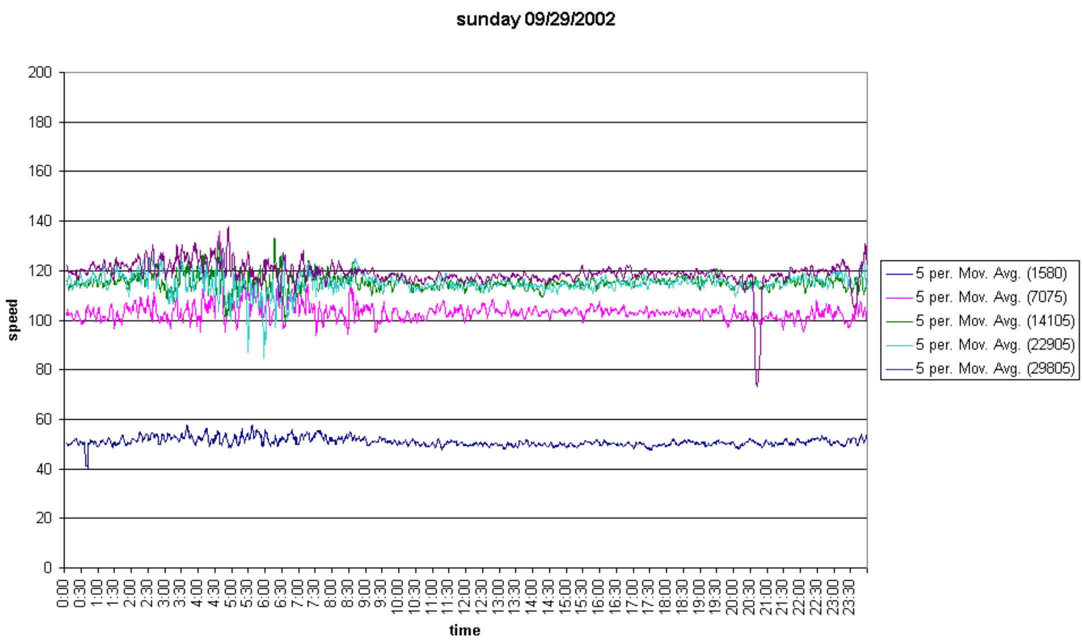


Figure B.23 Traffic on Sunday 29th of September 2002

Appendix C

C.1 The 2-1-1-topology

C.1.1 Zoetermeer

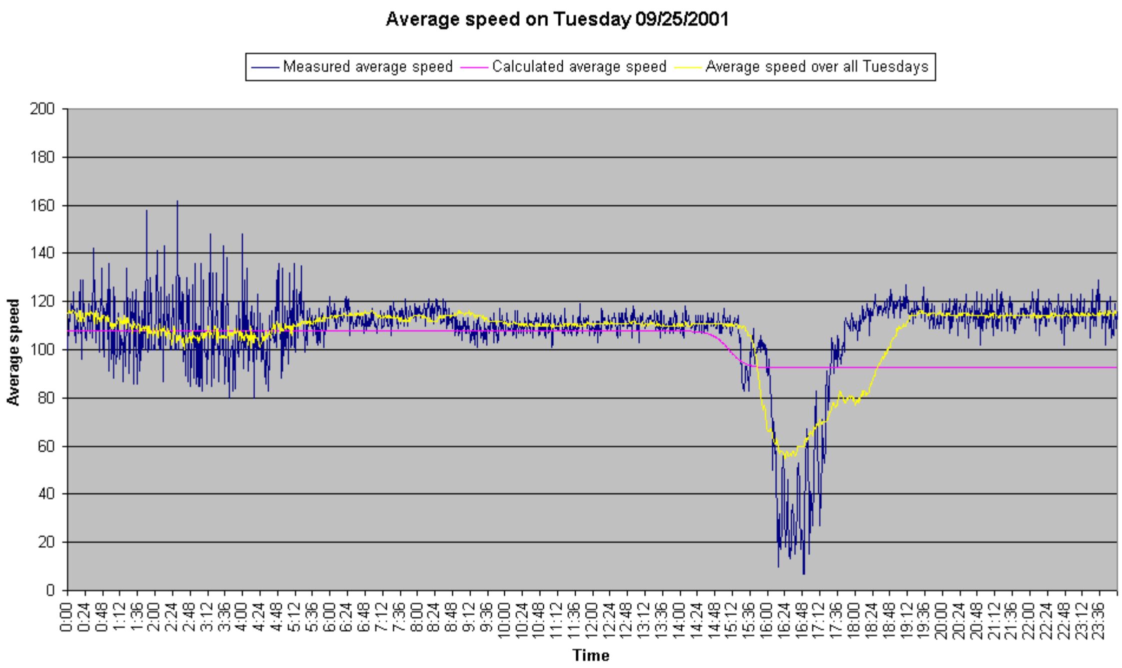


Figure C.1 The average speed on 25th of September 2001 near Zoetermeer with the 2-1-1-topology

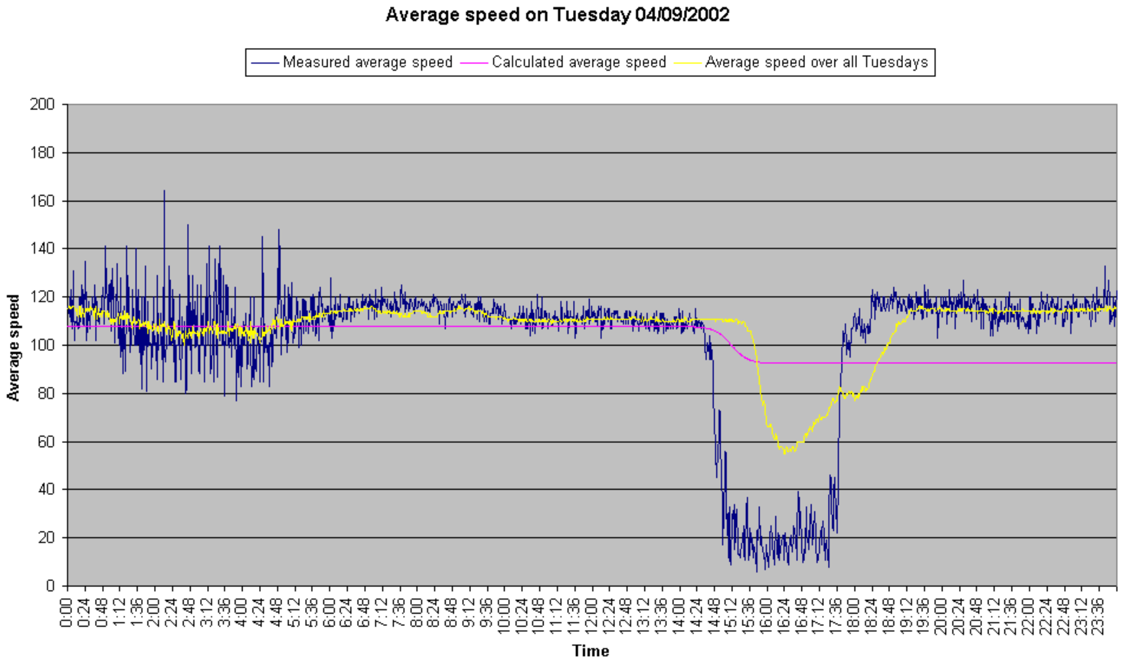


Figure C.2 The average speed on 9th of April 2002 near Zoetermeer with the 2-1-1-topology

C.1.2 Gouda

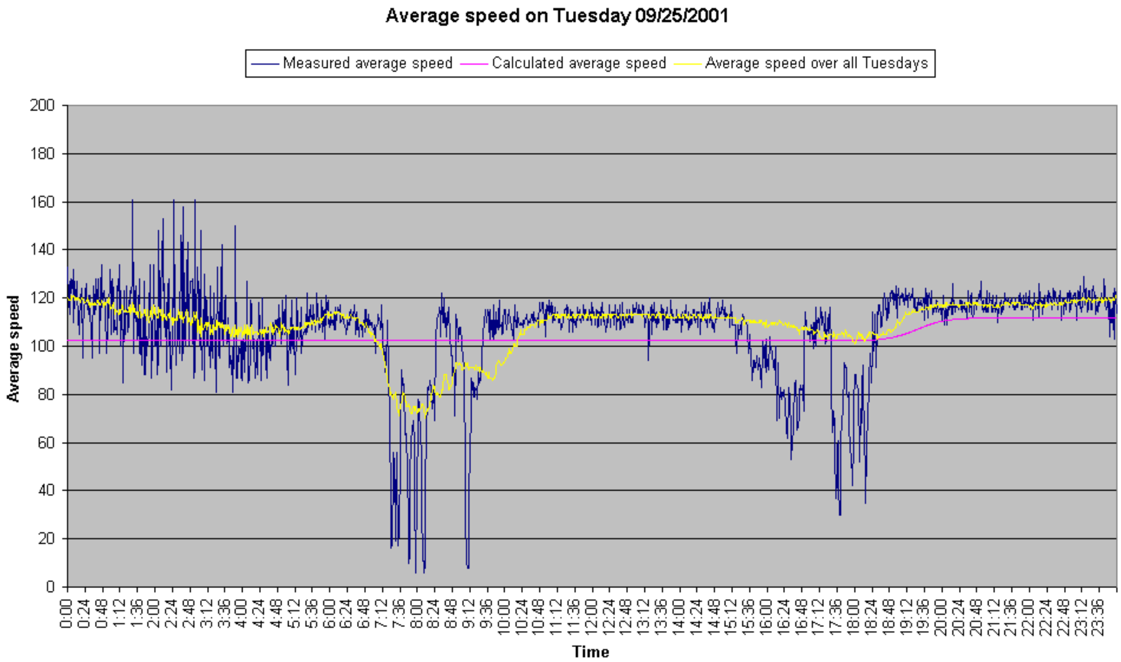


Figure C.3 The average speed on 25th of September 2001 near Gouda with the 2-1-1-topology

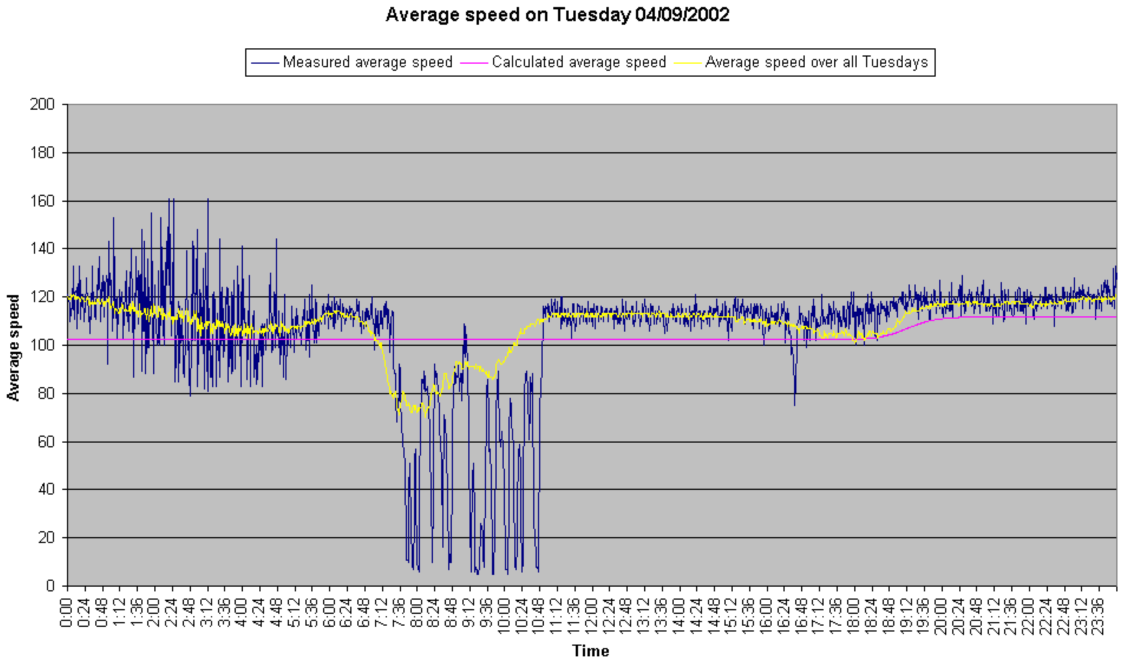


Figure C.4 The average speed on 9th of April 2002 near Gouda with the 2-1-1-topology

C.2 The 2-2-1-topology

C.2.1 Zoetermeer

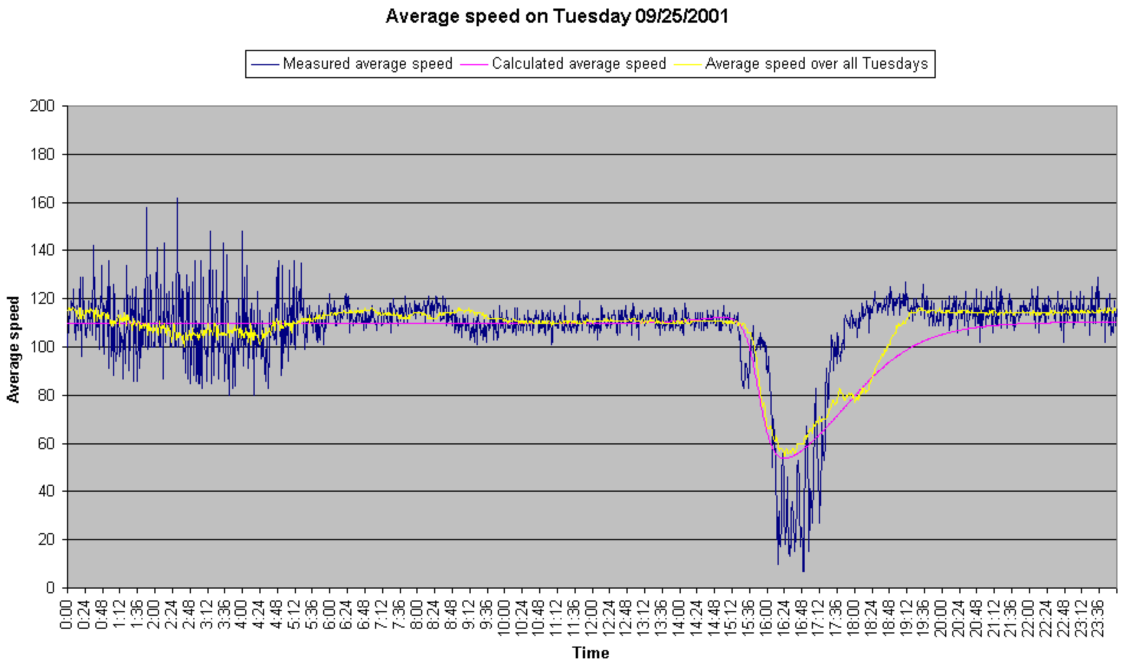


Figure C.5 The average speed on 25th of September 2001 near Zoetermeer with the 2-2-1-topology

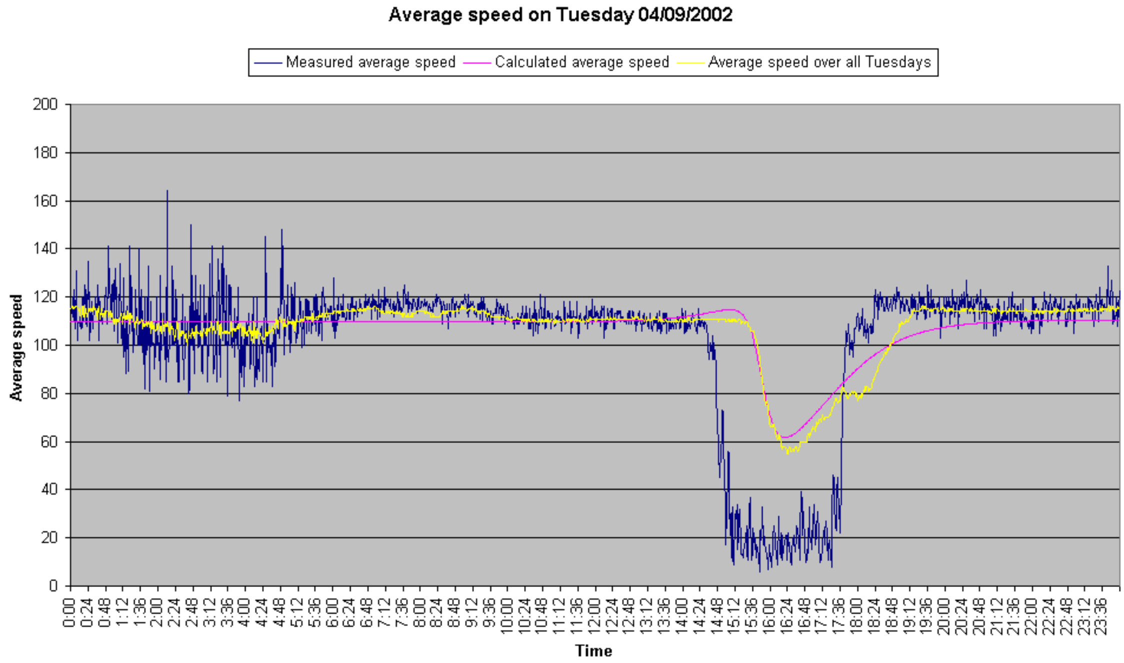


Figure C.6 The average speed on 9th of April 2002 near Zoetermeer with the 2-2-1-topology

C.2.2 Gouda

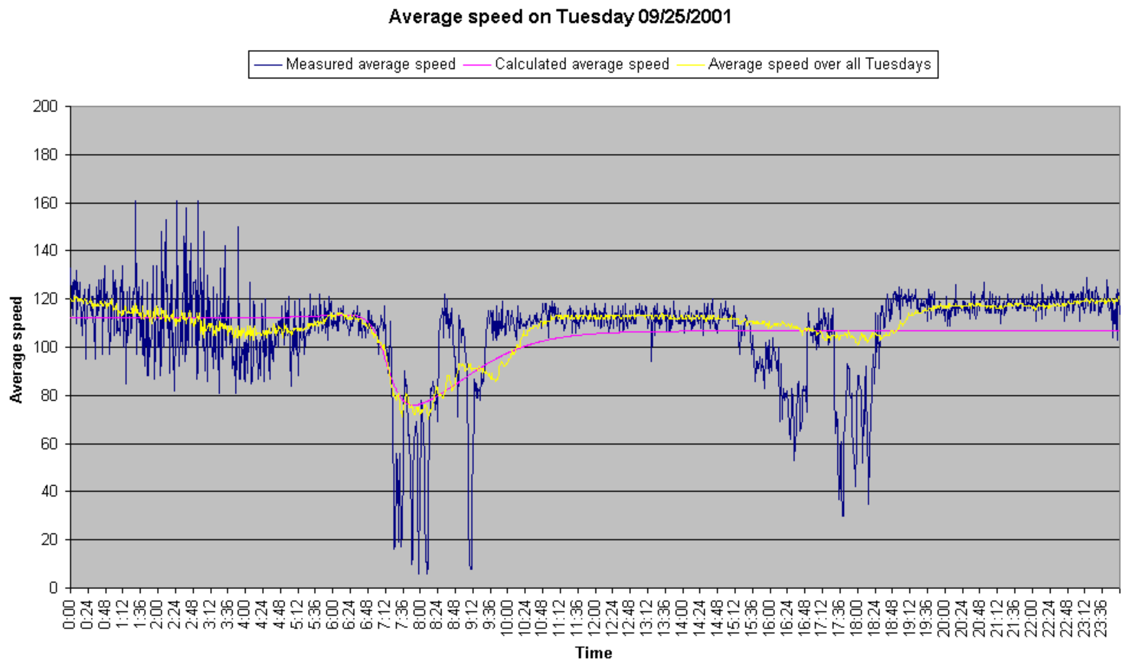


Figure C.7 The average speed on 25th of September 2001 near Gouda with the 2-2-1-topology

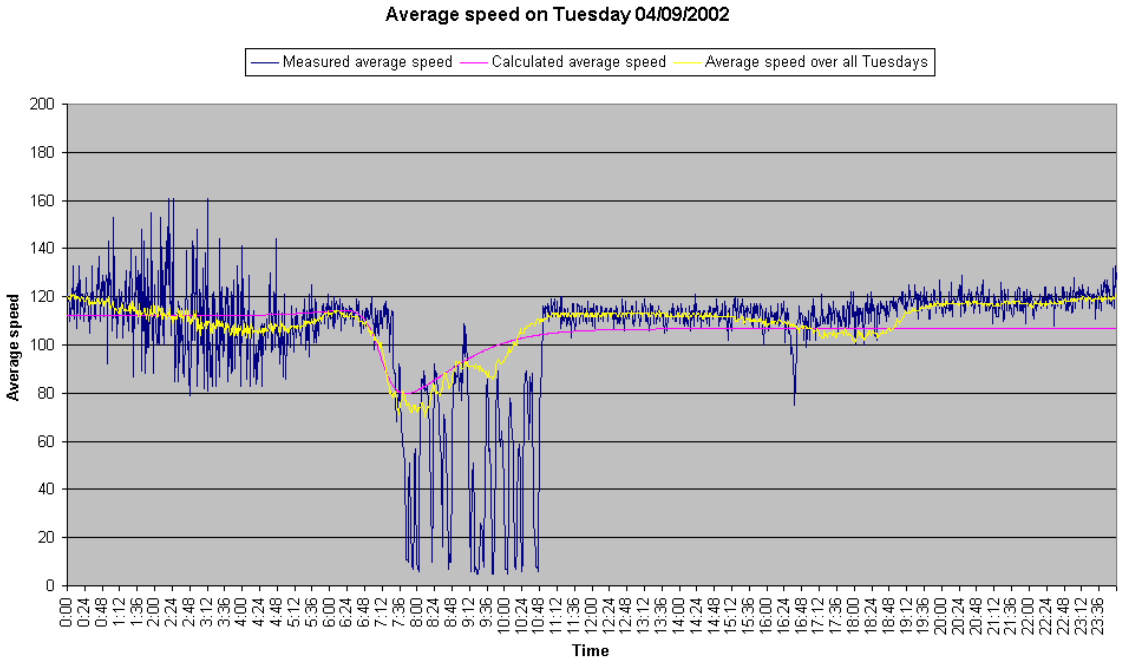


Figure C.8 The average speed on 9th of April 2002 near Gouda with the 2-2-1-topology

C.3 The 2-3-1-topology

C.3.1 Zoetermeer

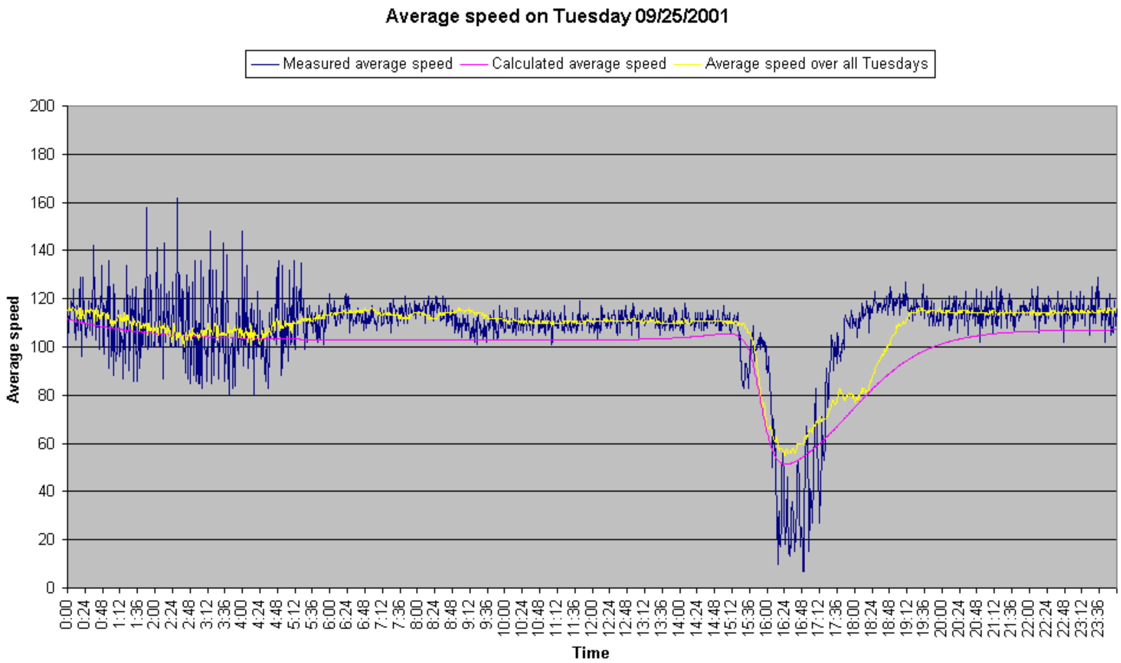


Figure C.9 The average speed on 25th of September 2001 near Zoetermeer with the 2-3-1 -topology

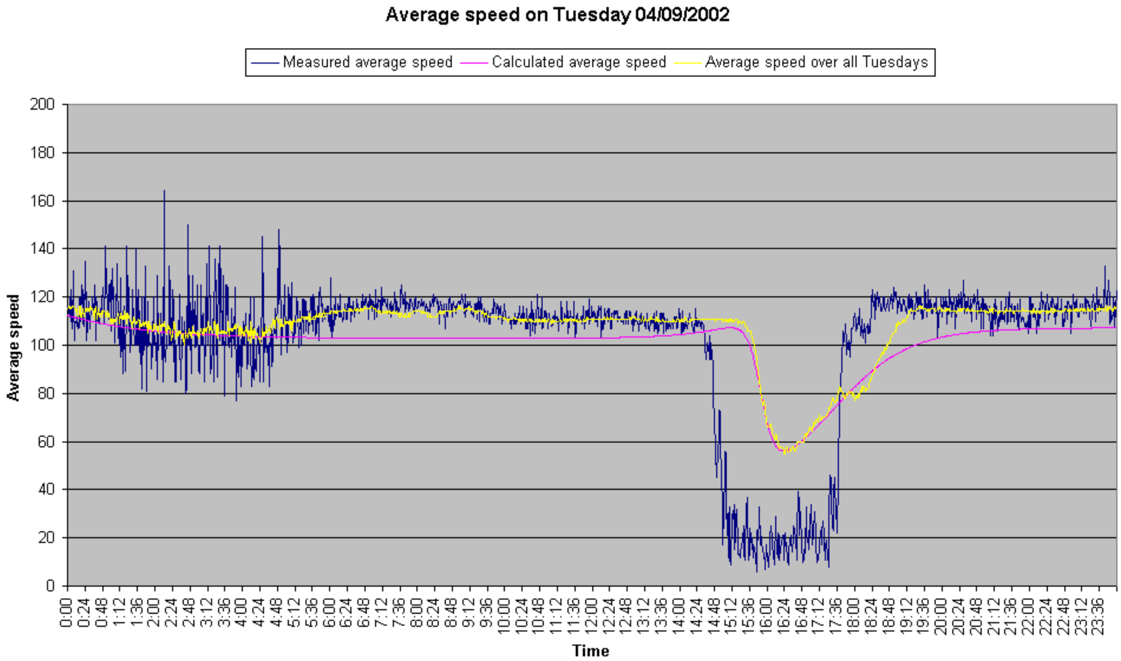


Figure C.10 The average speed on 9th of April 2002 near Zoetermeer with the 2-3-1-topology

C.3.2 Gouda

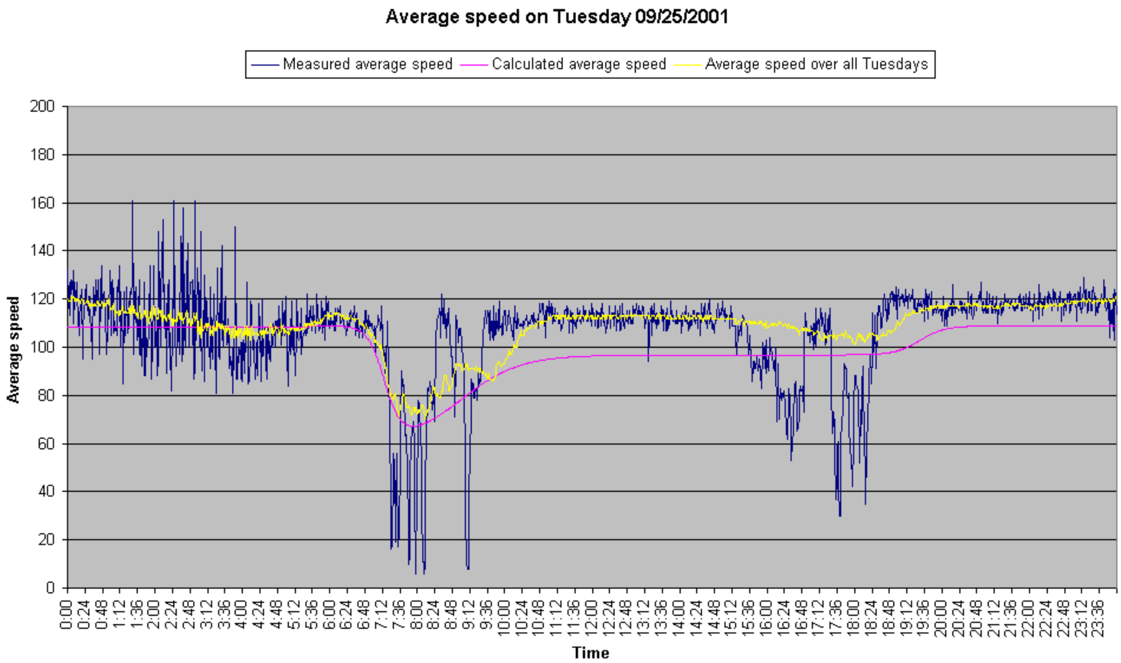


Figure C.11 The average speed on 25th of September 2001 near Gouda with the 2-3-1-topology

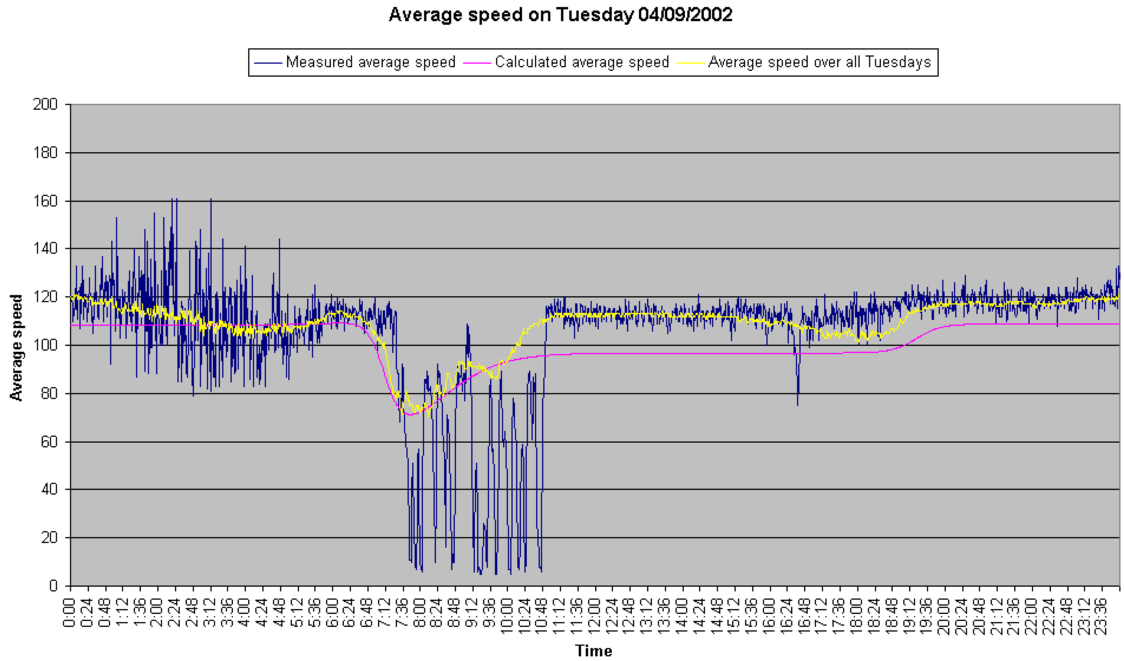


Figure C.12 The average speed on 9th of April 2002 near Gouda with the 2-3-1-topology

C.4 The 2-4-1-topology

C.4.1 Zoetermeer

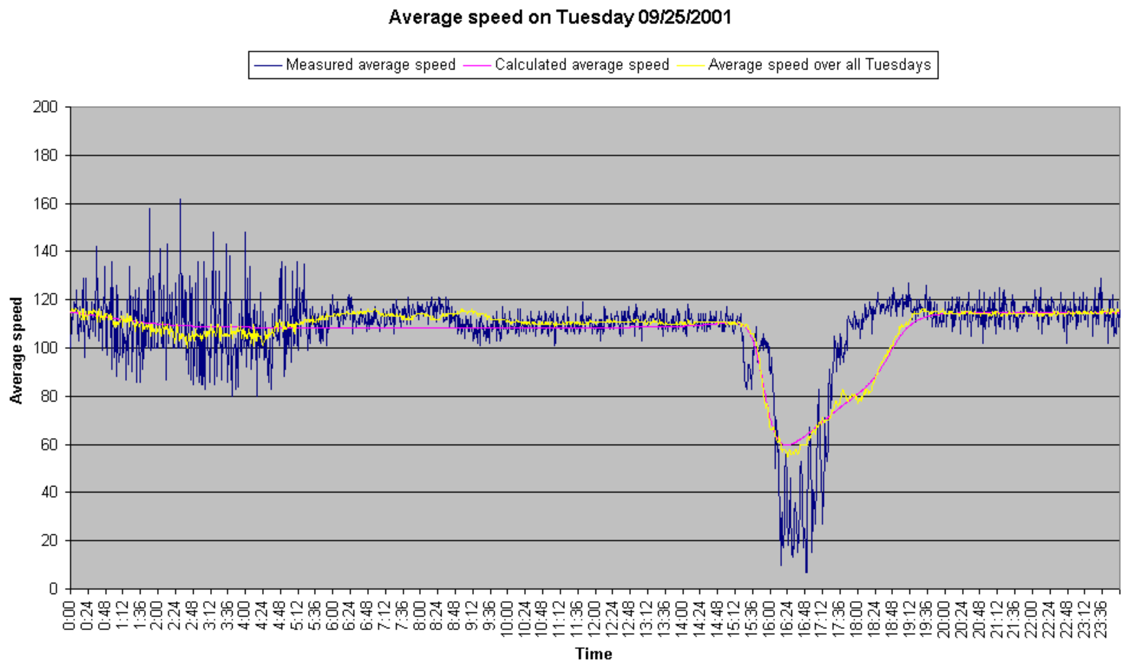


Figure C.13 The average speed on 25th of September 2001 near Zoetermeer with the 2-4-1-topology

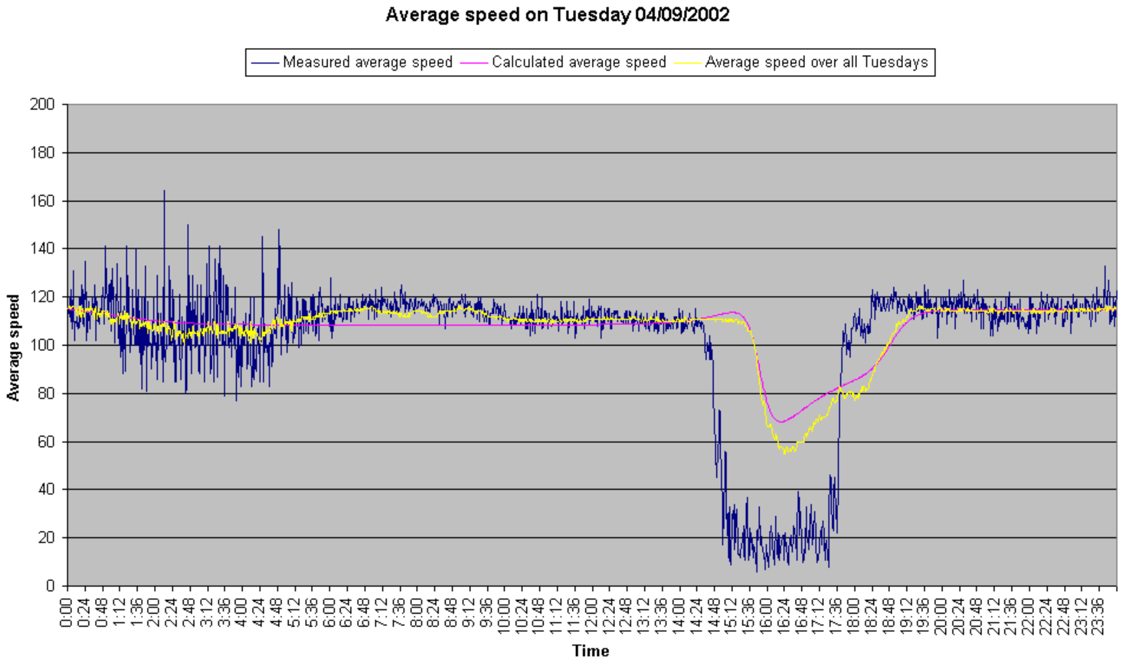


Figure C.14 The average speed on 9th of April 2002 near Zoetermeer with the 2-4-1-topology

C.4.2 Gouda

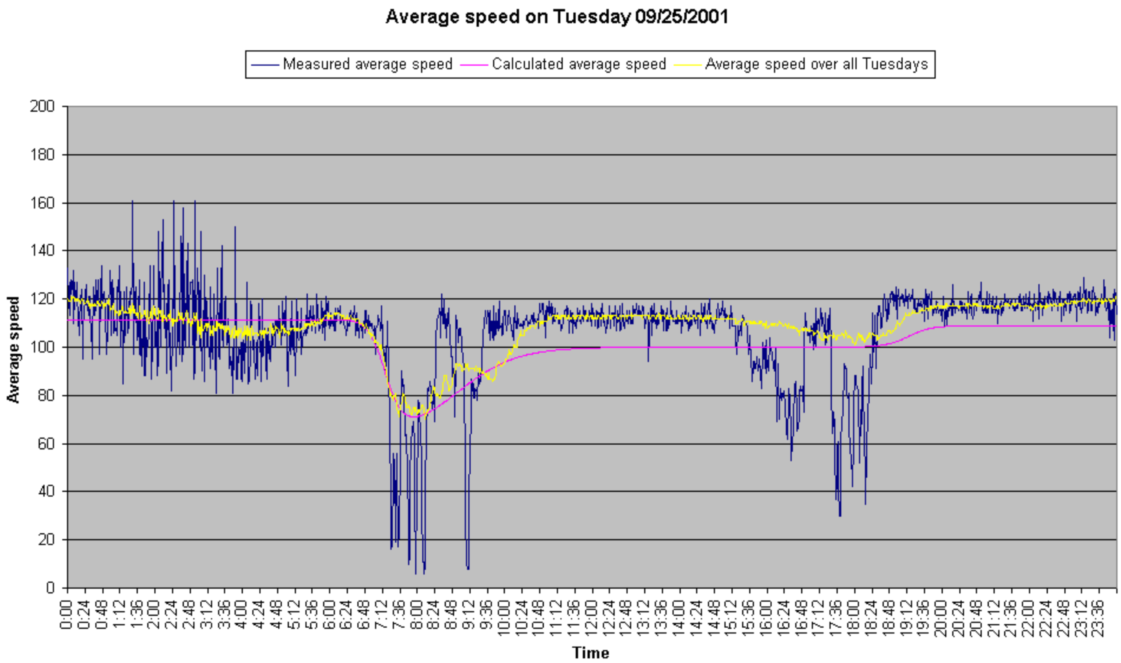


Figure C.15 The average speed on 25th of September 2001 near Gouda with the 2-4-1-topology

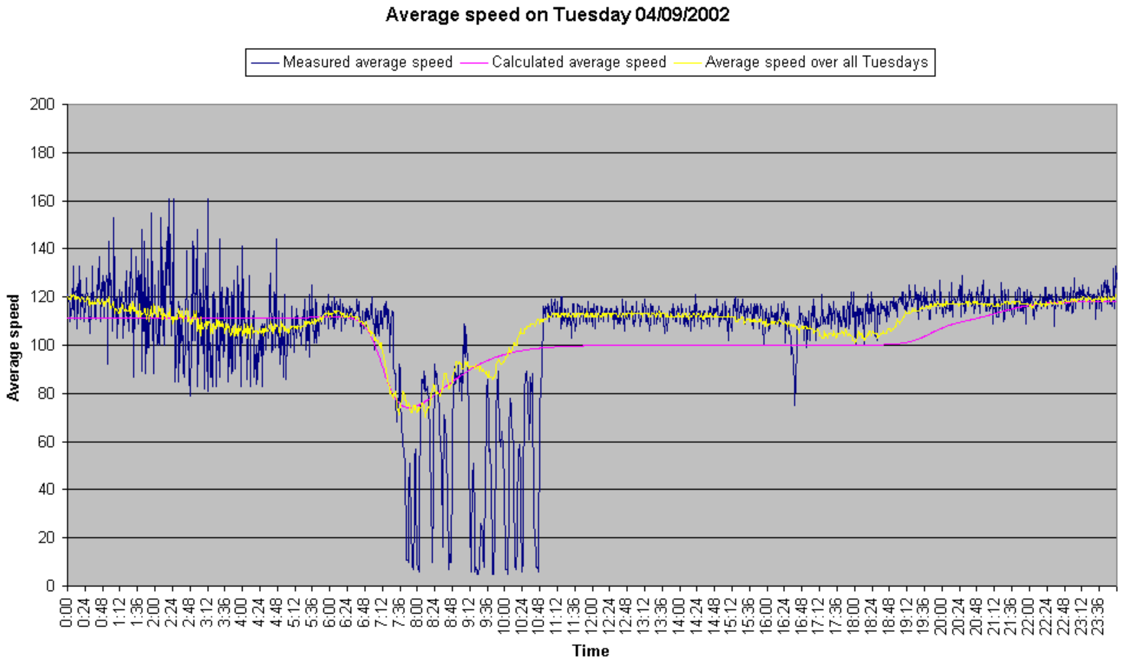


Figure C.16 The average speed on 9th of April 2002 near Gouda with the 2-4-1-topology

C.5 The 2-5-1-topology

C.5.1 Zoetermeer

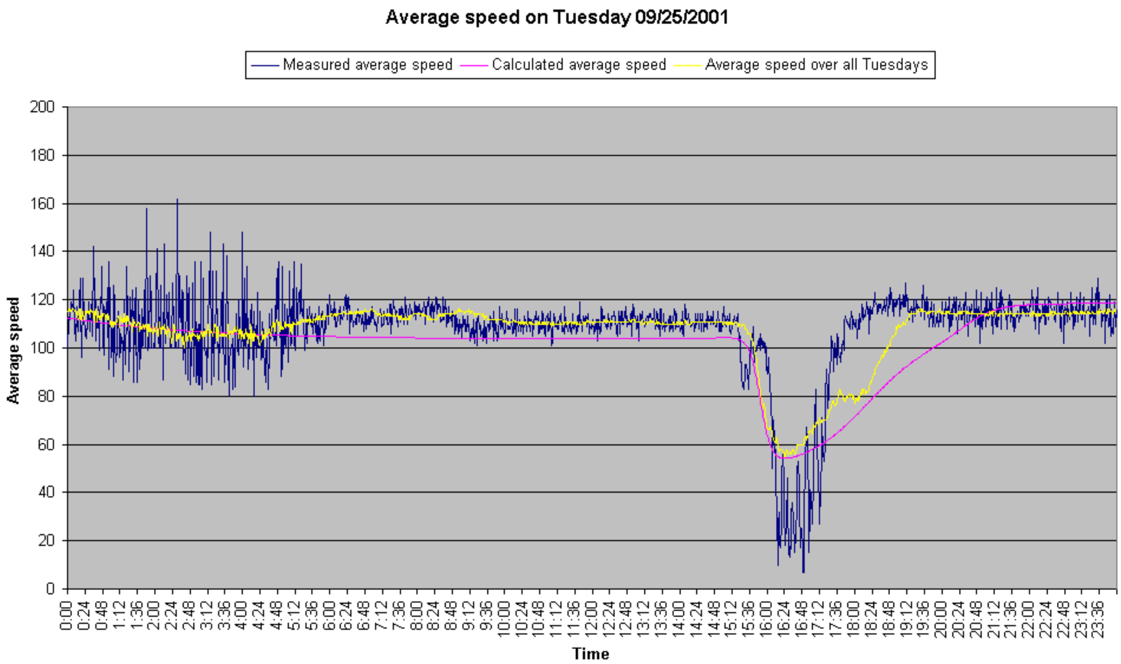


Figure C.17 The average speed on 25th of September 2001 near Zoetermeer with the 2-5-1-topology

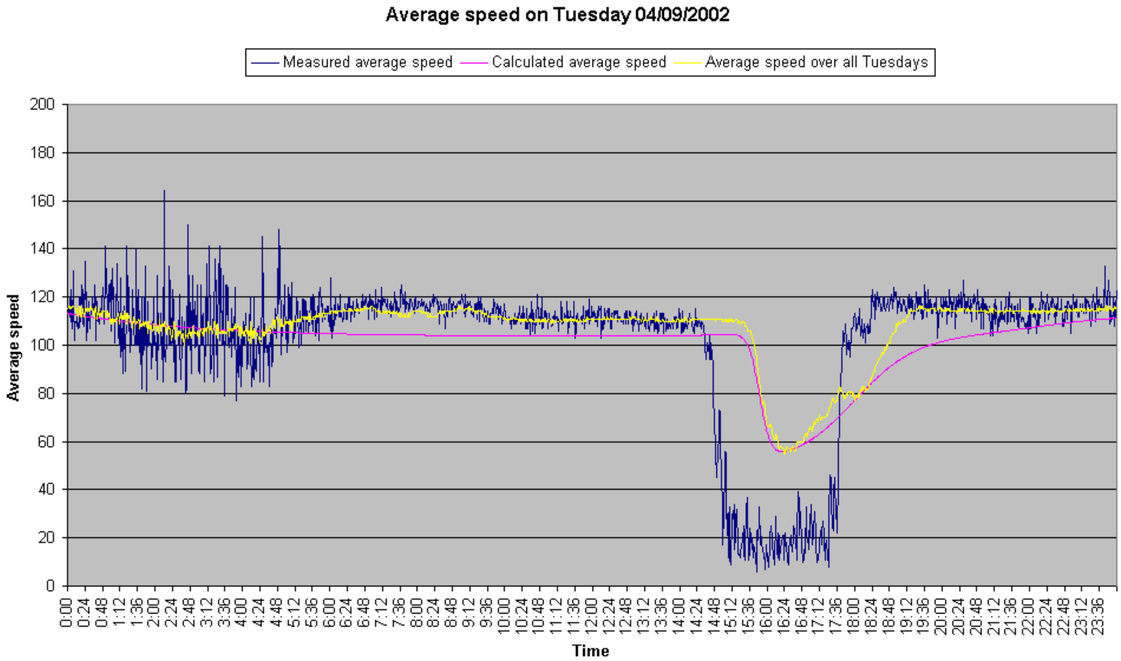


Figure C.18 The average speed on 9th of April 2002 near Zoetermeer with the 2-5-1-topology

C.5.2 Gouda

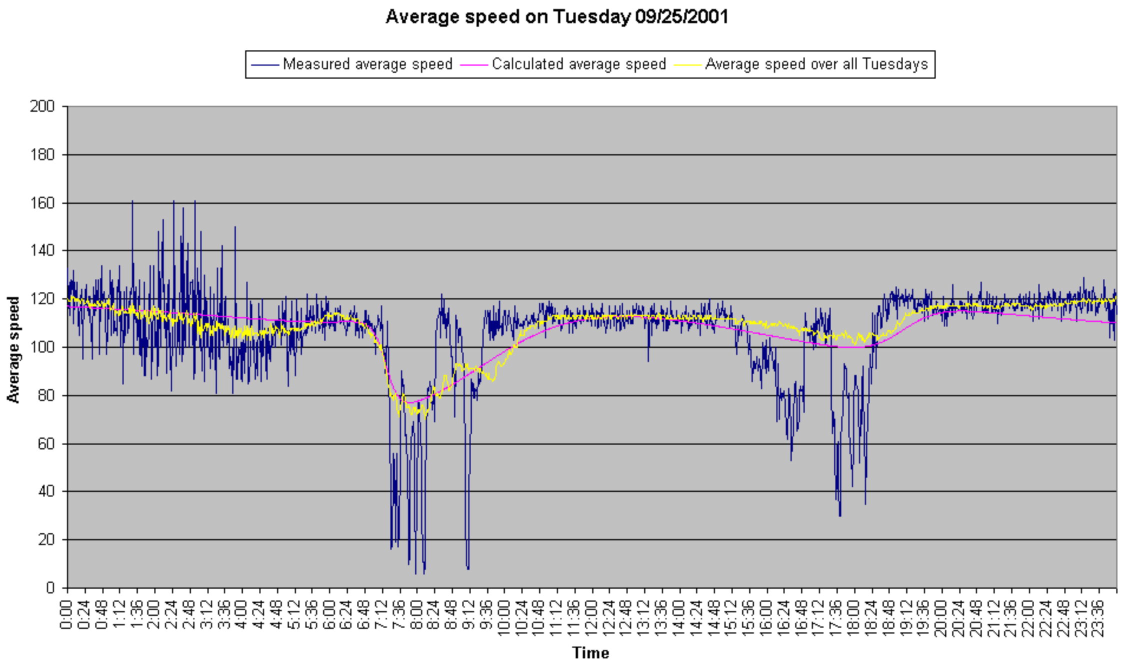


Figure C.19 The average speed on 25th of September 2001 near Gouda with the 2-5-1-topology

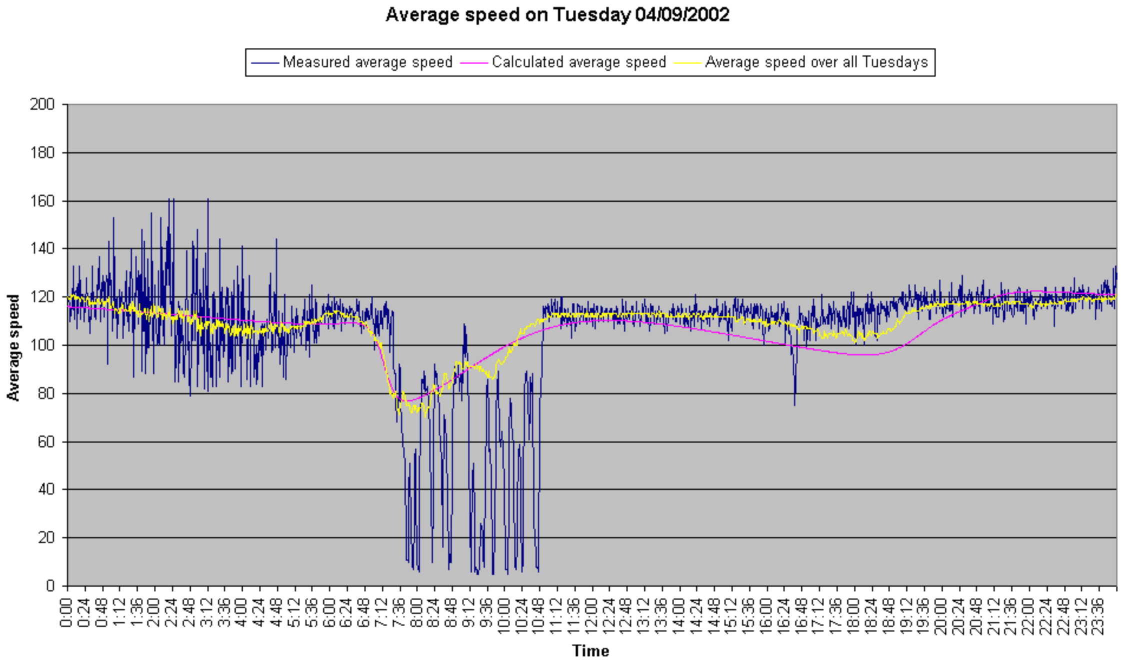


Figure C.20 The average speed on 9th of April 2002 near Gouda with the 2-5-1-topology

C.6 The 2-6-1-topology

C.6.1 Zoetermeer

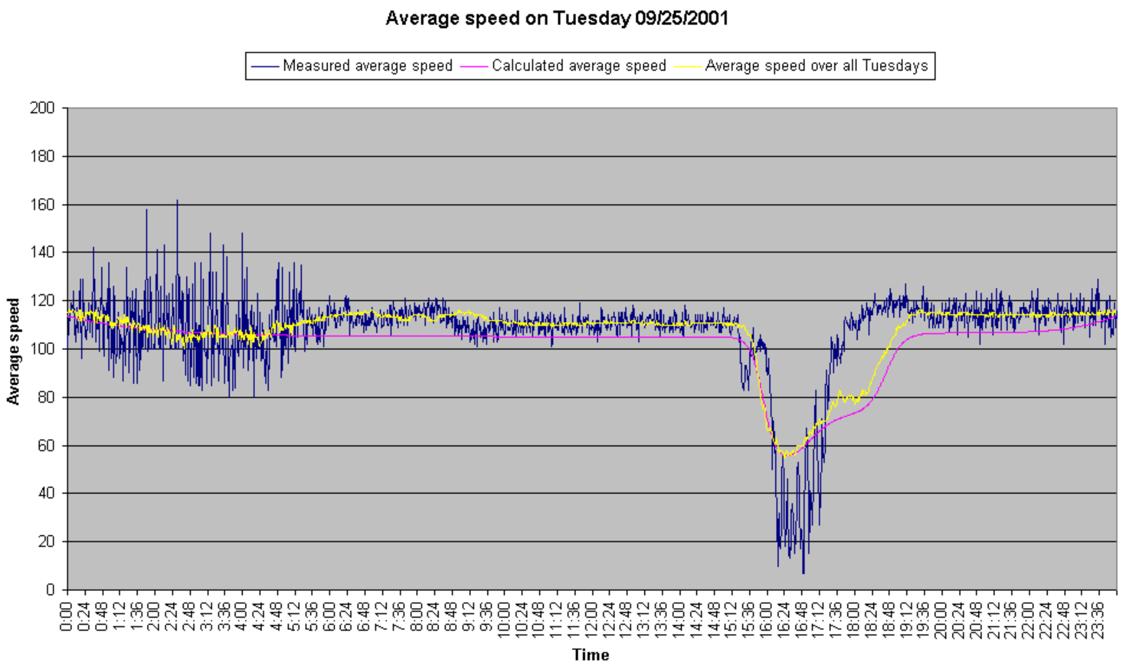


Figure C.21 The average speed on 25th of September 2001 near Zoetermeer with the 2-6-1-topology

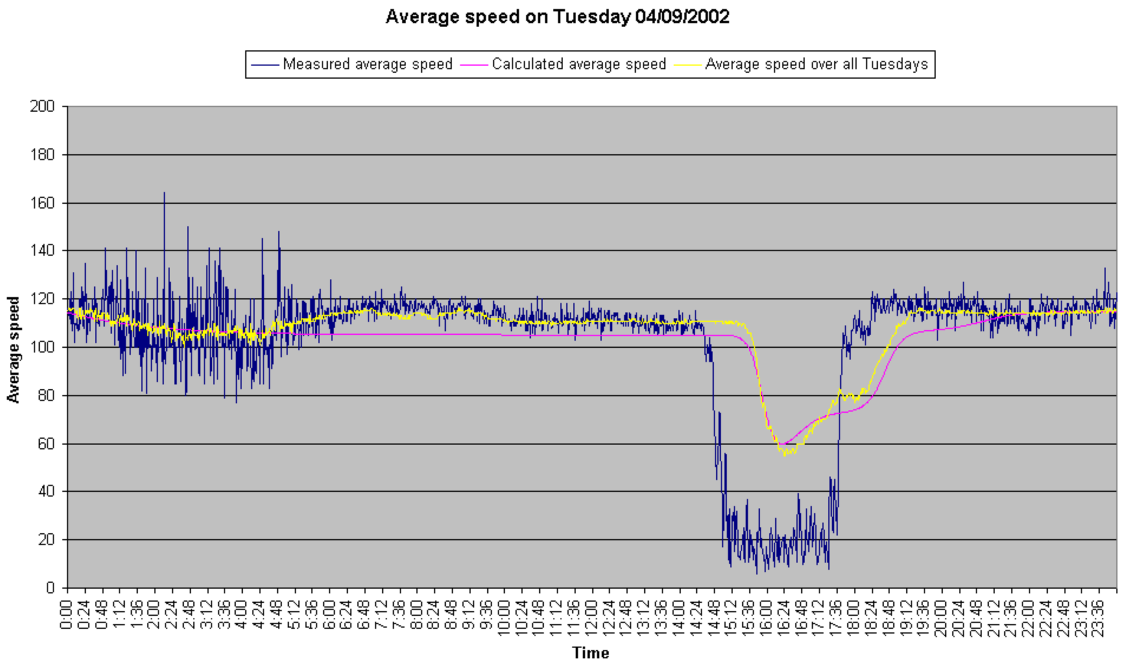


Figure C.22 The average speed on 9th of April 2002 near Zoetermeer with the 2-6-1-topology

C.6.2 Gouda

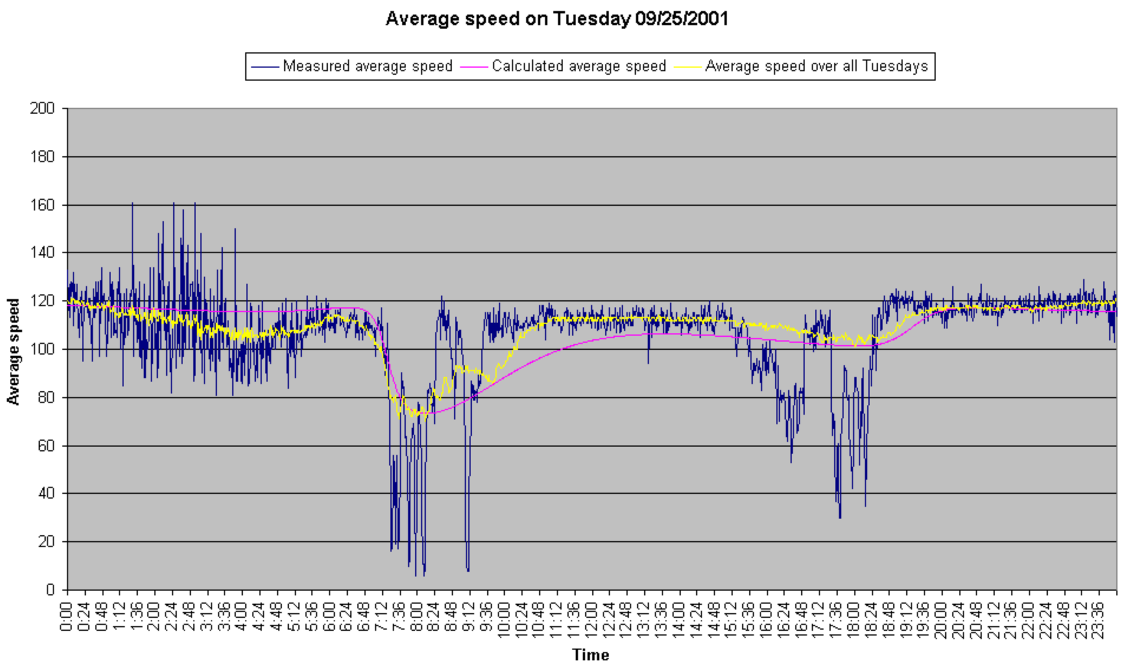


Figure C.23 The average speed on 25th of September 2001 near Gouda with the 2-6-1-topology

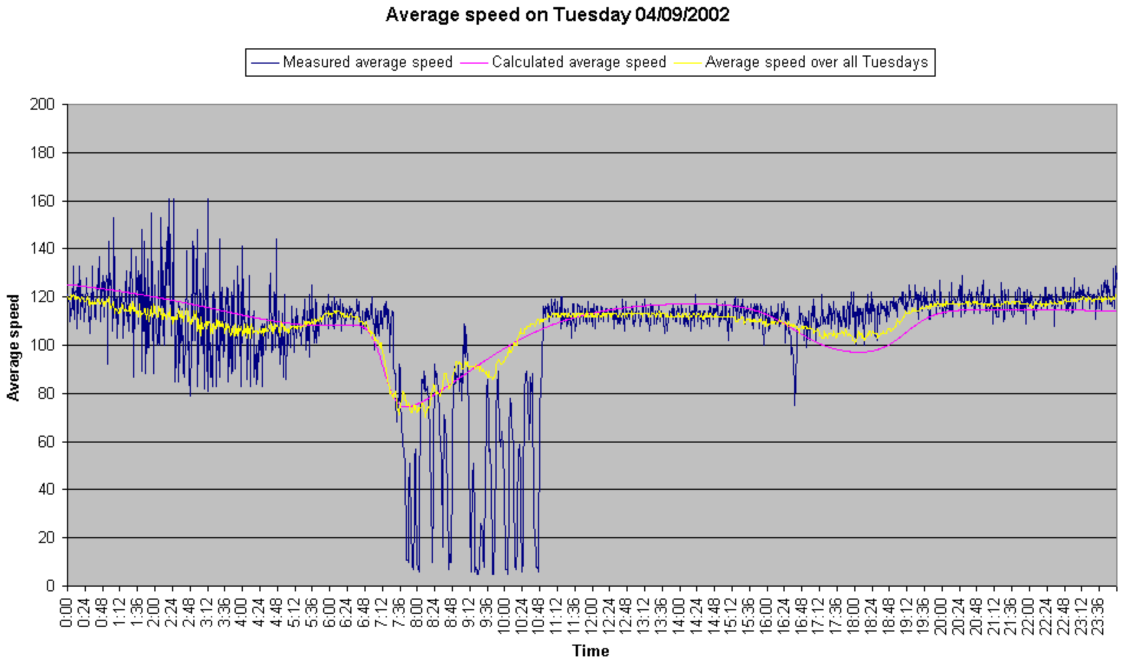


Figure C.24 The average speed on 9th of April 2002 near Gouda with the 2-6-1-topology

C.7 The 5-1-4-topology

C.7.1 Zoetermeer

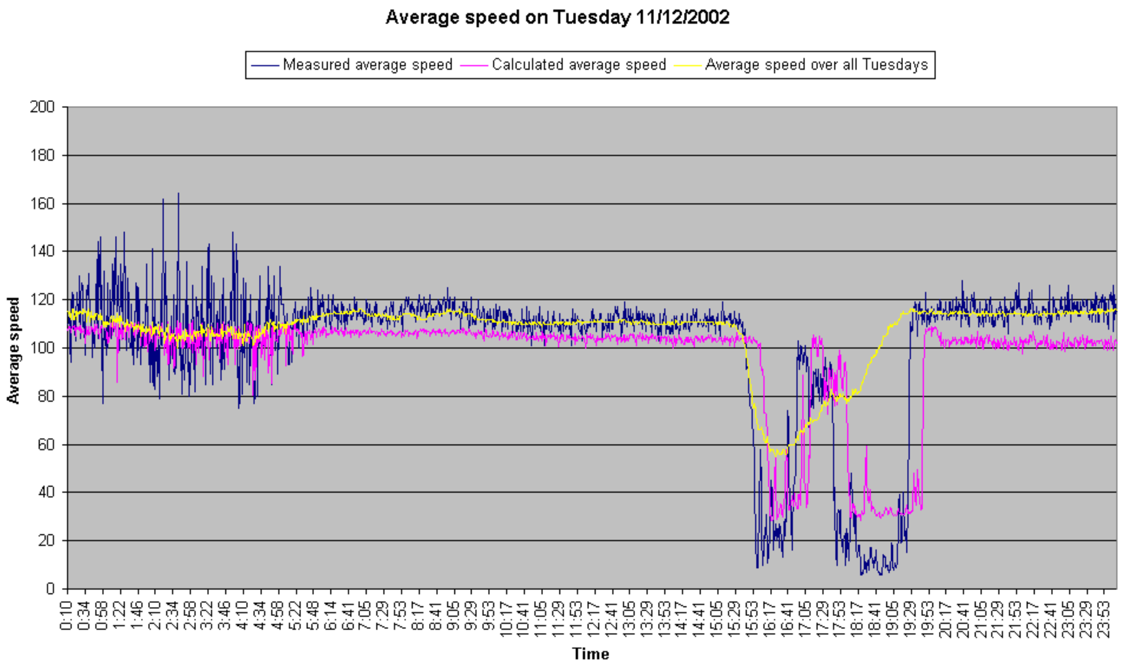


Figure C.25 The average speed on 12th of November 2002 near Zoetermeer with the 5-1-4-topology 20 minutes ahead

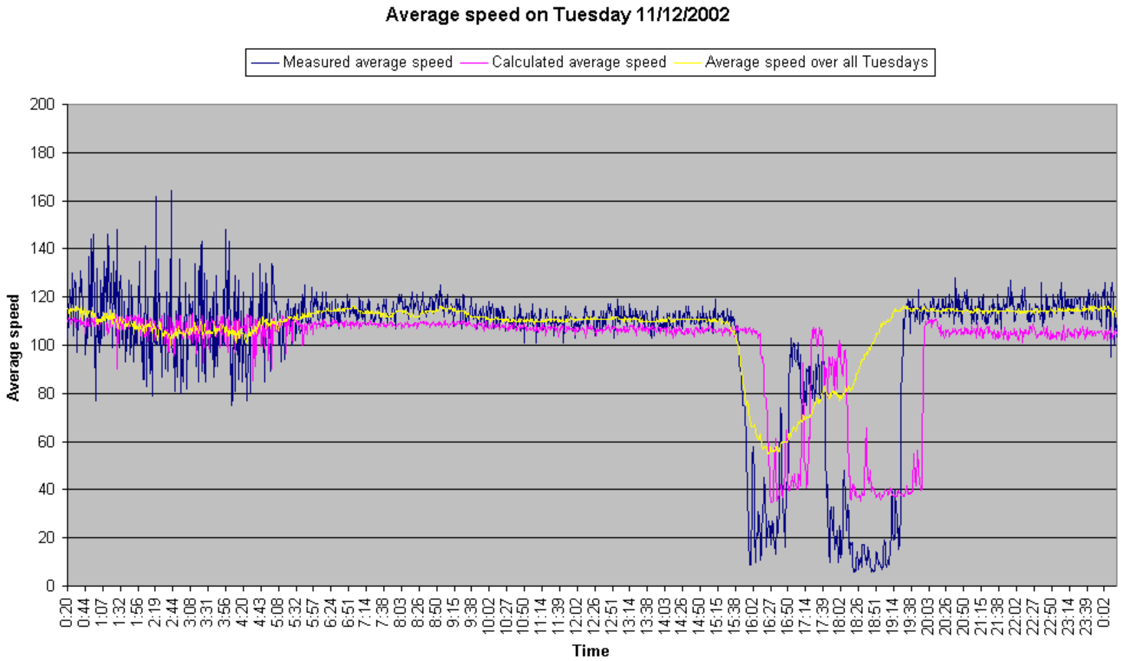


Figure C.26 The average speed on 12th of November 2002 near Zoetermeer with the 5-1-4-topology 30 minutes ahead

C.7.2 Gouda

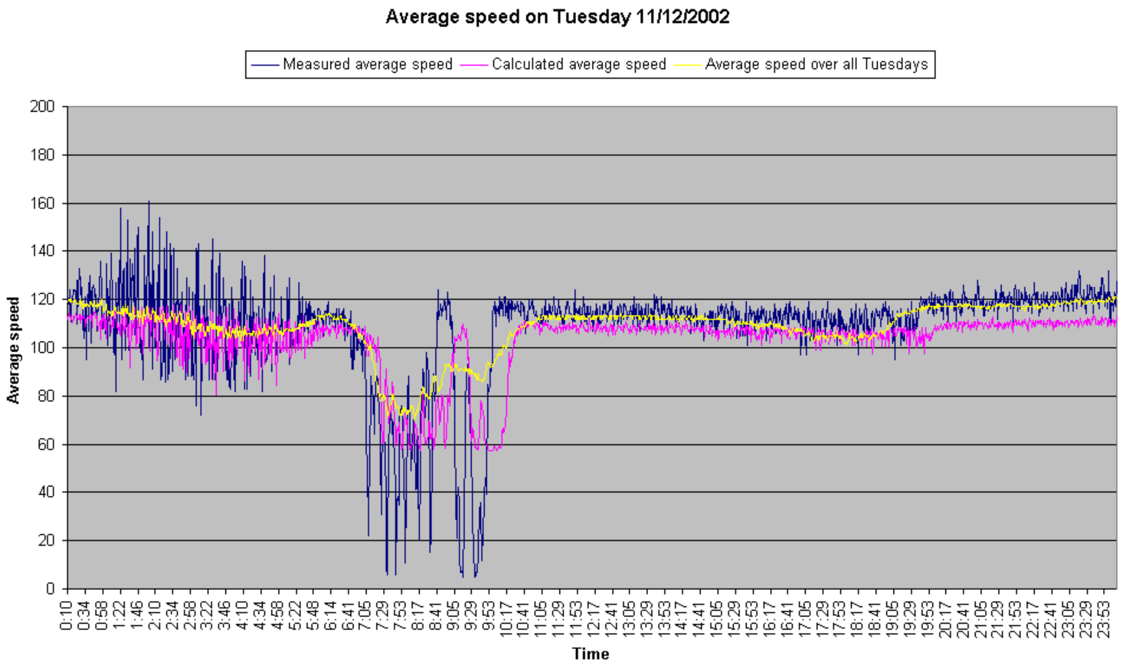


Figure C.27 The average speed on 12th of November 2002 near Gouda with the 5-1-4-topology 20 minutes ahead

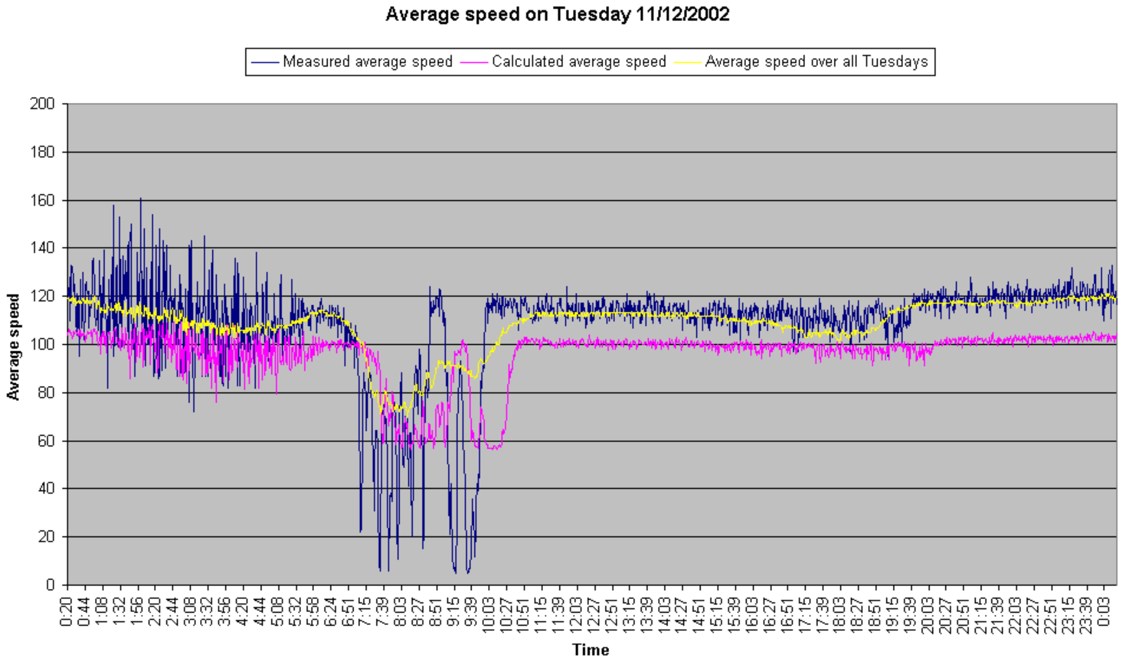


Figure C.28 The average speed on 12th of November 2002 near Gouda with the 5-1-4-topology 30 minutes ahead

C.8 The 5-2-4-topology

C.8.1 Zoetermeer

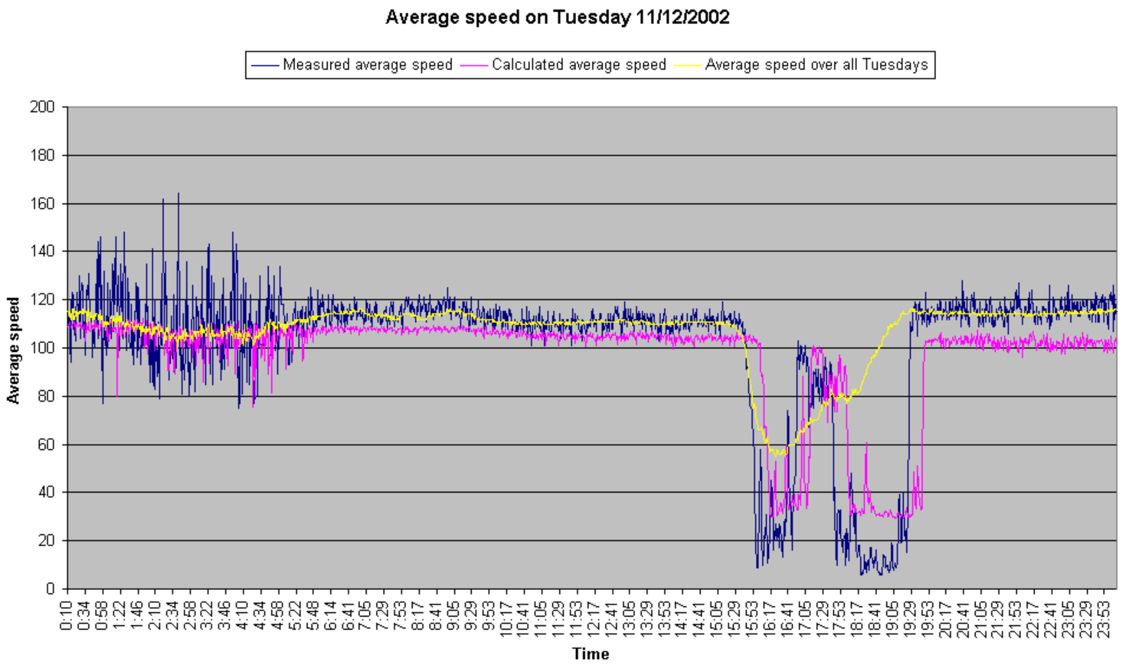


Figure C.29 The average speed on 12th of November 2002 near Zoetermeer with the 5-2-4-topology 20 minutes ahead

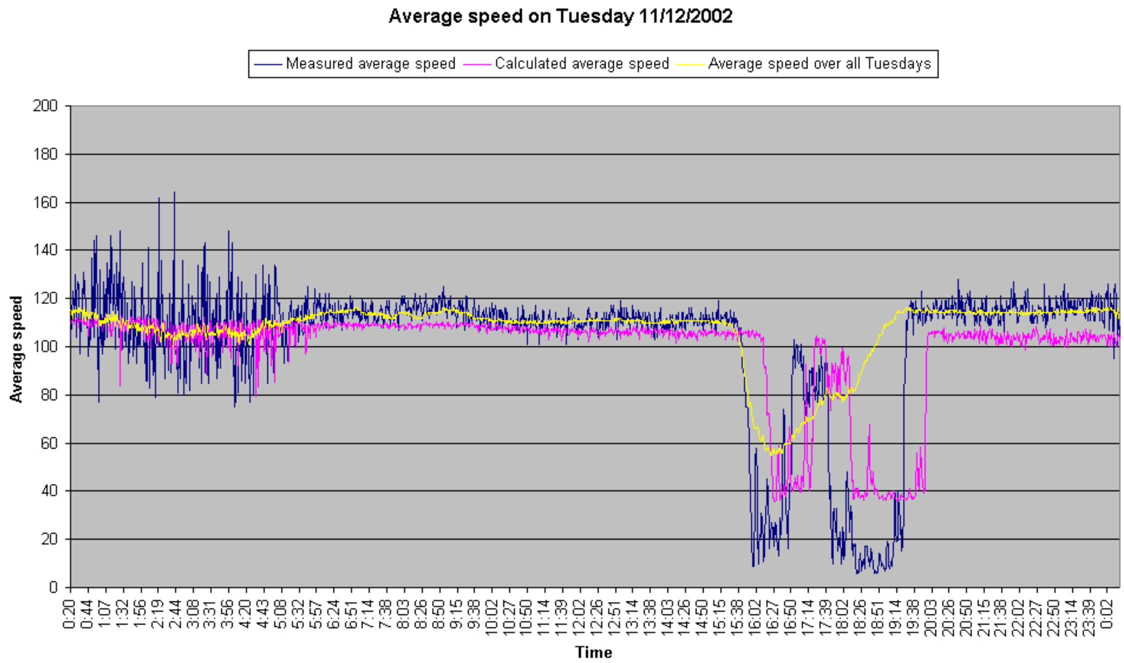


Figure C.30 The average speed on 12th of November 2002 near Zoetermeer with the 5-2-4-topology 30 minutes ahead

C.8.2 Gouda

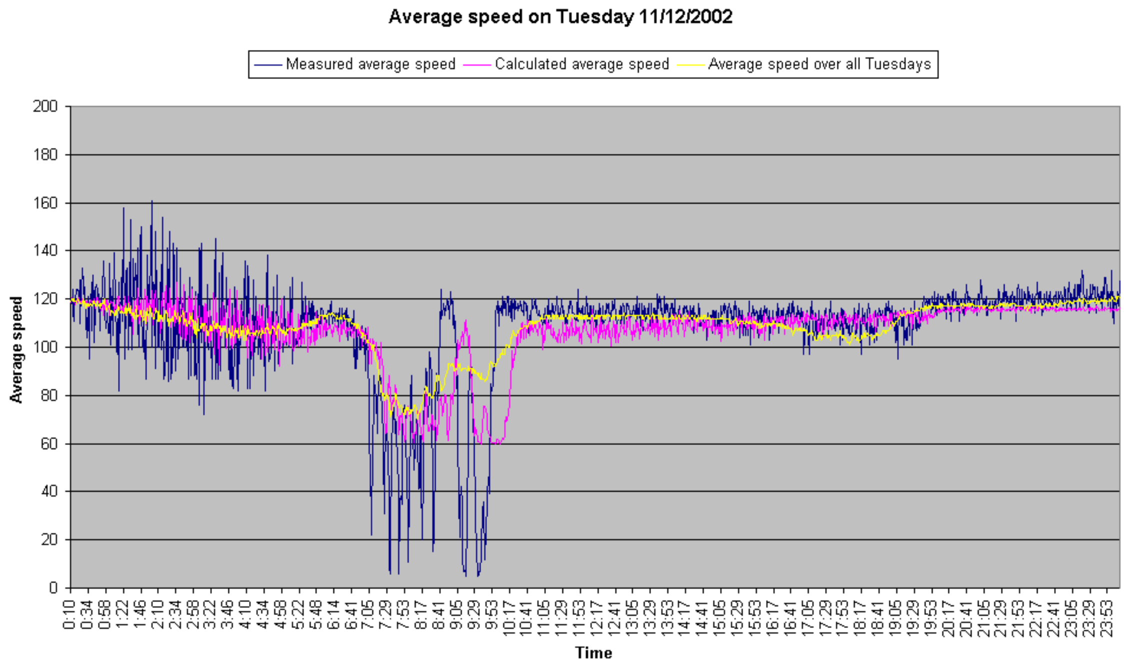


Figure C.31 The average speed on 12th of November 2002 near Gouda with the 5-2-4-topology 20 minutes ahead

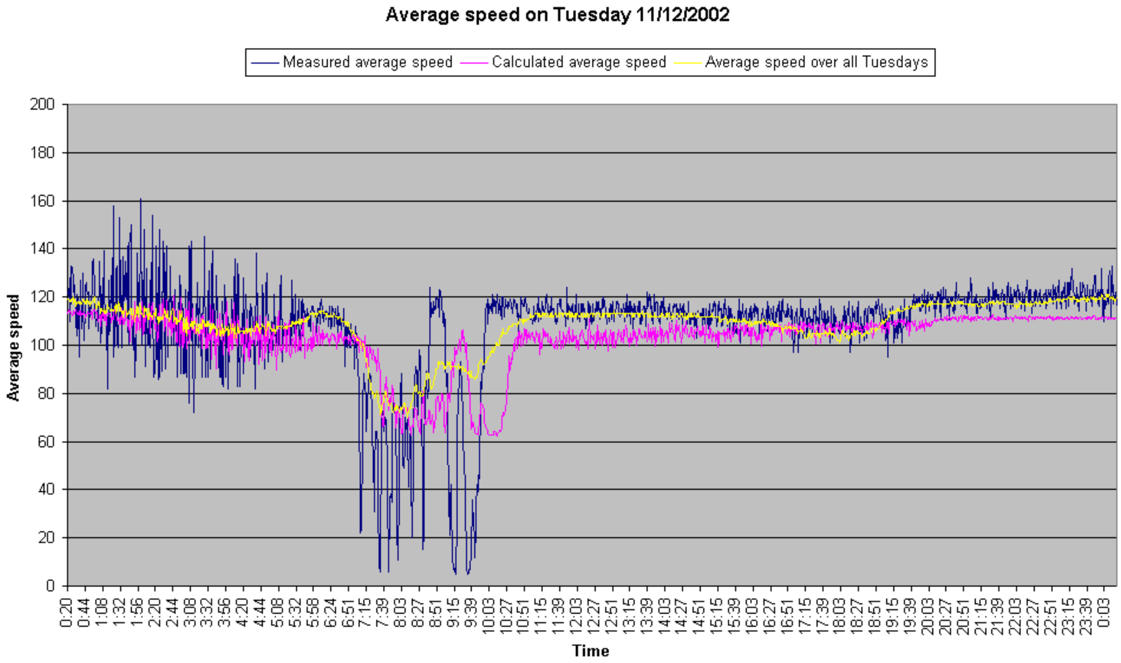


Figure C.32 The average speed on 12th of November 2002 near Gouda with the 5-2-4-topology 30 minutes ahead

C.9 The 5-4-4-topology

C.9.1 Zoetermeer

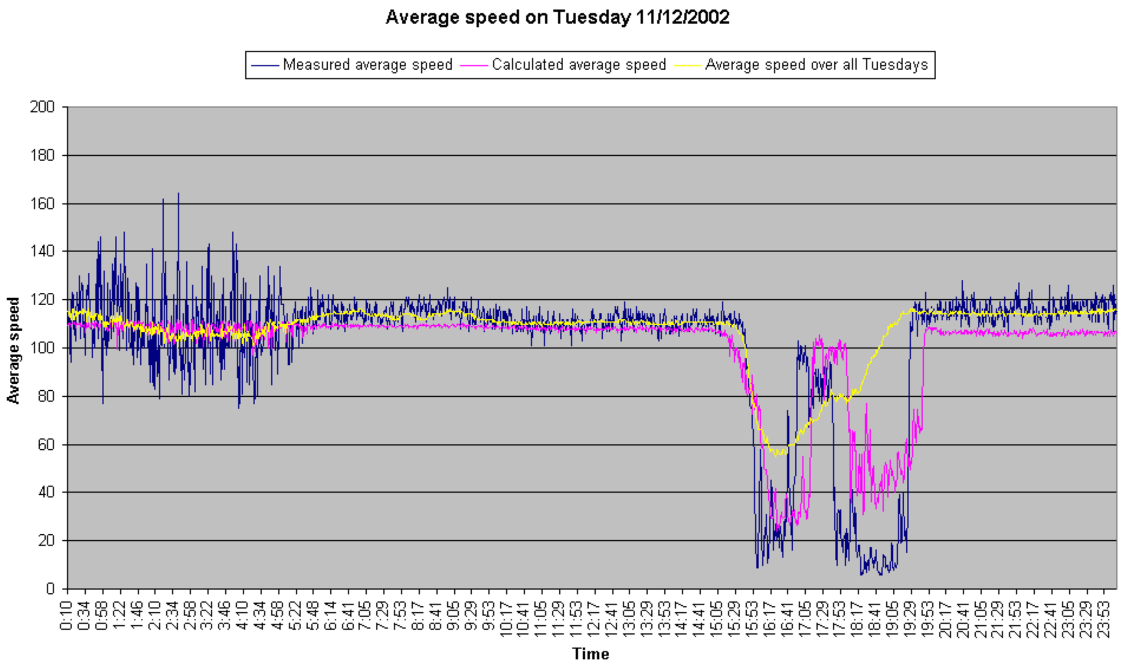


Figure C.33 The average speed on 12th of November 2002 near Zoetermeer with the 5-4-4-topology 20 minutes ahead

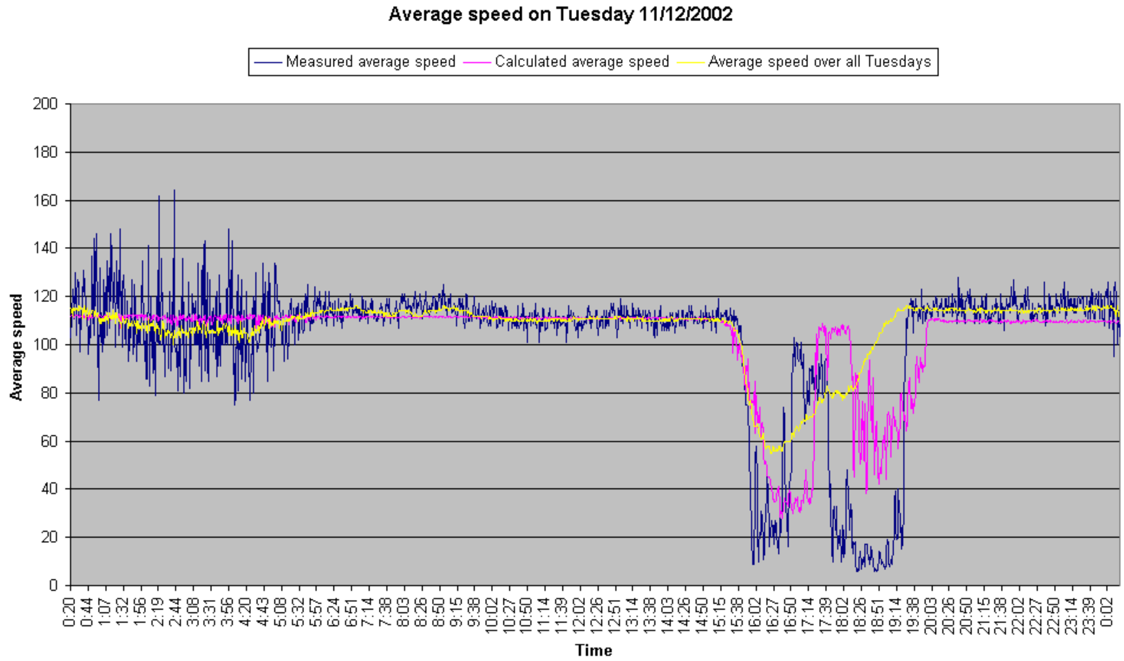


Figure C.34 The average speed on 12th of November 2002 near Zoetermeer with the 5-4-4-topology 30 minutes ahead

C.9.2 Gouda

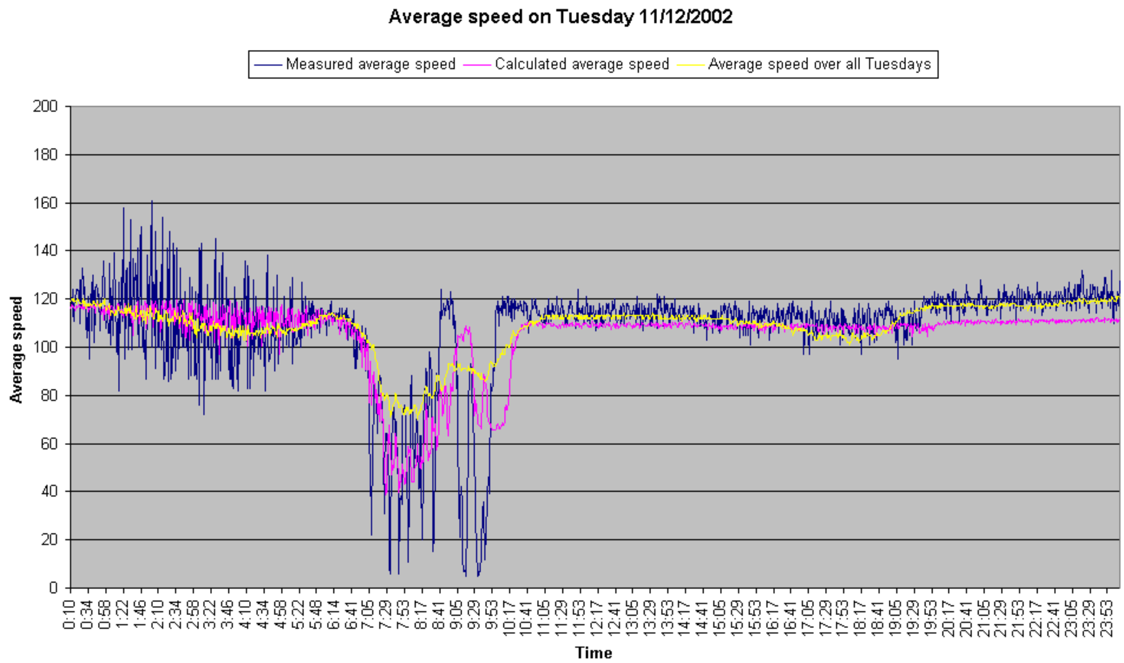


Figure C.35 The average speed on 12th of November 2002 near Gouda with the 5-4-4-topology 20 minutes ahead

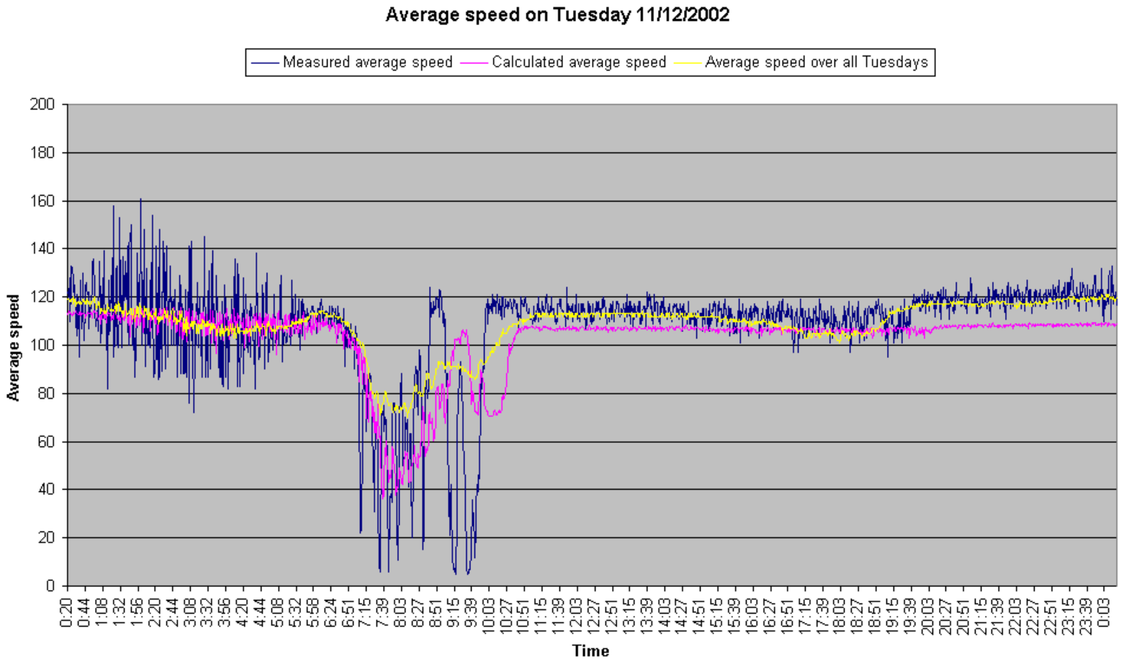


Figure C.36 The average speed on 12th of November 2002 near Gouda with the 5-4-4-topology 30 minutes ahead

C.10 The 5-5-4-topology

C.10.1 Zoetermeer

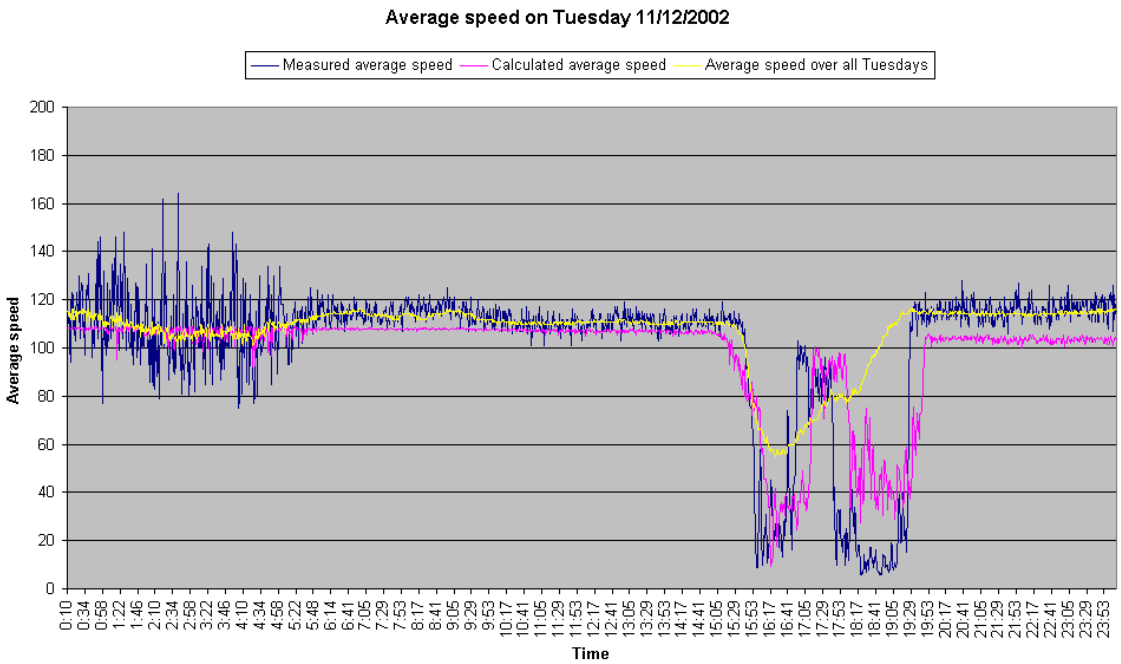


Figure C.37 The average speed on 12th of November 2002 near Zoetermeer with the 5-5-4-topology 20 minutes ahead

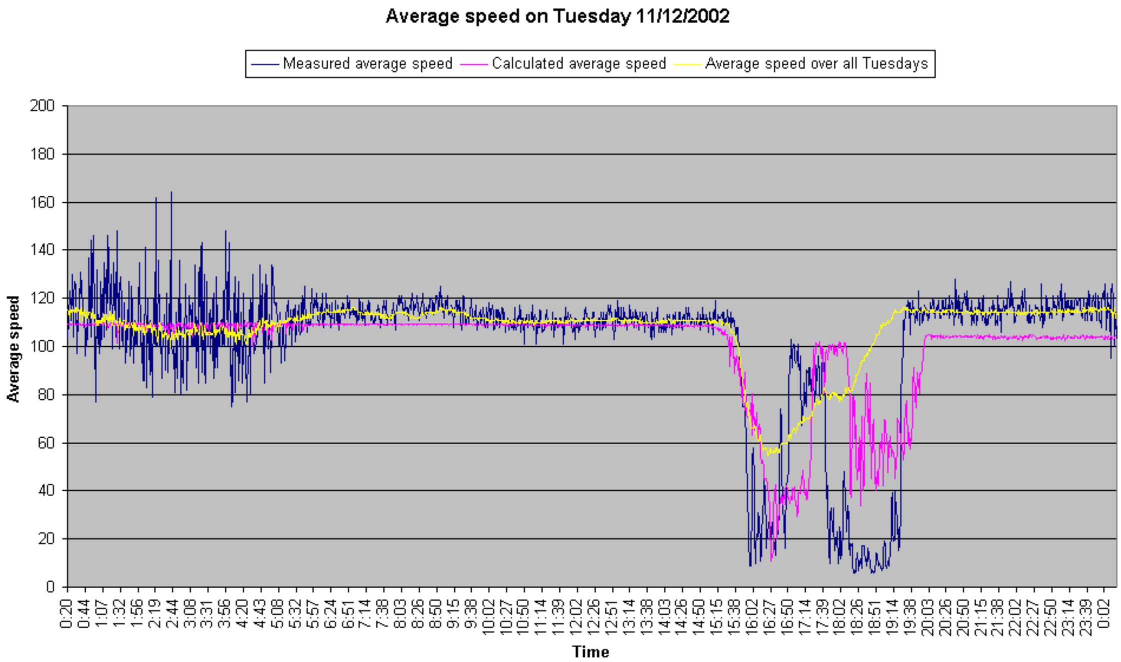


Figure C.38 The average speed on 12th of November 2002 near Zoetermeer with the 5-5-4-topology 30 minutes ahead

C.10.2 Gouda

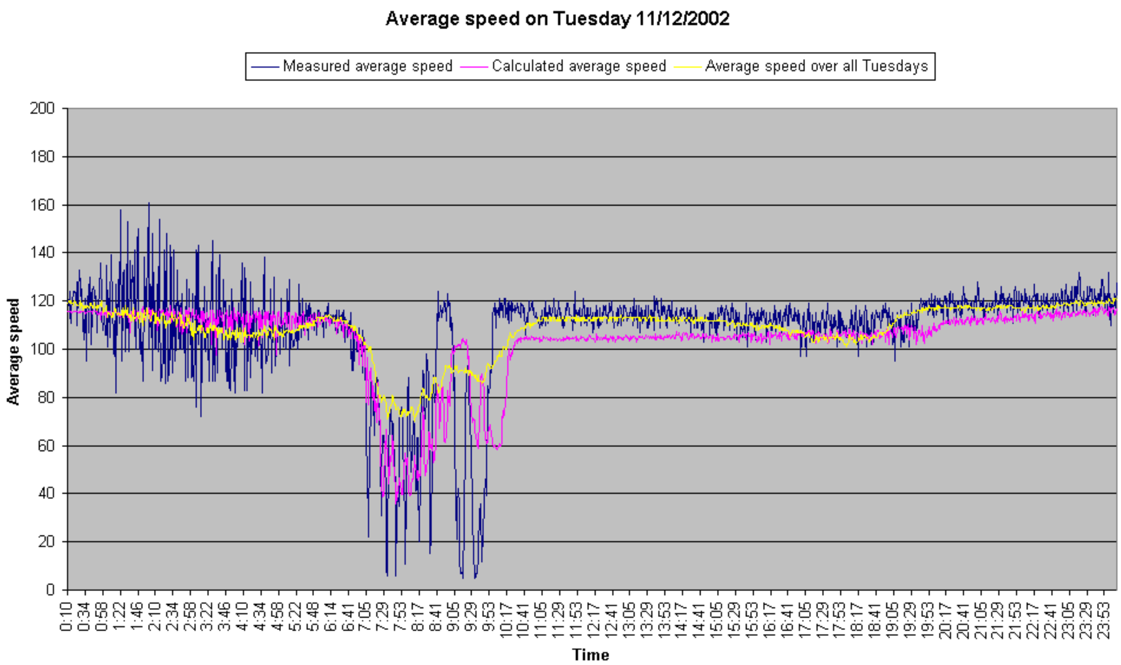


Figure C.39 The average speed on 12th of November 2002 near Gouda with the 5-5-4-topology 20 minutes ahead

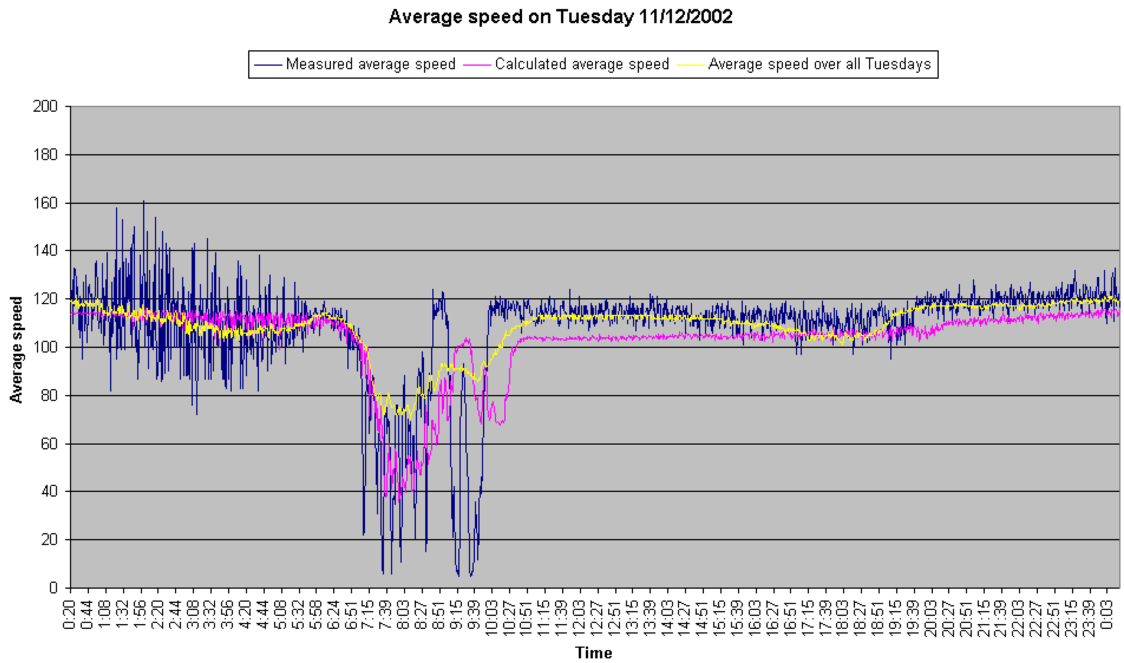


Figure C.40 The average speed on 12th of November 2002 near Gouda with the 5-5-4-topology 30 minutes ahead

C.11 The 5-8-4-topology

C.11.1 Zoetermeer

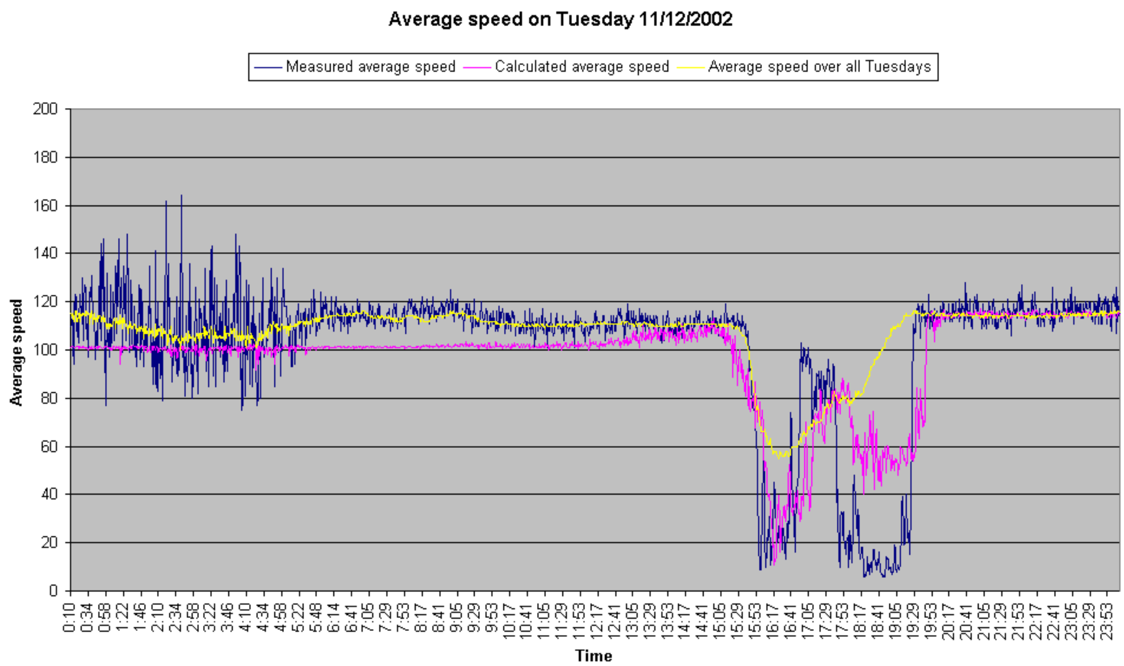


Figure C.41 The average speed on 12th of November 2002 near Zoetermeer with the 5-8-4-topology 20 minutes ahead

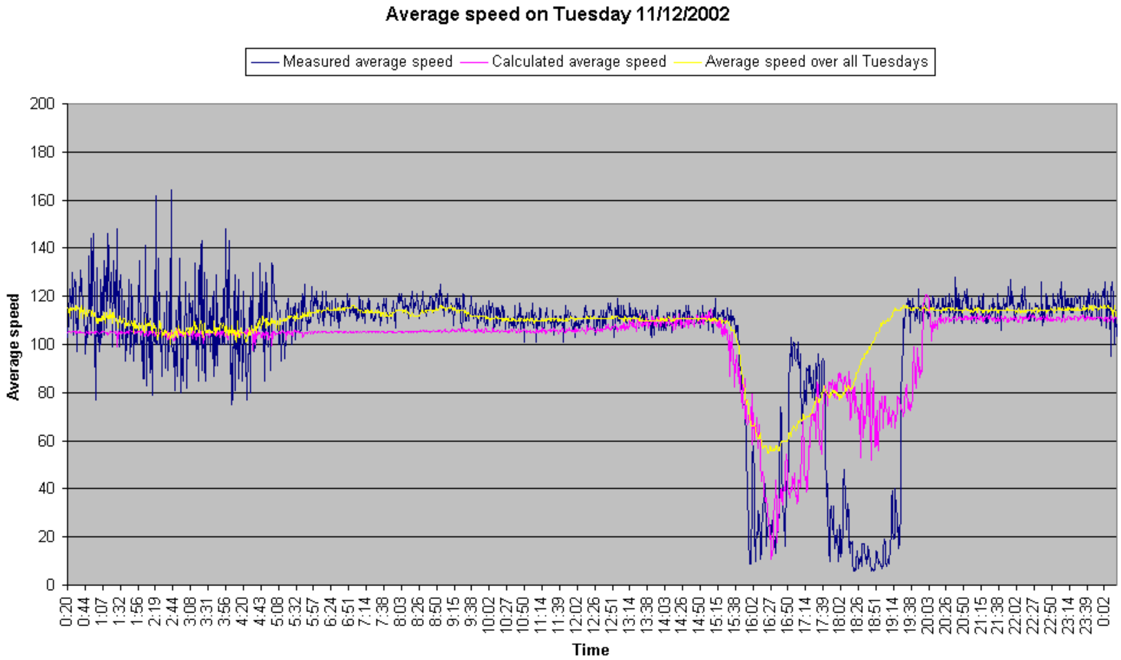


Figure C.42 The average speed on 12th of November 2002 near Zoetermeer with the 5-8-4-topology 30 minutes ahead

C.11.2 Gouda

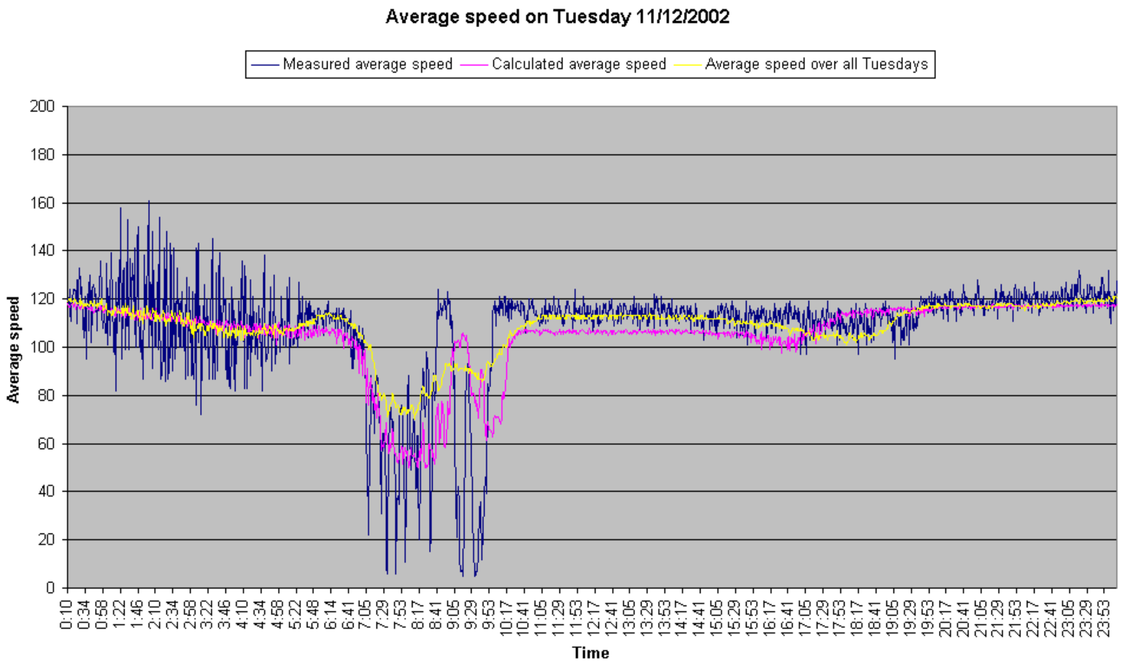


Figure C.43 The average speed on 12th of November 2002 near Gouda with the 5-8-4-topology 20 minutes ahead

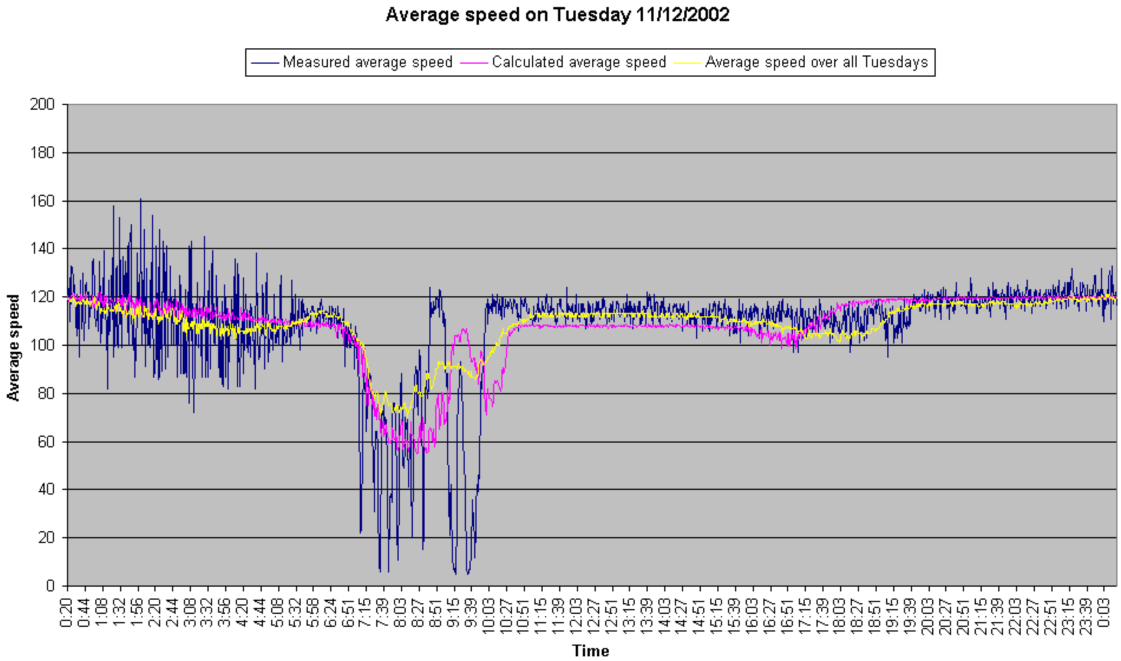


Figure C.44 The average speed on 12th of November 2002 near Gouda with the 5-8-4-topology 30 minutes ahead

C.12 The 5-9-4-topology

C.12.1 Zoetermeer

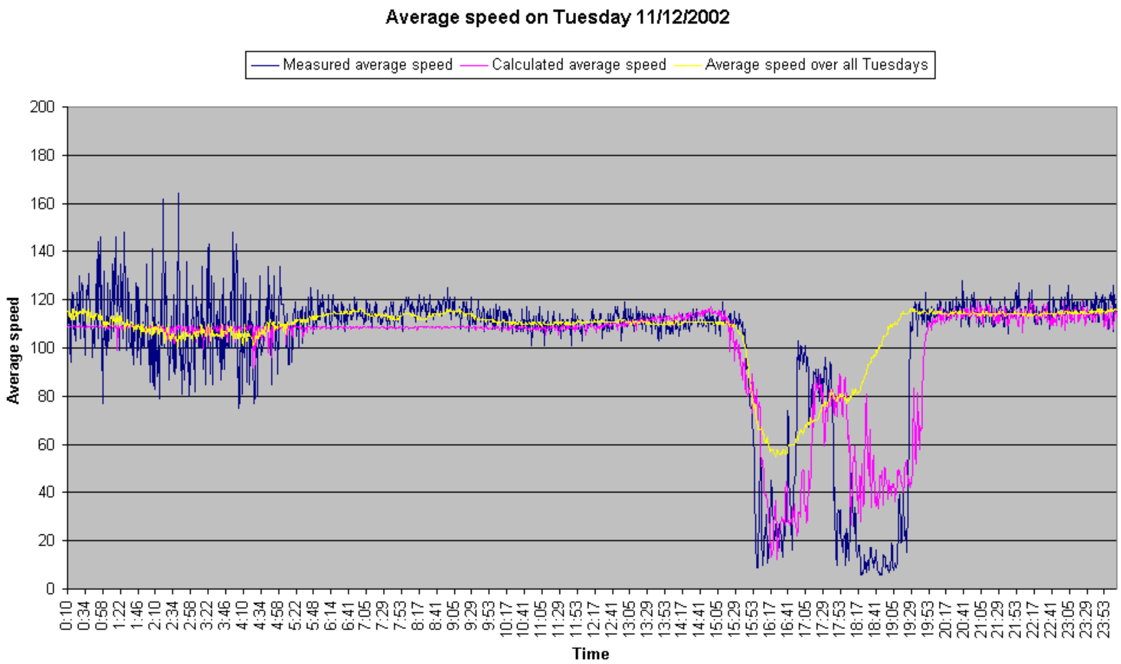


Figure C.45 The average speed on 12th of November 2002 near Zoetermeer with the 5-9-4-topology 20 minutes ahead

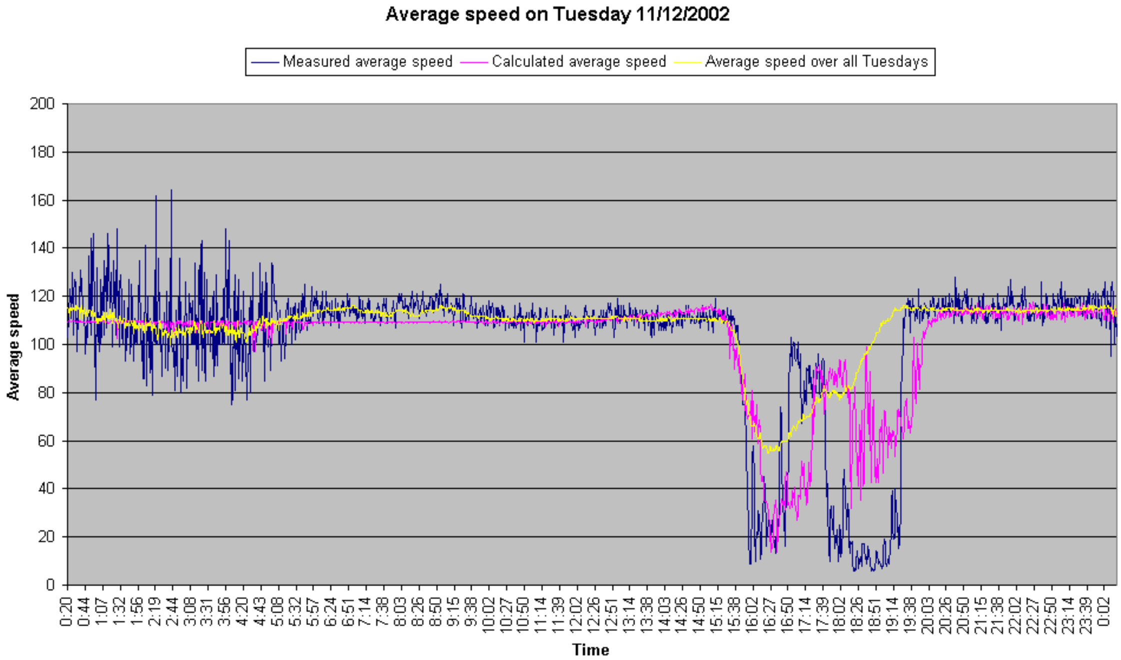


Figure C.46 The average speed on 12th of November 2002 near Zoetermeer with the 5-9-4-topology 30 minutes ahead

C.12.2 Gouda

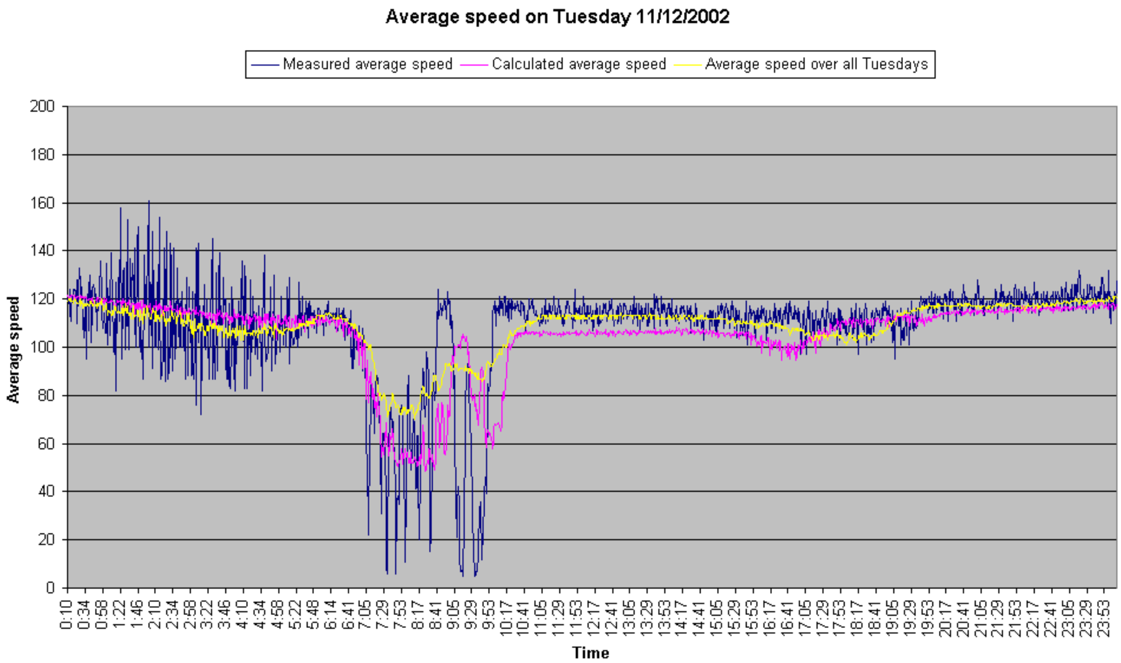


Figure C.47 The average speed on 12th of November 2002 near Gouda with the 5-9-4-topology 20 minutes ahead

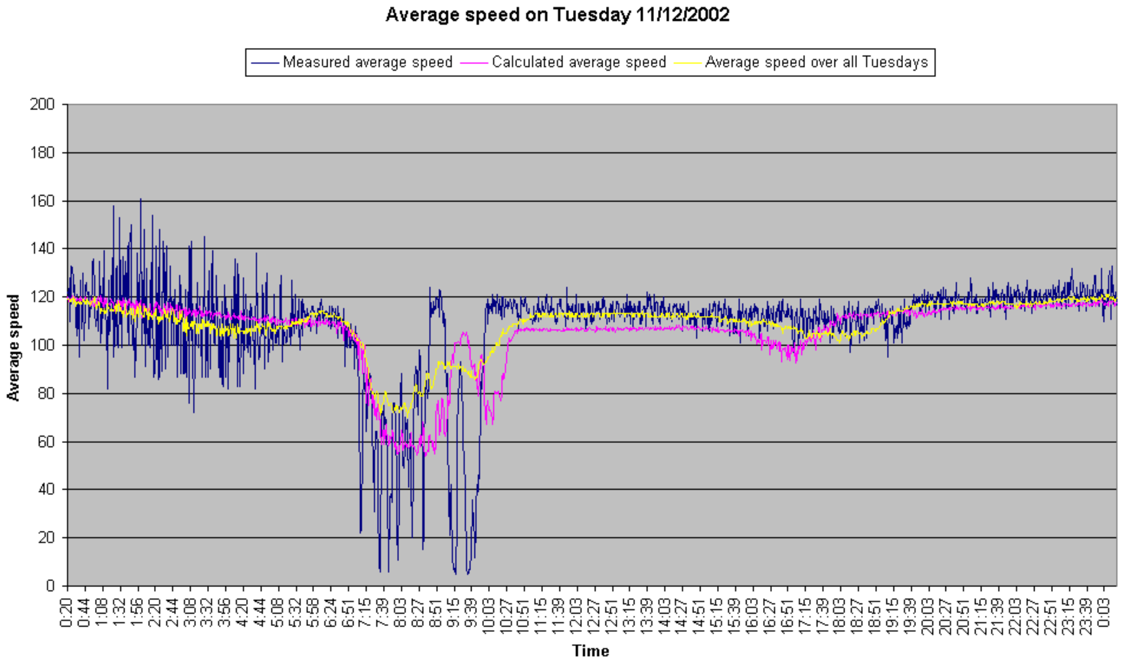


Figure C.48 The average speed on 12th of November 2002 near Gouda with the 5-9-4-topology 30 minutes ahead

C.13 The 5-10-4-topology

C.13.1 Zoetermeer

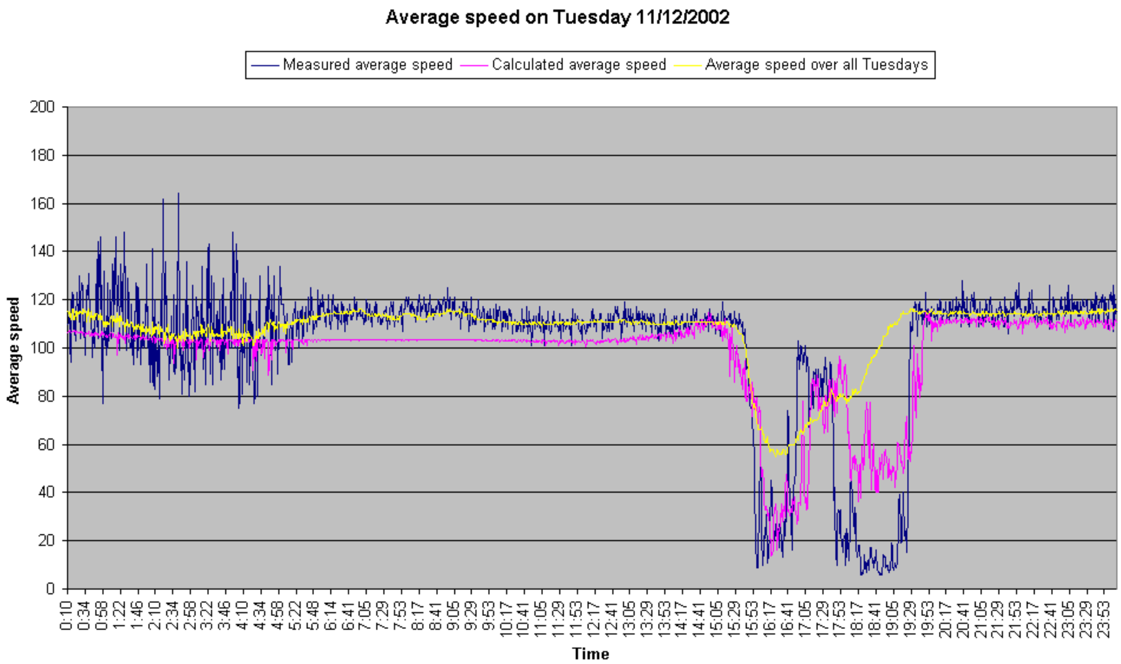


Figure C.49 The average speed on 12th of November 2002 near Zoetermeer with the 5-10-4-topology 20 minutes ahead

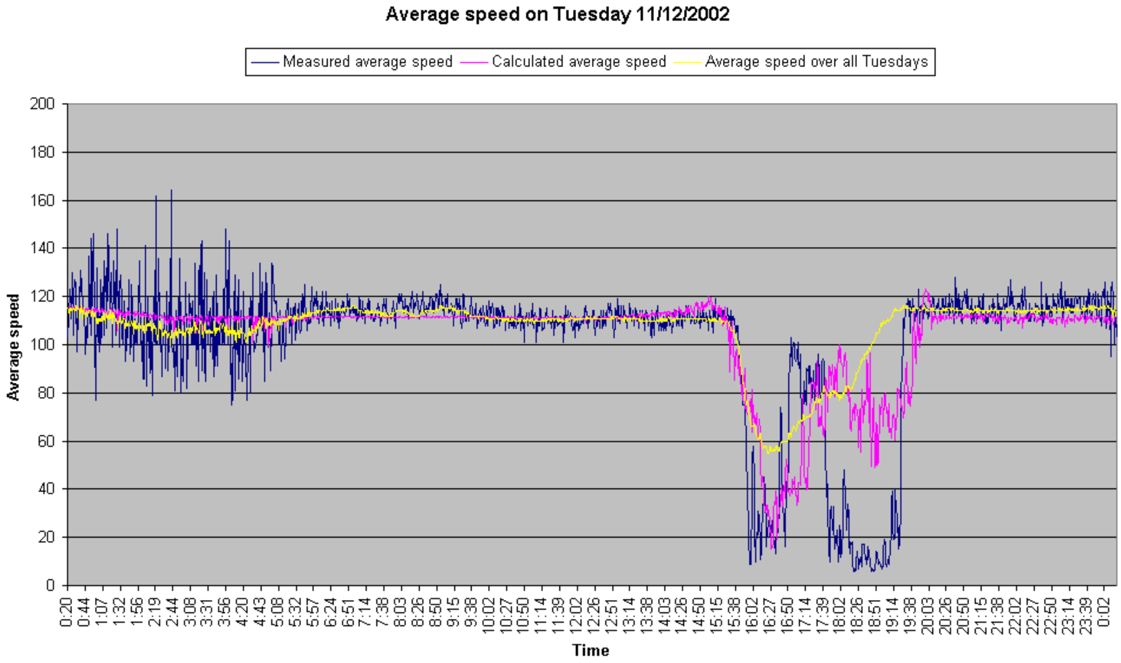


Figure C.50 The average speed on 12th of November 2002 near Zoetermeer with the 5-10-4-topology 30 minutes ahead

C.13.2 Gouda

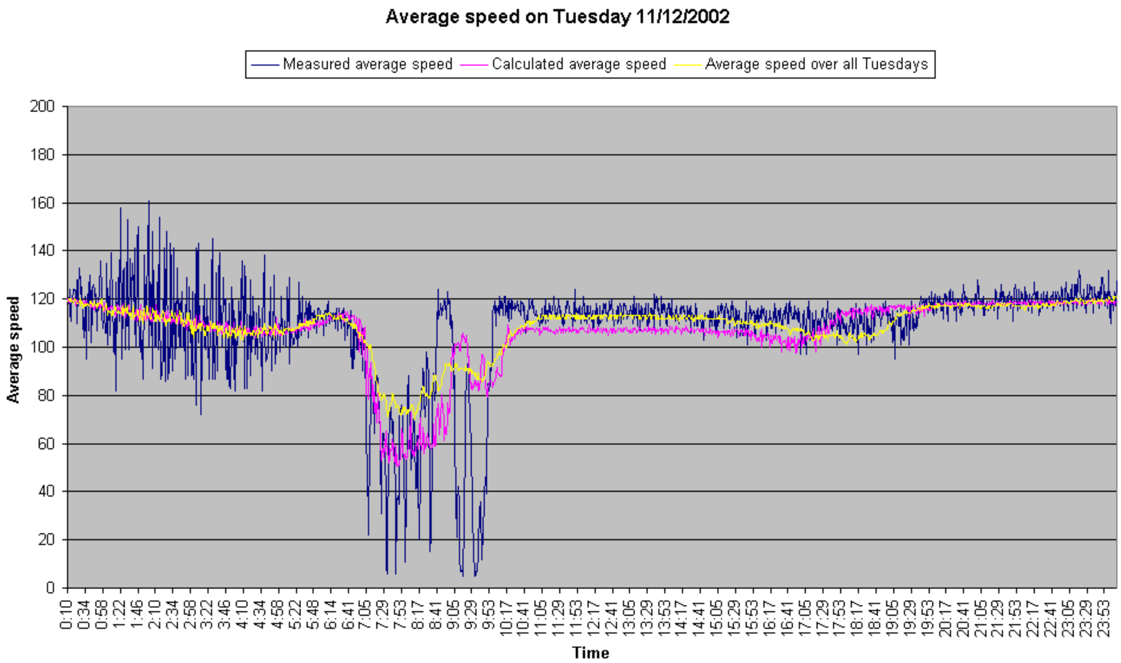


Figure C.51 The average speed on 12th of November 2002 near Gouda with the 5-10-4-topology 20 minutes ahead

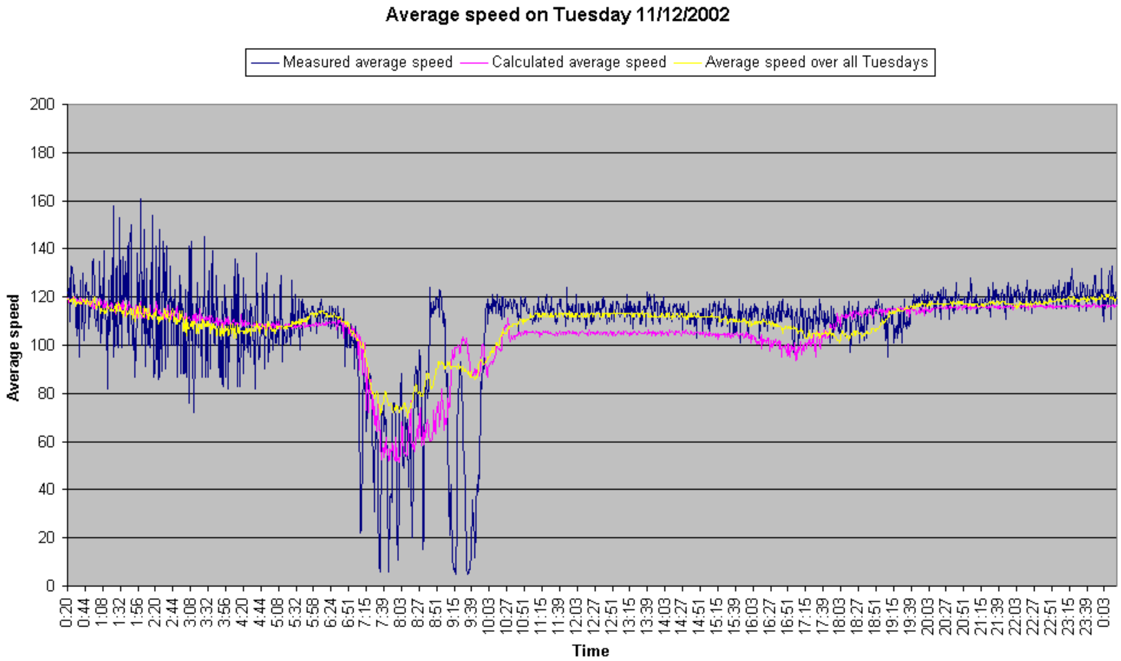


Figure C.52 The average speed on 12th of November 2002 near Gouda with the 5-10-4-topology 30 minutes ahead

C.14 The 5-11-4-topology

C.14.1 Zoetermeer

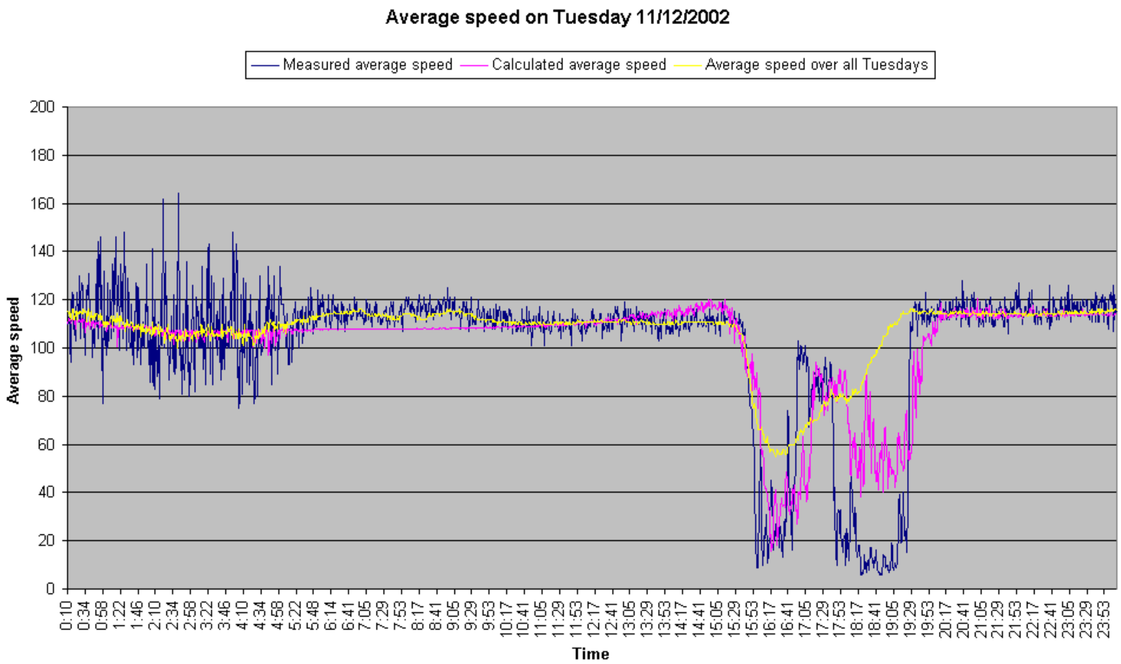


Figure C.53 The average speed on 12th of November 2002 near Zoetermeer with the 5-11-4-topology 20 minutes ahead

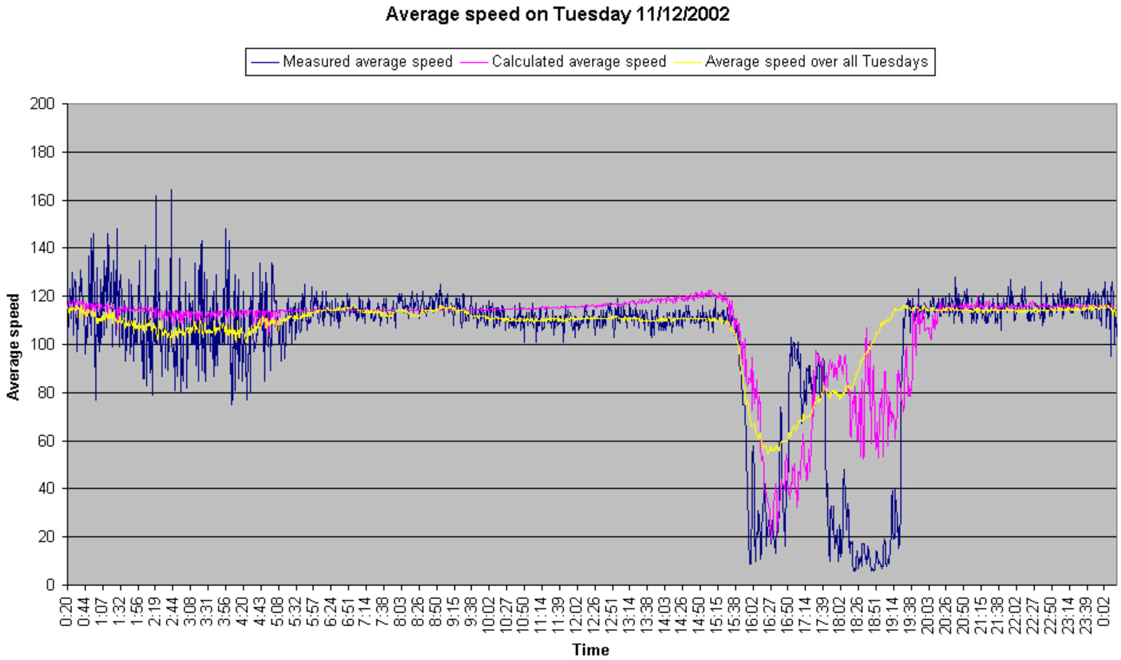


Figure C.54 The average speed on 12th of November 2002 near Zoetermeer with the 5-11-4-topology 30 minutes ahead

C.14.2 Gouda

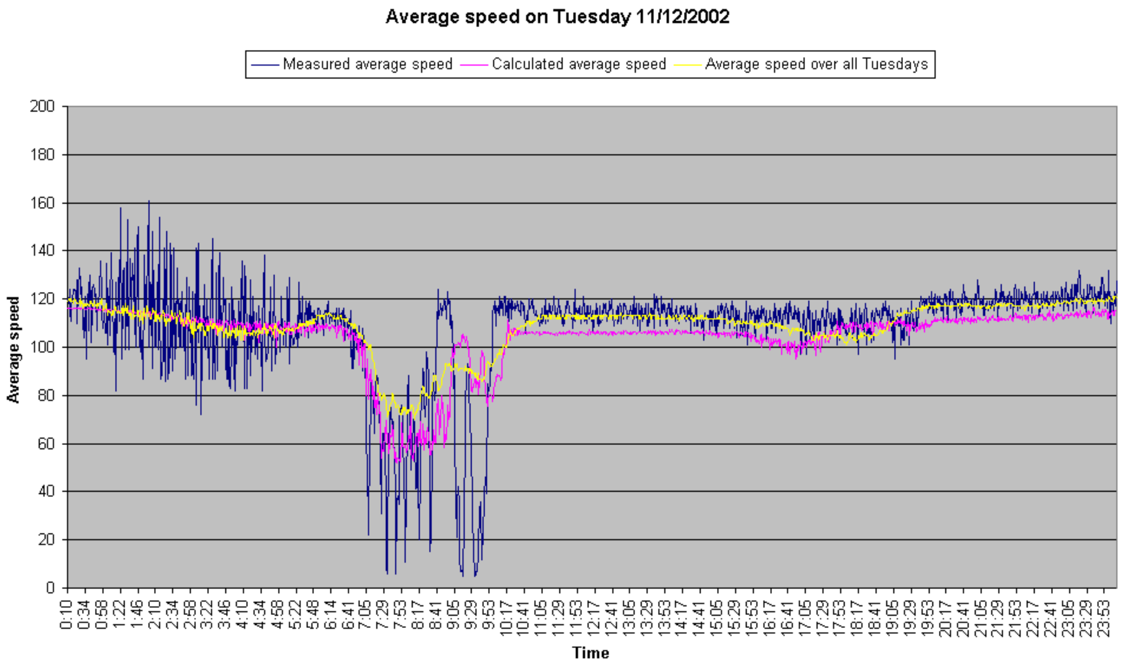


Figure C.55 The average speed on 12th of November 2002 near Gouda with the 5-11-4-topology 20 minutes ahead

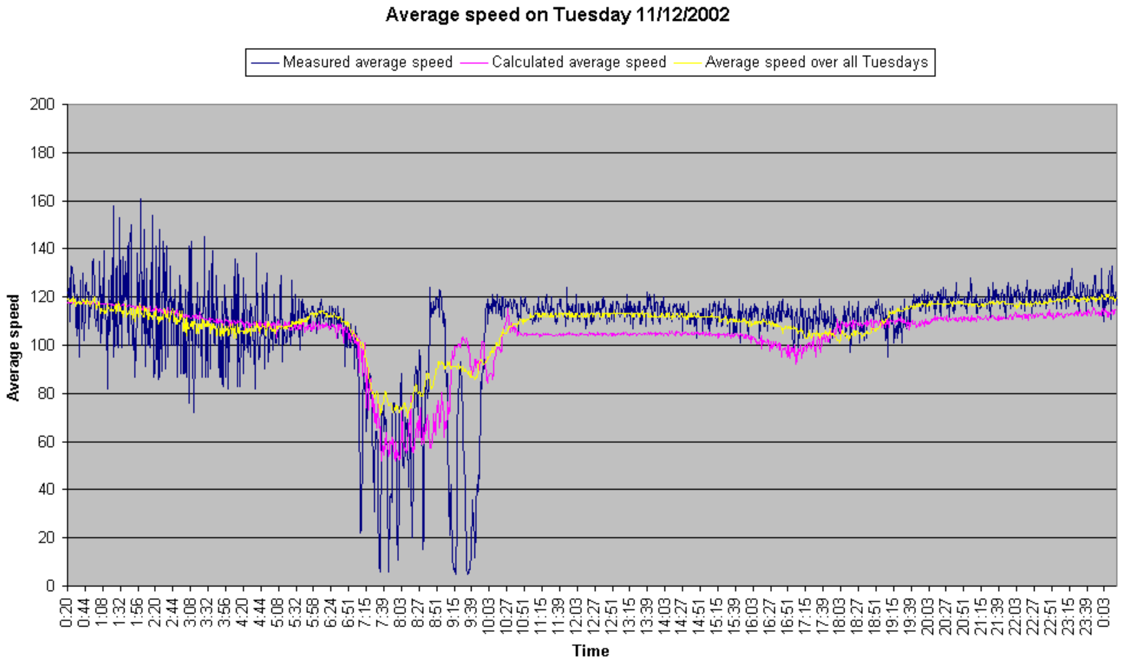


Figure C.56 The average speed on 12th of November 2002 near Gouda with the 5-11-4-topology 30 minutes ahead

C.15 The 5-15-4-topology

C.15.1 Zoetermeer

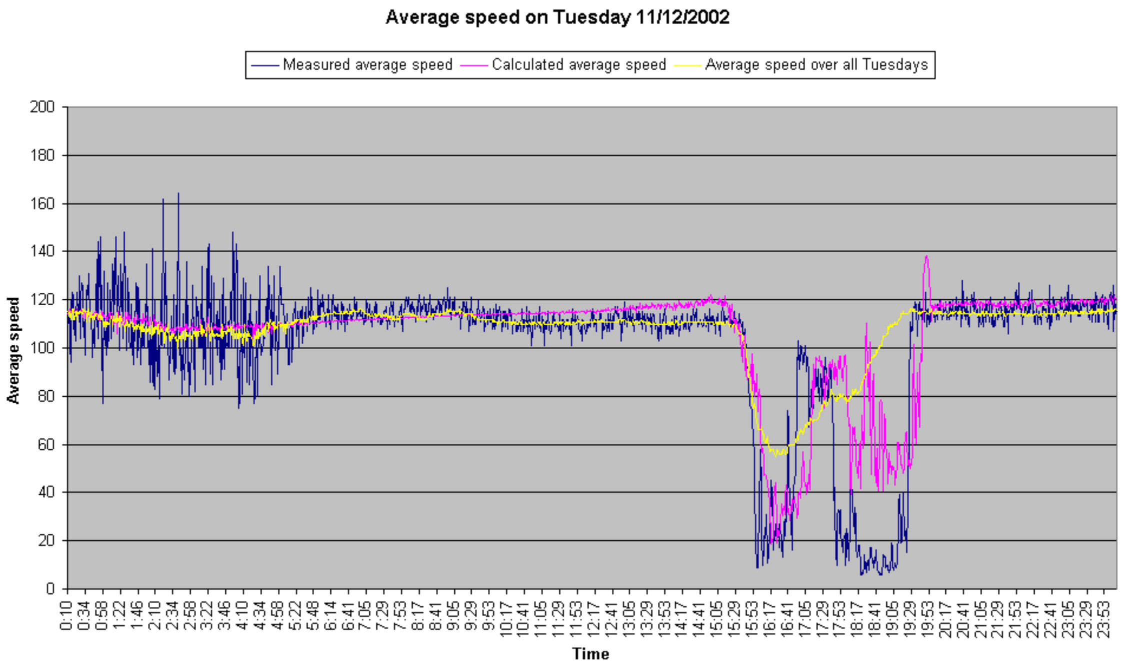


Figure C.57 The average speed on 12th of November 2002 near Zoetermeer with the 5-15-4-topology 20 minutes ahead

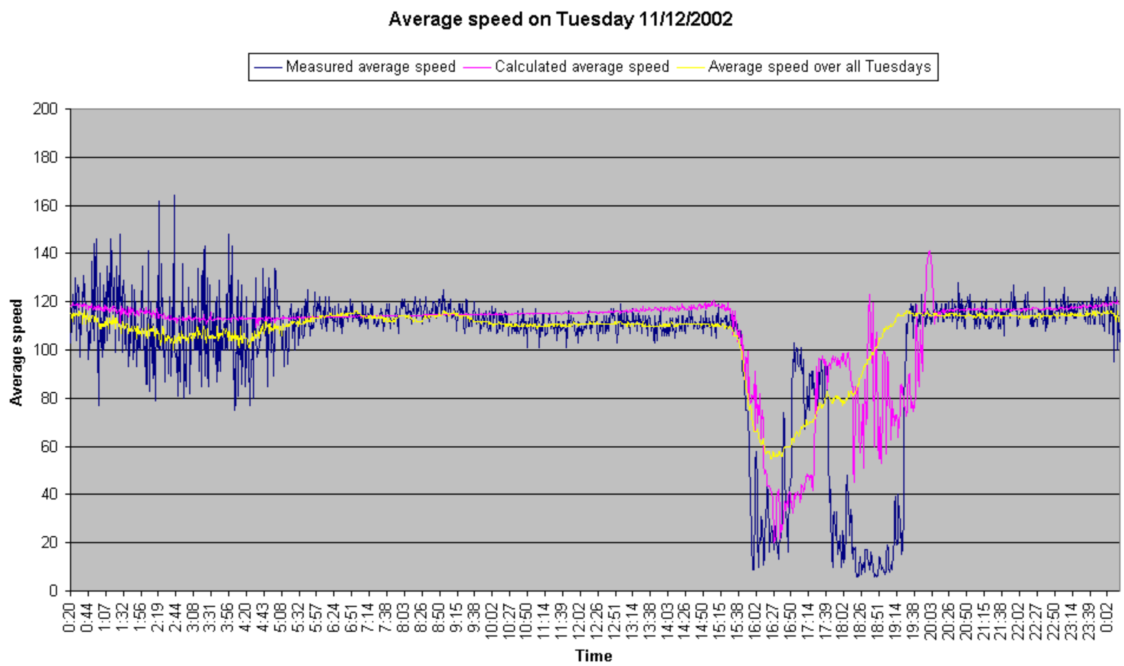


Figure C.58 The average speed on 12th of November 2002 near Zoetermeer with the 5-15-4-topology 30 minutes ahead

C.15.2 Gouda

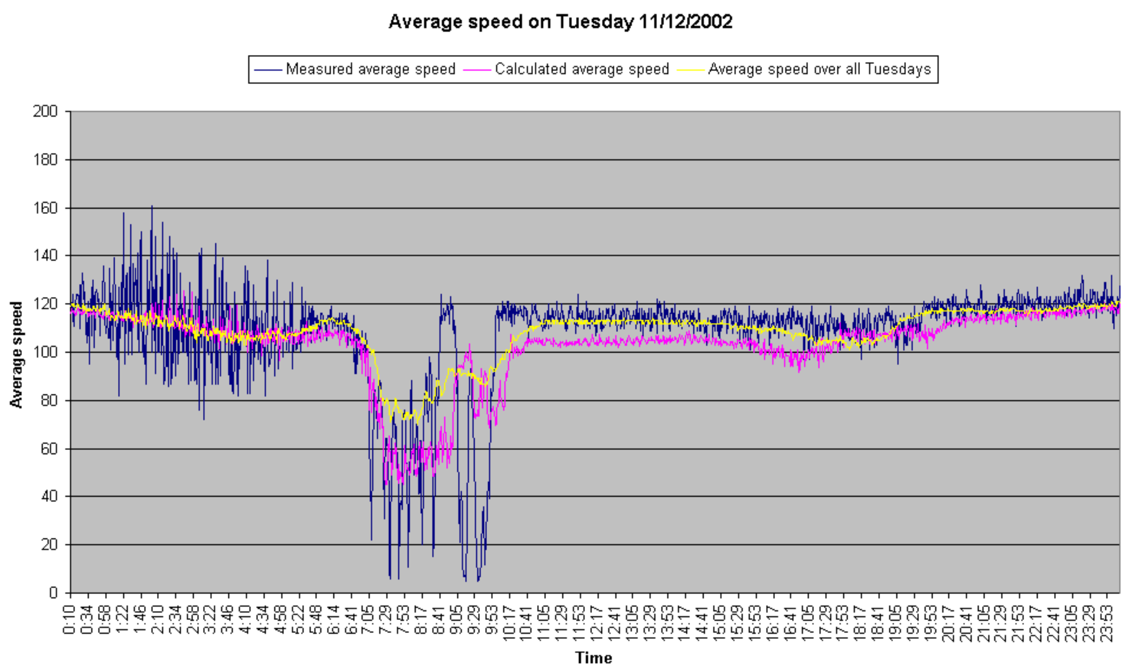


Figure C.59 The average speed on 12th of November 2002 near Gouda with the 5-15-4-topology 20 minutes ahead

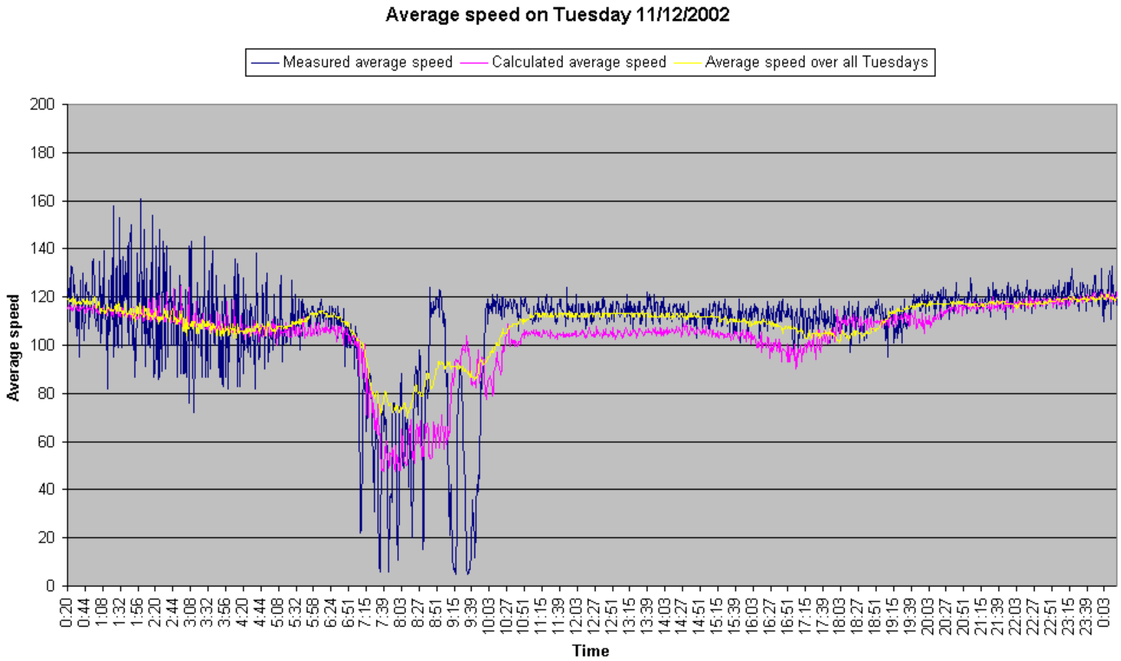


Figure C.60 The average speed on 12th of November 2002 near Gouda with the 5-15-4-topology 30 minutes ahead