

# **Theory and Application of Advanced Traffic Forecast Methods**

Vom Fachbereich Physik

der

Universität Duisburg-Essen (Campus Duisburg)

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigte Dissertation

von

Roland Chrobok

aus

Mülheim an der Ruhr

Referent: Prof. Dr. rer. nat. Michael Schreckenberg

Korreferent: Prof. Dr. rer. nat. Kai Nagel

Tag der mündlichen Prüfung: 11. Juli 2005



# Meinen Eltern



# Abstract

In the present work the problem of forecasting traffic states is investigated. Different traffic forecast algorithms are developed and applied in an advanced traffic information system.

Therefore, first an overview about existing traffic forecast algorithms is given. Four basic approaches can be distinguished: Parametric regression, nonparametric regression, neural networks, and heuristics. The results of the former works show, that a comparison between these methods is very difficult because in the works very different data are used.

To decide what is basically essential for a traffic forecast algorithm, the data of 4,480 inductive loop detectors of the motorway network of North Rhine-Westphalia are empirically analysed for the period from 2000/10/05 until 2004/09/31. Because there are many sources for possible measurement errors, several methods are proposed to avoid misinterpretations. The basic traffic data that are measured with the loop detectors are the traffic flow, velocity, and the occupancy. Thereby, for flow and velocity a distinction between lorries and passenger cars is possible.

In statistically analysing this data in detail it becomes clear, that it strongly depends on the forecast horizon which method works best. For horizons of a few minutes parametric regressions of the most recently measured traffic data perform best. With a growing forecast horizon it becomes more and more important to use for the forecast the knowledge about the causes that have an effect on the traffic at the particular time.

In order to put this knowledge in the forecast algorithm, a cluster method is proposed leading to classified days with which the forecast traffic time series can be calculated. This classified traffic time series is then combined with parametric and nonparametric regression techniques that consider the most recently measured data to enhance the forecast results for different forecast horizons.

Finally, an application using the new traffic forecast algorithms is presented. Therefore, the forecast data is combined with a spatial temporal traffic model to transform the point information of the loop detectors into a line information a network wide traffic state is consisting of. This can be seen in the traffic information system OLSIM, that offers among other information a 30 and a 60 minute traffic forecast.



# Zusammenfassung

Im Rahmen der vorliegenden Arbeit wird das Problem der Verkehrsprognose untersucht. Verschiedene Prognosealgorithmen werden entwickelt und in einem Verkehrsinformationssystem zur Anwendung gebracht.

Hierzu wird zunächst ein Überblick über bestehende Verkehrsprognosealgorithmen gegeben. Vier grundsätzliche Ansätze können unterschieden werden: parametrisierte Regression, Regression ohne Parameter, neuronale Netze und Heuristiken. Eine Analyse der früheren Arbeiten zeigt, dass ein objektiver Vergleich der Methoden schwierig ist, da in jeder Arbeit jeweils sehr unterschiedliche Verkehrsdaten benutzt werden.

Daher werden zunächst Daten von 4.480 Induktionsschleifen welche auf den Autobahnen in Nordrhein-Westfalen installiert sind im Zeitraum vom 2000/10/05 bis zum 2004/09/31 analysiert. Verschiedene Möglichkeiten werden entwickelt, um potentielle Messfehler zu vermeiden. Von den acht Datensätzen die übermittelt werden beinhalten der Verkehrsfluss, die Geschwindigkeit und die Belegung die grundsätzlichen Informationen. Hierbei können der Fluss und die Geschwindigkeit nach Pkw und Lkw unterschieden werden.

Eine umfangreiche statistische Analyse der Daten zeigt klar, dass die richtige Methode stark vom Prognosehorizont abhängt. Für Horizonte von einigen wenigen Minuten liefern parametrisierte Regressionen aus den zuletzt gemessenen Daten die besten Resultate. Mit wachsendem Prognosehorizont wird es wichtiger, Ursachen welche eine Wirkung auf den Verkehrszustand haben können in der Prognose zu berücksichtigen.

Um dieses Wissen über Ursache und Wirkung in den Prognosealgorithmus einfließen zu lassen, wird eine Cluster Methode benutzt. Diese hat klassifizierte Tage zum Ergebnis, aus denen sich Prognoseganglinien berechnen lassen. Diese klassifizierten Prognoseganglinien werden mit Regressionsmodellen mit und ohne Parameter verknüpft, um Wissen über den aktuellen Verkehrszustand in das Prognosemodell einfließen zu lassen und die Prognoseresultate für verschiedene Prognosehorizonte zu verbessern.

Schließlich wird eine Anwendung der neuen Prognosealgorithmen vorgestellt. Hierzu werden die Prognoseredaten der Induktionsschleifen mit einem räumlich zeitlichen Verkehrsmodell verknüpft, um aus den Punktdaten einen netzwerkweiten Verkehrszustand zu berechnen. Dieser kann schließlich im Verkehrsinformationssystem OLSIM als 30 oder 60 Minuten Prognose abgerufen werden.



# Contents

<b>1</b>	<b>Preface</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Outline . . . . .	3
<b>2</b>	<b>Basics</b>	<b>5</b>
2.1	Measured Observables . . . . .	5
2.2	Statistical Concepts . . . . .	6
2.3	Quality Measures . . . . .	8
<b>3</b>	<b>Methods of Forecasting Traffic: State of the Art</b>	<b>11</b>
3.1	Parametric Regression . . . . .	12
3.1.1	Introduction to ARIMA Modelling . . . . .	12
3.1.2	Special Cases of Parametric Regression . . . . .	15
3.1.3	ARIMA Modelling for Traffic Forecast . . . . .	17
3.1.4	Kalman Filter . . . . .	20
3.2	Nonparametric Regression . . . . .	24
3.2.1	Introduction to Phase Space Embedding . . . . .	24
3.2.2	Nonparametric Regression for Traffic Forecast . . . . .	26
3.3	Neural Networks . . . . .	32
3.3.1	Introduction to Neural Networks . . . . .	32
3.3.2	Neural Networks for Traffic Forecast . . . . .	35
3.4	Heuristics and Knowledge Based Systems . . . . .	38
3.5	Combined Methods . . . . .	48
3.5.1	Combining Short and Long Term Forecasts . . . . .	48
3.5.2	Other Combinations . . . . .	56
3.6	Comparisons . . . . .	57
<b>4</b>	<b>Traffic Data of Inductive Loop Detectors</b>	<b>63</b>
4.1	Acquisition of Traffic Data . . . . .	63
4.2	Measurement Errors . . . . .	69
4.2.1	Missing Data . . . . .	69
4.2.2	Errors Detected by the Sub Central . . . . .	70
4.2.3	Rollback Error . . . . .	72
4.2.4	Obvious Errors . . . . .	74

4.2.5	Roundoff Errors . . . . .	75
4.2.6	Scaling Errors . . . . .	76
<b>5</b>	<b>Empirical Data Analysis</b>	<b>83</b>
5.1	Secular Trend . . . . .	83
5.2	Seasonal Differences . . . . .	90
5.3	Daily Differences . . . . .	99
5.4	Special Events . . . . .	104
5.5	Random Noise . . . . .	114
5.6	Interpretation . . . . .	126
5.6.1	Data to Forecast . . . . .	126
5.6.2	Methods Depending on the Forecast Horizon . . . . .	130
<b>6</b>	<b>New Forecast Algorithm</b>	<b>135</b>
6.1	Classification with Cluster Analysis . . . . .	135
6.1.1	<i>K</i> -Medoid Cluster Algorithm . . . . .	136
6.1.2	Clustering of Days According to Attributes . . . . .	138
6.2	Calculating the Forecast . . . . .	152
6.3	Considering Recently Measured Data . . . . .	160
6.4	Forecast Results . . . . .	166
<b>7</b>	<b>Application</b>	<b>173</b>
7.1	Traffic Information System OLSIM: Motivation and Background .	173
7.2	Principle . . . . .	174
7.3	Applied Forecast Model . . . . .	175
7.4	Spatial Temporal Traffic Model . . . . .	176
7.5	Network Topology . . . . .	176
7.6	Visualisation . . . . .	178
<b>8</b>	<b>Conclusions and Outlook</b>	<b>181</b>
8.1	Conclusions . . . . .	181
8.2	Outlook . . . . .	183
<b>Abbreviations</b>		<b>187</b>
<b>Erklärung</b>		<b>189</b>
<b>Curriculum Vitae</b>		<b>191</b>
<b>Acknowledgements</b>		<b>193</b>
<b>Bibliography</b>		<b>195</b>

# Chapter 1

## Preface

“Prediction is difficult, especially the future.”

**Niels Bohr**  
(1885 – 1962), Winner of the 1922 Nobel Prize in physics

Frequent road traffic congestion is a global issue and causes significant economic damage. Especially in densely populated regions the capacity of the road network is often at its limits and it is hardly possible or even socially untenable to extend the road network. Many proposed solutions to this problem like *dynamic traffic management* or *advanced traffic information systems* need traffic forecasts or at least approximations about future traffic states. In the following an introduction to the problem of forecasting traffic states is given.

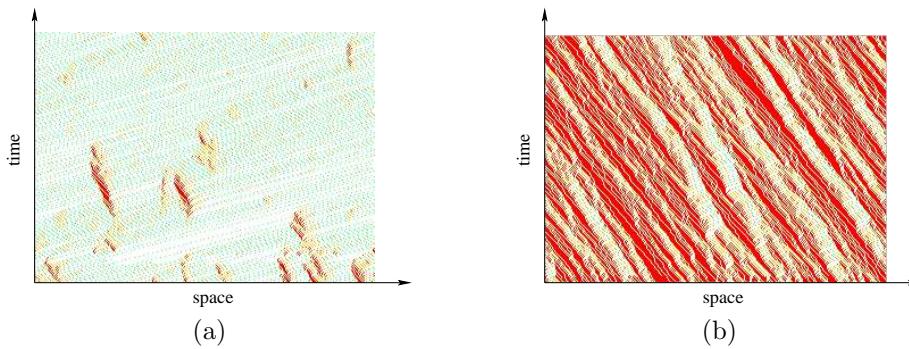
### 1.1 Introduction

From the beginning of the history the knowledge about the future is a value that humanity has been looking for, but for millenia, no feasible methods to make statements about the future could be found. To handle this difficulty, resourceful persons, at this times mostly priests, evolved methods of ambiguity to a high standard. A very popular example is the Oracle at Delphi, which told the king Kroisos of Lydia (~ 595 – 547 B.C.) if he crossed the river, a great kingdom would be destroyed. The kingdom that finally was destroyed was his own, so the Oracle had been right. The “quality” of the first forecast methods was improved with the range of possible interpretations. Even today many people dispute the content of the innumerable biblical prophecies or the verses of Nostradamus (1503 – 1566, actually Michele de Nostredame), the most famous prophet in the middle ages. Nowadays, the evolution of modern physics provides feasible statements about the future in many fields. Essential is the understanding of the environment and the development of a physical model. With this model and a given initial state (for instance, the present), in certain limits future states can be estimated. The limits can be caused by the finite accuracy with which the model maps the reality,

as well as by a finite predictability in general. A famous example of this comes with quantum mechanics.

Thus, physical models are also developed for a better understanding and for forecasts of traffic states. As easily can be seen from the extensive literature (see, for instance, [Fukui et al., 2003, Helbing et al., 2000, Schreckenberg and Wolf, 1998, Wolf et al., 1996]), the main focus is to investigate the dynamics of spatial temporal traffic models. Thereby it is no rareness, that the investigations are only done from a theoretical point of view and without any empirical data. It should not be forgotten that the validation of the model with the reality is a crucial task of every physical investigation.

A class of models that are often validated with different kinds of empirical data are based on the cellular automaton model of Nagel and Schreckenberg [1992] (Nagel-Schreckenberg model). This model with its enhancements during the recent years offer a powerful tool for short term traffic forecasts in that sense, that a vehicle trajectory can be estimated from a statistical point of view. An example of trajectories of vehicles simulated with the Nagel-Schreckenberg model on a single lane roundabout with periodic boundary conditions can be seen in Fig. 1.1 (a) and (b) for different densities.



**Fig. 1.1.** Space time plot of the cellular automaton traffic model according to [Nagel and Schreckenberg, 1992] with relative densities of (a) 0.15 and (b) 0.52. Shown are trajectories of vehicles on a single lane roundabout (periodic boundary conditions) that are green for high velocities and red for low velocities. Whereas the spatial temporal traffic model makes estimates about the dispersion of the traffic congestions, the basic parameter that influences the presence or absence of a congestion is the traffic demand that has to be externally set up using the density. The task of the present work is to make estimates about this external variable.

Whereas such traffic models provide information about the spatial temporal behaviour of vehicles, they are not able to make statements about the general traffic demand. In the example of Fig. 1.1 this demand has to be adjusted externally using the density. But as can be seen, information about the general traffic demand is very essential. With a relative density of 0.15 in Fig. 1.1 (a) there are

only a few vehicles with a low velocity (red), with a relative density of 0.52 the whole roundabout is congested as can be seen in Fig. 1.1 (b).

In reality there is of course no single lane roundabout with periodic boundary conditions. In fact there are many different sections between traffic lights, on- or off-ramps, or intersections that act as *sources* or *sinks*, that means that permanently vehicles appear or disappear. It depends on the general volume of traffic that differs on timescales that are larger than the time a vehicle needs to pass the particular section. Thus, the general task is to forecast the traffic data without up- or downstream traffic data.

Forecasting traffic data with and without measured up- or downstream information is in former works not always clearly distinguished. Thereby, it is a completely different task. With upstream information a traffic model can approximate the trajectory of a vehicle. With this trajectory the time the vehicle is arriving at a downstream section can be estimated. In a similar manner with downstream information the time a congestion is arriving at an upstream section can be approximated.

Without up- or downstream information future traffic has to be estimated with models that use other information. The development of these models is the basic task of this work.

For the application, the traffic information system OLSIM (**OnLine Traffic SIMulation**) that is presented at the end of this work, the methods of forecasting the data of the loop detectors are combined with a spatial temporal traffic model to obtain information about the overall traffic state in a section. This makes sense for two reasons. First, in reality loop detectors or any kind of detecting devices are not installed all over the road network but at certain points. A realistic simulation is the only way to obtain information about the traffic state apart from this points. Secondly, the simulation provides better results on very short timescales because it also uses up- and downstream information.

Before an application is ready for operation a lot of work has to be done. An outline of this thesis is presented in the following.

## 1.2 Outline

In Chapter 2 an introduction to the basics of the physical background of traffic engineering, basic statistics, and some quality measures is given. The next chapter gives an overview about former works that investigate the problem of traffic forecasting. The four basic approaches that can be distinguished are proposed as well as combinations and comparisons. Chapter 4 presents the acquisition of the empirical minute aggregated data of about 4,500 loop detectors that are used in this work. A very important point that is unfortunately often neglected is the handling of errors in measurement coming automatically in such mass data. An extensive discussion how many of them can be detected and avoided is the main focus of this chapter. In Chapter 5 these data are statistically analysed in regard

to different effects on different timescales. For the secular trend even another data source is used. Chapter 6 describes the new forecast method that provides solutions to open questions that appear in former works. An application that uses the new results is presented in Chapter 7. Finally, the conclusions and an outlook are given in Chapter 8.

# Chapter 2

## Basics

“Mathematik ist eine Bedingung aller exakten Erkenntnis.”

**Immanuel Kant**  
(1724 – 1804), German philosopher

This chapter gives an introduction to the basics of the physical background of traffic engineering, basic statistics, and quality measures. The focus lies on those terms that are frequently used in this work. More special techniques and algorithms developed in this work are described in the particular chapter itself.

### 2.1 Measured Observables

Many people are quite common with expressions like “free flow”, “traffic congestion”, or “stop-and-go”. To quantify traffic states we need certain measurable observables that describe the traffic. Those that are needed for this work are proposed in the following.

The traffic flow  $J$  gives the number  $N$  of vehicles that pass a cross section in a given time interval  $T$ :

$$J = \frac{N}{T}. \quad (2.1)$$

The density  $\rho$  is the number of vehicles  $N$  that are present in a certain part of the road  $\Delta x$ :

$$\rho = \frac{N}{\Delta x}. \quad (2.2)$$

The velocity of a single vehicle  $n$  is the fraction of the distance  $\Delta x$  passed by a vehicle and the travel time  $\Delta t$  the vehicle needs to pass  $\Delta x$ :

$$v_n = \frac{\Delta x}{\Delta t}. \quad (2.3)$$

To be precise this is the mean velocity for the distance  $\Delta x$ . To get the current velocity  $\Delta x$  must be chosen infinitesimal small.

The mean velocity for  $N$  vehicles can be calculated using Eq. 2.8 accordingly. The traffic flow  $J$  of  $N$  vehicles, their mean velocity  $\bar{v}$ , and the density  $\rho$  are connected with the well known hydrodynamic relationship

$$J = \rho \bar{v}, \quad (2.4)$$

that holds, if a homogeneous traffic state is assumed. In traffic theory the flow-density relationship is called the *fundamental diagram*.

The occupancy  $\rho_{\text{rel}}$  is a measurement for how long a part of the road with the length  $L$  is covered by a vehicle  $n$  with length  $l_n$  and a velocity of  $v_n$ :

$$\rho_{\text{rel}} = \frac{1}{T} \sum_{t \leq t_n \leq t+T} \Delta t_n = \frac{1}{T} \sum_{t \leq t_n \leq t+T} \frac{L + l_n}{v_n}, \quad (2.5)$$

thereby is  $\Delta t_n$  the time the part of the road is covered in the time interval  $T$ . The occupancy can be related to the density under the assumption that the occupancy is proportional to it. The observations of [Koshi et al., 1981] show, that this is nearly the case. An occupancy of one means that vehicles are standing bumper to bumper which leads to a density  $\rho_{\text{max}}$  of

$$\rho_{\text{max}} = \frac{N}{\sum_{n=1}^N l_n} = \frac{1}{\bar{l}}, \quad (2.6)$$

with the mean length  $\bar{l}$  of  $N$  vehicles.

Then, the relation between the occupancy and the density is quite easy:

$$\rho = \rho_{\text{rel}} \rho_{\text{max}}. \quad (2.7)$$

Two approximations for  $\rho_{\text{max}}$  are  $\rho_{\text{max}} \approx 133\frac{1}{3}$  veh/min [Nagel and Schreckenberg, 1992] or  $\rho_{\text{max}} \approx 140$  veh/min [Neubert, 2000].

## 2.2 Statistical Concepts

The statistical analysis of the loop detector data is the basis of this work. In the following a short introduction is given for the most common methods and terms. Note that all of the relations are estimations of the proper terms that hold for discrete and finite time series. To go more into detail the reader is referred, for instance, to [Clarke and Cooke, 1978, Gottinger, 1980, Sachs, 1984, Terrell, 1999]. Given  $N$  measurement values  $x_1, x_2, \dots, x_N$  the arithmetic mean or average  $\bar{x}$  is calculated with

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n. \quad (2.8)$$

In particular the numerical mean is an approximation of the expected value  $E[x]$ . The variance  $\sigma^2(x)$  is the second moment of the variables  $x$  about the mean  $\bar{x}$

$$\sigma^2(x) = \frac{1}{N-1} \sum_{n=1}^N [x_n - \bar{x}]^2. \quad (2.9)$$

Its square root, the standard deviation or dispersion  $\sigma(x)$ , is a measure of the average deviation of the measurements  $x$  from the mean  $\bar{x}$  or the estimated value:

$$\sigma(x) = \sqrt{\frac{1}{N-1} \sum_{n=1}^N [x_n - \bar{x}]^2}. \quad (2.10)$$

The covariance  $\sigma(xy)$  between two paired measurement values  $x$  and  $y$  represents a measure of the correlation between  $x$  and  $y$ . The covariance is

$$\sigma(xy) = \frac{\sum_{n=1}^N [x_n - \bar{x}] [y_n - \bar{y}]}{N-1} = \frac{\sum_{n=1}^N x_n y_n - N \bar{x} \bar{y}}{N-1}. \quad (2.11)$$

Note, that if  $x$  and  $y$  are independent hold

$$\bar{xy} = \bar{x}\bar{y} \text{ and } \sigma(xy) = 0. \quad (2.12)$$

The correlation coefficient  $r_{xy}$  is a standardized measure of the correlation between two paired variables  $x$  and  $y$ . This is found by the relation

$$r_{xy} = \frac{\sigma(xy)}{\sigma(x)\sigma(y)}. \quad (2.13)$$

The correlation coefficient always lies within the interval  $[-1,1]$ . From Eq. 2.12 directly follows that  $r_{xy} = 0$  if  $x$  and  $y$  are independent. If  $x$  and  $y$  both tend to increase simultaneously, then  $r_{xy} > 0$ , and if  $x$  increases when  $y$  tends to decrease, then  $r_{xy} < 0$ . A perfect relation between  $x$  and  $y$  occurs when  $r_{xy} = 1$  or  $r_{xy} = -1$ .

Note that the correlation coefficient is only a mathematical measure for the existence of correlation. But there can be different reasons for this. Two measurement rows  $x$  and  $y$  can be correlated because of the following physical reasons:

1.  $x$  influences  $y$ .
2.  $y$  influences  $x$ .
3.  $x$  and  $y$  are mutually influenced.
4.  $x$  and  $y$  are externally influenced.
5. The correlation is random.

## 2.3 Quality Measures

To evaluate the forecast a measure is needed to describe the performance. By the forecast of metric values the difference between the forecast value  $y_n$  and the measured value  $x_n$  is called the forecast error  $e_{\text{Forecast}} = y_n - x_n$  as well as the absolute difference is called the absolute error  $e_{\text{AE}} = |y_n - x_n|$ . To be more able to rate the deviation, often the relative error  $e_{\text{RE}} = \frac{y_n - x_n}{x_n}$  or the absolute relative error  $e_{\text{ARE}} = \frac{|y_n - x_n|}{x_n}$  are used. In technical works that are often given as percent error  $e_{\text{PE}} = e_{\text{RE}} \times 100\%$  or absolute percent error  $e_{\text{APE}} = e_{\text{ARE}} \times 100\%$ , respectively.

Because the result is strongly influenced by random fluctuations, a single measurement is not enough to rate the outcome. So several measurements calculating several forecast errors are needed to get statistical relevance. In a lot of works either the frequency of the error in a given interval or its mean is denoted.

Let  $x_1, x_2, \dots, x_N$  and  $y_1, y_2, \dots, y_N$  be two series of measured values to be compared. In the following the most common measures are described.

- The *mean absolute error* ( $e_{\text{MAE}}$ ) is the mean of all absolute differences:

$$e_{\text{MAE}} = \frac{1}{N} \sum_{n=1}^N |x_n - y_n|. \quad (2.14)$$

- The *mean square error* ( $e_{\text{MSE}}$ ) is the mean of all squared differences and is more sensible to large outliers:

$$e_{\text{MSE}} = \frac{1}{N} \sum_{n=1}^N [x_n - y_n]^2. \quad (2.15)$$

- The *root mean square error* ( $e_{\text{RMSE}}$ ) is the square root of the  $e_{\text{MSE}}$ :

$$e_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{n=1}^N [x_n - y_n]^2}. \quad (2.16)$$

- The *mean relative error* ( $e_{\text{MRE}}$ ) is the mean absolute difference in relation to one of the values. In general it makes sense to rate the difference to the absolute value, especially if many different large values occur within the measurement series. Indeed it will be very large or infinite if values appear near or equal to zero, what is often the case within traffic time series. The  $e_{\text{MRE}}$  is calculated:

$$e_{\text{MRE}} = \frac{\sum_{n=1}^N \frac{|x_n - y_n|}{x_n}}{N}. \quad (2.17)$$

- Similar facts hold for the *relative root mean square error* ( $e_{\text{RRMSE}}$ ):

$$e_{\text{RRMSE}} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left[ \frac{x_n - y_n}{x_n} \right]^2}. \quad (2.18)$$

- The *root mean square error proportional* ( $e_{\text{RMSEP}}$ ) is the  $e_{\text{RMSE}}$  divided by the mean of the measures and avoids the problem of infinite values of the  $e_{\text{MRE}}$ :

$$e_{\text{RMSEP}} = \frac{\sqrt{N \sum_{n=1}^N [x_n - y_n]^2}}{\sum_{n=1}^N x_n}. \quad (2.19)$$

- The *maximum error* ( $e_{\text{MAXE}}$ ) is the upper bound of all differences:

$$e_{\text{MAXE}} = \max(|x_n - y_n|). \quad (2.20)$$

- The *mean error* ( $e_{\text{ME}}$ ) makes sense if positive and negative fluctuations equalise each other:

$$e_{\text{ME}} = \frac{\sum_{n=1}^N [x_n - y_n]}{N}. \quad (2.21)$$

- The *mean absolute percent error* ( $e_{\text{MAPE}}$ ) is the  $e_{\text{MRE}}$  in percent:

$$e_{\text{MAPE}} = \frac{\sum_{n=1}^N \frac{|x_n - y_n|}{x_n}}{N} \cdot 100 \ %. \quad (2.22)$$

- The *equality coefficient* ( $C_{\text{equal}}$ ) is a value between 0 and 1. For total equality it is 1:

$$C_{\text{equal}} = 1 - \frac{\sqrt{\sum_{n=1}^N [x_n - y_n]^2}}{\sqrt{\sum_{n=1}^N x_n^2} + \sqrt{\sum_{n=1}^N y_n^2}}. \quad (2.23)$$



## Chapter 3

# Methods of Forecasting Traffic: State of the Art

“The most reliable way to forecast the future is to try to understand the present.”

**John Naisbitt**  
(\*1930), American trend analyst and futurist

In this chapter important former works are introduced that have been done in regard to traffic forecast. What distinguishes traffic forecasting from the more traditional forecasts of transportation planning is the length of the forecast horizon. While planning models use socioeconomic data and trends to forecast over a period of years, traffic state forecasting models act on timescales from seconds to month.

Traffic forecasting is an essential part of dynamic traffic control and traffic management systems. To inform upstream drivers about the downstream situation using variable message signs often the so called *one-step-ahead forecast* is required. This means a forecast of the next discrete point in the traffic time series. For instance, for minute aggregated data this means a one minute forecast. With a forecast of half an hour up to one or two hours it is possible to act just in time before incidents, for instance, with dynamic route guidance systems. Forecasts of days are used for planning of, for instance, events or construction areas.

The methods, that are applied up to now in the field of traffic forecasting, can be distinguished into four groups: parametric regression, nonparametric regression, neural networks, and heuristics. In the following, first the methods are proposed. Thereafter some of the works are mentioned, in which the forecast methods are used. Sometimes also combinations or comparisons are used.

### 3.1 Parametric Regression

A large class of parametric regression models are the so called *ARIMA* (Auto Regressive Integrated Moving Averages) methods. Many other well known algorithms like exponential smoothing or moving averages are special cases of these methods. In the following a short introduction to ARIMA modelling is given. To go deeper inside the reader is referred to common literature about time series analysis and forecasting, for instance, [Brockwell and Davis, 1996, Kantz and Schreiber, 1999, Thomopoulos, 1980] or the original work [Box and Jenkins, 1976], where the *Box-Jenkins* techniques of ARIMA modelling are introduced.

#### 3.1.1 Introduction to ARIMA Modelling

It is an important fact, that ARIMA modelling only works with stationary time series. A stationary time series is one, where all the data are in equilibrium with a common mean and a common variance. This fact is already a large restriction for possible forecasts. Nevertheless, the method is often successfully used especially for short term forecasts of traffic data.

Two basic models are encountered: the *autoregressive* (AR) and the *moving average* (MA) model. In the AR model, the current entry  $x_t$  of a time series is related in a linear manner to its  $p$  most recent entries  $(x_{t-1}, x_{t-2}, \dots, x_{t-p})$  and to a Gaussian distributed white noise  $e_t$  with mean zero and variance  $\sigma^2$  by the relation

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t. \quad (3.1)$$

Thereby  $(\phi_1, \dots, \phi_p)$  are parameters that have to be chosen.

In the MA model, the current entry  $x_t$  is related to the  $q$  most recent one-step-ahead forecast errors  $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$  and the current noise  $e_t$ :

$$x_t = -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t. \quad (3.2)$$

Thereby  $(\theta_1, \dots, \theta_q)$  are parameters that have to be chosen.

This approach is based on the theorem known as the *Wold Decomposition* [Anderson, 1971], that holds for stationary time series if the  $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$  are Gaussian distributed with mean zero and variance  $\sigma^2$ .

At this point it is important to mention, that in spite of the same name Eq. 3.2 has only little to do with the moving averages of Eq. 3.13, Eq. 5.16, or whenever at other places in this work the expression *moving averages* is used.

In the mixed *Autoregressive Moving Average* (ARMA( $p, q$ )) model, the entry  $x_t$  is related to the  $p$  most recent entries  $(x_{t-1}, x_{t-2}, \dots, x_{t-p})$ , the  $q$  most recent forecast errors  $(e_{t-1}, e_{t-2}, \dots, e_{t-q})$  and the current noise  $e_t$  in the following way:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t. \quad (3.3)$$

Using the backshift operator  $B^\alpha$  defined by

$$B^\alpha z_t = z_{t-\alpha}, \quad (3.4)$$

and the conventions

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad (3.5)$$

$$\theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q, \quad (3.6)$$

Eq. 3.3 can be written in the following form:

$$\phi(B)x_t = \theta(B)e_t. \quad (3.7)$$

Equation 3.7 describes together with the conventions in Eqs. 3.4 – 3.6 the ARMA( $p,q$ ) model. Because in most cases the time series  $x_t$  is not stationary, there are several approaches to get them nearly stationary, for instance, by differencing:

$$\begin{aligned} y_t &= [1 - B]^d x_t \\ \phi(B)y_t &= \theta(B)e_t. \end{aligned} \quad (3.8)$$

With the conventions mentioned above Eq. 3.8 describes an *Autoregressive Integrated Moving Average* (ARIMA( $p,d,q$ )) model.

For time series that show strong seasonal patterns, like the weekly recurring demand in traffic time series, a better stationarity can be achieved by a *seasonal* ARIMA process (ARIMA( $p,d,q$ )( $P,D,Q$ )<sub>S</sub>), that is defined by

$$\begin{aligned} y_t &= [1 - B]^d [1 - B^S]^D x_t \\ \phi(B)\Phi(B^S)y_t &= \theta(B)\Theta(B^S)e_t, \end{aligned} \quad (3.9)$$

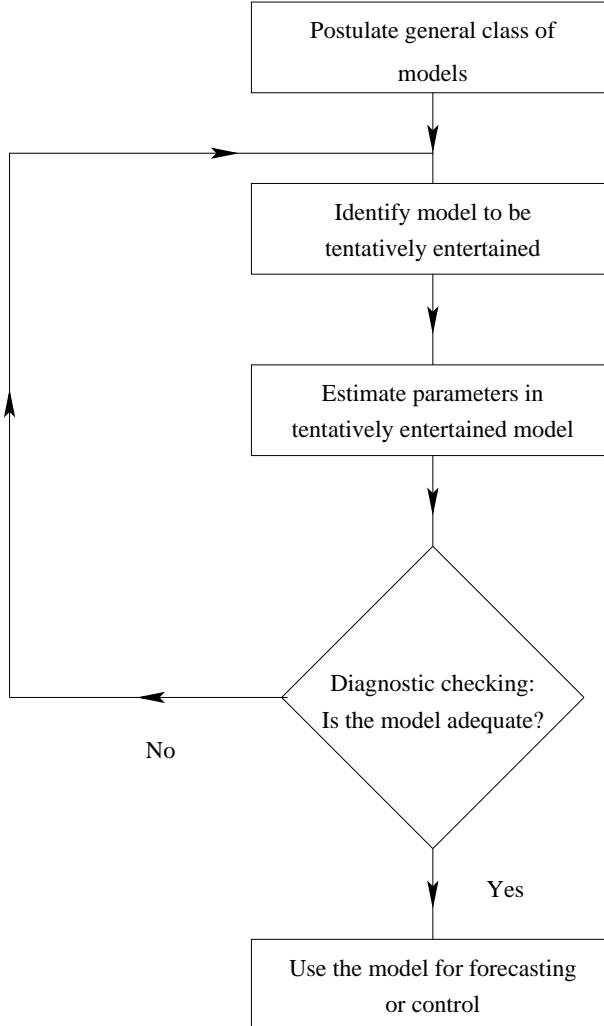
with the following conventions for  $\Phi$  and  $\Theta$ :

$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_p z^P, \quad (3.10)$$

$$\Theta(z) = 1 - \Theta_1 z - \dots - \Theta_q z^Q. \quad (3.11)$$

It is obvious, that there are a lot of parameters that have to be assigned to fit an ARMA( $p,q$ ), an ARIMA( $p,d,q$ ), or a seasonal ARIMA( $p,d,q$ )( $P,D,Q$ )<sub>S</sub> model to the measured data. Depending on the model that is used the ways to estimate the parameters may be different. One well known scheme that has been proposed by Box and Jenkins [1976] is the three stage iterative procedure. A scheme of the procedure is shown in Fig. 3.1. ARIMA models are fitted to an observed series by: identification, estimation, and diagnostic checking.

At the identification stage, the tentative models to fit to the data, that is, values of  $p$  and  $q$ , are determined by inspecting the autocorrelation and partial autocorrelation functions of the time series and its differences. Then they are compared with those of some basic stochastic processes. In general the autocorrelation function of a moving average process of order  $q$  has a cutoff after lag  $q$ , while its partial autocorrelation function tails off. Conversely, the autocorrelation of an autoregressive process of order  $p$  tails off in the form of damped exponentials or damped sine waves, while its partial autocorrelation function has a cutoff after lag  $p$ . For mixed processes, both the autocorrelations and partial autocorrelations



**Fig. 3.1.** Iterative procedure for ARIMA model development proposed by Box and Jenkins [1976]. After the general class of models is postulated, a three stage iterative procedure is applied. During identification the general number of parameters of  $p$  and  $q$  are determined. In the estimation stage the parameters are approximated. At last, in diagnostic checking, it is verified whether the model is adequate. If this is not the case, the three stage procedure starts from the beginning.

tail off. Failure of the autocorrelation function to die out rapidly suggests that differencing is needed ( $d > 0$ ).

At the estimation stage, maximum likelihood estimates are obtained for each model parameter. An efficient method for computing the maximum likelihood estimates of the model parameters has been proposed by Akaike [1973, 1974].

Finally, the fitted model is diagnosed to ensure that the estimated model residuals are white noise deviations, otherwise the model should be redesigned by repeating

the three stages of model construction.

Note, that this way to find an adequate model is extensive. Furthermore the proposed way to visually inspect the autocorrelation and partial autocorrelation function is somewhat arbitrary and difficult to automate. Because it is a well known fact (see, for instance, [Ahmed and Cook, 1979, Ahmed, 1983]), that there are no remarkable long term autocorrelations in traffic time series data often some different low order ARIMA methods are used for forecasting and the results are compared using quality measures like proposed in Sec. 2.3. Often they are also compared with other parametric regression models. Those that are often used are proposed in the following. Note, that many of them are special cases of low order ARIMA models.

### 3.1.2 Special Cases of Parametric Regression

In the *naive forecasting* model the estimates of the future data are taken to be the same as the most recently measured. The forecast  $\hat{x}_t$  of time  $t$  for every future data is just

$$\hat{x}_t = x_t. \quad (3.12)$$

Although this forecast model is easy to apply and requires minimal data, the forecasts generated are poor and of small practical value. On the other hand it often serves as a comparison to other, more refined forecasting models.

In the *moving average* model  $\mu_t$ , the average data of the  $N$  most recently time steps is calculated:

$$\mu_t = \frac{x_t + x_{t-1} + \dots + x_{t-N+1}}{N}. \quad (3.13)$$

The number of  $N$  must be selected by the forecaster. Like the naive model this is a horizontal one and the value  $\mu_t$  is used for all future data values. Note, that Eq. 3.13 can also be calculated recursive:

$$\mu_t = \mu_{t-1} + \frac{x_t - x_{t-N}}{N}. \quad (3.14)$$

As already mentioned, in spite of the same name the moving average model or just *moving averages* have only little to do with the MA model that is described by Eq. 3.2. Whenever the expression moving averages is used in this work, either Eq. 3.13 is meant if they are used for forecasts, or Eq. 5.16 if they are only used to smooth a curve.

The *double moving average* model  $\mu_{\text{double},t}(\tau)$  is an extension of the moving average model and is used for curves that follow a trend pattern. The trend adjustment is received by adjusting the values for  $\tau$  steps in future with

$$\mu_{\text{double},t}(\tau) = 2\mu_t - \mu_t^* + \left[ \frac{2}{N-1} \right] [\mu_t - \mu_t^*] \tau, \quad (3.15)$$

whereby

$$\mu_t^* = \frac{\mu_t + \mu_{t-1} + \dots + \mu_{t-N+1}}{N}, \quad (3.16)$$

or recursive

$$\mu_t^* = \mu_{t-1}^* + \frac{\mu_t - \mu_{t-N}}{N}. \quad (3.17)$$

One of the most common forecasting models in use today is the *single smoothing* model  $s_t$ . Often it is also called *exponential smoothing*. To apply the single smoothing model, the forecaster selects a smoothing parameter  $\alpha$  with  $0 \leq \alpha \leq 1$ . Then the forecasts become

$$\begin{aligned} s_1 &= x_1, & \text{for the first value,} \\ s_t &= \alpha x_t + [1 - \alpha] s_{t-1}, & \text{for each other.} \end{aligned} \quad (3.18)$$

Similar to the naive and the moving average model the single smoothing model is a horizontal one for all future values. Because exponential smoothing is used in operation in many different fields, adaptive approaches for adjusting the smoothing constant  $\alpha$  have been suggested by many authors. One approach that is often used is introduced by Trigg and Leach [1967]. In the *Trigg and Leach adaptive* model  $\alpha_t$  is set equal to a tracking signal  $\alpha_t = \text{SE}_t/\text{SAE}_t$ , whereby  $\text{SE}_t$  is the smoothed one-step-ahead forecast error  $e_t = s_{t-1} - x_t$ :

$$\begin{aligned} \text{SE}_t &= 0, & \text{for the first value,} \\ \text{SE}_t &= \gamma e_t + [1 - \gamma] \text{SE}_{t-1}, & \text{for each other,} \end{aligned} \quad (3.19)$$

$\text{SAE}_t$  the smoothed one-step-ahead absolute error  $|e_t|$ :

$$\begin{aligned} \text{SAE}_t &= 0, & \text{for the first value,} \\ \text{SAE}_t &= \gamma |e_t| + [1 - \gamma] \text{SEA}_{t-1}, & \text{for each other,} \end{aligned} \quad (3.20)$$

and  $\gamma$  another smoothing constant with  $0 \leq \gamma \leq 1$ . In general it holds  $-1 < \alpha_t < 1$ , but initial conditions have to be chosen as  $\alpha_t = 0$  until  $\text{SAE}_t \neq 0$  to avoid dividing by zero.

The *double smoothing average* model  $s_{\text{double},t}(\tau)$  is an extension of the single smoothing model for trend patterns. It is calculated for  $\tau$  steps in future by

$$\begin{aligned} s_{\text{double},1}(\tau) &= x_1 \\ b_1 &= 0 \end{aligned} \quad \left. \begin{aligned} s_{\text{double},t}(\tau) &= x_t + [1 - \alpha]^2 e_t + [b_{t-1} - \alpha^2 e_t] \tau \\ b_t &= b_{t-1} - \alpha^2 e_t \end{aligned} \right\} \quad \begin{aligned} &\text{for the first value,} \\ &\text{for each other.} \end{aligned} \quad (3.21)$$

Thereby  $e_t = s_{\text{double},t-1}(1) - x_t$  is the one-step-ahead forecast error and it holds  $0 \leq \alpha \leq 1$ .

Another trend model that applies smoothing techniques is called *single smoothing with linear trend*  $s_{\text{trend},t}(\tau)$ . As before, a smoothing parameter  $\alpha$  with  $0 \leq \alpha \leq 1$  is chosen and is used to assign weights to the past data values to forecast  $\tau$  steps in future:

$$\begin{aligned} s_{\text{trend},1}(\tau) &= x_1 \\ b_1 &= 0 \end{aligned} \quad \left. \begin{aligned} s_{\text{trend},t}(\tau) &= \alpha x_t + [1 - \alpha] s_{\text{trend},t-1}(1) + \tau b_t \\ b_t &= \alpha^2 [x_t - s_{\text{trend},t-1}(1)] + b_{t-1} \end{aligned} \right\} \quad \begin{aligned} &\text{for the first value,} \\ &\text{for each other.} \end{aligned} \quad (3.22)$$

A similar method works with two parameters  $\alpha$  and  $\beta$ . Within the *single smoothing with trend adjustment* model  $s_{\text{trad},t}(\tau)$  the value for  $\tau$  steps in future is calculated by

$$s_{\text{trad},t}(\tau) = s_t + \Delta s_t \tau, \quad (3.23)$$

with

$$\begin{aligned} \Delta s_1 &= 0, & \text{for the first value,} \\ \Delta s_t &= \beta [x_t - s_{t-1}] + [1 - \beta] \Delta s_{t-1}, & \text{for each other.} \end{aligned} \quad (3.24)$$

The value  $s_t$  is calculated with Eq. 3.18 and  $0 \leq \beta \leq 1$ .

As it has been mentioned above, many of the parametric regression models are special cases of low order ARIMA models if the parameters are chosen in the right way. In this manner the the naive model is equivalent to ARIMA(0,1,0), the single smoothing corresponds to ARIMA(0,1,1), and double smoothing to ARIMA(0,2,2). More examples for ARIMA forecast equations are given in [Thomopoulos, 1980].

### 3.1.3 ARIMA Modelling for Traffic Forecast

Ahmed and Cook [1979] and Ahmed [1983] use Box–Jenkins techniques to compute one-step-ahead forecasts of traffic flow and occupancy data. As input serve a total set of 166 time series representing about 27,000 min of detector measured data. The data are observations of three freeway surveillance systems in Los Angeles, Minneapolis, and Detroit during the afternoon peak periods. The Los Angeles data are 20 s aggregated data of traffic flow and occupancy per lane and the data from Minneapolis and Detroit are flow and occupancy aggregated over all lanes at 30 s and 60 s intervals, respectively.

Analysing the autocorrelation and partial autocorrelation functions the authors conclude, that an ARIMA(0,1,3) model is reasonable. After fitting the model parameters a comparative evaluation of the forecasting performance is presented comparing the model with the moving average model, the double smoothing average model, and the Trigg and Leach adaptive model. In evaluating the four forecasting models, the mean absolute error  $e_{\text{MAE}}$  and the mean square error  $e_{\text{MSE}}$  are used as evaluation criteria.

The ARIMA(0,1,3) model is seen to be superior, the values are stated by the authors from 1.30 to 6.50 for  $e_{\text{MAE}}$  and as 2.80 to 91.41 for  $e_{\text{MSE}}$ . Unfortunately at this points no units are given and the results for flow or occupancy are not considered separately. The results of the other models are given in ratio to the ARIMA model. The ratio to ARIMA of the moving average model varies between 1.00 and 1.27 for  $e_{\text{MAE}}$  and between 1.00 and 1.45 for  $e_{\text{MSE}}$  for a number of  $N = 5$  values in the moving average. The results indicate that both  $e_{\text{MAE}}$  and  $e_{\text{MSE}}$  increase with increasing  $N$ .

The best results of the double smoothing average model are associated with values of the smoothing constant  $0.1 \leq \alpha \leq 0.3$  for that the ratio to ARIMA range from 1.00 to 1.64 for  $e_{\text{MAE}}$  and from 1.00 to 1.43 for  $e_{\text{MSE}}$ .

Whereas the performances of the moving average and the double smoothing average model are comparable to the ARIMA(0,1,3) model, the Trigg and Leach adaptive model does not improve the forecasts. The authors conclude with a discussion how the model can be used in applications and the parameters can be updated in real time.

According to these results in [Ahmed and Cook, 1981, 1983] this one-step-ahead forecast is used for motorway incident detection. Therefore the ARIMA(0,1,3) model is applied on minute aggregated occupancies. If the measured occupancy lies outside confidence limits of two standard deviations from the forecast value, an incident is detected. The author states, that using a total set of 1692 min of occupancy observations associated with 50 traffic incidents in the Lodge Freeway in Detroit, USA, all of the 50 incidents could be detected with this method with a false alarm rate of 0.014. The authors give the false alarm rate as the rate of wrong incident messages out of the total incident messages generated by the algorithm. Supposedly there are usually more than one incident message during a traffic incident so the result must be taken with care. This is also, because the idea of incident detection based on ARIMA time series forecasting is again formulated later in a little bit different form by [Ferrari, 1988a]. Validating this approach in [Hoops et al., 2000], a false alarm rate of at least 0.45 is observed, whereby only a rate of 0.25 of the incidents have been detected. The reason for this are random fluctuations of large magnitude (see also Sec. 5.5).

Davis et al. [1990] try to use this method for occupancies and storage rates at on- and off-ramps not only to detect, but also to forecast congestions. In doing so the authors come upon the challenging problem, that linear time series analysis works quite well forecasting mean values but not so well for those extremes corresponding to the onset of congestion. Finally they use adjacent storage rates to forecast congestions.

In [Williams, 2001] the ARIMA model is enhanced by using also the data of upstream sensors for the forecast of downstream locations. Therefore Eq. 3.9 is modified in the following way:

$$\phi(B)\Phi(B^S)[1-B]^d[1-B^S]^D \left[ V_t - \sum_{i=1}^m \frac{\omega_i(B)}{\delta_i(B)} B^{k_i} U_{i,t} \right] = \theta(B)\Theta(B^S) e_t, \quad (3.25)$$

where  $V_t$  is the traffic flow time series at the downstream forecast location and  $U_{i,t}$  the  $i = (1, 2, \dots, m)$  upstream data series at time  $t$ . Equation 3.25 describes the so called ARIMAX model.

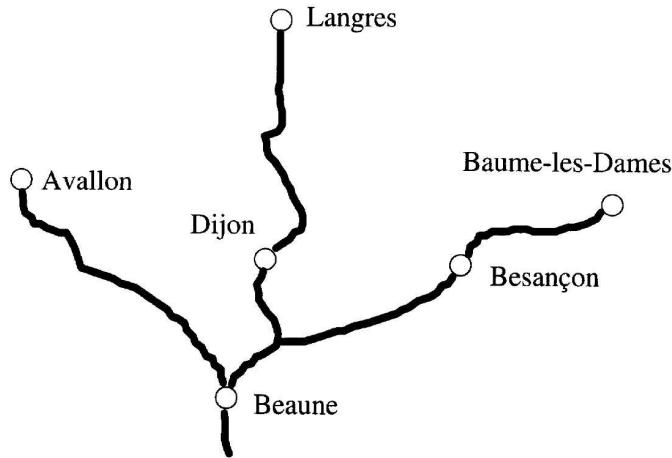
The pure delay of the  $i^{\text{th}}$  upstream traffic flow series is indicated by  $k_i$ ,  $\omega_i(B)$  is the numerator transfer function polynomial and  $\delta_i(B)$  the denominator transfer function polynomial for the  $i^{\text{th}}$  series.

The investigated traffic data is used in many other works according to traffic forecast, thus at this point a short description of the test field is given. The data are stemming from four counting stations near Beaune, France, whereby three of them are located approximately 90 km upstream from the forecast location in

Beaune (see Fig. 3.2):

- Beaune - Southbound flows on motorway A6 at kilometer point 305.3,
- Avallon - Southbound flows from the direction of Paris on motorway A6 at kilometer point 211.8,
- Langres - Southbound flows from the direction of Nancy on motorway A31 at kilometer point 98.8, and
- Beaume-Les-Dames - Southbound flows from the direction of Mulhouse on motorway A36 at kilometer point 91.0.

The data are 30 min flow rates during the months July and August of the years 1984 until 1990.



**Fig. 3.2.** Geographical relationship of data collection points used within the ARIMAX, the ATHENA (cp. Sec. 3.4 on page 43), and the KARIMA (cp. Sec. 3.5.2 on page 56) method. The picture is stemming from [van der Voort et al., 1996]. The principal site is in Beaune, where the flow along three motorways converges onto a single motorway. The three upstream measurement points are at Avallon, Baume-les-Dames, and Langres, about 90 km north of the Beaune counting station.

In a first step a seasonal ARIMA(2,0,2)(0,1,1)<sub>48</sub> model based on the parameter estimates for the years 1984 until 1988 is tested on the data of the year 1990 for comparison with two other traffic forecast techniques that use the same traffic data for forecast, the ATHENA (Sec. 3.4 on page 43) and the KARIMA (Sec. 3.5.2 on page 56) method. This is done by calculating the particular percent error  $e_{PE}$  and give the ratio of all  $e_{PE}$  that lie in certain limits, for instance,  $-5\% \leq e_{PE} \leq 5\%$ . Then, these rates are compared among the methods.

In doing so, the ARIMA method provides the highest percentage of small forecast errors with  $-5\% \leq e_{PE} \leq 5\%$ . Nevertheless, the authors do not state, that it

outperforms the other methods, because the smallest percentage of negative outliers ( $ePE < -25\%$ ) is performed by the ATHENA model, and the smallest percentage of positive outliers ( $25\% < ePE$ ) by the KARIMA method.

In the second step the parameters for the ARIMAX model are determined and it is shown, that this model outperforms the seasonal ARIMA method. To do this, first, an univariate ARIMA(4,0,0)(0,1,1)<sub>48</sub> model is fitted to the response series and each of the input data series. Then, cross correlations of the input series against the response series are calculated to determine the response delay  $k_i$  and the preliminary order of the transfer functional polynomials  $\omega_i(B)$  and  $\delta_i(B)$ . Any parameters, that are not statistically significant, are dropped from the final transfer function model. As a result the transfer functional polynomials depend on the investigated location and are of an order from 0 to 2 what corresponds to travel times up to one hour from the input data location to the response data location.

Using auto correlation functions and partial auto correlation functions the final form of the ARIMA model is identified as an ARIMA( $p,0,q$ )(0,1,1)<sub>48</sub> where  $1 \leq p + q \leq 5$ . Finally, the authors state the ARIMA(2,0,0)(0,1,1)<sub>48</sub> as the best one. Using it and fitting the parameter the authors state, that the ARIMAX model consistently result in root mean square error  $e_{RMSE}$ , that is about 15 veh/hour lower than the corresponding univariate ARIMA model.

This result is not surprising, because what is done with the enhancement of the univariate ARIMA model, is to use also upstream traffic data, that are available in this special case. A similar approach is already done a few years earlier in [van Arem et al., 1997] for forecasting link travel times. As stated in the introduction of this work, adjacent traffic data without any on- or off-ramps between the measurement stations is not the general case and moreover the task of a spatial temporal traffic model.

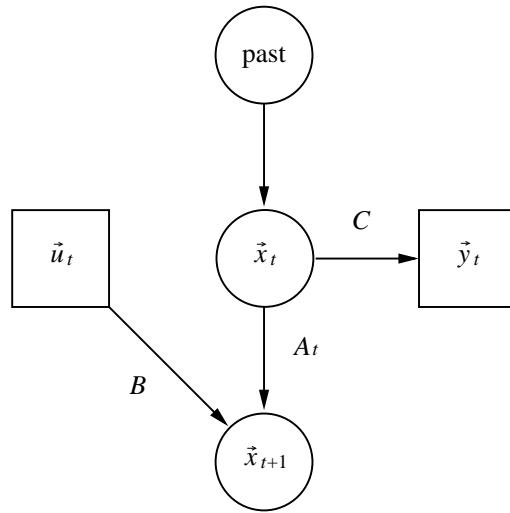
Nevertheless the work of Williams [2001] is a good example, that for proper forecasts the model must be enhanced with external information. This is the reason, why parametric regression techniques, especially those of Sec. 3.1.2, often are used in combined methods for very short term forecasts like the one-step-ahead forecast. Also the results of the works of Ahmed and Cook [1983] and Davis et al. [1990] have shown, that it is very difficult to forecast basic changes only with ARIMA modelling. The basic difficulty lies in the fact, that the ARMA process needs a stationary time series. Nevertheless in the following another often used parametric regression technique, the Kalman Filter, is proposed.

### 3.1.4 Kalman Filter

An especially in the field of signal detection and automatic control often used parametric regression technique is the so called *Kalman Filter*. The basic difference in regard to the techniques mentioned up to now is, that the state of the system and the observation process is considered separately. The Kalman Filter is named after its inventor Rudolf Emil Kalman, who published in 1960 his famous

paper [Kalman, 1960]. In the following, only a short introduction can be given. To go deeper inside the reader is referred to the existing extensive literature (see, for instance, [Harvey, 1990, Haykin, 1990, Zarchan and Musoff, 2000]).

The Kalman filter model is based on the assumption, that at each discrete time point  $t$  an observation vector  $\vec{y}_t$  is related to the state vector  $\vec{x}_t$  through an observation equation. The state process is an unobserved first order Markov process and is specified by a transition equation. On the basis of past observations  $(\vec{y}_1, \vec{y}_2, \dots, \vec{y}_t)$ , and the current observation  $\vec{y}_{t+1}$  an optimal estimation of  $\vec{x}_{t+1}$  is required. It turns out, that the filter depends on the observations only through the previous estimate of the state  $\vec{x}_t$  and the current estimation error, and so allows a recursive computation.



**Fig. 3.3.** Independence graph of the Kalman Filter model.

The basis of the model are the transition and observation equations:

$$\begin{aligned}\vec{x}_{t+1} &= A_t \vec{x}_t + B \vec{u}_t + \vec{q}_t, \\ \vec{y}_t &= C \vec{x}_t + \vec{r}_t,\end{aligned}\tag{3.26}$$

where

- $t$  index of time,
- $\vec{x}_t$   $m \times 1$  vector of the states,
- $\vec{y}_t$   $n \times 1$  vector of the observations,
- $\vec{u}_t$   $l \times 1$  vector of the exogenous variables,
- $A_t$   $m \times m$  matrix relating the previous state to the current one,
- $B$   $m \times l$  matrix, that relates the exogenous variables to the state,
- $C$   $n \times m$  matrix, that relates the state to the measurement,
- $\vec{q}_t$   $m \times 1$  vector, that represents the process noise,
- $\vec{r}_t$   $n \times 1$  vector, that represents the measurement noise.

The random vectors  $\vec{q}_t$  and  $\vec{r}_t$  are assumed to be independent of each other, with zero mean and an  $m \times m$  covariance matrix  $Q$  and an  $n \times n$  covariance matrix  $R$  given by

$$\begin{aligned} Q_{ij} &= \delta_{ij} \vec{q}_i \vec{q}_j, \\ R_{ij} &= \delta_{ij} \vec{r}_i \vec{r}_j, \end{aligned} \quad (3.27)$$

with the Kronecker Delta  $\delta_{ij}$ .

The structure of Eq. 3.26 is held in the conditional independence graph (Fig. 3.3) in which its missing edges indicate independences.

After an initialisation of an  $m \times 1$  state vector  $\hat{\vec{x}}_0$  and an  $m \times m$  error covariance matrix  $P_0$  the recursion, that determines the Kalman filter is

$$K_{t+1} = [A_t P_t A_t^T + Q] C^T [C [A_t P_t A_t^T + Q] C^T + R]^{-1}, \quad (3.28)$$

$$\hat{\vec{x}}_{t+1} = A_t \hat{\vec{x}}_t + B \vec{u}_t + K_{t+1} [\vec{y}_{t+1} - C [A_t \hat{\vec{x}}_t + B \vec{u}_t]], \quad (3.29)$$

$$P_{t+1} = [I - K_{t+1} C] [A_t P_t A_t^T + Q]. \quad (3.30)$$

The index  $T$  means the transpose of the matrix. The  $m \times n$  matrix  $K_t$  is the *Kalman gain* or *blending factor*, that minimises the error covariance matrix  $P_t$ . Now, an estimation of the state ahead  $\hat{\vec{x}}_{t+1}^*$  can be forecast with

$$\hat{\vec{x}}_{t+1}^* = A_t \hat{\vec{x}}_t + B \vec{u}_t. \quad (3.31)$$

To forecast more than one state ahead,  $\hat{\vec{x}}_t$  can be replaced by the particular estimation  $\hat{\vec{x}}_t^*$ . Because the details of the computational and probabilistic origins of the Kalman filter extend the scope of this work, the interested reader is referred to [Brown and Hwang, 1992, Jacobs, 1993, Maybeck, 1979].

Whittaker et al. [1997] present a forecast system, that applies a Kalman Filter on a motorway network around Rotterdam in The Netherlands. It contains about 50 on- and off-ramps and 500 loop detectors, which measure minute aggregated data of, among others, flow, velocity, and occupancy. These monitoring stations devide the network in about 500 motorway sections called links.

The state vectors  $\vec{x}_t$  are chosen as approximately  $1000 \times 1$  vectors consisting of the two variables flow  $J_{s,t}$  and density  $\rho_{s,t}$  of each link  $s$  at time  $t$ . The monitoring stations are represented by an approximately  $500 \times 1$  observation vector  $\vec{y}_t$ , that holds the information of the loop detectors, for instance, flow or occupancy. The exogenous input  $\vec{u}_t$  is chosen as a  $50 \times 1$  vector of flows from monitoring stations on the on-ramp entrances.

The transition matrix  $A_t$  carries both the network topology and the traffic dynamics. In the network topology three cases are distinguished:

- a single section  $s$  is followed by a single section  $s + 1$ ,
- a single section  $s$  is followed by two sections  $s_1$  and  $s_2$  (fork, for instance, at off-ramps),

- two sections  $s_1$  and  $s_2$  are followed by a single section  $s$  (join, for instance, at on-ramps).

The section to section transitions are obtained using three equations.

The first is a balance equation for the density  $\rho_{s,t}$  that is based on vehicle conservation

$$\rho_{s,t+1} = \rho_{s,t} + d_s^{-1} [J_{s-1,t} - J_{s,t}], \quad (3.32)$$

whereby  $d_s$  is the area (lanes  $\times$  length) of link  $s$ .

The second is a flow equation based on Eq. 2.4:

$$J_{s,t+1} = \min(v_{s,t}, v_{s+1,t})\rho_{s,t}. \quad (3.33)$$

The third is a generic relationship between velocity and density with which the velocities  $v_{s,t}$  and  $v_{s+1,t}$  are computed. Unfortunately the authors do not clearly state how they cope with measured velocities. Similar equations are used for sections to joins. For fork to sections turning fractions are performed from the measurements.

Because the observed flow can just be assumed as the real one the matrix  $C$  is directly obtained as block diagonal. To obtain estimates for the matrices  $A_t$  and  $B$  a model

$$x_t - \phi_1 x_{t-1} = \theta_1 u_{t-1} + e_t, \quad (3.34)$$

which is a very simple form of the ARIMAX model (Eq. 3.25), is applied to the measured data. In a 60 min window the parameters are re-estimated through the whole day.

Describing all the details would extend the scope of this work and the interested reader is referred to the original paper. But it can be understood, that applying the approach to the whole network requires a lot of computer resources. It is especially unfeasible for real time applications. Thus, the authors investigate a subnetwork consisting of five links, that give a total of 29 state variables within the state vector. As results the authors present figures of five-step ahead forecasts for the occupancy and a comparison of two-step ahead forecasts with the naive forecasting model for different noise variance matrices  $Q$  and  $R$ . Obviously the naive forecast is outperformed, but note, that also within this application adjacent sections are used.

As other parametric regression methods the Kalman Filter alone is not able to forecast certain differences in the daily traffic data and is therefore often only used in combination with other methods. A comparison between several parametric regression methods such as different ARIMA methods, Kalman filter, but also multiple regression techniques with adjacent detectors is investigated in [Saito et al., 1997] for traffic signal control. Thereby it is an indisputable fact, that the right choice of the parametric regression method strongly depends on the particular investigated traffic data, that is again strongly dependent on the traffic situations that can appear.

## 3.2 Nonparametric Regression

The basic difference between the parametric and nonparametric regression lies in the fact, that for the latter a whole function has to be computed instead of a set of parameters. Most of the nonparametric regression techniques used in traffic forecasting are based on the idea of *phase space embedding* and finding the *nearest neighbourhood in phase space*. In the following a short introduction to phase space embedding and related techniques is given. An overview about these nonlinear time series analysis can be found, for instance, in [Kantz and Schreiber, 1999].

### 3.2.1 Introduction to Phase Space Embedding

In a purely deterministic system, once its present state is fixed, the states at all future times are determined as well. Thus, a vector space (called a *state space* or *phase space*) can be established for the system so that specifying a point in this space specifies the state of the system, and vice versa. Then the dynamics of the system can be studied by studying the dynamics of the corresponding phase space points.

The concept of the state of a system is powerful even for nondeterministic systems. A large class of systems can be described by a set of states and some kind of transition rules which specify how the system may proceed from one state to the other. Prominent members of this category are the stochastic Markov processes for which the transition rules are given in the form of a set of transition probabilities and the future state is selected randomly according to these probabilities. What is often observed in an experiment is not a phase space object but a time series. Therefore the observations have to be converted into state vectors. This is the important problem of *phase space reconstruction* or *phase space embedding*, which is technically solved by the method of delays or related constructions.

Given a sequence of scalar measurements  $s_n$ , a *delay reconstruction* in  $m$  dimensions is formed by the vectors  $\vec{s}_n$  given as:

$$\vec{s}_n = (s_{n-[m-1]\nu}, s_{n-[m-2]\nu}, \dots, s_{n-\nu}, s_n). \quad (3.35)$$

The time difference in number of samples  $\nu$  (or in time units,  $\nu\Delta t$ ) between adjacent components of the delay vectors is referred to as the *lag* or *delay time*. If there are more than one observable, the vector can be multivariate and can consist of different types of data at different time steps. Finding a suitable embedding, especially the minimal embedding dimension, is an essential problem in nonlinear time series analysis techniques.

One approach that solves this problem is the method of *false nearest neighbours* (FNNs), proposed by Kennel et al. [1992]. For each point  $\vec{s}_i$  in the time series the nearest neighbour  $\vec{s}_j$  in an  $m$  dimensional phase space is found, that means, the distance  $d_{i,j} = \|\vec{s}_i - \vec{s}_j\|$  in regard to a certain norm  $\|\cdot\|$  is minimum. Then both

points are iterated and

$$R_i = \frac{|s_{(i+1)\nu} - s_{(j+1)\nu}|}{d_{i,j}} \quad (3.36)$$

is computed. If  $R_i$  exceeds a given heuristic threshold  $R_h$ , this point is marked as having a FNN. The criterion that the embedding dimension is high enough is, that the fraction of points for which  $R_i > R_h$  is zero or at least sufficiently small. As norm usually the  $p$ -norm is used:

$$\|\vec{x}\|_p = \left[ \sum_{i=1}^n |x_i|^p \right]^{\frac{1}{p}}. \quad (3.37)$$

The most widely used are

- the *Manhattan metric*  $\|\cdot\|_1$ ,
- the *Euclidean distance*  $\|\cdot\|_2$ , and
- the *maximum norm*  $\|\cdot\|_\infty = \max_{1 \leq i \leq n}$ .

Finally a forecast algorithm can be formulated. Given a scalar time series  $(s_1, \dots, s_N)$ , a delay time  $\nu$  and an embedding dimension  $m$  has to be chosen (in most cases the delay time follows directly from the discrete measurement, for instance, for minute aggregated traffic data  $\nu = 1$  min). In order to forecast  $\Delta\nu$  steps ahead of  $N$ , a parameter  $\epsilon$  of the order of the resolution of the measurements is chosen to form the  $\epsilon$ -neighbourhood  $\mathcal{N}(\vec{s}_N)$ . This  $\epsilon$ -neighbourhood  $\mathcal{N}(\vec{s}_N)$  are the  $k$  state space vectors  $\vec{s}_j$  for that hold  $\epsilon > d_{N,j}$ . For the  $k$  state space vectors  $\vec{s}_j \in \mathcal{N}(\vec{s}_N)$  the individual “forecasts”  $s_{j+\Delta\nu}$  are looked up. For  $j + \Delta\nu \leq N$  they are available because these are all points in the past.

The finally accepted forecast is then a combination of all individual forecasts. That can be the arithmetic mean. In this case the forecast can also be formulated as follows:

$$s_{N+\Delta\nu} = k^{-1} \sum_{\vec{s}_j} \text{Ind} \left( 1 - \frac{d_{N,j}}{\epsilon} \right) s_{j+\Delta\nu}. \quad (3.38)$$

Here  $\text{Ind}(x)$  is the indicator function

$$\text{Ind}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{else.} \end{cases} \quad (3.39)$$

and

$$k = \sum_{\vec{s}_j} \text{Ind} \left( 1 - \frac{d_{N,j}}{\epsilon} \right). \quad (3.40)$$

Often the individual forecasts are even weighted, for instance, in regard to their distance  $d_{N,j}$ . Equation 3.38 formulates the so called  $k$ -nearest neighbour ( $k$ -NN) model. An alternative to choosing the  $\epsilon$ -neighbourhood is to just choose the

number  $k$  of the nearest neighbours, that are taken into account. This approach avoids the problem of quantifying the size of  $\epsilon$  and especially the case, that there are no neighbours in the  $\epsilon$ -neighbourhood.

### 3.2.2 Nonparametric Regression for Traffic Forecast

One of the first approaches in forecasting traffic flow with nonparametric regression techniques is done by Davis and Nihan [1991]. The authors use the  $k$ -NN method for calculations of minute aggregated traffic flow  $J(t)$  and occupancy  $\rho_{\text{rel}}(t)$  data of time step  $t$  at one particular cross section on the Interstate 5 motorway between King and Snohomish counties in Washington State, USA.

The multivariate phase space vector to calculate  $J(t)$  and  $\rho_{\text{rel}}(t)$  is formed by

$$\begin{aligned} \vec{s}_n = & (J(t-1), \rho_{\text{rel}}(t-1), J_{\text{u}}(t-1), \rho_{\text{rel},\text{u}}(t-1), J_{\text{u}}(t-2), \rho_{\text{rel},\text{u}}(t-2), \\ & J_{\text{or-d}}(t), J_{\text{or-u1}}(t), J_{\text{or-u2}}(t-1)). \end{aligned} \quad (3.41)$$

Thereby  $J_{\text{u}}(t)$  and  $\rho_{\text{rel},\text{u}}(t)$  denote the flow and occupancy at an upstream section,  $J_{\text{or-d}}(t)$  the flow at an downstream, and  $J_{\text{or-u1}}(t)$  and  $J_{\text{or-u2}}(t)$  the flow at two upstream on-ramps. The choice of the time delays 0, 1, and 2 for the upstream and downstream cross sections comes from their distance to the cross section whose data is to be forecast. Albeit the selection of this vector is somewhat arbitrary. The results are not discussed here in detail, because there is a very sparse database of about one and a half hours of traffic data at the authors disposal. The interested reader is referred to the original work.

An advanced nonparametric regression technique is used by Faouzi [1996]. The author especially investigates how different nearest neighbourhood data can be reasonably combined. Therefore he uses a method that is called *kernel estimation*. To calculate a forecast at the  $N^{\text{th}}$  time step  $\Delta\nu$  time steps ahead, all individual forecasts  $s_{j+\Delta\nu}$  of all possible delay vectors  $\vec{s}_j$ , that are of an embedding dimension  $m$ , are weighted using a *kernel function*  $K(\vec{x})$ . Possible are the vectors  $\vec{s}_j$  for that hold  $m \leq j \leq N - \Delta\nu$ . The forecast  $\hat{s}_N(\Delta\nu)$  is calculated in the following way:

$$\hat{s}_N(\Delta\nu) = \sum_{j=m}^{N-\Delta\nu} \theta_j s_{j+\Delta\nu}, \quad (3.42)$$

where the weights  $\theta_j$  are defined as:

$$\theta_j = \frac{K([\vec{s}_N - \vec{s}_j] / w_n)}{\sum_{i=m}^{N-\Delta\nu} K([\vec{s}_N - \vec{s}_i] / w_n)}. \quad (3.43)$$

Thereby, the bandwidth  $w_n$  is a parameter that has to be chosen. It determines how fast the influence of further neighbours decreases. The choice of this parameter is of great importance. When  $w_n$  increases very stable forecasts with low

variances are obtained but at the expense of flexibility. This problem is known as the *bias/variance dilemma* [German et al., 1992].

For the kernel function  $K(\vec{x})$  it is important, that it has desirable properties. The simplest class of kernels consists of probability density functions that satisfy

$$K(\vec{x}) \geq 0, \quad \int_{-\infty}^{\infty} K(\vec{x}) d^m x = 1. \quad (3.44)$$

The kernel function that is used in [Faouzi, 1996] is the Gaussian kernel on  $\mathbb{R}^m$ :

$$K(\vec{x}) = [2\pi]^{-m/2} \exp\left(-\frac{\|\vec{x}\|_2^2}{2}\right), \quad \vec{x} \in \mathbb{R}^m. \quad (3.45)$$

To get a trade-off between bias and estimator variance the parameter  $w_n$  is chosen as follows:

$$w_n = \frac{\sigma}{n^{1/[m+4]}}, \quad (3.46)$$

whereby  $\sigma$  is the empirical standard deviation of the  $n$  observations before forecast. Deheuvels [1977] has shown, that this choice limits the kernel influence on the results and guarantees an optimal rate of convergence with respect to the mean integrated square error criterion.

The embedding dimension  $m$  is found using a cross validation like procedure. This means, that the mean square error of a test data set is minimised varying different dimensions  $m_x$ :

$$m = \arg \min_{m_x} \left( m_x^{-1} \sum_{j=N-\Delta\nu-m_x+1}^{N-\Delta\nu} [s_{j+\Delta\nu} - \hat{s}_N(\Delta)]^2 \right). \quad (3.47)$$

Thereby the operator  $\arg$  means the argument of the operator  $\min$ , in this case  $m_x$ . Unfortunately it is not given in [Faouzi, 1996], which embedding dimension finally is used.

The algorithm is tested for one-step-ahead forecasts of five minute aggregated traffic flow data collected at one single location in a test site area at Toulouse, France. The data are collected from 1993/11/16 to 1993/11/19. The forecasts are compared with an ARIMA(0,1,0) model that is completely outperformed by the kernel estimator. The mean relative error of the prognosis results is stated by the author as  $e_{MRE} = 0.042$ .

Sun et al. [2003] use a similar algorithm called *local linear regression* (see also [Fan and Gijbels, 1996]). This approach is based on the assumption, that a univariate response  $y$  depends on multivariate covariates  $\vec{x} = (x_1, \dots, x_d)$  with an unknown function  $y = f(\vec{x})$ . Given  $n$  observations  $(\vec{x}_j, y_j)$  with  $j = 1, \dots, n$ , and the observations  $\vec{x}_c$ , the forecast  $y_c$  is estimated approximating the measurements  $y_j = f(\vec{x}_j)$  with the firsts parts of a Taylor series around  $\vec{x}_c$ :

$$y_j = y_c + \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\vec{x}_c) [x_{ji} - x_{ci}]. \quad (3.48)$$

The prognosis value  $y_c$  as well as the derivatives  $\frac{\partial}{\partial x_i} f(\vec{x}_c)$  are approximated using the past measurements. For convenience let denote  $\beta_0 = y_c$  and  $\beta_i = \frac{\partial}{\partial x_i} f(\vec{x}_c)$ . An estimator for the  $\vec{\beta} = (\beta_0, \dots, \beta_d)$  is calculated minimising the function:

$$\sum_{j=1}^n \left[ y_j - \beta_0 - \sum_{i=1}^d \beta_i [x_{ji} - x_{ci}] \right]^2 K_{\mathbf{B}}(\vec{x}_j - \vec{x}_c). \quad (3.49)$$

It can be calculated straight forward, that the  $\vec{\beta}$  that fulfills this condition is

$$\vec{\beta} = [\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d]^{-1} \mathbf{X}_d^T \mathbf{W} \vec{y}, \quad (3.50)$$

where

$$\mathbf{X}_d = \begin{pmatrix} 1 & x_{11} - x_{c1} & \dots & x_{1d} - x_{cd} \\ 1 & x_{21} - x_{c1} & \dots & x_{2d} - x_{cd} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} - x_{c1} & \dots & x_{nd} - x_{cd} \end{pmatrix}, \quad (3.51)$$

and

$$\mathbf{W} = \text{diag}(K_{\mathbf{B}}(\vec{x}_j - \vec{x}_c)), \vec{y} = (y_1, \dots, y_n). \quad (3.52)$$

Thereby is  $K_{\mathbf{B}}(\vec{x}) = \frac{1}{|\mathbf{B}|} K(\mathbf{B}^{-1} \vec{x})$ , where  $K(\vec{x})$  is a kernel function with the properties of Eq. 3.44.  $T$  indicates in this case the transpose of the matrix. Like in the kernel estimation of Faouzi [1996], also Sun et al. [2003] use a Gaussian kernel.  $\mathbf{B} = w_n \mathbf{I}$  is the bandwidth matrix and  $|\mathbf{B}|$  its determinant.

The multivariate covariate vectors act as the delay vectors and the embedding dimension is calculated using a cross validation like procedure as in the kernel estimation, but in opposite to that this method is also used to choose the bandwidth  $w_n$ .

Note, that the forecast is finally calculated in Eq. 3.50. For the case that the matrix  $\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d$  is nearly singular it is changed to

$$[\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d]^* = \mathbf{X}_d^T \mathbf{W} \mathbf{X}_d + \text{diag}(\lambda_0, \dots, \lambda_d). \quad (3.53)$$

The  $\lambda_i$  are calculated as  $\sqrt{n_0}/3$  where  $n_0$  are the mean number of observations. Finally this method is applied to forecast 5 min aggregated velocity data one- to five-steps-ahead and the results are compared with the kernel regression method and the  $k$ -NN model. The data are collected on Houston US-290 Northwest Freeway, USA, during the morning peak hours from 6:00 to 10:00 at 32 days from February 2002 to July 2002. At a segment of a length of 1.55 miles the travel times are measured using an automatic vehicle identification system. From that the velocities are calculated.

The authors state that the local linear regression method performs best, but the database is very small and the velocity has very strong fluctuations. In fact the kernel regression method for the one-step-ahead forecast is with a mean relative error of  $e_{\text{MRE}} = 8.91$  very similar to the local linear regression method ( $e_{\text{MRE}} =$

8.46). The authors conclude, that better comparisons shall be done with larger databases.

Before using the  $k$ -NN method for forecast, Shin et al. [1999] use a Daubechies wavelet transformation (see [Daubechies, 1988]) to smooth travel time data between bus stops from July 1996 to September 1996 of public buses in the Seoul metropolitan area. The data is obtained by wireless communication facilities which communicate between the buses and the stations. The data is collected in different intervals, and after the noise is removed the nearest neighbour in regard to a delay vector is used for one-, two-, and three-step ahead forecasts. Two models are used: one with a time lag of 18 min and an embedding dimension of 5, and one with a time lag of 5 min and an embedding dimension of 6.

The reason for this work is a comparison of the state space forecasting model with and without smoothing the raw data with the wavelet transformation. As a quality measure they use the correlation coefficient  $r$  between the forecast and the measured data. This is with  $r = 0.9989$  remarkable higher with the smoothing than without ( $r = 0.9835$ ). The authors conclude, that especially on short timescales fluctuations appear, that are mostly composed of unpredictable measurement noise. Every kind of traffic time series that has to be forecast first should be cleaned up of this noise. In the empirical data later in Chapter 5 can clearly be seen, that indeed noise in traffic data is a very complex theme because the amplitude of the noise is of a large magnitude. The question of random noise is discussed in detail in Sec. 5.5. At this point it has also to be mentioned, that a direct measurement of the travel time even nowadays is an exception. As discussed later in Sec. 5.6.1 in most cases the travel time is calculated from other forecast data.

Otokita and Hashiba [1998] even use the nearest neighbourhood method to estimate the travel time. Therefore the authors use measured flow and occupancy data. Because for this method the measured data of the point in time that is forecast is needed, it is more a simulation or calculation than a forecast of travel times and the work is not discussed in detail at this point.

Nair et al. [2001] use nearest neighbourhood techniques to analyse traffic data and use the results to forecast traffic flow. The experiment is performed for inductive loop detector data on an 11 kilometer stretch of the San Antonio freeway system in Texas, USA, during the months of January 1999 through June 1999. The double loops are placed on the freeway every half a mile, for both the main lanes and the on- and off-ramps.

The raw data are screened for errors and then aggregated into 5 min estimates of average speed, flow, and average occupancy. For the analysis only the peak hour, 6:00 to 10:00 is used. The delay vectors  $\vec{s}_n$  for the velocity time series are formed and the optimum delay embedding  $m$  is found with the FNN method. Then the  $\epsilon$ -neighbourhood according to the Euclidean distance is found. The parameter  $\epsilon$  is chosen on a trial-and-error basis. This means, the performance of the predictor is observed for different  $\epsilon$  and depending on the forecast result the radius  $\epsilon$  is chosen. For each phase space point  $\vec{s}_j$  in the  $\epsilon$ -neighbourhood of  $\vec{s}_n$  the  $\Delta\nu^{\text{th}}$  next

phase space point  $s_{j+\Delta\nu}$  is found. The first samples of these phase space points are then averaged to obtain an estimate for the  $\Delta\nu$ -step-ahead forecast  $s_{n+\Delta\nu}$ . The results are compared to a multilayer feedforward neural network (see Sec. 3.3). The authors state, that although the forecast error is larger than that for the neural network the results ascertain the deterministic nature of the time series and motivate further analysis.

An approach where the state space is not only defined with a delay vector but also with historical data is proposed by Williams and Smith [1999] and Smith et al. [2002]. The authors describe a nearest neighbour nonparametric regression model that is used to forecast traffic flow one-step-ahead. Because as database act 15 min aggregated traffic flow, the forecast horizon is 15 min. The data are stemming from two loop detector locations on the London Orbital Motorway M25 collected from 1996/09/04 to 1996/11/30.

After a test with different numbers of lagged values the authors conclude, that state vectors with only lagged observations may not contain enough information about the state of the system. In contrast to the models mentioned up to now, a hybrid state vector  $\vec{s}(t)$  is defined as

$$\vec{s}(t) = (J(t), J(t-1), J(t-2), J_{\text{hist}}, J_{\text{hist}}(t+1)). \quad (3.54)$$

Here  $J(t)$  is the traffic flow at the time of day  $t$  and  $J_{\text{hist}}(t)$  is the historical average flow according to the weekday and to the time of day  $t$ .

Then, for the current state space vector  $\vec{s}_c(t)$  the potential neighbourhood is calculated using the Euclidean distance up to a certain number of  $k$  nearest neighbours  $\vec{s}_i(t)$ , whereby  $k$  is chosen between 5 and 40 inclusive in increments of five. From the  $k$  nearest neighbours follow directly the  $k$  values  $J_i(t+1)$  for the time  $t+1$  of all  $i = 1, 2, \dots, k$  nearest neighbours.

At this point it must be mentioned, that calculating the historical average flow  $J_{\text{hist}}$  according to the weekday fills external knowledge into the model. The analysis of traffic time series according to recurring time dependent patterns are in general the subject of Sec. 3.4. To be precise, both methods are combined in this approach, but because the general approach is the phase space embedding technique, it is not discussed as the other combinations in Sec. 3.5.

Also in this work the authors do some considerations how the  $k$  nearest neighbourhood data can be combined to a single forecast value  $\hat{J}_c(t+1)$ . Therefore six different methods are investigated:

- The *straight average* approach takes just the mean of the  $k$  values  $J_i(t+1)$ :

$$\hat{J}_c(t+1) = \frac{1}{k} \sum_{i=1}^k J_i(t+1). \quad (3.55)$$

- The *weighted by inverse of distance* approach assumes, that selected neighbours that are relatively closer in regard to the Euclidean distance  $d_i$  to the

current state provide a better indication of the future condition:

$$\hat{J}_c(t+1) = \frac{\sum_{i=1}^k d_i^{-1} J_i(t+1)}{\sum_{i=1}^k d_i^{-1}}. \quad (3.56)$$

- The *adjusted by  $J(t)$*  approach assumes that the output elements will provide better information, if they are adjusted by the ratio of the flow  $J_c(t)$  of the current state space vector  $\vec{s}_c(t)$  to the corresponding element  $J_i(t)$  in the selected neighbour vectors  $\vec{s}_i(t)$  before averaging:

$$\hat{J}_c(t+1) = \frac{1}{k} \sum_{i=1}^k J_i(t+1) \frac{J_c(t)}{J_i(t)}. \quad (3.57)$$

- In the same manner in the *adjusted by  $J_{\text{hist}}(t+1)$*  approach the output elements are adjusted by the ratio of the historical flow  $J_{\text{hist},c}(t+1)$  of the current state space vector  $\vec{s}_c(t)$  to the corresponding element  $J_{\text{hist},i}(t+1)$  in the selected neighbour vectors  $\vec{s}_i(t)$  before averaging:

$$\hat{J}_c(t+1) = \frac{1}{k} \sum_{i=1}^k J_i(t+1) \frac{J_{\text{hist},c}(t+1)}{J_{\text{hist},i}(t+1)}. \quad (3.58)$$

- The *adjusted by both  $J(t)$  and  $J_{\text{hist}}(t)$*  combines the previous two adjustments by averaging the ratios:

$$\hat{J}_c(t+1) = \frac{1}{2k} \sum_{i=1}^k J_i(t+1) \left[ \frac{J_c(t)}{J_i(t)} + \frac{J_{\text{hist},c}(t+1)}{J_{\text{hist},i}(t+1)} \right]. \quad (3.59)$$

- The *adjusted by both  $J(t)$  and  $J_{\text{hist}}(t+1)$  and weighted by inverse of distance* approach combines the previous three adjustments:

$$\hat{J}_c(t+1) = \frac{\sum_{i=1}^k [2d_i]^{-1} J_i(t+1) \left[ \frac{J_c(t)}{J_i(t)} + \frac{J_{\text{hist},c}(t+1)}{J_{\text{hist},i}(t+1)} \right]}{\sum_{i=1}^k d_i^{-1}}. \quad (3.60)$$

In summary, the latter five calculation methods are heuristic attempts to improve the *straight average* forecast by taking into account the relative distance of the neighbours from the forecast point or the relation of key neighbour state elements to the corresponding forecast point elements.

As a forecast performance the results are compared to the forecast

$$\hat{J}_c(t+1) = \frac{J_c(t)}{J_{\text{hist},c}(t)} J_{\text{hist},c}(t+1) \quad (3.61)$$

and the seasonal ARIMA(1,0,1)(0,1,1)<sub>672</sub> parametric model using the mean relative error  $e_{\text{MRE}}$ . Although the *adjusted by  $J(t)$*  method is the best performing  $k$ -NN forecast method ( $e_{\text{MRE}} = 0.0947$ ), it does not match the performance of the seasonal ARIMA forecasts ( $e_{\text{MRE}} = 0.0883$ ). Moreover the *adjusted by  $J_{\text{hist}}(t + 1)$*  ( $e_{\text{MRE}} = 0.105$ ), *straight average* ( $e_{\text{MRE}} = 0.1007$ ), and *weighted by inverse of distance* ( $e_{\text{MRE}} = 0.0996$ ) approaches are outperformed by the naive model ( $e_{\text{MRE}} = 0.0995$ ).

### 3.3 Neural Networks

*Neural networks* is a broad term covering many different architectures. The operation of these architectures can vary enormously. However, all neural networks share some basic common features. In the following only a short introduction to some basics of neural networks is given. Going more into detail would extend the scope of this work, so the interested reader is referred to the extensive existing literature, such as [Bishop, 1995, Hecht-Nielsen, 1990, Reed and Marks, 1999, Ripley, 1996], and Dougherty [1995] gives a short review about Neural Networks applied to transport problems.

#### 3.3.1 Introduction to Neural Networks

Neural networks are composed of a number of very simple processing elements, known as *neurons* (see Fig. 3.4). These neurons take in data from a number of sources and compute an output dependent in some way on the values of the inputs using an internal *transfer function*. The neurons are joined by weighted connections. Data flow along these connections and are scaled during transmission according to the values of the weights  $w_{i,m}$ .

In general terms the relationship between the inputs  $(x_0, x_1, \dots, x_N)$  of neuron  $m$  and its output  $y_m$  is given by

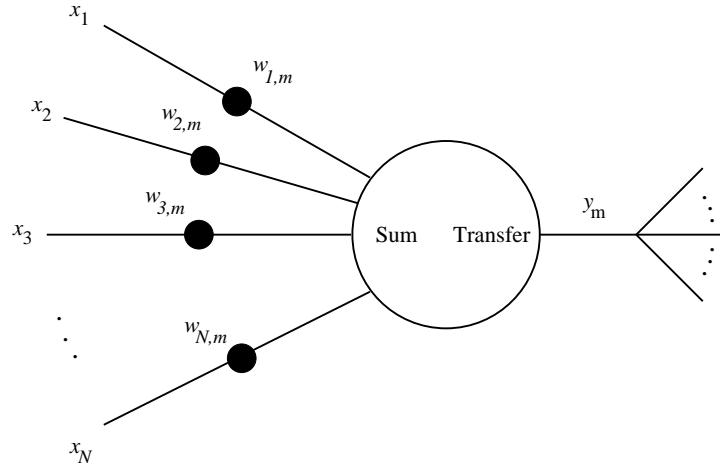
$$a_m = \sum_{i=0}^N w_{i,m} x_i, \quad (3.62)$$

$$y_m = f(a_m). \quad (3.63)$$

The transfer function  $f(a)$  is typically a nonlinear function such as a sigmoid. Note that the output can be binary or continuous.

The output of a particular neuron is a part of the input received by another. To communicate with the outside world some connections take data in from an external source, whilst others pass data back out. The neural network's functionality is bound to the values of the connection weights, which can be updated over time, causing the neural network to adapt or learn.

Although the network is basically described by Eqs. 3.62 – 3.63, several simplifications are needed to work with neural networks. Some of them are introduced in the following.

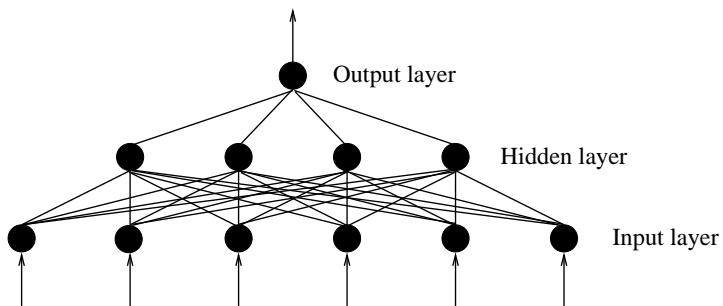


**Fig. 3.4.** A neuron  $m$ . Each input  $x_i$  of other neurons are weighted with the weights  $w_{i,m}$ . After a summation a transfer function computes the output  $y_m$ .

The neurons are arranged in layers with the existence or absence of a connection between two neurons. The input layer is connected to the output layer with one or more hidden layers. A typical scheme of a feedforward network can be seen in Fig. 3.5.

The connection weights have a minimum and a maximum strength. Furthermore all the neurons within a layer, or often the entire network behave in the same way. This means, that all neurons use the same formula to compute an output from the weighted inputs.

Finally, a rule is defined which determines how and when the weights are changed. Neural networks can be categorised according to the type of learning rule employed.



**Fig. 3.5.** A typical feedforward neural network with an input, a hidden, and an output layer.

In *supervised learning*, the computed output

$$y_m = f \left( \sum_j x_j w_{j,m} \right) \quad (3.64)$$

is compared with a desired one  $d_m$ , and a global error function is computed, which updates the weights. An easy example is the perceptron learning rule for single layer networks:

$$\Delta w_{i,m} = \eta \left[ d_m - f \left( \sum_j x_j w_{j,m} \right) \right] x_i, \quad (3.65)$$

with the learning rate  $\eta$  (the choice of learning rate  $\eta$  just changes the scaling of  $w_{j,m}$ ).

In a similar manner works the *delta learning* rule, also known as the *least-mean-squares* rule, the *Widrow-Hoff* rule, or the *ADALINE* rule (for **AD**aptive **L**INear **E**lement). It has the same weight update equation 3.65 but holds for networks with a continuous transfer function  $f(a_m)$ .

It is shown mathematically that single layer perceptrons do not converge to a solution for a problem which is not linearly separable [Minsky and Papert, 1969]. A method that solves this problem is the *back-propagation algorithm*, which works for multi layer networks.

The back-propagation algorithm calculates how the error changes as each weight is increased or decreased slightly. Therefore the error derivative of the weights  $\frac{\partial}{\partial w_{i,m}} e$  is computed. This can be done by first computing the rate  $r_m$  at which the error  $e$  changes as the output of a neuron  $m$  is changed. For the output layer  $\mathcal{L}$  this is just the difference between the current and the desired output. To compute it for a neuron  $n$  in the hidden layer just before the output layer  $\mathcal{L}$ , first all weights  $w_{n,i}$  between that neuron and the output layer neurons  $i$  have to be multiplied by  $r_i$  and then the products are summed up:

$$r_n = \sum_{i \in \mathcal{L}} w_{n,i} r_i. \quad (3.66)$$

After calculating  $r_n$  for all neurons  $n$  in the hidden layer just before the output layer all other rates can be computed in the same manner moving from layer to layer in a direction opposite the way the activities propagate through the network. This is what gives back propagation its name. Once the  $r_m$  have been computed for all neurons, it is straight forward to compute the derivative for each incoming connection of the neuron. It is just the product of  $r_m$  and the activity through the incoming connection.

Note that the majority of the applications used for traffic forecasts are back-propagation neural networks, also known as multi layer perceptrons (MLP).

In *reinforcement learning* the network is just trained by inform whether it performed well or badly for each iteration. If the neural network performed well,

processing elements within the network which are outputting an active signal are examined, and the strength of the most active input connections to these neurons is increased. Often this is done on a *winner takes all* basis, that means only the connection providing the strongest input is updated. Conversely, if the neural network performed badly the strength of some connections are weakened. Reinforcement learning is often known as Kohonen learning [Kohonen, 1995].

*Self-organising networks* operate purely on input data. All criteria for updating the weights are determined internally. The main use of such networks are classification problems where it is not certain beforehand what the definition of the classes should be. An example of this type of network is that of the adaptive resonance theory (see, for instance, [Grossberg, 1976]).

Finally there are *combined networks* that use different learning methods for the connections between different layers.

### 3.3.2 Neural Networks for Traffic Forecast

The interesting point in the work of Florio and Mussone [1996] is the chosen input data. In opposite to most other investigations of traffic forecasts using neural networks and even other algorithms, the authors do not only use flow, density, and velocity, but also the percentage of heavy good vehicles, the brightness from 1 (darkness) to 6 (bright light), the visibility  $s$  from  $s = 0$  m to  $s > 520$  m, three different weather conditions (clear, rain, snow/ice), and the state of variable message signs (absence, accident, queue, fog).

The data collection environment is employed on the three-lane motorway section Padua-Mastre in Venice, Italy. It runs for a distance of 11 km from the Dolo tollgate to Mestre tollgate. The data are grouped according to detection sections and are subdivided according to the type of information detected by each station. The difference essentially consists of the different sampling period, which ranges from 30 s to 120 s for the flow, the density, and the velocity. Because meteorological data, brightness, and messages of variable message signs occur only in certain stations, for all other detector stations the data from the closest section are assigned to them.

The data are preprocessed and filled into a 2-hidden-layer feedforward neural network. The input layer consists of 9 neurons where the data mentioned above and the forecast horizon are filled in. The first hidden layer consists of 8, the second of 4 neurons. As output the flow, the density, and the velocity are generated. The transfer function is a sigmoid one.

The authors state, that the best results are received forecasting the velocity with a mean relative error of  $e_{MRE} = 0.05$ , whereas density and traffic flow are consisting of unpredictable measurement noise that lead to  $e_{MRE} = 0.30$ . Nevertheless, the work is a good example for the attempt to put any kind of available data in a neural network to forecast traffic data.

Chen et al. [1997] use a MLP with two hidden layers to investigate the effects of detector spacing on the short term traffic flow forecasting performance. The

transfer function is a sigmoid one. Different network architectures are used, but unfortunately details are not given in the article.

At first the authors use simulated data generated by the AIMSUN [Ferrer and Barceló, 1993] traffic simulation software calibrated to the M2 motorway in Kent corridor, Great Britain. Simulated loop detectors are placed at 500 m intervals and also on on-ramps. The data are averaged over 5 min periods. To generate one-step-ahead forecasts of one detector station a back-propagation neural network with a single output is used. The output is the flow at time step  $t + 5$  of the forecast loop detector. As input serve the traffic flow off all other loop detectors at time step  $t$ . As the experiments progress, detectors are gradually removed from the set of inputs and the effect on forecasting performance is noted.

To train the network four different simulated flow profiles are generated that range from quite light flow to very heavy flow. Three of them are used for training a neural network and the randomly selected remaining one is used for testing. Forecasting performance is observed by plotting the simulated and forecast values over a set of input-output pairs. In order to obtain an optimal minimum, each neural network model is trained by ten different initial weights and the best is saved.

For detector spacings from 500 m to 3,000 m the relative root mean square error  $e_{RRMSE}$  of the forecasts is calculated. In case of 500 m detector spacing hold  $e_{RRMSE} = 0.25$  and for 3,000 m  $e_{RRMSE} = 0.3$ . At this point it has to be mentioned, that it is always difficult to compare forecast results of different kind of data, in this case even simulated traffic data. Nevertheless it can be said, that for five minute aggregated data a relative error of about 0.25 is very high, especially with information about upstream data. This means, that the neural network is not able to reproduce the spatial temporal traffic model that is used by the simulation.

For completeness it should be mentioned, that the work is extended to use real traffic data from the M25 in London, Great Britain. Loop detectors are placed at 500 m intervals and measure among others minute aggregated data of traffic flow. As input for the neural network only the upstream and downstream detectors are used. Flow data between 06:00 to 21:00 of 1996/10/01 to 1996/10/24 are used for training the network and data from 1996/10/25 to 1996/10/31 for testing. The authors conclude, that although the real world data is very noisy the forecasting performance is of “reasonable accuracy”.

Dougherty and Cobbett [1997] use a MLP to forecast flow, velocity, and occupancy. The analysis is conducted with loop detector data at about 17 points along inter-urban motorways in the Rotterdam/The Hague/Gouda triangle, The Netherlands. Data are available in 1 min intervals for each lane, but the authors state that initial experiments have not been very successful because of the large minute by minute stochastic variation. The data are therefore combined across the carriageway and averaged over a 5 min period.

During a large part of the work the authors deal with the problem of the input selection. Because the network is more effective, if data from previous time steps

are available, this gives hundreds of possible input data points. On the other hand the required computational effort rises approximately with the square of the number of inputs to the neural network. As a trade-off between the accuracy of the model and the required computational effort, the authors decide to use a MLP with 40 inputs and a single hidden layer with 10 neurons. To select the inputs a method called “elasticity analysis” [Dougherty and Joint, 1992] is used. Finally, the root mean square error proportional  $e_{RMSEP}$  of one-, three-, and six-step ahead forecasts of four different detector stations is calculated for all three measures: flow, velocity, and occupancy.

A comparison with the naive forecasting model is deflating. Forecasting the velocity and the occupancy the neural network model is outperformed. Only forecasting the traffic flow the neural network is able to keep up with the naive forecasting model. Considering, that also in this case adjacent data is used for forecasting, the results are poor. The authors conclude, that more tests are needed with different neural network architectures.

This is later done in [Dia, 2001], where four different neural network architectures are used to generate one-step-ahead forecasts of 20 s aggregated velocity data of four loop detector stations placed on a 1.5 km section of the Pacific Highway between Brisbane and the Gold Coast in Queensland, Australia. The raw data are collected over a 5 hour period on two days in April 1995 and consist of 5,000 observations. 3,000 of them are used for network training, 500 for cross validation and 1,500 for testing. The velocity measurements from the current time interval at a given detector station serve as input and the output comprises the velocity measurements at the same station at the future time interval.

A number of experiments are conducted to determine the best neural network architecture. Apart from MLP, a *recurrent network*, a *time lagged recurrent network* (TLRN), and a *hybrid network* is tested. The recurrent network feedbacks the hidden layer to itself, the TLRN is a MLP that is extended with short term memory structures, and the hybrid networks tested in the study comprise a combination of a TLRN and a method that is called *Principal Component Analysis*. See [Principe et al., 2000] for a detailed description of the mentioned components. Each of the above models are trained on the training data set for a maximum of 1,000 cycles. At the end of each cycle, the trained model is tested on the cross validation set and the mean square error  $e_{MSE}$  of the forecast is computed. If  $e_{MSE}$  decreases, the model is saved and the training is continued. The training is stopped, when  $e_{MSE}$  on the cross validation data increase by a specified threshold. Comparing the  $e_{MSE}$  during training, cross validation, and testing, the authors decide to use the TLRN for forecasts from 20 s to 15 min. The mean relative error  $e_{MRE}$  ranges from 0.06 for a forecast horizon of 20 s to 0.16 for 15 min. In a similar manner travel times between two detector stations are estimated using the TLRN. Therefore the velocity and the flow at time  $t$  of the two loop detectors serve as input of the network. The output is the travel time at  $t + n$  where  $n = 20$  s to 15 min. To train the network the outputs are compared with the historical travel time values which are determined by video taping the sections. The  $e_{MRE}$

of the forecast travel times ranges from 0.05 (20 seconds) to 0.07 (15 min). As already mentioned above, comparing the forecast results is very difficult. But in regard to the effort that come with this approach and according to the fact, that it has not been mentioned in the work, at least in which limits a certain velocity breakdown is possible to be forecast, a large advantage of the neural network cannot be established.

In spite of this, neural networks are in general often considered in the context of traffic forecast or even traffic control. This comes, because pattern recognition and the assignment to certain events is a basic task in traffic forecasting as can later be seen in Chapter 5. In the next section several approaches are proposed where traffic time series are manually analysed to develop classifications that help to forecast recurring structures of traffic time series. The idea using neural networks is, to automatise the cognitive abilities that are required for this analysis. Unfortunately it seems, that large improvements are needed before neural networks can be applied in operation for traffic forecast applications.

Nevertheless, recent interesting and also promising works with artificial neural networks should be mentioned at this point. A performance evaluation of neural network models in traffic flow forecasting can be found in [Yun et al., 1998]. Of basic interest is often the impact of bottlenecks, in that manner in [Ledoux, 1997] queue lengths at urban intersections and in [Vemuri et al., 1998] traffic delays in motorway construction zones are forecast with artificial neural networks. In [McFadden et al., 2001] the focus is to forecast velocities on two lane rural motorways. Qiao et al. [2001] investigates the problem of forecasting traffic flow dispersion what is related to the field of traffic simulation. Murat [1999] even applies an artificial neural network to forecast long term traffic volumes over years. Short term traffic forecasts are investigated of urban traffic flow in [Yin et al., 2002], of freeway traffic flow in [Abdulhai et al., 1999], and of link travel times in [Innاما, 2001, Park and Rilett, 1999, Rilett and Park, 2001, van Lint, 2004].

### 3.4 Heuristics and Knowledge Based Systems

In the original sense, heuristics means a strategy, that forms a sensible way to the solution. What is meant here is a method, that rests on the assumption, that there are similar traffic conditions on similar days. The approach is as follows: if all influences at a certain date are known and if there are data that are measured on days with the same influence, this data is used as an approximation of the new day.

First systematic approaches of analysing historical values for traffic flow forecast are described in [Leutzbach and Siegner, 1974]. The authors use hourly aggregated data of two loop detectors. Their basic approach of long term forecasts is based on the observation that the traffic flow strongly depends on the day of the week and the time of day. Thus, the traffic flow time series of each detector is

subdivided in  $7 \times 24 = 168$  partial time series. Each partial time series  $J(t)$  is estimated as the sum  $\hat{J}(t)$  of a trend  $J_T(t)$  and a periodic part  $J_P(t)$ :

$$\hat{J}(t) = J_T(t) + J_P(t), \quad (3.67)$$

whereby  $J_T(t)$  is estimated by a linear regression and  $J_P(t)$  is described by a Fourier series:

$$J_P(t) = a_0 + \sum_k \left[ a_k \cos \frac{2\pi k t}{p} + b_k \sin \frac{2\pi k t}{p} \right]. \quad (3.68)$$

The period  $p = 52$  can be interpreted as the interval of one year. The authors state, that the significance of the coefficients  $a_k$  and  $b_k$  is tested statistically and that normally no more than 2 coefficients of  $a_k$  and  $b_k$  are needed to estimate the periodic components. The forecast is just the extrapolation of  $\hat{J}(t)$ .

Strong deviations are interpreted as special events. The authors distinguish between chronological predictable events like holidays and chronological unpredictable events like extreme weather conditions or accidents. The predictable events, that are considered in the forecast, are further distinguished in events that occur on a fixed weekday (like the holidays of Pentecost and Corpus Christi that occur on a Thursday) and such, that occur on a fixed date (like labour day at the 1st of May or Christmas)<sup>1</sup>.

To adjust the time series, the authors first forecast the values with the model  $\hat{J}(t)$  and then manually pick out the days with bad forecast results. If that are days with special events, the values are replaced with adjacent values. With those adjusted time series new models  $\hat{J}_n(t)$  are calculated.

Finally, the special events are considered. Therefore, for the events that are on fixed weekdays, the differences between the measured traffic flow at these days and the model  $\hat{J}_n(t)$  are averaged over all special days. For the events at a fixed date just the historical mean of these days is calculated whereby the trend is added.

In their example they find 48 special days in one year. The long term forecast results are partially bad because in the meanwhile the infrastructure has been changed. In [Siegener and Schmitt, 1980] the relative root mean square error  $\epsilon_{RMSE}$  dithers between 0.221 and 0.3 using this method. Note, that in opposite to short term forecasts this method can be used for arbitrary forecast horizons. In [Krause, 1988] an approach of forecasting incidents is investigated with the objective to relieve the overload using dynamic traffic assignment. The facility is placed around Dernbach and Koblenz and separates the traffic going north to the motorways BAB 3 and BAB 61. The traffic data are stemming from 30 junctions fully equipped with loop detectors at the BAB 3, BAB 48, and BAB 61

---

<sup>1</sup>Note that such investigations strongly depend not only on the country but also on the particular region and even on the time. The 17<sup>th</sup> of June, for instance, has been a holiday at this time, but since the German Unification the holiday is displaced with the 3<sup>rd</sup> of October.

in a region around Cologne, Bonn, and Koblenz, Germany. To forecast the traffic conditions and to detect the incidents in advance a forecast model along the time axis and a spatial temporal forecast is evolved.

For the spatial temporal forecast the author uses a traffic model to deduce travel times from estimated traffic flow time series. The basic assumption is, that up to a point of overloading the traffic flow propagates undisturbed through the network. The traffic flow is the only value that is really forecast. The density and velocity is calculated using a spatial temporal traffic model.

The spatial temporal model is based on the continuity equation. The density is calculated using the difference of the traffic flow estimated at the beginning and at the end of a section divided by the length of the section. Then an empirical relation between the density and the travel time is used to deduce the travel time. Because details of spatial temporal traffic models are not part of this work the interested reader is referred to the original work.

What is of basic interest here, is the traffic flow forecast at each single section without the influence of the adjacent sections. The basis of the traffic flow forecast at the different sections  $s_i$  are classified traffic flow time series  $J_{\mathcal{G}_{D_n}, s_i}(t)$ . Thereby is  $\mathcal{G}_{D_n}$  a group of certain days  $d \in \mathcal{G}_{D_n}$  that are described by  $D_n$ . Three types  $D_n$  are used: working days, Saturdays, and Sundays.

To enhance the forecast results a feedback model is proposed. This feedback model is build up the principle, that each forecast value is compared with a measured value after passing the time period of the forecast horizon. Therefore the forecast error  $\epsilon_{t_0}$  is calculated at the time step  $t_0$  the forecast is generated. This is done at each time step. Using these values a weighted forecast error  $\epsilon_{t_0}^*$  is calculated:

$$\epsilon_{t_0}^* = \beta \epsilon_{t_0} + \beta [1 - \beta]^1 \epsilon_{t_{0-1}} + \dots + \beta [1 - \beta]^{m-1} \epsilon_{t_{0-m+1}}. \quad (3.69)$$

The choice of the parameter  $\beta$  and the number of analysis intervals  $m$  is from essential relevance. The cumulative sum of the weights is smaller than 1 for finite  $m$ . On the other hand the author does not want to choose too large  $m$  because too old forecast errors should not influence the result. At this point the author states, that for  $\beta$  should hold  $0.2 < \beta < 0.4$  and for  $m$ :  $6 \leq m \leq 12$ . Note that the cumulative sum of the weights is 0.74 for  $\beta = 0.2$  and  $m = 6$ , so  $\epsilon_{t_0}^*$  is systematically a factor 0.26 too small.

The first idea of the author is now to adjust the forecast traffic time series  $J_{\mathcal{G}_{D_n}, s_i}(t_p)$  for each future time step  $t_p$  with the weighted forecast error  $\epsilon_{t_0}^*$  at the point in time  $t_0$  the forecast is generated:

$$J_{\mathcal{G}_{D_n}, s_i, t_0}(t_p) = J_{\mathcal{G}_{D_n}, s_i}(t_p) + \epsilon_{t_0}^*. \quad (3.70)$$

Some trials show that thereby oscillations could appear. The forecast values fluctuate around the measured values. If the forecast values are too small for a long period, a correction addend  $\epsilon_{t_0}^*$  is calculated that increases the forecast traffic flow. In doing so, it is now possible, that the new forecast value is too high. Until

this value is considered in  $\epsilon_{t_p}^*$  a certain time interval, the forecast horizon  $t_p - t_0$ , has to be passed.

This is why the author finally introduces a damping coefficient  $\eta_{s_i, t_0, t_p}$  for each cross section  $s_i$  and future time step  $t_p$ :

$$\eta_{s_i, t_0, t_p} = [1 - \eta_{s_i, t_0}]^{[t_p - t_0]}. \quad (3.71)$$

Here is  $\eta_{s_i, t_0}$  the section based damping coefficient. For completeness it has to be mentioned, that a damping coefficient is also calculated for the process of the propagating traffic flow.

Then the forecast traffic flow  $J_{G_{D_n}, s_i, t_0, a}(t_p)$  results as:

$$J_{G_{D_n}, s_i, t_0, a}(t_p) = J_{G_{D_n}, s_i}(t_p) + \epsilon_{t_0}^* \eta_{s_i, t_0, t_p}. \quad (3.72)$$

The investigations are done with hardcopies of 15 min aggregated traffic flow values at the carriageway and the differences at on- and off-ramps. The loop detectors at on- and off-ramps measure just the traffic flow, the loop detectors on the travelled way also the velocity.

Because the time series forecast is embedded in the spatial temporal model it is hard to separate the effect of the feedback and of the damping from the results given in the work. The author states that in general the feedback model improves the results of the forecasts during congestions but at the expense of forecast results during free flow situations. The damping coefficient improves the results in free flow conditions in opposite to the results in congested situations. The author states that for  $\eta_{s_i, t_0}$  should hold  $0.05 < \eta_{s_i, t_0} < 0.2$ .

An approach in forecasting travel times is done by Janko [1994]. He uses historical travel times  $\Delta t_{s, m}$  of certain sections  $s$  at certain times  $m$  to calculate classified travel time time series  $\overline{\Delta t}_{s, m}$  and uses them for forecast. To make the model adaptive he applies the single smoothing model to the historical time series in the following way:

$$\overline{\Delta t}_{s, m, i} = \alpha \Delta t_{s, m, d_c} + [1 - \alpha] \overline{\Delta t}_{s, m, i-1}. \quad (3.73)$$

Thereby is  $\overline{\Delta t}_{s, m, i}$  the new smoothed classified travel time for the time interval  $m$  on the section  $s$  calculated with certain days with a similar influence on traffic. Accordingly,  $\overline{\Delta t}_{s, m, i-1}$  is the same value calculated without the travel time data  $\Delta t_{s, m, d_c}$  of the most recently day  $d_c$  with a similar influence on traffic. The parameter  $\alpha$  is just the smoothing coefficient according to Eq. 3.18.

The finally forecast travel time  $\Delta t_{s, m}^*$  is adjusted to the ratio of the most recently measured travel time  $\Delta t_{s, m-1, d_c}$  (that is, this of the time interval  $m-1$ ) to its historical value  $\overline{\Delta t}_{s, m-1, i}$  as follows:

$$\Delta t_{s, m}^* = \overline{\Delta t}_{s, m, i} \frac{\Delta t_{s, m-1, d_c}}{\overline{\Delta t}_{s, m-1, i}}. \quad (3.74)$$

The idea of exponential smoothed classified traffic time series is later again used by Zackor et al. [1996]. The authors propose a method for forecasting traffic flow

up to 48 hours ahead using classified traffic flow time series and current traffic flow data. They divide the days in six groups of days with a similar influence on traffic (Monday, TWT, Friday, Saturday, Sunday, and holidays, whereby TWT means Tuesdays until Thursdays).

The forecast traffic flow  $\hat{J}_{d,s}(i)$  at the cross section  $s$  for the particular group  $d$  in the time interval  $i$  consists of the following factors and one addend:

$$\hat{J}_{d,s}(i) = \overline{\text{MDT}}_s \cdot b_{d,s}(i) \cdot f_{d,s} \cdot f_{m,s} \cdot f_{\text{hol},s}(i) \cdot f_{U,s} + \Delta q_{\text{SE},s}. \quad (3.75)$$

Here  $\overline{\text{MDT}}_s$  is the mean daily traffic flow of the cross section  $s$ , that means an overall average traffic flow for all days.

The factor  $b_{d,s}(i)$  is the ratio of the traffic flow  $J_{d,s}(i)$  in the time interval  $i$  of the cumulative sum of the traffic flow measured during the days of the group  $d$ :

$$b_{d,s}(i) = \frac{J_{d,s}(i)}{\sum_{i=1}^{i_{\text{day}}} J_{d,s}(i)}. \quad (3.76)$$

Here  $i_{\text{day}} = 1440\Delta t^{-1}$  min is the number of time intervals of the length  $\Delta t$ . To get the factor  $b_{d,s}(i)$  adaptive it is exponentially smoothed with every new day belonging to the same group  $d$ , so far this day is not influenced by one of the following exceptional events:

- begin or end of school holidays,
- begin or end of long weekend,
- big events like fairs, exhibitions, or sporting events,
- known construction areas, and
- time periods when variable message signs indicate a detour.

With the daily factor  $f_{d,s}$  the differences from the groups of days to the weekly average are taken into account. It is the distribution of the mean daily traffic flow  $\text{MDT}_{d,s}$  of group  $d$  to the whole mean daily traffic flow  $\overline{\text{MDT}}_s = 6^{-1} \sum_d \text{MDT}_{d,s}$  during all the six groups  $d$ . In the same manner like  $b_{d,s}(i)$  it is adapted after every week using exponential smoothing, whereby weeks with a strong influence of holidays or big events are not considered.

In a similar manner the factor of the seasonal influence  $f_{m,s}$  reflects the monthly differences. It is calculated as

$$f_{m,s} = \frac{\sum_{l=1}^{N_m} \text{MDT}_{l,s}}{N_m \overline{\text{MDT}}_s}. \quad (3.77)$$

Here  $\text{MDT}_{l,s}$  is the mean daily traffic flow at day  $l$  and  $N_m$  the number of days during the month  $m$ . Because the time of the school holidays differs from year to

year and the traffic due to school holidays would influence this calculation, the effect is filtered with a denominator for the days during school holidays:

$$\text{MDT}_{l,s} = \sum_{i=1}^{i_{day}} \frac{J_{l,s}(i)}{f_{\text{hol},s}(i)}. \quad (3.78)$$

Like the other factors the seasonal factor  $f_{m,s}$  is exponentially smoothed.

The holiday factor

$$f_{\text{hol},s}(i) = \frac{b_{d,\text{hol},s}(i)}{b_{d,s}(i)} \quad (3.79)$$

describes the relationship of the hourly factor  $b_{d,\text{hol},s}(i)$  of day group  $d$  measured at a day that is influenced by school holidays to the normal hourly factor  $b_{d,s}(i)$ . Outside time periods with school holidays the holiday factor is one and it is only calculated for cross sections  $s$  with a significant influence due to holidays.

The factor  $f_{U,s}$  reflects the increasing traffic volume in regions with high spare time activities especially in times of good weather conditions. The authors propose, that until further data analyses admit a more exact classification, the factor should be  $f_{U,s} = 1.1$  in times of good weather conditions and  $f_{U,s} = 1.0$  else.

Finally, the addend  $\Delta q_{\text{SE},s}$  in Eq. 3.75 describes the extra distribution of traffic flow because of special big events like fairs or sport events. Because this value is strongly dependent on the number of the expected visitors of these events, the authors state, that it should be approximated analysing the increasing traffic in the spatio temporal environment of the big events.

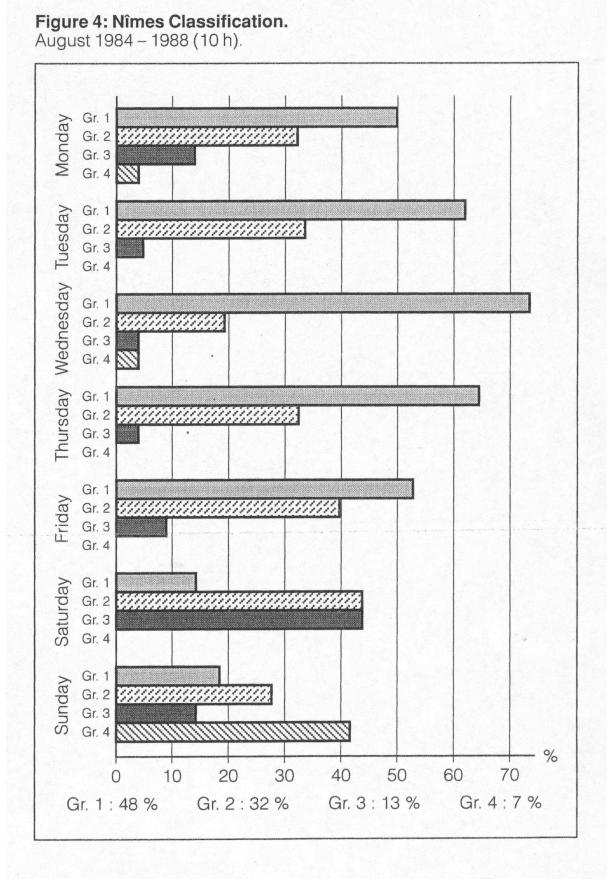
To consider the current traffic data the algorithm proposed in [Zackor, 1976] (see Sec. 3.5.1 on page 49) is used. For completeness it has to be mentioned, that the cross section based traffic flow forecast is combined with a spatio temporal traffic model.

The forecast method is tested on the motorway network around the cities Frankfurt and Stuttgart in Germany. The data are provided by 41 loop detectors. Those are analysed during 100 days of different time periods and traffic situations for the calibration of the model.

As results the authors state, that the mean relative error  $e_{\text{MRE}}$  is usually under 0.1, but only in case of suitable classified traffic time series. The distinction in 6 daily groups seems to be suitable, but under special conditions additional daily groups should be defined. In case of exceptional events one has to count on increasing forecast errors, especially the duration and intensity of a traffic incident cannot be forecast without external information.

A daily independent classification is proposed within the system ATHENA [Danech-Pajouh and Aron, 1991] to calculate one-step-ahead forecasts of half hourly traffic flow data. The data used are flow rates for southbound traffic during the months July and August of the years 1984 until 1990 from four counting stations near Beaune, France. The data are the same that are used by the ARIMAX method (see Sec. 3.1.3 on page 18) and later by the KARIMA method (see Sec. 3.5.2 on page 56). The geographical relationship can be seen in Fig. 3.2.

The traffic data are transformed in profile vectors, containing the half hourly traffic flow divided by the total cumulative traffic flow since midnight. Thereafter the profile vectors are divided in exact four classes using the  $k$ -means cluster algorithm [McQueen, 1967]. Then, for every class and every half an hour a regression model linking the traffic flow at hour  $t + 1$  to the flow at the hours  $t$ ,  $t - \frac{1}{2}$ ,  $t - 1$ , and so on, is proposed. For example, if the forecast period is 8:00 to 18:00 (an amplitude of ten hours) and the number of classes created is 4 per hour, then the number of models for a given month is equal to  $10 \times 2 \times 4 = 80$ . Unfortunately it is not given how the regression model works on detail.



**Fig. 3.6.** Day class histogram within the ATHENA model for a particular profile vector for 10:00 and the month August from 1984 - 1988 (picture from [Danech-Pajouh and Aron, 1991]).

To forecast the traffic flow the most recently measured traffic profile vectors are compared with these of the classes. The closest class to the observations in terms of the route mean square error  $e_{RMSE}$  is chosen to be the appropriate one for forecast. Then the regression model is applied. The method is tested on several

freeways in France and the authors state the  $e_{RMSE}$  to lie between 0.05 and 0.20. The interesting part of the work is the automatic daily independent classification of the traffic flows. In other works that deal with the problem of historical traffic time series the classification is done manually. In Fig. 3.6 the results can be seen for a particular detector location. A reason for the number of exactly 4 classes is not given, so it could make sense to investigate other classifications.

In [Lamboley et al., 1997, van Iseghem and Danech-Pajouh, 1999] a traffic forecast method for urban traffic is presented. The test network is composed of the east half of the Boulevard Peripherique and a part of the Paris urban network in France. The data of the years 1992 to 1994 are used to forecast several days of the year 1995. To characterise the traffic four indicators are introduced:

- indicator of mean flow,
- indicator of travel time,
- indicator of fluidity of the velocity in regard to free flow traffic,
- indicator of delay of the velocity in regard to free flow traffic.

All of the indicators are composed of the traffic values flow and the occupancy. By using these indicators standard days are identified. After studying the hourly variation in traffic annual and daily periods are specified. A distinction is made between three groups of days: weekdays (outside summer holidays), Saturdays (outside summer holidays), and Sundays (outside summer holidays). On the basis of the daily and hourly variations in flow curves the authors decide to divide the year into three periods:

- school term time: September to March,
- school term time with many disturbances due to a large number of public holidays,
- summer holidays (July and August).

The day is divided into five periods:

- night: 21:00 to 4:00,
- build up: 5:00 to 7:00,
- morning peak: 8:00 to 11:00,
- afternoon: 12:00 to 15:00,
- evening peak: 16:00 to 20:00.

In the first step only those days without exogenous events (bad weather, strikes, demonstrations, or public holidays) are investigated. To forecast the traffic flow on the basis of the daily and hourly variations in flow curves the following multiple regression equation is proposed to forecast the traffic flow one  $\hat{J}_{p,j+1,t}$  and two  $\hat{J}_{p,j+2,t}$  days in advance, respectively:

$$\hat{J}_{p,j+1,t} = a + bJ_{p,j-1,t} + cJ_{p,j-6,t} + dS + eD + fL + gH + hE, \quad (3.80)$$

$$\hat{J}_{p,j+2,t} = a + bJ_{p,j-1,t} + cJ_{p,j-5,t} + dS + eD + fL + gH + hE, \quad (3.81)$$

where

- $\hat{J}_{p,j+1,t}$  forecast flow at the day  $j$  for the day  $j + 1$  for the periods  $p$  and  $t$ ,
- $\hat{J}_{p,j+2,t}$  forecast flow at the day  $j$  for the day  $j + 2$  for the periods  $p$  and  $t$ ,
- $J_{p,j-1,t}$  measured flow of the previous day  $j - 1$  during the periods  $p$  and  $t$ ,
- $J_{p,j-5,t}$  measured flow five days before ( $j - 5$ ) during the periods  $p$  and  $t$ ,
- $J_{p,j-6,t}$  measured flow six days before ( $j - 6$ ) during the periods  $p$  and  $t$ ,
- $p$  period of the year (three modalities),
- $t$  period of the day (five modalities),
- $j$  day the forecast is calculated,
- $S$  binary variable for Saturday,
- $D$  binary variable for Sunday,
- $L$  binary variable for Monday,
- $H$  binary variable for all public holidays,
- $E$  binary variable for the day before and the day after all public holidays and the period between 1st and 14th of August.

The parameters  $a, b, c, d, e, f, g, h$  are coefficients in the equation that have to be empirically determined.

In a second step the effects of exogenous events like weather and strikes are studied but it is not clear, how the results are implemented in the forecast procedure. The authors give the mean relative error  $e_{\text{MRE}}$  forecasting the traffic flow to be frequently less than 0.1, whereas it is 0.15 to 0.2 in forecasting the indicator of delay.

To fill forecast traffic data into a macroscopic traffic simulation model, Meißner [1998] needs forecast traffic values of the flow and the velocity. For this two models are used. The first one is a linear regression method. A regression line is calculated through the most recent measured data minimizing the distances to the data points. For future values the line is used as a forecast.

The second one is a heuristic based method. The days are classified in certain groups with similar traffic flow. Because for this work very few data are available, the average flow and velocity of only three working days is used as a typical traffic time series for working days. Furthermore the resulting time series is smoothed intensely, but the author does not state over which time interval.

The more interesting part of this work are considerations about pattern matching in regard to changes of certain peaks within a special class. The point in time

of special phenomena like morning peak or afternoon rush hour can differ from day to day. In the same manner traffic time series with similar characteristics can differ in their absolute values.

This is why an approach is presented that should move the classified time series along the abscissa and the ordinate in such a way, that it is in good agreement with the recently measured values. Therefore, for each time series a regression line is calculated at each time step  $t$  for the  $N$  most recent values. The slope of this line is taken as the slope at the time step  $t_{1/2} = t - N/2$ . As measure for the agreement of the classified with the current traffic time series acts the absolute difference of the slopes at time step  $t_{1/2}$ . This difference is minimised moving the classified time series along the abscissa, that means along the time. The author states, that if this movement is too large, the classified time series is not typical for the particular day. The proposed limit for this method is given as approximately one hour and thirty minutes.

If the minimum is found the classified traffic time series is shifted along the ordinate until the traffic values are in agreement. For traffic forecasts this shifted classified traffic time series is used.

The author compares the two methods calculating one hour forecast in regard to the real time. Unfortunately the number of values used for each regression is not given. To compare the forecast traffic time series with the measured one the mean absolute error  $e_{MAE}$  is used for both flow and velocity and the mean error  $e_{ME}$  is used for the flow. Using these quality measures, the heuristic based method seems to be superior, but as the author embeds both methods in his simulation framework, the linear regression method seems to be more adequate. The author ascribes this to the effect, that the current traffic time series does not follow the classified one in case of unexpected strong deviations in the measured values.

For this reason, in many works classified traffic time series are combined with regression techniques. It is quite remarkable, that in most cases the historical traffic time series are only used for long term forecasts, whereas for short term forecasts those methods are preferred, that use the most recent traffic data. A large part of the following section describes works dealing with this problem.

For completeness, some more interesting works about heuristics and knowledge based systems in regard to traffic forecast should be mentioned. Classified traffic time series from past experiences are also presented in [Chadenas and Kirschfink, 1999, Czuka et al., 2000], in [Leerkamp, 1999] for urban traffic, for link travel times in [Hounsell and Ishtiaq, 1997, Rozemond, 1997], and for the special case of construction areas in [Beckmann and Zackor, 2001, Laffont et al., 1996, Ober-Sundermeier, 2003]. In [Davis and Yang, 2001, Yang and Davis, 2002] theoretical investigations about the uncertainty in forecasting traffic with monthly, weekly, and day of week changes are made.

### 3.5 Combined Methods

In the previous sections many forecast methods are mentioned. So, the idea to combine those methods to enhance the forecast result is straight forward. Whereas in [Faouzi, 1999] a fundamental combination of different forecasts for the same data is discussed, the forecast results of different works have shown, that the feasibility of a forecast method depends on the forecast horizon. Whereas low order ARIMA time series models are feasible for short term forecasts, heuristics in form of historical traffic time series are the only way to forecast the traffic state with a horizon of hours up to days. In the next section works are described that deal with that problem. Thereafter works are mentioned, where several methods are combined in a different way.

#### 3.5.1 Combining Short and Long Term Forecasts

The basic idea to combine the current value with a historical value is already mentioned at the end of the work of Leutzbach and Siegener [1974]. The authors shortly describe a method for short term forecasts up to one hour. Starting from the long term forecast value they add a correction value. This correction value is calculated from the most recently measured values, whereby a linear and an exponential model is compared.

This approach in combining long term estimates with actual measures is again later used by Siegener and Schmitt [1980]. The authors use the traffic flow data of three loop detectors in the Rhein-Main Region, Germany. They analyse the data during the years 1973 until 1976 to forecast the interval 1977/06/01 until 1977/09/30.

The forecast value  $\hat{J}_{h,w}(t)$  of the  $t^{\text{th}}$  weekday  $w$  and hour  $h$  is calculated as the sum of the long term estimate  $J_{h,w}^*(t)$  and a correction addend  $d_{h,w}(t)$ :

$$\hat{J}_{h,w}(t) = J_{h,w}^*(t) + d_{h,w}(t). \quad (3.82)$$

The long term estimate  $J_{h,w}^*(t)$  is directly taken from the model of Leutzbach and Siegener [1974] (see Sec. 3.4 on page 38). The correction addend  $d_{h,w}(t)$  is calculated using the differences of the most recently measured data  $J_{i,w}(t)$  to the long term estimate  $J_{i,w}^*(t)$  for all hours  $i$  of the forecast day up to the hour  $h-1$  ( $i = 1, 2, \dots, h-1$ ):

$$d_{i,w}(t) = J_{i,w}(t) - J_{i,w}^*(t). \quad (3.83)$$

Then,  $d_{h,w}(t)$  is calculated using the exponentially smoothed values  $d_{i,w}(t)$  (see Eq. 3.18) with the smoothing constant  $\alpha$ . To find the optimal value for  $\alpha$  the authors calculate several forecasts with different  $\alpha$  and chose the one ( $\alpha = 0.3$ ) with the smallest relative root mean square error  $e_{\text{RRMSE}}$ . They state that deviations of  $\alpha$  only lead to a small influence of  $e_{\text{RRMSE}}$ .

The authors give the  $e_{\text{RRMSE}}$  for one week forecasts from 1977/06/01 until 1977/09/30 between approximately 0.2 until 0.25. Furthermore they state, that

the mean forecast errors have not always been smaller than the long term estimates.

Finally, the model is used to generate short term forecasts of an interval of 5, 15, 30, and 60 min. At this point it has to be mentioned, that in the work even the long term and the one week forecasts are calculated for 5 min and 15 min intervals as well. But therefore only a linear approximation between the hourly values is used. The authors give the  $e_{RRMSE}$  for the forecasts between 0.129 and 0.244.

Zackor [1976] develops an algorithm for traffic control facilities, that includes among others a model for cross section based short term traffic forecasts. As the forecast value acts the traffic density  $\rho$ . The forecast is based on an analysis of the traffic density in a time interval  $T_A$ . Within this interval the density is detected. Then a line  $\rho(t)$  is calculated using linear regression through the measurement points. The regression line can be described with a few parameters:

$$\rho(t) = \rho_2 + \frac{\rho_2 - \rho_1}{T_A}t, \quad (3.84)$$

with  $-T_A \leq t \leq 0$  and

- $t$  time,
- $\rho_1$  traffic density at the beginning of the regression line,
- $\rho_2$  traffic density at the end of the regression line.

The parameters  $\rho_1$  and  $\rho_2$  have to be quantified using the regression and comply to estimations of the densities at the beginning  $t_0 - T_A$  and at the end  $t_0$  of the analysis time interval:

$$\rho(t = -T_A) = \rho(t = t_0 - T_A) = \rho_1 \quad (3.85)$$

$$\rho(t = 0) = \rho(t = t_0) = \rho_2 \quad (3.86)$$

Because the author states that the function  $\rho(t)$  should fulfill the following constraints, if it is used for forecast ( $t > t_0$ ):

- continuously differentiable transition at the point  $t_0$ ,
- convergence against an expectation  $\rho_E$ ,

the following function for  $t > t_0$  is proposed:

$$\rho(t) = \rho_1^*(t) + \rho_2^*(t) + \rho_3^*(t), \quad (3.87)$$

that consists of the following three addends:

$$\rho_1^*(t) = [\rho_1 - \rho_2] e^{-\frac{t}{T_A}}, \quad (3.88)$$

$$\rho_2^*(t) = 2\rho_2 - \rho_1, \quad (3.89)$$

$$\rho_3^*(t) = [\rho_E + \rho_1 - 2\rho_2] \left[ 1 - e^{-\lambda t^2} \right]. \quad (3.90)$$

Here  $\lambda$  is a damping coefficient.

This forecast function is calculated for  $m$  time steps  $t_{p,j}$ :

$$t_j = t_0 + \frac{jT_P}{m}, \quad (3.91)$$

with  $j = 1, 2, \dots, m$ , the length of the whole forecast interval  $T_P$ , the number of all single forecast intervals  $m$ , and the index  $j$ .

The expectation  $\rho_E$  of the traffic density is calculated from the expectation of the traffic flow using a fundamental diagram. If the density is too high, a congestion occurs and the traffic states are adjusted. The algorithm is tested in the Rhine-Main region in Germany and the author recommends for the analysis time interval  $T_A = 10$  min, the whole forecast interval  $T_P = 30$  min, and the partition into  $m = 10$  intervals each.

By Stephanedes et al. [1981] the following approaches to combine the historical traffic flow  $J_{\text{hist},t}$  with the current measurement  $J_{\text{cur},t}$  to calculate one-step-ahead forecasts  $\hat{J}_t$  for time step  $t$  of minute aggregated urban traffic flows are investigated:

$$\begin{aligned} \hat{J}_t = & J_{\text{hist},t} + \gamma [J_{\text{hist},t-1} - J_{\text{cur},t-1}] + [1 - \alpha] \sum_{s=0}^{t-1} \alpha^s [J_{\text{cur},t-s-1} - J_{\text{hist},t-s-1}] \\ & + \gamma [1 - \alpha] \sum_{s=0}^{t-2} \alpha^s [J_{\text{cur},t-s-2} - J_{\text{hist},t-s-2}], \end{aligned} \quad (3.92)$$

$$\hat{J}_t = \gamma_1 J_{\text{cur},t} + [1 - \gamma_1] \left[ \hat{\mu}_0 \alpha^t + [1 - \alpha] \sum_{s=0}^{t-1} \alpha^s J_{\text{cur},t-s-1} \right], \quad (3.93)$$

$$\hat{J}_t = a_0 + a_1 J_{\text{cur},t} + a_2 [J_{\text{cur},t} - J_{\text{cur},t-1}] + a_3 \left[ \sum_{k=1}^N J_{\text{cur},t-k} / N \right]. \quad (3.94)$$

Here  $\gamma, \gamma_1, \alpha, \alpha^s, \hat{\mu}_0, a_0, a_1, a_2, N$  are several parameters that have to be chosen according to the particular model. The authors call the methods UTSC-2 (Eq. 3.92), UTSC-3 (Eq. 3.93), and Eq. 3.94 is the approach *proposed* by the authors. What is meant with *historical data* and how  $J_{\text{hist},t}$  is calculated is not explained in detail. The authors test traffic flow time series collected at six locations in the Minneapolis-St. Paul metropolitan area from October to December 1979. One minute and five minute forecasts are investigated, whereby for the five minute forecasts the one minute data is just aggregated to five minutes and the algorithms are the same. Forecasts are calculated and compared with the measured data using the mean square error  $e_{\text{MSE}}$  and the mean absolute error  $e_{\text{MAE}}$  for several parameters. The results are compared among the combinations in Eqs. 3.92 – 3.94 as well as with the historical value  $J_{\text{hist},t}$  and the current value  $J_{\text{cur},t}$  alone.

As a result the authors state, that for the 5 min forecast the historical value as well as the UTSC-2 algorithm provide the best forecasts. For the one minute forecast the Eq. 3.94 with the parameters ( $a_0, a_1, a_2 = 0, a_3 = 1$ ), what is exactly

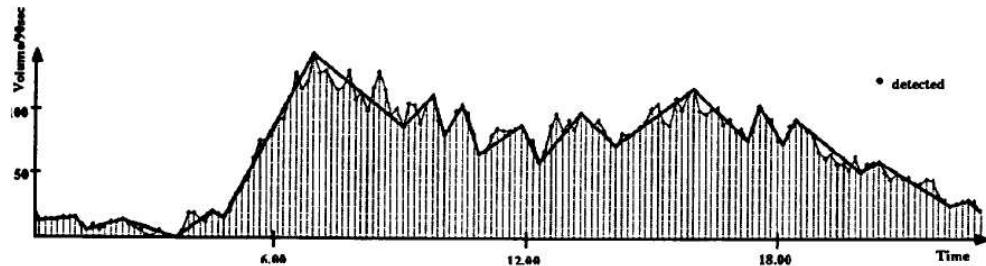
the moving average model (Eq. 3.13), provides the best forecast results. The approaches presented here show, that simple time series models provide suitable results especially for very short forecast horizons like the one-step-ahead forecast for minute aggregated traffic data. With a growing forecast horizon, in this case for 5 min, historical values improve the results.

In [Kim, 1994] two forecast models are developed and evaluated: a link based model and a network based model. The link based forecast model has two components. One component is an ARIMA time series model based on the most recently measured traffic data. The other component is the smoothed historical traffic flow for the same period as obtained from previous days. These two components are combined to represent the dynamic fluctuations in the traffic flow behaviour. The combination is designed to calculate the forecast traffic flows for a maximum forecast horizon of 30 min, divided into 6 min time intervals.

The second model is a network based forecast method that combines current traffic, historical average, and upstream traffic. Three strategies are developed for traffic forecast: a combination of historical average and upstream traffic, a combination of current traffic and upstream traffic, and a combination of all three variables.

The three models are evaluated through regression analysis and the author found the third model as the most applicable. The developed models are applied to real freeway data in 15 min time interval measured by regular inductive loop detectors. The author concludes, that the forecast models are capable of producing reliable and accurate forecasts under congested traffic condition and that the system perform better in the 15 min range than in ranges of 30 min or 45 min.

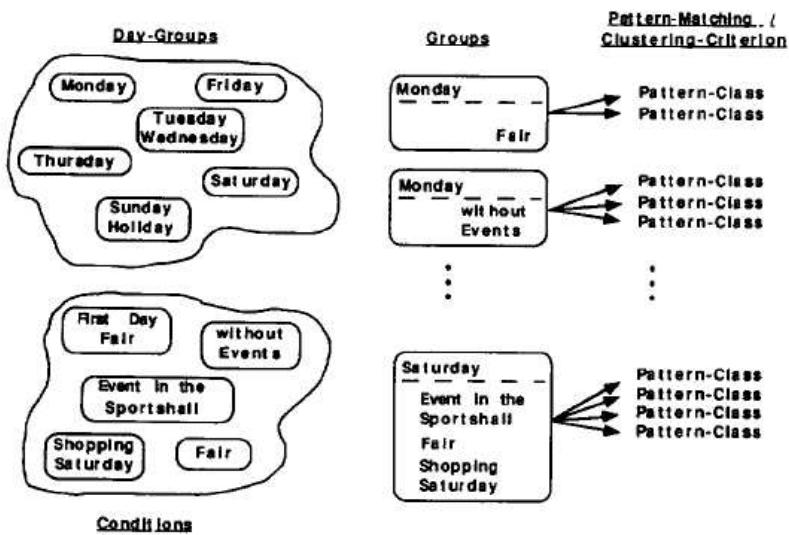
Wild [1997] describes a forecasting method for traffic flow at cross sections (see also [Wild, 1996]). As database act 16 inductive loop detectors placed in Cologne, which measure 90 s aggregated data of the traffic flow. The data are recorded at 53 days of the years 1993 and 1994. The measured raw data are transformed into 9 min time series by adding up 6 values of the 90 s cycles. Missing values are interpolated using an average of the adjacent values. The resulting 9 min time series is transformed using the following procedure.



**Fig. 3.7.** To filter only the crucial peaks out of a noisy traffic time series, in [Wild, 1997] it is transformed into a polyline. Thereafter each peak is considered as an object.

Assuming that the critical patterns from the point of view of forecast are trend changes and peak values, first all extrema of the time series are determined and are interpreted as piece wise linear trend lines. In subsequent passes unimportant short and steep peaks are cut. Therefore, starting at the beginning of the traffic time series, each three adjacent values are considered and it is tested, whether the median can be smoothed. As criterion act a horizontal and a vertical limit. The results of the transformation algorithm can be seen in Fig. 3.7.

The resulting polyline or pattern is represented as a doubly connected list of objects. Each object corresponds to one change point of the time series and is connected to the previous and the next change point. After this, the resulting polylines are arranged into groups (see Fig. 3.8) with several clusters. Therefore the polylines are subdivided considering the different days of the week patterns and site specific information like *fair*, *shopping Saturday*, or *event in the sportshall*.



**Fig. 3.8.** The pattern classification with day groups and conditions from the Cologne field trial. The picture is stemming from [Wild, 1997], where the groups are contained with a pattern matching algorithm.

In doing so, for each polyline is tested, whether a group of the same day and conditions already exists. If such a group does not exist, it will be created and a new cluster will be filled with this archetypal pattern. If a group exists, the new pattern will be compared with all present patterns from the group. This pattern matching is based on the maximal vertical difference between patterns. If the maximal difference to the best matching pattern is small enough to satisfy the clustering criterion, the patterns belong to the same cluster. Then, some characteristics of the new pattern are included in the objects of the representative pattern for the cluster and the new pattern is discarded. If there are significant

differences, a new cluster is created with the new pattern as a representative. In order to forecast historical pattern, the detected and transformed patterns are matched with the historical patterns in the data base. Therefore current conditions are used, coming either from the operator or from an event calendar. Then, the same similarity based pattern matching as used for classification performs the final choice.

Finally, the current and the historic data are both used to forecast the future traffic flow. Thereby it is considered, that the current data have a large influence for short forecast horizons, whereby with a growing horizon the influence of the historical data increases. The following equation fulfills this condition:

$$J_{\text{prog}}(t_p) = J_{\text{hist}}(t_p) + k\Delta J(t_0), \quad (3.95)$$

with

$$\begin{aligned} \Delta J(t_0) &= J_{\text{cur}}(t_0) - J_{\text{hist}}(t_0), \\ k &= \begin{cases} \eta \left[ 1 - \frac{t_h}{t_{\text{hmax}}} \right], & \text{if } 0 < t_h \leq t_{\text{hmax}}, \\ 0, & \text{if } t_h > t_{\text{hmax}}, \end{cases} \\ t_h &= t_p - t_0. \end{aligned}$$

Thereby is

$$\begin{aligned} J_{\text{prog}}(t_p) &\triangleq \text{forecast traffic flow,} \\ J_{\text{hist}}(t_p) &\triangleq \text{historical traffic flow,} \\ J_{\text{cur}}(t_0) &\triangleq \text{most recently measured traffic flow,} \\ t_0 &\triangleq \text{point in time the forecast is calculated,} \\ t_p &\triangleq \text{time of the forecast traffic flow,} \\ \eta \text{ and } t_{\text{hmax}} &\triangleq \text{parameters to be chosen.} \end{aligned}$$

To evaluate this method, the database is filled up with data of 6 loop detectors of 42 days of the years 1993 and 1994. Then the traffic data of 10 days in March 1994 is forecast for three forecasting horizons (270 s, 900 s, and 3,600 s). To quantify the difference between the forecast and the real data, different quality measures are used and the results are compared with the 3 value moving average model and the naive forecasting model. The results can be seen in Table 3.1.

Unfortunatly the database is very sparse, but it can clearly be seen, that for longer forecast horizons the combined method outperforms the other ones. For very short term horizons the naive forecast model and the moving averages outperform the proposed method. So it has to be investigated, whether the method in Eq. 3.95 can be improved or the parameters could be chosen in another way.

A similar method in combining short term traffic forecasts with historical data is used by Matsumura et al. [1998]. The authors propose a travel time forecast method that is based on statistical values and their differences to current measured link travel times. The forecast travel time  $\hat{T}(t, n)$  of a link at the  $n^{\text{th}}$  point ahead from the present time of day  $t$  is calculated as follows:

$$\hat{T}(t, n) = T_s(t + n) + F_n(D(t)), \quad (3.96)$$

horizon	Wild	$\mu_t$	naive
$e_{\text{MAE}}$ [veh/min]			
270 s	1.56	1.47	1.38
900 s	2.05	2.07	2.04
3,600 s	2.66	3.73	3.69
$e_{\text{RMSE}}$ [veh/min]			
270 s	2.07	1.92	1.84
900 s	2.67	2.74	2.72
3,600 s	3.49	5.11	5.07
$e_{\text{MRE}}$			
270 s	0.1460	0.1344	0.1276
900 s	0.1870	0.1798	0.1773
3,600 s	0.2363	0.2723	0.2694
$e_{\text{RMSEP}}$			
270 s	0.15	0.14	0.13
900 s	0.19	0.18	0.18
3,600 s	0.24	0.33	0.33

**Table 3.1.** Comparison of the forecast results for three forecast horizons using the model proposed by Wild [1997] (Wild) with results of the 3 value moving average model ( $\mu_t$ , see Eq. 3.13) and the naive forecasting model (naive, see Eq. 3.12).

with

$$F_n(D(t)) = \begin{cases} \frac{[C_{\text{set}}-n]D(t)}{C_{\text{set}}}, & \text{if } C_{\text{set}} - n > 0, \\ 0, & \text{else,} \end{cases}$$

$$D(t) = T_c(t) - T_s(t),$$

thereby is

- $T_s(t + n)$  statistical travel time of the link at the  $n^{\text{th}}$  point ahead from present time of day  $t$ ,
- $C_{\text{set}}$  forecast point from which the added or subtracted value of the difference begins to become 0,
- $T_c(t)$  most recently measured travel time of the link at present time of day  $t$ .

As the statistical travel time  $T_s(t)$  act the arithmetic mean of the travel times of the link at a time corresponding to the present time of day  $t$ .

Travel time forecasts are calculated in 5 min intervals for a test area around Osaka, Japan. The travel times are collected at 12 weekdays between 1997/09/09 and 1997/10/02 for the statistical analysis and at 8 weekdays between 1997/10/06 and 1997/10/17 for the verification of the forecast algorithm.

To measure the travel times three methods are used:

- estimating the travel time of each link based on data obtained from presence type vehicle detectors,

- calculating the difference in the passage time of one and the same vehicle at multiple points by using a video detection and image processing system,
- calculating the difference in the passage time of one and the same vehicle at multiple points by using the two way communication function of infrared beacons.

The authors use mainly the vehicle detector data because there is an insufficient number of vehicles with a two way communication function and there is also a limited number of video detection devices. On each day travel time data are collected from about 6:30 to about 20:00. The authors state the mean relative error  $e_{MRE}$  to be 0.114 and conclude, that this method is “adequately practicable as an input to dynamic root guidance systems”.

Iwasaki and Saito [1999] combine an autoregressive model, whose parameters are estimated using a Kalman Filter, with historical traffic patterns to forecast velocities up to an hour. The traffic data is collected from an east bound of the Tomei Expressway section located in the western suburb of Tokyo, Japan. At this motorway section, vehicle detector stations were set up about each 2 km that measure 5 min aggregated data of flow, mean velocity and occupancy on each lane.

The proposed equation that combines both methods is the convex combination

$$v(t+1) = [1 - \beta] v_{AR}(t+1) + \beta v_{hist}(t+1). \quad (3.97)$$

The forecast velocity  $v(t+1)$  at time step  $t+1$  is the sum of the velocity obtained using the autoregressive model  $v_{AR}(t+1)$  weighted with  $1 - \beta$  and an historical velocity  $v_{hist}$  that is weighted with  $\beta$ . The historical velocity is obtained using the mean of data collected for about 3 years. The data is classified by each day of the week for each detector. Nevertheless the authors state that the result of statistical tests advise a 3 type classification. Those are weekday, Saturday, and holiday patterns.

The authors apply the model on forecast horizons from 10 min to 60 min in 10 min steps and for parameters  $\beta$  from 0 to 1 in steps of 0.1. The results are compared among each other using the root mean square error  $e_{RMSE}$ . The lowest value of  $e_{RMSE}$  is reached for the 10 min forecast with  $\beta = 0.1$ . With an increasing forecast horizon the better results are received with an increasing  $\beta$ . For a forecast horizon of 60 min best forecasts are received with  $\beta = 0.6$ .

In a same manner Rice and van Zwet [2001] use a linear combination of current and historical travel time data for one-step-ahead and 60 min forecasts of 5 min aggregated traffic data. From the distance and the velocity data of 116 single loop detectors along 48 miles of the Interstate 10 near Los Angeles, USA, the travel times are approximated.

The work affirms again that for short horizons a combination of the most recent measured values provides suitable results, whereas for longer horizons models that include historical data work better.

### 3.5.2 Other Combinations

In [van der Voort et al., 1996] a hybrid method called KARIMA is proposed for one-step-ahead forecasts of 30 min aggregated flow rates for southbound traffic during the months July and August of the years 1984 until 1990 from four counting stations near Beaune, France. The data are the same that are used by the ATHENA model (see Sec. 3.4 on page 43) and later by the ARIMAX model (see Sec. 3.1.3 on page 18). The test field can be seen in Fig. 3.2.

The difference between the KARIMA and a usual ARIMA method is, that the KARIMA model applies the ARIMA method not to all the data, but first clusters the data into several groups like days of the week. Then an ARIMA model is applied to each cluster on its own.

At first the authors try to arrange the cluster manually dividing the data sets in the months July and August as well as in workingdays and weekends. Comparing the results with those of the ATHENA model it comes out, that the ATHENA model is the superior one. Thus, the authors conclude that a better way of clustering is needed.

This clustering is fulfilled with a Kohonen self-organizing map [Kohonen, 1995], which is a special kind of a neural network. It consists of just two layers: a linear input layer which is wholly connected to the second layer, the map, via weighted connections. When a vector of input data is fed into the input layer, each of the neurons in the map is stimulated and the neuron with the highest activity is awarded the status of *winner* and its connection weights are increased.

First the authors test a  $20 \times 20$  square grid structure of neurons. The input vector contains the following 16 items of data:

- four values for the most recent half hour flow measurements at the four sites,
- four values for the difference in flow from the previous values of flow,
- a time variable to indicate the half hour,
- seven binary variables, each indicating a different day of the week.

In doing so, the trained map exhibits seven very strong clusters representing the seven days of the week. Because the authors are hoping to find a more sophisticated clustering criteria, a new data set is created without the binary variables. The map produced with this data set shows two large clusters and one smaller cluster. Different methods of splitting up the map into several data sets taking into account the spatial information and the strength of activity and fitting separate ARIMA models for each data set fail to improve the forecast of the ATHENA model.

As an alternative to a square grid hexagonal maps are tried as an alternative. The self-organising maps used have 15 rows of 20 nodes each. Again various combinations of input variables are tested and finally the input vectors contain 12 variables:

- four values for the most recent half hourly flow measurements at the four sites,
- a time variable to indicate the half an hour,
- seven binary variables, each indicating a different day of the week.

Unlike the square map, the hexagonal map does not converge on the solution of producing a cluster for each day of the week. Instead of this there is a cluster with low activity among a lot of nodes with high activity, so the authors decide to fit several ARIMA( $p,0,q$ ) models to the two groups and find that the best models have a value of 2 or 3 for  $p$  and a value of 1 or 2 for  $q$ .

Because the ARIMA model fit better to the group with the low activity neurons, the authors divide manually the highly activated coordinates into three different regions of data sets which show better results. Using this distinction into four classes an ARIMA model is fitted to each class and used for forecast. A 60 min forecast is computed by just summing two half hourly data points and using the same method.

The forecast results seem to be as feasible as those of the ATHENA model, but as the authors conclude, it is not possible to completely automate the process of fitting the best model, as a manual interpretation of the clusters is needed. Furthermore the authors point out that the model has been tested only on two months of data which is not enough to explore long term robustness.

What is basically tried with this method is to replace the  $k$ -means clustering of the ATHENA model, which is a basic approach for automatically classifying several weekdays, with a neural network. But, as the authors stated, the cognitive abilities of the neural network are neither good enough to fully automate the process nor could it basically enhance the classification.

Ma et al. [2001] combine six different forecasting methods with a multi-agent system. The best three models are picked up and with them the forecasting value is calculated. Describing the multi-agent system would extend the scope of this work and especially because also adjacent detectors are used and the results are not considerably better than with other models, the interested reader is referred to the original paper. An application of a travel time forecasting model based on day of week changes that are further analysed with nonparametric regression techniques is presented in [You and Kim, 2000].

## 3.6 Comparisons

Most of the methods that are presented so far are compared not only with the real world measurements but also with other forecast methods. The problem with this comparisons is, that they are mostly done to show the reliability of the model evolved by the author. This leads always to the risk of a subjective interpretation. Furthermore, in most cases a more complex model that is introduced is compared to one or more simple models such as the naive forecast. In this section those

works are mentioned, in that the comparison of complex forecast methods is the focus of the work itself. Unfortunately the number of such works is very limited, and as we will see the results are not always helpful.

Smith [1995] compares one-step-ahead forecasts of nonparametric regression, historical averages, ARIMA models, and neural networks. Therefore he uses 15 min aggregated traffic flow data. They are stemming from two sites on the Capital Beltway in Northern Virginia, USA. The author states, that the nonparametric regression significantly outperforms the others and uses it for multiple interval forecasts. The author describes the nonparametric regression model is found to perform well.

A comparison between ARIMA models, the ATHENA method, and neural networks used for one-step-ahead forecasts of half hourly traffic flows is presented in [Kirby et al., 1997]. The authors use the same half-hourly data of the months July and August as in the original ATHENA model (see Sec. 3.4 on page 43). The objective is to forecast the traffic flow at Beaune in Fig. 3.2.

The neural network that is used for forecast is a back-propagation network with one hidden layer. The input layer is of varying width, depending on the number of input parameters. The output layer has just the forecast traffic flow as neuron. The number of nodes in the hidden layer is determined by experiment. Furthermore a sigmoid transfer function and the delta learning rule is used.

Forecasting technique		Frequency of $e_{APE}$ [%]	
		0–20	20+
30 min	neural nets	91.5	8.5
	ARIMA (Jul)	97.0	3.0
	ARIMA (Aug)	97.1	2.9
	Univariate	95.4	4.6
	Multiple	95.5	4.5
	neural nets	84.8	15.2
60 min	ARIMA (Jul)	93.6	6.4
	ARIMA (Aug)	94.8	5.2
	Univariate	90.5	9.5
	Multiple	93.5	6.5
	ATHENA	97.5	2.5
	ATHENA	89.1	10.9

**Table 3.2.** Summary of the main performance results of a comparison between the ARIMA, the ATHENA, and a neural network model for traffic flow forecast. Shown is the frequency in % with which the absolute percentage error lies within zero and 20 %, and beyond 20 % (table from [Kirby et al., 1997]).

The authors state that the use of six historical patterns is found in practise to be the optimum as the improvement using a longer historical look back is negligible. A description of the approach used to determine the optimum input set is given in [Dougherty and Cobbett, 1997]. That means that six preceding values of both the three upstream and the principal measurement site yielding 24 data points

act as the input layer.

The ARIMA models used in this comparison rely in their simplest form. Unfortunately the exact parameters are not given in [Kirby et al., 1997], but it is described that two models are used. The first is given by a linear function of the previous Beaune flows through time. It is a univariate procedure. The second includes a function of historical flows at Beaune together with historical flows from other upstream points and is a multivariate procedure. Furthermore the authors decide to treat the months July and August separately to minimise the seasonal effect within the model.

The results from the models are compared using the frequency of absolute percentage errors  $e_{APE}$  (see Table 3.2). As can be seen, neural networks perform less well than ARIMA time series models when forecasting 30 min ahead, and less well than both, ARIMA and ATHENA models, when forecasting 60 min ahead. The ATHENA model provides the best results forecasting 60 min ahead. Furthermore only little improvement is achieved by including the upstream data for the ARIMA fits.

The authors state, that the training data used is particularly thin for models of a neural network type so the poor performance is not surprising. They conclude that it is not possible to generalise from the results of this study to suggest which approach may be best in all circumstances because the arguments may be affected by characteristics of the chosen data set. Nevertheless, a neural network is again outperformed by other techniques.

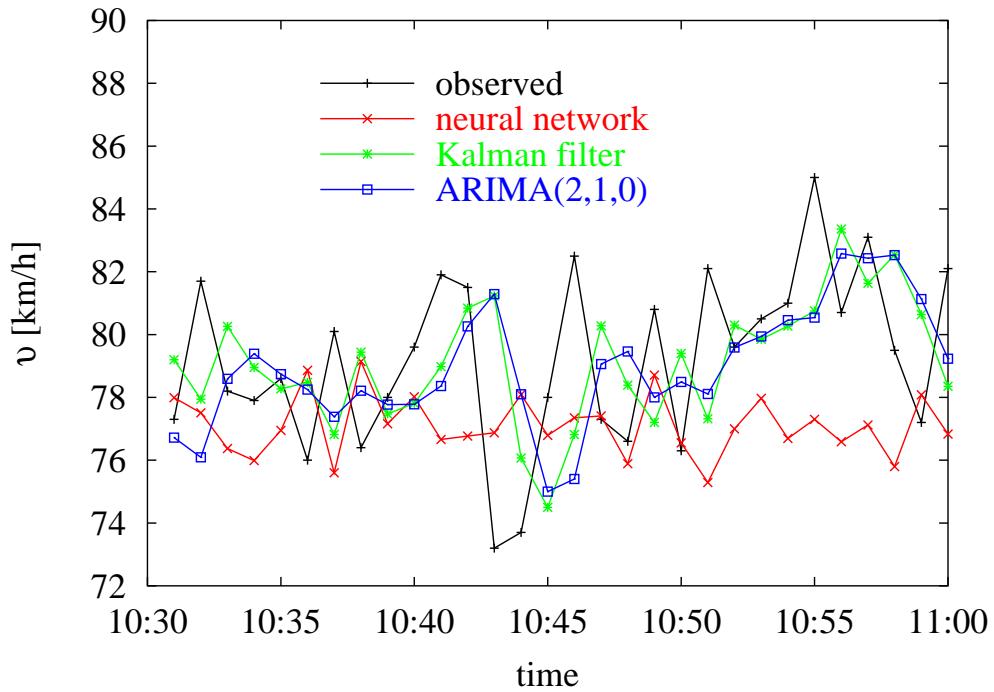
That especially for short term forecasts complex models are not able to show improvements in comparison to simple parametric regression methods show the results of Lee and Choi [1998]. The authors compare an ARIMA model, a neural network, and a Kalman filter to generate one-step-ahead forecasts of minute aggregated velocity data of the Olympic expressway in South-Korea. The sample data are collected by an image detector during 10:00 – 12:00 on the 1998/05/17 at the Dongho-Hannam Bridge. The data consist of velocity, flow, and occupancy. After the process of identification, parameter estimation, and diagnostic checking the authors decide to use an ARIMA(2,1,0) model. The neural network used in this work is a feedforward network with a sigmoid transfer function. It consists of one input layer with one neuron, two hidden layers with 6 neurons each and one output layer with the forecast velocity. As input unit serve the velocity, flow, and occupancy data for the last 30 minutes. To train the network a back propagation delta learning rule is used.

For the comparison the mean relative error  $e_{MRE}$ , the mean square error  $e_{MSE}$ , and the equality coefficient  $C_{equal}$  is used. The results can be seen in Table 3.3. Although the authors state, that the Kalman filtering model shows slightly superior results to the other techniques, it is very bold to speak about an improvement of the forecast. The observed time interval is very short, and the observed velocity is fluctuating around a nearly constant value, as can be seen in Fig. 3.9.

In this work several raw data are given that can be seen in Table 3.3. Thus, the forecast performance is compared with the naive forecasting model (Eq. 3.12), the

time	$v_{\text{observed}}$ [km/h]	$v_{\text{forecast}}$ [km/h]					
		neur. network	Kalman	ARIMA	naive	$s_t$	$\mu_t$
10:31	77.30	77.99	79.20	76.72	—	—	—
10:32	81.70	77.51	77.94	76.09	77.30	77.30	77.30
10:33	78.20	76.37	80.26	78.59	81.70	78.18	79.50
10:34	77.90	75.99	78.95	79.39	78.20	78.18	79.07
10:35	78.60	76.95	78.27	78.74	77.90	78.13	78.78
10:36	76.00	78.86	78.48	78.25	78.60	78.22	78.74
10:37	80.10	75.60	76.82	77.38	76.00	77.78	78.28
10:38	76.40	79.14	79.44	78.21	80.10	78.24	78.54
10:39	78.00	77.16	77.47	77.77	76.40	77.87	78.28
10:40	79.60	78.02	77.81	77.78	78.00	77.90	78.24
10:41	81.90	76.66	78.98	78.36	79.60	78.24	78.38
10:42	81.50	76.77	80.84	80.26	81.90	78.97	78.84
10:43	73.20	76.87	81.23	81.29	81.50	79.48	78.82
10:44	73.70	78.09	76.07	78.10	73.20	78.22	78.32
10:45	78.00	76.79	74.50	75.00	73.70	77.32	77.90
10:46	82.50	77.35	76.82	75.40	78.00	77.45	77.84
10:47	77.30	77.41	80.28	79.06	82.50	78.46	78.49
10:48	76.60	75.89	78.39	79.46	77.30	78.23	78.21
10:49	80.80	78.71	77.20	78.00	76.60	77.90	78.23
10:50	76.30	76.55	79.40	78.49	80.80	78.48	78.51
10:51	82.10	75.29	77.32	78.11	76.30	78.05	78.18
10:52	79.60	77.00	80.30	79.59	82.10	78.86	78.20
10:53	80.50	77.97	79.85	79.94	79.60	79.01	78.01
10:54	81.00	76.69	80.27	80.46	80.50	79.30	78.74
10:55	85.00	77.30	80.75	80.54	81.00	79.64	79.47
10:56	80.70	76.59	83.36	82.58	85.00	80.72	80.17
10:57	83.10	77.12	81.63	82.43	80.70	80.71	79.99
10:58	79.50	75.80	82.54	82.53	83.10	81.19	80.57
10:59	77.20	78.08	80.64	81.13	79.50	80.85	80.86
11:00	82.10	76.83	78.35	79.23	77.20	80.12	80.50
Forecast Errors							
$e_{\text{MRE}}$	0.0392	0.0339	0.0321	0.0386	0.0292	0.0303	
$e_{\text{MSE}}$	13.8230	9.9265	10.2922	12.9924	8.1247	8.1255	
$C_{\text{equal}}$	0.9762	0.9801	0.9797	0.9773	0.9820	0.9820	

**Table 3.3.** Comparison of different forecasting models for one-step-ahead forecasts of velocities. The observed velocities  $v_{\text{observed}}$  as well as the forecasting results of a four layer feedforward neural network, a Kalman filter, and an ARIMA(2,1,0) model are stemming from [Lee and Choi, 1998]. The forecasts of the other techniques are starting at 10:32, because the observed data from 10:00 until 10:30 are not given in [Lee and Choi, 1998]. All of the three complex methods are outperformed by an exponential smoothing model with  $\alpha = 0.2$  ( $s_t$ ) and an 11 minute moving average model  $\mu_t$ . Even the naive forecasting model shows better results than the neural network.



**Fig. 3.9.** Comparison of a four layer feedforward neural network, a Kalman filter, and an ARIMA(2,1,0) model as one-step-ahead velocity forecast methods by Lee and Choi [1998]. It seems that the velocity fluctuates around a mean value and that no model stands out in forecasting those fluctuations.

exponential smoothing model (Eq. 3.18) with  $\alpha = 0.2$ , and the moving average model (Eq. 3.13) with  $N = 11$ . The results are shown in Table 3.3 as well. As one can see, the moving average as well as the exponential smoothing model clearly outperform the more complex forecast methods. Furthermore, the naive forecast model shows better results than the neural network. Note, that the more complex forecast methods also use the data from 10:00 until 10:31 that are not given in [Lee and Choi, 1998] and thus could not be used here for the other techniques (thus, there is even no forecast for 10:31 in Table 3.3).

For completeness it has to be mentioned, that Lee et al. [1998] present a similar comparison of a multiple regression, ARIMA, Kalman filtering, and neural network model for velocity forecasts with different forecast horizons. Unfortunately many details are not given, so it is hardly possible to reconstruct the method in detail especially because they also use adjacent detectors.



## Chapter 4

# Traffic Data of Inductive Loop Detectors

“The whole thing is a number.”

**Pythagoras of Samos**  
(~ 569 – 475 B.C.), Greek mathematician and philosopher

In this chapter some main aspects of the traffic data that are used in this work are described. The understanding of the data source, how the data are acquired, how they are aggregated, and how they are transmitted to the final database from where they are extracted for further analyses is an essential part of the scientific work. Otherwise measurement errors as well as artifacts that are generated by the control unit can lead to physical misinterpretations.

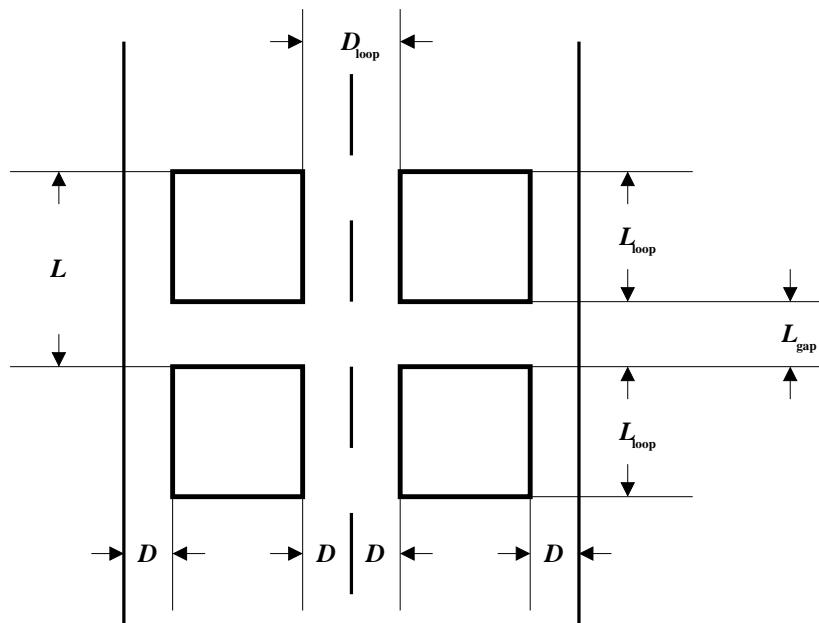
### 4.1 Acquisition of Traffic Data

The data used in this work are stemming from so called inductive loop detectors, often also called presence detectors, in North Rhine-Westphalia (NRW), a federal state of Germany. A loop detector is the inductive part of a series LC-oscillator buried in one lane of the street. In NRW there are two different kinds of loop detectors, detector type 1 and detector type 2. The measures of the different types can be seen in Table 4.1, a schematic sketch of the detector geometry is given in Fig. 4.1.

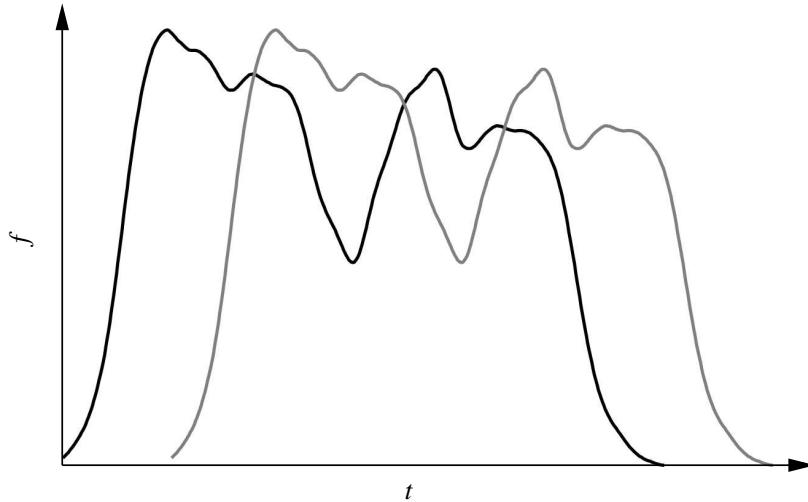
If a vehicle passes the circuit the inductance decreases and the frequency of the circuit increases. A signal can be measured, as it is illustrated in Fig. 4.2. In this way the traffic flow follows directly from the number of signals in a given interval, in this case 1 min. With two loop detectors in a row the velocity can be calculated with the time between two peaks and the distance of the loop detectors. From the time the detector is covered in one minute directly follows the occupancy. The characteristics of the curve give information about the kind of the vehicle.

detector type	1	2
$L_{\text{loop}}$ [mm]	$2500 \pm 5$	$1000 \pm 5$
$L_{\text{gap}}$ [mm]	$1500 \pm 5$	$1500 \pm 5$
$L$ [mm]	$4000 \pm 10$	$2500 \pm 10$
$D_{\text{loop}}$ [mm]	$1600 \pm 5$	$700 \pm 5$
$D$ [mm]	$800 \pm 5$	$350 \pm$

**Table 4.1.** Measures of the two types of inductive loop detectors in NRW.



**Fig. 4.1.** Geometry of the inductive loop detectors in NRW.



**Fig. 4.2.** Signal  $f$  of a lorry with trailer at two loop detectors in a row. The traffic flow follows directly from the number of signals in a given time interval. From the time between two peaks and the distance of the loop detectors the velocity can be calculated. The occupancy follows from the time the detectors are covered.

For details about the mode of operation see the manual of the manufacturer [Weiss-Electronic, 2001]. An electronic control unit interprets and aggregates the data in intervals of 1 min. For all the data this calculation is straight forward except for the velocity. From the distance of the heads of the loops  $L$  and the time between two peaks  $t_i$  there are two different possibilities to calculate the mean velocity  $\bar{v}$ .

The first is to use the arithmetic mean of the single velocities:

$$\bar{v}_{v_i} = \frac{1}{N} \sum_{i=1}^N v_i = \frac{1}{N} \sum_{i=1}^N \frac{L}{t_i}, \quad (4.1)$$

the second is to use the distance  $L$  divided by the mean time:

$$\bar{v}_{t_i} = \frac{L}{\frac{1}{N} \sum_{i=1}^N t_i}. \quad (4.2)$$

Because  $\frac{1}{N} \sum_{i=1}^N \frac{1}{t_i} \geq \left[ \frac{1}{N} \sum_{i=1}^N t_i \right]^{-1}$  it holds

$$\bar{v}_{v_i} \geq \bar{v}_{t_i}. \quad (4.3)$$

The difference between both definitions is that  $\bar{v}_{v_i}$  is the mean velocity in regard to the number of measured vehicles  $N$ , whereby  $\bar{v}_{t_i}$  is the mean velocity in regard

to the period of the measurement. This definition is equivalent to the sum of the single velocities weighted by the part of the time the vehicle is located between the two detectors, as it is shown in the following:

$$\begin{aligned}\bar{v}_{t_i} &= \frac{L}{\frac{1}{N} \sum_{i=1}^N t_i} = \frac{NL}{\sum_{i=1}^N t_i} = \frac{t_1}{\sum_{i=1}^N t_i} \frac{L}{t_1} + \dots + \frac{t_N}{\sum_{i=1}^N t_i} \frac{L}{t_N} \\ &= \frac{t_1}{\sum_{i=1}^N t_i} v_1 + \dots + \frac{t_N}{\sum_{i=1}^N t_i} v_N.\end{aligned}\quad (4.4)$$

That provides an obvious interpretation of Eq. 4.3. A low velocity has more influence on  $\bar{v}_{t_i}$  because the vehicle needs more time to pass the detector. Because  $\bar{v}_{t_i}$  is more sensible to measurement errors for small velocities,  $\bar{v}_{v_i}$  is used as the minute aggregated velocity.

All eight minute aggregated values that are provided by the electronic control unit can be seen in Table 4.2. The traffic data fulfill the standard according to the TLS-norm [Bundesministerium für Verkehr, 2000].

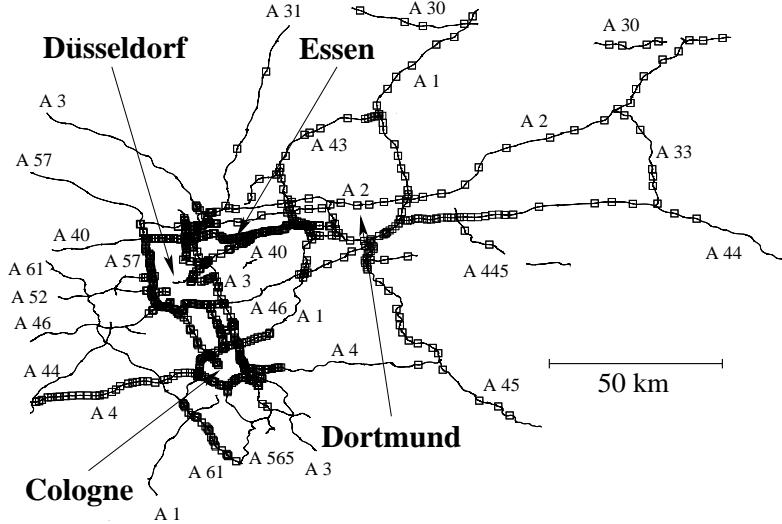
data	description	unit and resolution	systematic error	in interval
$J_{\text{veh}}$	flow of all vehicles	[veh/min]	$\Delta J_{\text{veh}} < 20 \%$	$J_{\text{veh}} \leq 10 \text{ veh/min}$
			$\Delta J_{\text{veh}} < 10 \%$	$J_{\text{veh}} > 10 \text{ veh/min}$
$J_{\text{lor}}$	flow of lorries	[veh/min]	$\Delta J_{\text{lor}} < 35 \%$	$J_{\text{lor}} \leq 10 \text{ veh/min}$
			$\Delta J_{\text{lor}} < 20 \%$	$J_{\text{lor}} > 10 \text{ veh/min}$
$v_{\text{pc}}$	velocity of passenger cars	[km/h]	$\Delta v_{\text{pc}} < 3 \text{ km/h}$	$v_{\text{pc}} \leq 100 \text{ km/h}$
			$\Delta v_{\text{pc}} < 3 \%$	$v_{\text{pc}} > 100 \text{ km/h}$
$v_{\text{lor}}$	velocity of lorries	[km/h]	$\Delta v_{\text{lor}} < 3 \text{ km/h}$	$v_{\text{pc}} \leq 100 \text{ km/h}$
			$\Delta v_{\text{lor}} < 3 \%$	$v_{\text{pc}} > 100 \text{ km/h}$
$t_h$	time headway	[ $10^{-1}$ s]	—	—
$\rho_{\text{rel}}$	occupancy	[ $10^{-2}$ ]	—	—
$\sigma(v)$	standard deviation of the velocity	[km/h]	—	—
$v_{\text{exp}}$	exponential smoothed velocity	[km/h]	—	—

**Table 4.2.** Traffic data provided per lane and per minute by the loop detectors in NRW with a systematic relative error in a given interval. For all data one byte is used. The value 255 indicates an measurement error. This standard is given according to the TLS-norm [Bundesministerium für Verkehr, 2000]

These traffic values offer minute per minute information about the current traffic state. Nevertheless not all of the values are of the same use. The time headway  $t_h$  is the time that lies between two vehicles that pass a detector. This time can be calculated in an easy way from the flow  $J_{\text{veh}}$  and the occupancy  $\rho_{\text{rel}}$ :

$$t_{h,\text{calc}} = \frac{1 - \rho_{\text{rel}}}{J_{\text{veh}}}.\quad (4.5)$$

So the additional transmission of  $t_h$  only helps to avoid some roundoff errors calculating  $t_h$  and to detect some measurement errors as can be seen in Sec. 4.2.6. The exponential smoothed velocity  $v_{\text{exp}}$  is calculated each minute using Eq. 3.18 with two different smoothing parameters depending on the current value of  $v_{\text{exp}}$  [Bundesministerium für Verkehr, 2000]. So it is only a different method of averaging the velocity.



**Fig. 4.3.** Schematic sketch of the freeway network of NRW. The locations of the inductive loop detectors are marked with an open rectangle. The density of the detection concentrate on densely populated regions that are additionally marked. For a better view the abbreviations “A” instead of the official name “BAB” are used for the name of the motorways.

Also the standard deviation of the velocity  $\sigma(v)$  (calculated using Eq. 2.10 with the velocities of each measured vehicle) offers no fundamental new information on the traffic state. The reason why it is provided each minute lies in the fact, that peaks of the variance indicate major changings in traffic states. This is why the variance or standard deviation has been proposed in the past to be used as a forecaster of traffic congestions [Kühne, 1984, 1987, 1991]. But it could not be affirmed, that variations in the velocity come before the formation of a congestion (see also [Ferrari, 1988a,b, 1989]) and algorithms that use this assumption are not able to detect congestions before they appear [Grunewald, 2001, Hoops et al., 2000].

So the traffic data that is essential for this work are the minute aggregated values of  $J_{\text{veh}}$ ,  $J_{\text{lor}}$ ,  $v_{\text{pc}}$ ,  $v_{\text{lor}}$ , and  $\rho_{\text{rel}}$ . From the given data the traffic flow of passenger cars

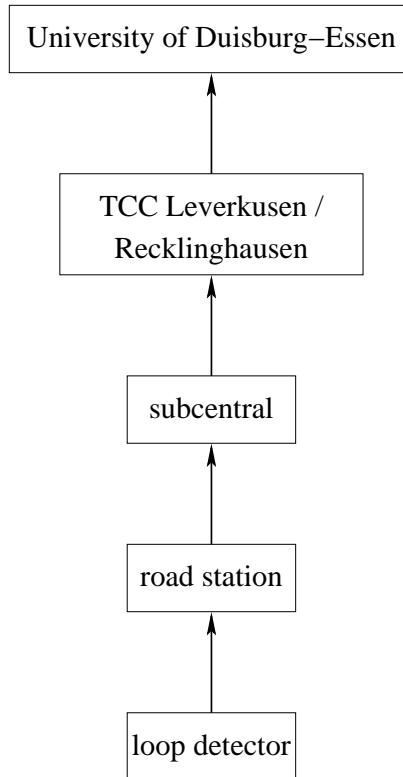
$$J_{\text{pc}} = J_{\text{veh}} - J_{\text{lor}} \quad (4.6)$$

can be calculated as easily as the mean velocity of all vehicles

$$v_{\text{veh}} = \begin{cases} \frac{J_{\text{pc}}v_{\text{pc}} + J_{\text{lor}}v_{\text{lor}}}{J_{\text{veh}}}, & \text{if } J_{\text{veh}} > 0 \text{ veh/min,} \\ \text{N/A,} & \text{else.} \end{cases} \quad (4.7)$$

Here, N/A means “Not Available” because no velocity can be measured in case of  $J_{\text{veh}} = 0$  veh/min. Data can be N/A for many reasons. Most of them are measurement errors, such as discussed in Sec. 4.2.

The data are collected systematically in the database that is used for this work from 2000/10/05 on. About 4,500 loop detectors are placed on German motorways in the state of NRW (see Fig. 4.3). The exact number changes during the years 2000 until 2004 because the network is in permanent extension. For the day of the 2004/05/31 holds, that 4,480 loop detectors are arranged in 2,440 direction depending cross sections.



**Fig. 4.4.** The data transfer from the loop detectors to the University of Duisburg-Essen. From the inductive loop detectors the data are transmitted via the road station to some of several sub centrals, from there to one of two traffic control centers in Leverkusen and Recklinghausen, and finally to the university Duisburg-Essen, where the data is stored in the database.

All the data are transmitted to some of several sub centrals, from there to one of two traffic control centers, and finally to the university Duisburg-Essen, where

the data is stored in the database, that is used for this work (see Fig. 4.4). For technical details about the transmission to the university see [Kaumann, 2000]. The name of each loop detector consists of the node (4 or 5 digits) the distance (3 digits) and the channel (3 digits). Each set of digits is separated by a hyphen. The meaning of the node and the distance is not important for this work. The channel implies the exactly placement on the road, for instance, lane and direction. In case of cross section based analyses the channel is not given in the text, but the driving direction. For instance, the loop detector 4092-072-002 is the detector on the left lane of the cross section 4092-072 in driving direction north of the motorway BAB 1.

Apart from the given systematic errors in Table 4.2 there are a lot of sources for measurement errors. The most frequent measurement errors, their impact and how their influence on further interpretations can be minimised is discussed in the following.

## 4.2 Measurement Errors

The systematic errors of the fully functional loop detectors given in Table 4.2 come from the imperfection of the loop detectors together with the electronic unit that interprets the signal. Apart from that, the geometry of the loops can cause inaccuracies because of the probability that vehicles pass the detector unmeasured [Kurkjian et al., 1980]. Since traffic data are measured, the problem how to deal with incorrect or incomplete data sets is investigated [Wright, 1974].

For this work such inaccuracies are no problem if they are known from the first and if they are not too large. But because of the large number of loop detector-data used in this work not every traffic data set can be manually screened for even obvious measurement errors. Nevertheless there are a lot of damaged loop detectors or detectors that are off due to construction areas. Furthermore the long way of the data from the signal to the database can cause different errors, for instance, because the network goes down or there is a blackout in a traffic central.

To avoid misinterpretations, in the following essential errors are discussed that have to be eliminated. Note, here it is only possible to analyse the symptoms, because there is no information available about the data before they are stored in the database. Furthermore it is not part of this work to optimise the measurement process. Nevertheless it is in most cases possible to state the source of the error.

### 4.2.1 Missing Data

If the network is slow or goes down, one of the two traffic control centers fails, or the computer with the database fails, data are missing for a while. Apart from the fact, that in some cases data of days with very interesting traffic conditions are missing, this is actually no problem. If it is important, that the traffic data are evenly spaced in time, the data can be checked for gaps before analysis. Then,

only time series without gaps are considered. In this case the number of possible time series is reduced. To avoid this, the gaps can be filled up with adjacent data values. Especially for a small number of gaps, this works quite well.

In other cases it is more important to maintain a large time series in a row, and it only must be ensured that not too many data values are missing, for instance, if the mean daily traffic is calculated over years like in Sec. 5.2. In this case the day can be checked for the quantity of missing data.

How errors due to missing data are avoided or in which way the algorithm is modified is mentioned at the particular calculation. Essential statistics about the number of days with  $N_{\text{miss}}$  min missing data are given in Table 4.3.

traffic control center	Leverkusen	Recklinghausen
total number of days	970	970
days with $N_{\text{miss}} = 0$ min missing data	679	478
days with $0 < N_{\text{miss}} \leq 720$ min missing data	261	451
days with $N_{\text{miss}} > 720$ min missing data	9	11
days with no data ( $N_{\text{miss}} = 1440$ min)	21	30

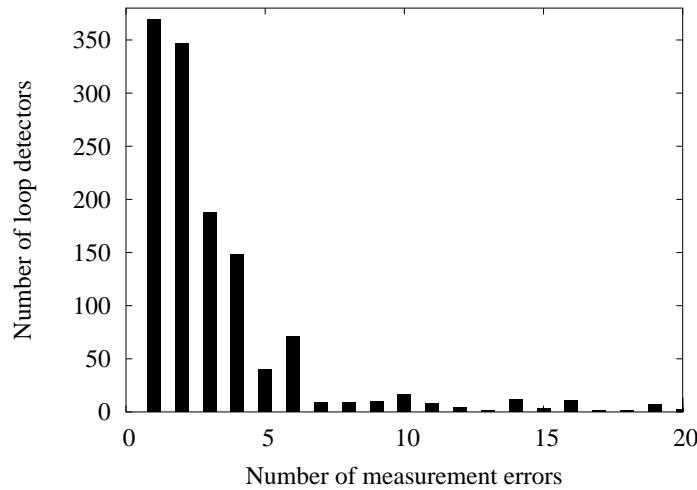
**Table 4.3.** Statistics about the number of days with  $N_{\text{miss}}$  min missing data from 2000/10/05 to 2004/05/31. In the statistics the data from the traffic control centers in Leverkusen and in Recklinghausen are distinguished.

#### 4.2.2 Errors Detected by the Sub Central

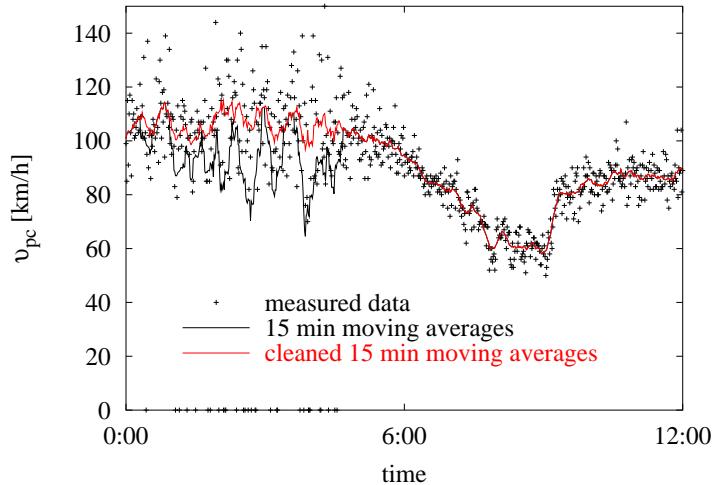
If for any reason a single value is not detected, the value is set to 255. Furthermore in each sub central there are some plausibility checks of each minute aggregated data row. According to [Klukas-Illen and Weiss, 2001] all values of a minute are set to 255 if

- $\mathcal{B}(J_{\text{veh}}) = 255$  or  $\mathcal{B}(J_{\text{lor}}) = 255$ ,
- $\mathcal{B}(J_{\text{lor}}) > \mathcal{B}(J_{\text{veh}})$ ,
- $\mathcal{B}(J_{\text{veh}}) - \mathcal{B}(J_{\text{lor}}) = 0$  and  $\mathcal{B}(v_{\text{pc}}) \neq 0$  and  $\mathcal{B}(v_{\text{pc}}) \neq 255$ ,
- $\mathcal{B}(J_{\text{lor}}) = 0$  and  $\mathcal{B}(v_{\text{lor}}) \neq 0$  and  $\mathcal{B}(v_{\text{lor}}) \neq 255$ ,
- $\mathcal{B}(J_{\text{veh}}) - \mathcal{B}(J_{\text{lor}}) > 0$  and  $\mathcal{B}(v_{\text{pc}}) = 0$ ,
- $\mathcal{B}(J_{\text{lor}}) > 0$  and  $\mathcal{B}(v_{\text{lor}}) = 0$ .

Thereby indicates  $\mathcal{B}$  the non dimensional value of the transmitted byte. Furthermore  $\mathcal{B}(\rho_{\text{rel}})$  is set to 255 if  $\mathcal{B}(\rho_{\text{rel}}) > 100$ . A validation of the data shows that none of the conditions above could be observed, except  $\mathcal{B}(\rho_{\text{rel}}) > 100$ . 162 detectors provide occupancies from time to time higher than 100 %. To deal with this circumstance the value of  $\rho_{\text{rel}}$  is checked for this whenever it is needed.



**Fig. 4.5.** Frequency distribution of errors detected by the sub central at the 2004/05/31. Together with 2,569 loop detectors that are completely free of this error more than 1,000 detectors with a small amount of errors offer reliable data. 611 detectors totally failed at this day.



**Fig. 4.6.** Systematically underestimation of the velocity of passenger cars  $v_{pc}$ . Although no vehicle is measured a velocity of  $v_{pc} = 0$  km/h is transmitted to the traffic control center. This can lead to systematical underestimations analysing the univariate traffic time series of the velocity (black line). Setting those values to N/A the time series can be cleaned and the underestimation avoided (red line).

In some cases, for instance, if a loop detector is cut off due to a construction area, the data is even set to 255 for a long time. In a similar way it sometimes happens, that a detector is replaced by another a few meters up- or downstream. Because the detector is identified by its location, this means that the former detector provide permanent values of 255. This happens also to each detector that is finally disconnected. For the 2004/05/31 holds that 611 loop detectors only provide values of 255 all day long, whereas 2,569 detectors are completely free of this error. In Fig. 4.5 the frequency of transmitted values of 255 at the 2004/05/31 can be seen. Because such errors are clearly detected, misinterpretations can be avoided in the same way as it is done with missing data.

A combination that can lead to misunderstandings and that can unfortunately not be avoided in the sub central is the transmission of the velocity  $v_{pc} = 0$  km/h or  $v_{lor} = 0$  km/h, respectively, whenever  $J_{pc} = 0$  veh/min or  $J_{lor} = 0$  veh/min. In Fig. 4.6 it can be seen that this leads to systematical underestimations analysing the univariate traffic time series of the velocity. Late at night, when no vehicle is measured, the velocity is also sometimes zero. This mistake can only be avoided, if the multivariate traffic time series is taken into account.

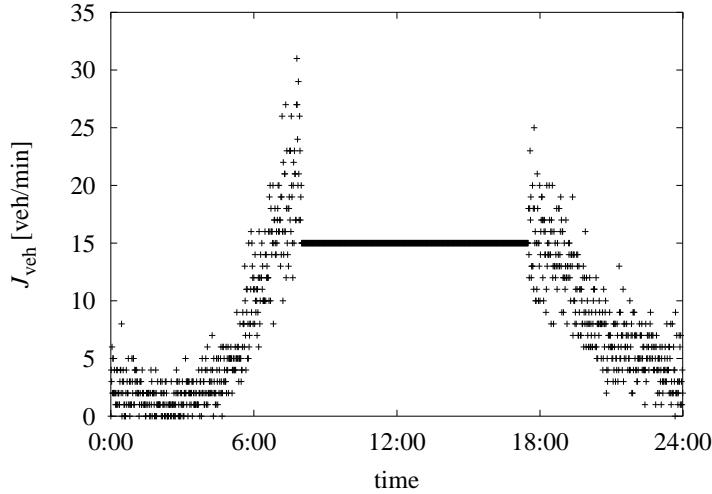
Then, whenever  $J_{pc} = 0$  veh/min or  $J_{lor} = 0$  veh/min, respectively, the corresponding velocities are set to the operator N/A that indicates, that they are not measured at this time. So the underestimation can be avoided.

#### 4.2.3 Rollback Error

The traffic data from the sub centrals update the data in the central every minute. Furthermore every minute an image of all data is sent from the two traffic control centers to the database at the university. In some cases, especially if there is an incident at the subcentral, the data at the traffic control center is not updated, and the same data is sent to the database. This leads to the case of the recurring of the 8 values several minutes in series and this is why the error is called *rollback error*.

The time for how long this error appears as well as the frequency can differ. Fortunately, the error influences the measurements in the same way it can be detected. A separate single rollback is nearly impossible to detect but influences the measurement only in the way that a gap in the time series is filled up with an adjacent value. Large rollback errors due to long term breakdowns of a sub central, like they are shown in Fig. 4.7, can be detected quite easily, because they influence the statistics of the time series in a dramatical manner, like it is shown in [Grunewald, 2001].

To avoid also misinterpretations of short term correlations and the results of the one-step-ahead predictor of the minute aggregated traffic data, the following



**Fig. 4.7.** Rollback error at the loop detector number 7243-003-001 at the 2002/09/30. Shown is the traffic flow  $J_{\text{veh}}$  during the day. This time series has a rollback function of  $R_{\text{rollback}}(\vec{q}, N) = 0.401$ , whereas for all further analysis a rollback function of  $R_{\text{rollback}}(\vec{q}, N) < 0.1$  is ensured.

rollback function  $R_{\text{rollback}}(\vec{q}, N)$  is defined:

$$R_{\text{rollback}}(\vec{q}, N) = \frac{\sum_{t=1}^N \mathcal{C}(\vec{q}(t), \vec{q}(t+1)) \mathcal{T}(\vec{q}(t), \vec{q}(t+1))}{\sum_{t=1}^N \mathcal{T}(\vec{q}(t), \vec{q}(t+1))}, \quad (4.8)$$

thereby is  $\vec{q}(t)$  the eight dimensional state space vector consisting of the measured bytes at time step  $t$  of a traffic time series  $\vec{q}$  with  $t = 1, 2, \dots, N$ . The test function  $\mathcal{T}$  indicates whether the data of two adjacent vectors are available, not zero, and not 255:

$$\mathcal{T}(\vec{q}(t), \vec{q}(t+1)) = \begin{cases} 0, & \text{if } \vec{q}(t) = \overrightarrow{\text{N/A}} \text{ or } \vec{q}(t+1) = \overrightarrow{\text{N/A}} \text{ or} \\ & J_{\text{veh}}(t) = 255 \text{ or } J_{\text{veh}}(t+1) = 255 \text{ or} \\ & (J_{\text{veh}}(t) = 0 \text{ and } J_{\text{veh}}(t+1) = 0), \\ 1, & \text{else.} \end{cases} \quad (4.9)$$

Thereby  $\overrightarrow{\text{N/A}}$  is the eight dimensional vector of N/A values. The identity function  $\mathcal{C}$  indicates that these vectors are identical:

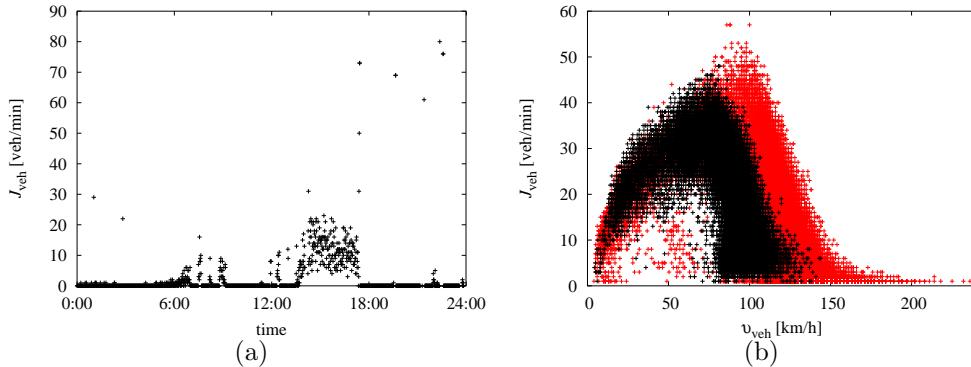
$$\mathcal{C}(\vec{q}(t), \vec{q}(t+1)) = \begin{cases} 1, & \text{if } \|\vec{q}(t) - \vec{q}(t+1)\|_2 = 0, \\ 0, & \text{else.} \end{cases} \quad (4.10)$$

So the rollback function is a measure for the frequency of two identical data sets in a row apart from other measurement errors. It is obvious that  $0 \leq$

$R_{\text{rollback}}(\vec{q}, N) \leq 1$  and the higher the rollback function the higher is the number of adjacent identical measurements and so the possibility for rollback errors. The data that can be seen in Fig. 4.7 has a rollback function of  $R_{\text{rollback}}(\vec{q}, N) = 0.401$ , whereas empirical investigations show that a daily traffic time series ( $N = 1440$ ) provides reliable results for  $R_{\text{rollback}}(\vec{q}, N) < 0.1$ . So this condition is ensured for all time series analysed in this work.

#### 4.2.4 Obvious Errors

Apart from the errors mentioned above some errors appear that are obvious when one is looking at the graph of a traffic time series, but they are nevertheless hard to detect among millions of measurements. In Fig. 4.8 (a) such a traffic time series of a particular loop detector at a day can be seen. Obviously the loop detector measures for any reason completely unreliable values. Because such unreliabilities have been a large problem analysing urban traffic data, in [Chrobok, 2000] some methods are proposed how to detect them.



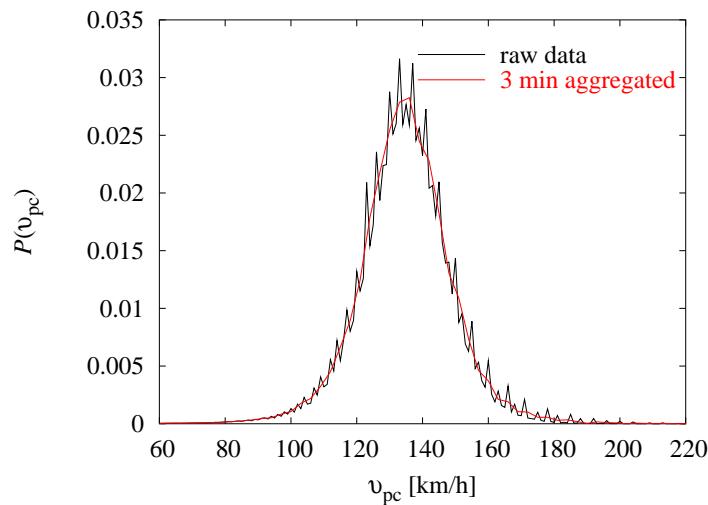
**Fig. 4.8.** (a) Measurement errors of the inductive loop detector number 6133-006-034 at the 2004/01/28. Although the errors are quite obvious when someone is looking at the graph, it is sometimes difficult to filter them among mass data. (b) Each intact loop detector measures only values up to the maximum flow of the lane. In this case the fundamental diagram of the cross section 4348-011 on the right lane (black crosses) and left lane (red crosses) of a three lane motorway during the second half of the year 2003 can be seen. To avoid the problem of identifying the maximum for each loop detector, a very high value of  $J_{\text{veh}} \geq 80$  veh/min is used as an error indicator.

One of those methods is to analyse the frequency distributions of measured data searching for a maximum value that can be measured. Then the traffic time series is scanned for data that lies beyond this limit. This method is based on the well known physical fact (see, for instance, [Leutzbach, 1998]), that a maximal traffic flow exists. This maximum depends on the capacity of the lane. It can be also seen in the fundamental diagram in Fig. 4.8 (b).

Optionally the method proposed in [Chrobok, 2000] can be already used in the sub central [Klukas-Illen and Weiss, 2001] and the data can be marked with 255 as explained in Sec. 4.2.2. Thereby is the maximum value a parameter that has to be manually adjusted. Unfortunately an analysis of the traffic data shows, that this method is not used in all sub centrals. Thus, in this work all traffic time series that are analysed are checked for the condition  $J_{\text{veh}} < 80 \text{ veh/min}$ , what is a very high limit also for lanes with high capacities.

#### 4.2.5 Roundoff Errors

In [Chrobok, 2000] there are also some investigations about motorway traffic data and also some of the data that are used in this work are analysed. Investigating the probability distributions of velocity functions it has been found, that they are in good agreement with Gaussian distributed functions. But also peaks have been found that are in no agreement with naturally probability distribution functions like the Gaussian distribution (see Fig. 4.9).



**Fig. 4.9.** Roundoff errors of the velocity at the loop detector number 4092-072-002. The probability distribution  $P(v_{\text{pc}})$  of measured passenger car velocities  $v_{\text{pc}}$  from the 2001/01/01 to the 2004/06/31 shows large peaks that can only be explained with roundoff errors in calculating the integer value of the velocity from the measured time a vehicle needs to pass the detector. Mistakes can be easily avoided aggregating the data in 3 min intervals. Then the peaks nearly completely disappear.

This peaks appear mainly for large velocities. Their appearance lies in the fact, that the velocity is theoretically calculated internally via the time  $\Delta t$  a vehicle needs to pass a detector of the given length  $L$  mentioned in Sec. 4.1. Then the velocity is given with an accuracy of 1 km/h. Because  $\Delta t$  is measured in milliseconds and no floating points for the calculation of the velocity are used, the velocities in unities km/h correspond to different quantities of  $\Delta t$  measurements.

For low velocities this does not matter, whereas for high velocities holds that the quantity of  $\Delta t$  measurements for adjacent values can differ drastically. Even more it can be, that some values could not be measured.

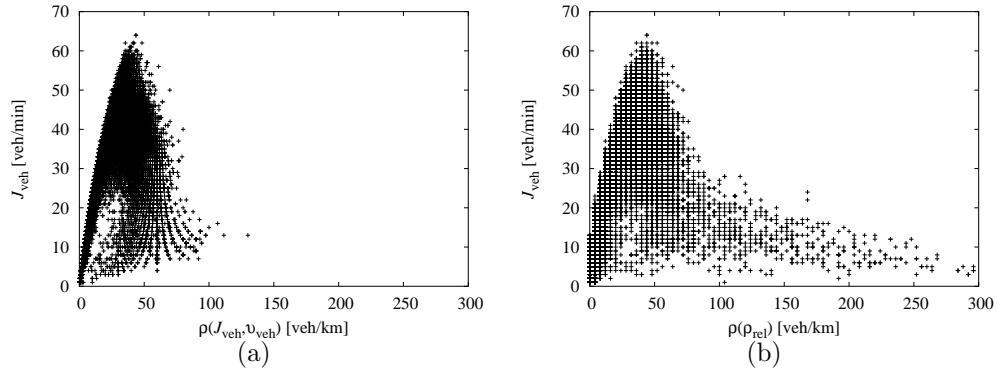
To make this more clear an example is given in the following. Imagine a detector length of  $L = 2.5$  m. Then, if a time  $\Delta t = 87$  ms is measured the velocity is calculated as  $v = \frac{2.5 \text{ m}}{0.087 \text{ s}} \frac{3.6 \text{ s} \cdot \text{km}}{\text{m} \cdot \text{h}} \approx 103.45 \text{ km/h}$ . So the integer value 103 km/h is measured. But the next faster velocity that can be measured belongs to the time  $\Delta t = 86$  ms and is  $v = \frac{2.5 \text{ m}}{0.086 \text{ s}} \frac{3.6 \text{ s} \cdot \text{km}}{\text{m} \cdot \text{h}} \approx 104.65 \text{ km/h}$ . Rounding off this value the integer value of the velocity is  $v = 105 \text{ km/h}$ . So there is no possibility of measuring  $v = 104 \text{ km/h}$ . The higher the velocity the higher is the number of integer values of the velocity in kilometres per hour that could not be measured. The only reason, why the value of  $v = 104 \text{ km/h}$  appears in the data set in spite of this is, that not each single velocity is transmitted but the minute aggregated velocity (for instance, when two cars of  $v_1 = 103 \text{ km/h}$  and  $v_2 = 105 \text{ km/h}$  pass the detector in a minute).

Because of the two possible different detector lengths  $L$  the peaks can appear at different values. Note that not all of the loop detectors behave in this way. It seems that newer ones use floating point operations internally. In any case this effect is no problem because the inaccuracy is under the limit given in Table 4.2. As can be seen in Fig. 4.9, if the data is aggregated in 3 min intervals a smooth curve appears as expected.

#### 4.2.6 Scaling Errors

Very hard to detect are measurements with a wrong scaling. This means, that one or more of the measured traffic values of a loop detector are systematically too low or too high. What makes the detecting even more complicate is when there is no constant scaling factor, but in a particular region the values are unreliable. One well known example of this are problems reliably detecting the number of vehicles that pass the detector in very densely traffic states. Furthermore the corresponding very low velocity is not detected. The reason therefore is the measurement technique that is based on the movement of the vehicle. That means, that a vehicle is only measured if it passes the loop detector completely. In every other case neither the very low velocity nor the vehicle itself is detected. The traffic flow is under-, the velocity overestimated.

The main effect of this phenomenon is discussed in detail in [Neubert, 2000] and can be seen in the fundamental diagram if the density  $\rho$  is calculated using the hydrodynamical relationship of Eq. 2.4 ( $\rho = \rho(J_{\text{veh}}, v_{\text{veh}})$ ). In case of a congestion, in the fundamental diagram very low flows  $J_{\text{veh}}$  at low  $\rho(J_{\text{veh}}, v_{\text{veh}})$  are found like it is shown in Fig. 4.10 (a), instead of ending at the maximum  $\rho_{\text{max}}$  if the occupancy  $\rho_{\text{rel}}$  and Eq. 2.7 ( $\rho = \rho(\rho_{\text{rel}})$ ) is used like in Fig. 4.10 (b). Using  $\rho(\rho_{\text{rel}})$  instead of  $\rho(J_{\text{veh}}, v_{\text{veh}})$  is also the only way to solve this problem. Nevertheless inaccuracies because of missing very slow vehicles cannot be avoided and in the extensive literature no method is known how to solve this problem, apart from



**Fig. 4.10.** (a) Errors because of movement based measurements. In case of congestion, the traffic flow  $J_{\text{veh}}$  is over-, the velocity  $v_{\text{veh}}$  underestimated and the fundamental diagram turns back to zero. (b) Using the occupancy  $\rho_{\text{rel}}$  in calculating the density  $\rho(\rho_{\text{rel}})$  provides the correct fundamental diagram. How the problem with the vehicles, that are not detected in the traffic flow can be solved is unfortunately not known.

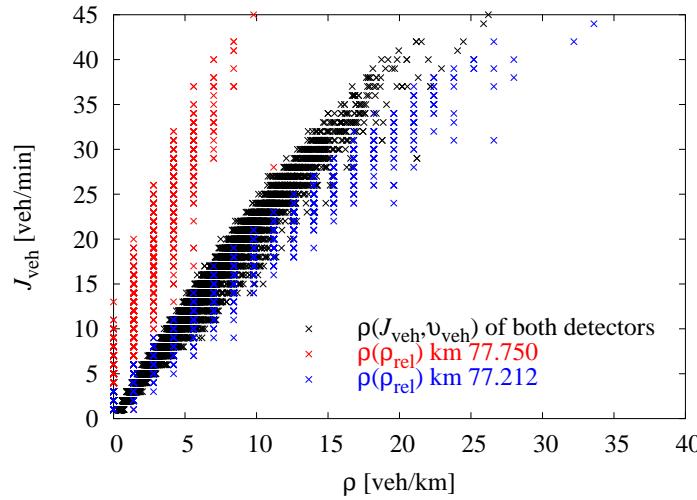
changing the detection device.

Unfortunately just the occupancy  $\rho_{\text{rel}}$  is the measurement value, that is the most difficult one to be measured uniformly. To explain this the Fig. 4.2 can help. Imagine a threshold beyond which the signal is interpreted as a vehicle that passes the detector. For the measuring of the flow or the velocity changings of a certain size of this threshold do not matter. But even small changings of the threshold influence the time, the detector is interpreted as being covered, and in the same way they influence the measured occupancy.

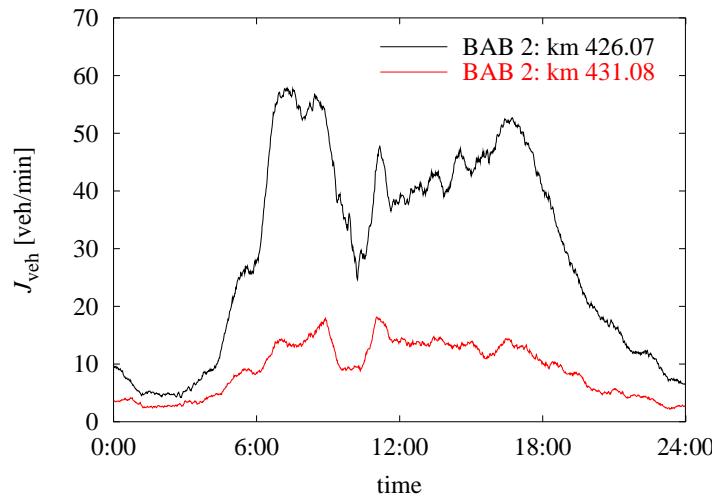
In Fig. 4.11 fundamental diagrams of two detectors with a distance of a 538 m and no on- or off-ramps between them are shown. Whereas the densities  $\rho(J_{\text{veh}}, v_{\text{veh}})$  calculated with Eq. 2.4 lie in the same region, the densities  $\rho(\rho_{\text{rel}})$  calculated with Eq. 2.7 are in two different regions.

The decision which detector measures wrong occupancies coheres with the choice of  $\rho_{\text{max}}$  in Eq. 2.7. So it is even hard to quantify the faultiness of each detector. For the same reason it is nearly impossible to obtain the real absolute value of the occupancy.

In many cases as well as in the present work the different scaling of the occupancy is no problem, because it is considered as a relative quantity. But if the occupancies of two detectors are compared or even spatial correlations are investigated as done in [Knospe, 2002], the different scaling can lead to completely wrong results. For all applications or investigations that need comparable occupancies, for instance, in controlling variable message signs, Grunewald [2001] proposes to rescale the occupancies according to the density that belongs to the maximum measured traffic flow in the fundamental diagram. This means that a rescaled value  $\rho_{\text{rescaled}}(l) = c_{\rho}(l)\rho_{\text{rel}}$  with the scaling factor  $c_{\rho}(l)$  that depends on the particular



**Fig. 4.11.** Scaling errors measuring the occupancy  $\rho_{\text{rel}}$  in the fundamental diagram of two loop detectors. Both detectors are placed at the motorway BAB 57, the detector 6253-021-002 at km 77.212 and the detector 6253-027-002 at km 77.750. Because the distance is only 538 m and there are no on- and off-ramps between them, similar traffic states are expected. Indeed the fundamental diagram correspond to each other, if the density  $\rho$  is calculated using the hydrodynamical relation  $\rho = \rho(J_{\text{veh}}, v_{\text{veh}})$ . But if the measured occupancy  $\rho_{\text{rel}}$  is used, the points lie in two different regions. This comes because the measurements of the occupancy  $\rho_{\text{rel}}$  are not uniform.

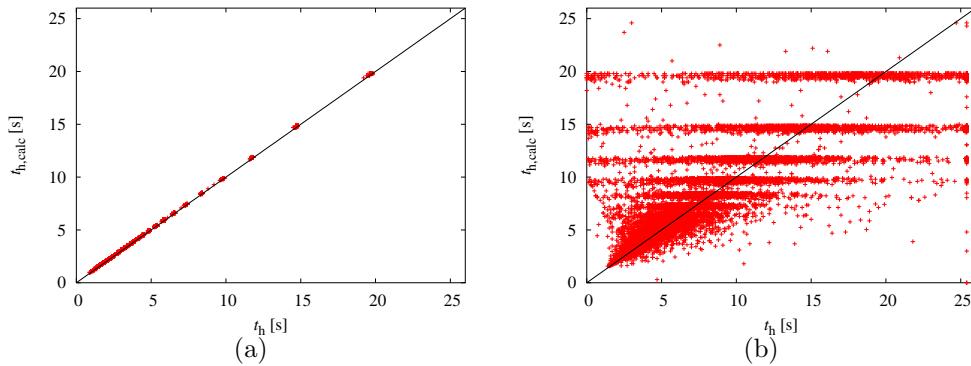


**Fig. 4.12.** Scaling errors of the traffic flow  $J_{\text{veh}}$ . Measuring  $J_{\text{veh}}$  at two adjacent cross sections of the motorway BAB 2 at the 2003/12/05 show that only a fraction amount of measured vehicles that pass the section at km 426.07 are measured about 5 kilometres downstream at km 431.08. In contrast to scaling errors of the occupancy, scaling errors of the traffic flow are not so frequent and it is quite clear which loop detector is defect.

loop detector  $l$  is used for all further calculations. For instance, the loop detectors that can be seen in Fig. 4.11 can be rescaled with  $c_\rho(6253-021-002) = 0.786$  and  $c_\rho(6253-027-002) = 2.857$ .

As mentioned above scaling errors of the velocity or the traffic flow are not as common as of the occupancy. But nevertheless they can appear. Like in the case of the occupancy, some obvious scaling errors can be found comparing adjacent loop detector data when there are no on- or off-ramps between them. Note that for comparing the traffic flow at different sections it is important to accumulate the flow for all possible lanes of the cross section. The example in Fig. 4.12 shows the traffic flow measurements of two adjacent cross sections with a distance of about 5 kilometres, but with no on- and off-ramps between them. In opposite to scaling errors of the occupancy it is in most cases quite clear which loop detector measure unreliable values. In this example the detectors at kilometre 431.08 fail measuring much too few vehicles. In fact, they nearly only measure the lorries, and a few month later they totally failed.

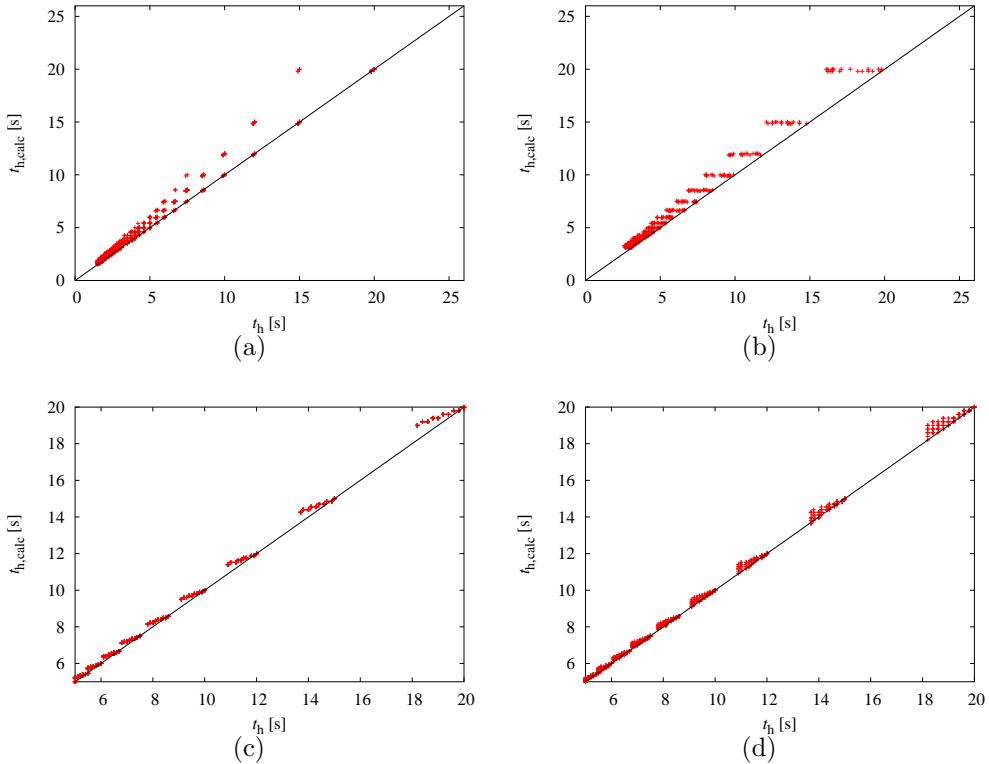
Comparing the loop detector data with data of adjacent detectors is very time consuming. For every pair of loop detectors it has to be ensured, that there is no possibility for leaving the motorway between them. Furthermore there are not everywhere adjacent loop detectors available that fulfill this condition.



**Fig. 4.13.** (a) Measured  $t_h$  and calculated  $t_{h,calc}$  time headways of the loop detector number 4098-063-002 during February 2002. The calculations come from the measured traffic flow  $J_{veh}$  and the occupancy  $\rho_{rel}$  using Eq. 4.5. (b) The same graph for the loop detector 2635-000-017 shows strong deviations from the bisecting line through the origin.

Another method to test the loop detectors for different scalings of the values is using Eq. 4.5 and compare the measured time headway  $t_h$  and the calculated one  $t_{h,calc}$ . Apart from some round off errors the values have to be exactly the same, if the measurements are completely consistent. This holds in general, for instance, for detector 4098-063-002 in Fig. 4.13 (a). But as can be seen in Fig. 4.13 (b) for the detector 2635-000-017 this condition is not satisfied.

Because of the roundings, the inconsistencies can be identified interpreting the



**Fig. 4.14.** Time headway testing using the analysis of different patterns that come with deviations of the measured  $t_h$  and the corresponding calculated  $t_{h,calc}$  time headways. Having  $t_h$  at the abscissa and  $t_{h,calc}$  at the ordinate, deviations of the bisecting line through the origin can be interpreted as follows. (a) Systematic errors  $\Delta J_{veh}$  in regard to the real traffic flow  $J_{veh}$  lead to points lying on the curve  $t_{h,calc} = \frac{J_{veh}}{J_{veh} + \Delta J_{veh}} t_h$ . (b) Most common are errors because of varying deviations of the time headway. These can be clearly identified because they cause straight lines at constant  $t_{h,calc}$ . In the same manner constant (c) and varying (d) inaccuracies measuring the occupancy can be seen as straight lines or scattering triangles that deviate from the bisecting line.

patterns in the graphs. In Fig. 4.14 (a) – (d) the patterns of different defects are shown. The most common error that is appearing in the data are fortunately unreliable measurements of the time headway, so in most cases the flow, occupancy, and velocity can be used for analysis.

This method of *time headway testing* of the reliability of the measured data is only possible if all of the three values are available. In this work *time headway testing* is always used, when a univariate or multivariate traffic time series of a single loop detector is needed. Note that identifying the particular pattern may be difficult, especially when more than one traffic value is faulty. Furthermore the magnitude of the deviation can differ.



# Chapter 5

## Empirical Data Analysis

“Look back to earlier times and you will know the things to come.”

Chinese proverb

In this chapter an empirical analysis of the traffic data is presented up to statistical results in regard to traffic forecast. The focus is to describe the different effects on different timescales, whether and how they can be described with classified traffic time series and how they can be forecast.

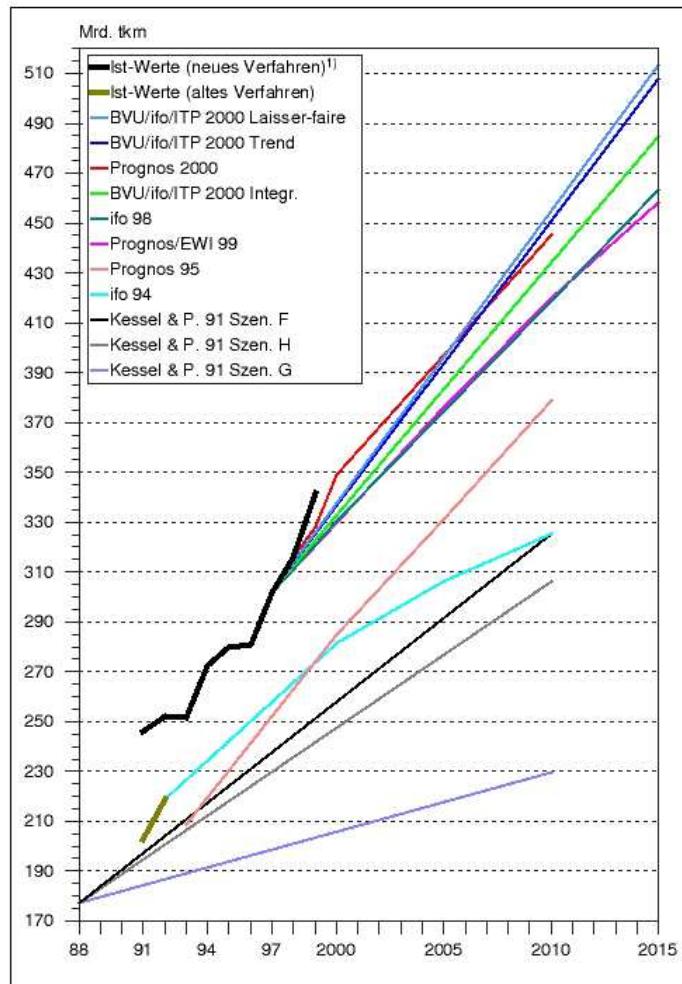
### 5.1 Secular Trend

The change of the traffic demand in future is an important factor not only for the evolution of the economic growth but also for infrastructure planning in any modern society. This is why there is a large interest of many institutions in forecasting future traffic demands in large timescales [Bundesministerium für Verkehr, 2003, Deutsche Shell AG, 1999]. The timescales of such forecasts are at least a month, but normally years.

Also for traffic information systems that usually work on smaller timescales, changings of long term conditions can influence the results of short term forecasts. For instance, the classification of mean daily traffic that is investigated in Sec. 5.3 and later used in Chapter 7 comprises sometimes analyses of days that occur only one time a year. In such a case long term changings of the traffic demand influence drastically the results. Therefore, in this section some aspects of the long term evolution of the traffic demand are analysed.

Investigating all aspects that could influence the traffic demand in years would extend the scope of this work. A basic problem that comes with this attempt is, that even some causes that will influence the traffic demand are not known yet what leads to large mistakes. And also the impacts of present conditions that are taken into account can be rated in a completely different way, what leads to a large range of different forecast results as can be seen in Fig. 5.1.

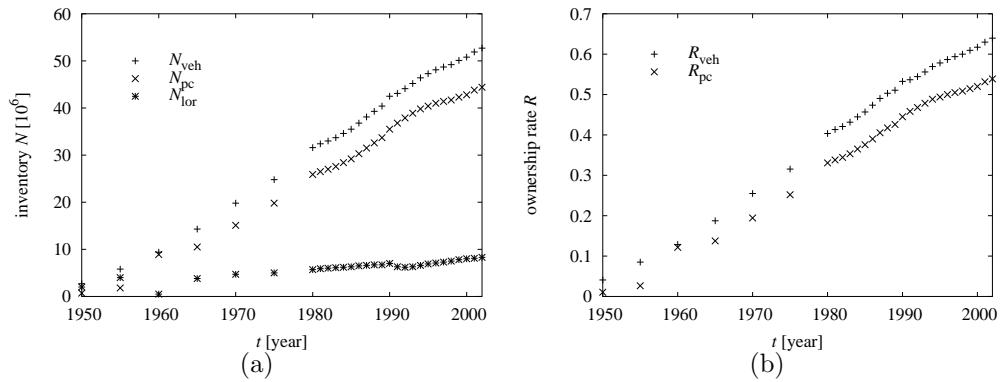
Abbildung 4-23: Vergangenheitsentwicklung und langfristige Prognosen der Transportleistung des gesamten Straßengüterverkehrs



**Fig. 5.1.** Forecasting of secular trends used by the government in Germany. On the abscissa is the year, on the ordinate the effort of freight transport measured in  $10^9$  tons · km. Interpreting the impacts in a different way causes a varying set of different forecasts with a large range. The figure is stemming from [Bundesministerium für Verkehr, 2003].

So the objective in this work can only be to give an upper and a lower limit for the long term evolution of the traffic demand. To investigate the secular trend the data of the loop detectors cannot be used, because those data are available since October 2000. Even other traffic counts are only available for particular sections or small time intervals. But what is needed is a measurement that is obtainable over a time interval of the last decades, most suitable would be a measurement value that reflects the traffic load from the first years of the invention of power driven vehicles up to now.

One measurement value that fulfills this condition is the vehicle inventory that is published every year by the Federal Ministry of Transport of Germany [Bundesministerium für Verkehr, 2002]. Therein traffic data are available from the year 1950 on. The total mileage would be an even more comparable measure, but unfortunately this is only given for the last recent years. In Fig. 5.2 (a) the inventory of all vehicles are shown as well as those of passenger cars and lorries. Unfortunately there seem to be two obvious inconsistencies: In the year 1960 there seems to be a wrong classification of passenger cars and lorries because almost all vehicles are identified as passenger cars. From the year 1989 to 1990 the vehicle inventory increases abnormally higher, from the year 1990 to 1991 abnormally lower than in other years. Corresponding to this the lorry inventory seems to decrease from 1990 to 1991. These inconsistencies may come due to faults in the data collection during the reunification of Germany, but in [Bundesministerium für Verkehr, 2002] there are no reasons given for that.



**Fig. 5.2.** (a) Inventory of passenger cars  $N_{\text{pc}}$ , lorries  $N_{\text{lor}}$  and all vehicles  $N_{\text{veh}}$  in Germany from 1950 up to now. (b) Vehicle ownership rate  $R_{\text{veh}}$  and passenger car ownership rate  $R_{\text{pc}}$  from 1950 up to now. The data are stemming from [Bundesministerium für Verkehr, 2002].

As mentioned above not all of the factors that influence the vehicle inventory can be taken into account. But because of changings in the population development during the last decades, it makes sense not to use the inventory, but to use the

passenger car ownership rate  $R_{pc}$ , that calculates the number of passenger cars  $N_{pc}$  in relation to the population  $N_{pop}$ :

$$R_{pc} = \frac{N_{pc}}{N_{pop}}. \quad (5.1)$$

With a little bit different interpretation of the expression “ownership rate” this can be done for all vehicles  $N_{veh} = N_{pc} + N_{lor}$  whereby  $N_{lor}$  means the number of lorries. The vehicle ownership rate  $R_{veh}$  can be calculated:

$$R_{veh} = \frac{N_{veh}}{N_{pop}}. \quad (5.2)$$

The interpretation differs in that sense, that most inhabitants really own the passenger cars, whereby the vehicle ownership rate can be interpreted as the sum  $R_{veh} = R_{pc} + R_{lor}$  with the lorry ownership rate  $R_{lor}$  as addend. But most of the inhabitants do not own a lorry. This is rather a measurement for a demand in mobility of goods that an inhabitant produces. Note that the definition of “lorry” in case of measurements with loop detectors is given by the characteristic of the signal, whereas the definition in [Bundesministerium für Verkehr, 2002] comes from the official declarations. In Fig. 5.2 (b)  $R_{pc}$  and  $R_{veh}$  are shown from 1950 up to now.

The forecasts that predict the highest growth of traffic come from a constant annual percentage increase, what leads obviously to an exponential growth:

$$R_{veh}(t) = R_0 \exp(c[t - y_0]). \quad (5.3)$$

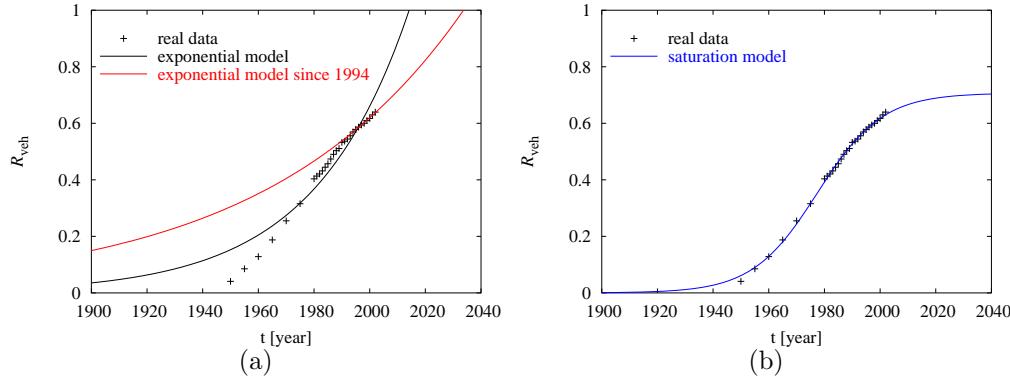
Thereby is  $t$  the time and the parameters  $c$  and  $y_0$  can be calculated approximating a least square fit through the past data.  $R_0 = 1$  veh/person is only needed for the unit <sup>1</sup>.

As can be seen in Fig. 5.3 (a) it makes no sense to use the fit through all the data. Obviously the timescales are already to large for a theory of an exponential growth. Using the data of the last ten years the function fits better with the values  $c = 0.01420 \text{ year}^{-1}$  and  $y_0 = 2033.7 \text{ year}$ . The corresponding function of the passenger car ownership rate fits best with  $c = 0.01193 \text{ year}^{-1}$  and  $y_0 = 2054.3 \text{ year}$ . Also these estimations can only serve as a model for an upper limit. It can be calculated in an easy way that the law leads to exorbitantly high values far in the future. In nature exponential growth only appears for a while until a limit is crossed and the law loses its validity.

Thus, to get the lower limit for approximations of  $R_{veh}(t)$ , a saturation model is proposed. Therefore the theory of an exponential growth is enhanced with the feasible assumption, that there exists a limited maximum number of vehicles and a saturation is assumed.

---

<sup>1</sup>Of course also the formula  $R_{veh,exp}(t) = R_0 \exp(at)$  can be fitted but that leads to very low numerical values.



**Fig. 5.3.** Future estimation of the vehicle ownership rates. (a) The most common present estimates come from an exponential growth. Whereas the fit through all the real data from the past leads to completely unreliable results, the fit through the data of the last ten years with the exponential model since 1994 may serve as an upper limit for further approximations. (b) The saturation model provides a lower limit for the evolution of the ownership rate. It comes from an assumption of a saturation using Eq. 5.8. The fit parameters can be seen in Table 5.1.

As a first approximation we are searching for a function that fulfills all of the following conditions:

- zero far in the past,
- exponential rising for a while,
- saturation after that,
- and converging to a limit.

An analytical function that fulfills the conditions is the hyperbolic tangent:

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \quad (5.4)$$

To be able to fit the function to the constraints the parameters  $a, b, S_{\text{veh}}, Y_{\text{veh}}$  have to be introduced. The vehicle ownership rate  $R_{\text{veh}}(t)$  can be written in the following form:

$$R_{\text{veh}}(t) = a \tanh(S_{\text{veh}} [t - Y_{\text{veh}}]) + b. \quad (5.5)$$

Thereby the product of the parameters  $a$  and  $S_{\text{veh}}$  is the the maximum slope of the curve and  $Y_{\text{veh}}$  is the inflection point, the year with the maximum slope. The parameters  $a$  and  $b$  can be substituted or eliminated, respectively. One of them can be directly eliminated due to the constraints, that there have been no vehicles far in the past. This means:

$$\lim_{t \rightarrow -\infty} R_{\text{veh}}(t) = 0 \Rightarrow 0 = -a + b \Rightarrow a = b. \quad (5.6)$$

Now let  $R_{\text{vehmax}}$  be the maximum of the vehicle ownership rate that can be reached in future. Then, one parameter can directly be substituted:

$$\lim_{t \rightarrow \infty} R_{\text{veh}}(t) = R_{\text{vehmax}} \Rightarrow R_{\text{vehmax}} = a + b. \quad (5.7)$$

Putting Eq. 5.6 and Eq. 5.7 in Eq. 5.5, the vehicle ownership rate  $R_{\text{veh}}(t)$  can be written in the following form:

$$R_{\text{veh}}(t) = \frac{R_{\text{vehmax}}}{2} [\tanh(S_{\text{veh}} [t - Y_{\text{veh}}]) + 1]. \quad (5.8)$$

In the same manner using Eqs. 5.5 – 5.8 the passenger car ownership rate  $R_{\text{pc}}(t)$  can be approximated as follows:

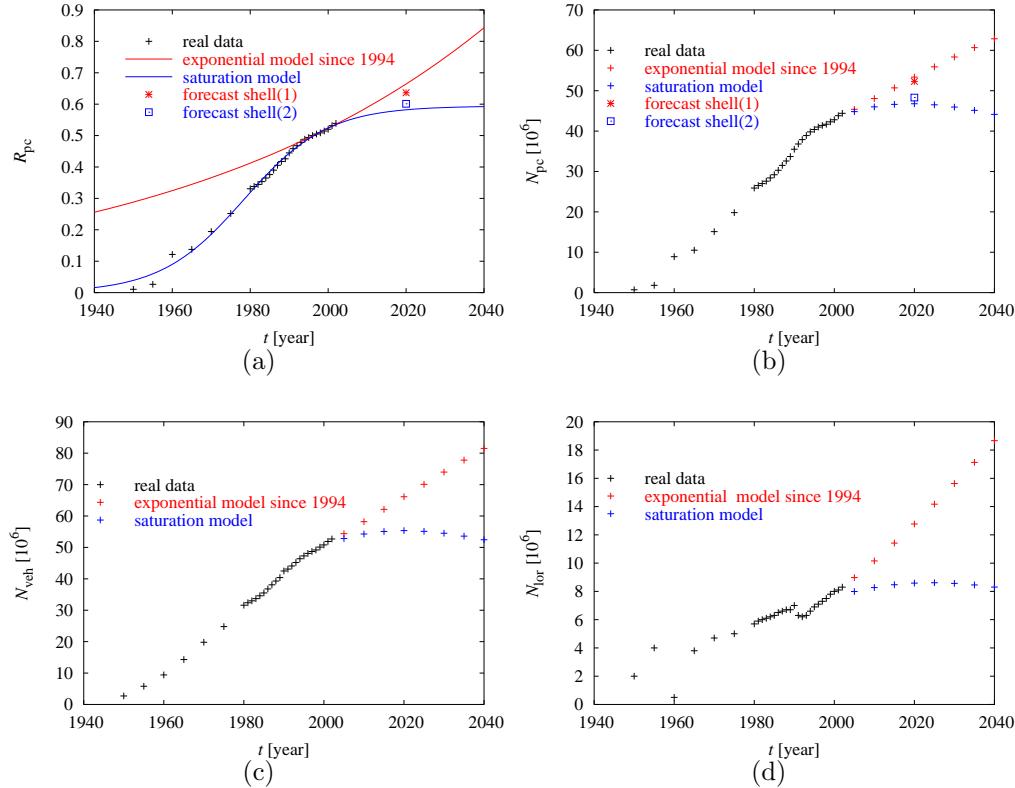
$$R_{\text{pc}}(t) = \frac{R_{\text{pcmax}}}{2} [\tanh(S_{\text{pc}} [t - Y_{\text{pc}}]) + 1]. \quad (5.9)$$

The parameters  $Y_{\text{pc}}$ ,  $Y_{\text{veh}}$ ,  $S_{\text{pc}}$ ,  $S_{\text{veh}}$ ,  $R_{\text{pcmax}}$  and  $R_{\text{vehmax}}$  have to be derived from the data. Therefore Eq. 5.8 is fitted to the measurement points using a least square fit. The results of the fitted vehicle ownership rate can be seen in Fig. 5.3 (b). Table 5.1 shows the approximations of all the parameters. According to this, the maximum vehicle ownership rate that is reached in the future is  $R_{\text{vehmax}} = 0.71$  veh/person, the maximum passenger car ownership rate  $R_{\text{pcmax}} = 0.59$  pc/person what can lead to the interesting interpretation that far in the future on average each person will own 0.59 passenger cars. The inflection points both already have been and are with  $Y_{\text{veh}} \approx 1977$  year and  $Y_{\text{pc}} \approx 1978$  year close to each other.

$Y_{\text{veh}}$ [year]	$S_{\text{veh}}$ [year $^{-1}$ ]	$R_{\text{vehmax}}$ [veh/person]	$\frac{1}{2}R_{\text{vehmax}}S_{\text{veh}}$ [veh/(person · year)]
1977.46	0.043	0.71	0.0152
$Y_{\text{pc}}$ [year]	$S_{\text{pc}}$ [year $^{-1}$ ]	$R_{\text{pcmax}}$ [pc/person]	$\frac{1}{2}R_{\text{pcmax}}S_{\text{pc}}$ [pc/(person · year)]
1978.43	0.047	0.59	0.0139

**Table 5.1.** Fit parameters of the vehicle and the passenger car ownership rates. According to this far in the future all persons will own on average 0.59 passenger cars. Interesting are also the inflection points during the years 1977 and 1978.

To compare these theories with another estimation a study in [Deutsche Shell AG, 1999] can be used. This study delivers two different forecast models of the future evolution of the traffic. The authors state that the first one, that is called here shell(1), is based on the assumption of an economic growth of  $0.018$  year $^{-1}$ , a stable population and an increased globalisation. Similar to this the second



**Fig. 5.4.** (a) Comparison of forecasts of the passenger car ownership rate  $R_{pc}$ . The forecasts of shell(1) and shell(2) lie between the exponential model that serves as an upper limit for approximations and the saturation model that serves as a lower limit. The same holds for the passenger car inventory  $N_{pc}$  that can be seen in (b). The same calculations can be done for the vehicle inventory  $N_{veh}$  (c), that is feasible for comparisons with the complete traffic demand, and for the lorry inventory  $N_{lor}$  that is shown in (d) for completeness in spite of the errors acquiring the data in the past.

variant shell(2) is based on the expectation of an economic growth of  $0.014 \text{ year}^{-1}$ , a decreasing population and an international interconnection that is limited by several governments.

Using the scenario shell(1) an increase of 8 million passenger cars up to the year 2020 is forecast, whereby the scenario shell(2) leads to an increase of 4 million passenger cars. Thus, based on the number of passenger cars of the year 1999 when the prognosis is made, shell(1) forecasts for the year 2020 a number of 52.3 million passenger cars and shell(2) 48.3 million passenger cars.

To compare these approaches with the passenger car ownership rate, the inventory has to be divided by the particular population that is expected for the year 2020. This is within shell(1) 82.2 million people because a constant population is expected, and within shell(2) 80.339 million people, what is the decrease expected in the second variant of Sommer [2000]. The comparison of the estimations can be seen in Fig. 5.4 (a). Obviously the estimations lie between the estimated upper and lower limits.

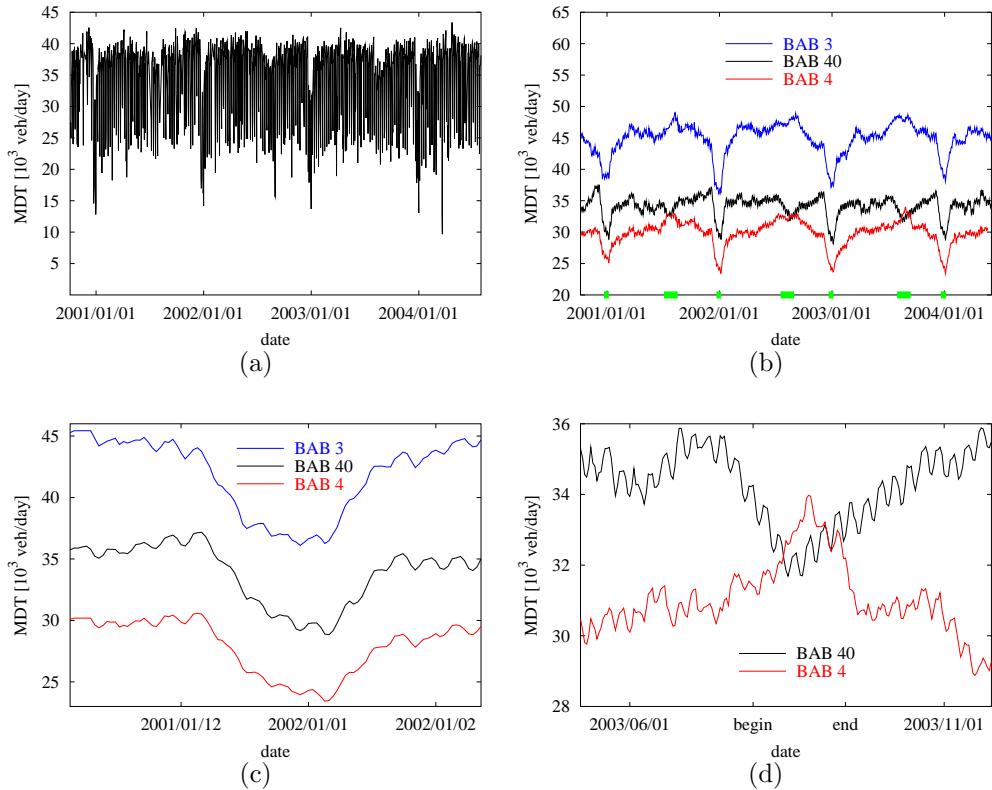
To complete considerations for the long term estimations not the ownership rates are needed but the inventories. Therefore the rates have to be multiplied by forecast populations. For this the estimations of the second variant of Sommer [2000] are used. The results can be seen in Fig. 5.4 (b) – (c).

## 5.2 Seasonal Differences

To investigate the seasonal differences the traffic data of the loop detectors from 2000/10/05 up to 2004/07/31 provide reliable information. Therefore the traffic flow of each lane is aggregated to a cross section based traffic flow and the minute based traffic data are accumulated to mean daily traffic (MDT). This MDT are exactly all vehicles that pass a cross section per day. To avoid errors only days with more than 1,200 minutes of correct measurements are taken into account.

In Fig. 5.5 (a) the MDT of the cross section 5877-019 of the motorway BAB 40 driving direction east can be seen. Unfortunately the MDT can differ from day to day and due to the fluctuations the graph is not very feasible for analysis. Better information offer the 31 days moving averages (for the calculation see Eq. 5.16) that can be seen in Fig. 5.5 (b) also for the cross section 4349-025 on the BAB 3 driving direction north and for the cross section 4488-096 on the BAB 4 driving direction west. During the yule tide and the corresponding school holidays the traffic demand clearly decreases on all motorways. In Fig. 5.5 (c) a higher resolution of this effect can be seen. The summer holidays lead to different effects that can be seen in Fig. 5.5 (d). They cause a decreasing traffic demand on the motorway BAB 40, which is a typical motorway used by commuters who go to work, whereas on the BAB 4 an increasing traffic is observed during the summer holidays.

Also the absolute quantity of vehicles can differ strongly from motorway to motorway. Nevertheless the characteristics of the MDT of all motorways stay in general



**Fig. 5.5.** (a) The MDT of the cross section 5877-019 on the motorway BAB 40 driving direction east from 2000/10/05 to 2004/07/31. As can be seen the MDT differs strongly from day to day. (b) The 31 days moving averages of the MDT of cross sections on the German motorways BAB 3, BAB 4, and BAB 40 is more feasible for comparisons. In general the characteristics stay the same. The biggest changes in the traffic demand are caused by the yule tide with the turn of the year and the summer holidays, that are both marked with the wide green bars on the abscissa. In (c) and (d) there are higher resolutions of these effects. Whereas the yule tide in (c) leads to similar patterns, even in these stable structures the same cause can have different effects. This can be seen around the summer holidays in (d). At the BAB 40, what is a typical motorway for commuters, the MDT decreases, whereas it increases on the BAB 4.

similar. From the Table 5.2 can be seen that the traffic on the large motorways are highly correlated. Those correlations are quite easy to interpret. The reasons that cause the differences in the traffic demand are strongly dominated by the calendar. In most cases the same cause lead to the same effect. On the one hand the summer holidays are an exception in the sense, that they lead to different effects. On the other hand that confirms the strong dominance of the calendar affecting the traffic demand.

1. cross section			2. cross section			r <sub>12</sub> (MDT)
number	motorway	direction	number	motorway	direction	
4135-015	BAB 1	south	4214-015	BAB 2	east	0.8656
4135-015	BAB 1	south	4349-025	BAB 3	north	0.8455
4135-015	BAB 1	south	4488-096	BAB 4	west	0.8716
4135-015	BAB 1	south	5877-019	BAB 40	east	0.7916
4214-015	BAB 2	east	4349-025	BAB 3	north	0.8155
4214-015	BAB 2	east	4488-096	BAB 4	west	0.8733
4214-015	BAB 2	east	5877-019	BAB 40	east	0.8990
4349-025	BAB 3	north	4488-096	BAB 4	west	0.8662
4349-025	BAB 3	north	5877-019	BAB 40	east	0.7560
4488-096	BAB 4	west	5877-019	BAB 40	east	0.8086

**Table 5.2.** Correlation coefficients r<sub>12</sub>(MDT) of the MDT on large German motorways of the year 2002. In spite of the space dependency the MDT of the large German motorways in different directions show high correlations which proves the large dominance of the calendar as external variable.

This strong influence of the calendar can also be seen in the power spectrum. Therefore the time dependent MDT(t) is disaggregated in sine and cosine shaped contributions of different frequencies  $f_n = \omega_n/2\pi$  and amplitudes  $a_n$  and  $b_n$ , respectively:

$$\text{MDT}(t) \equiv \widehat{F}(k\Delta T) = \frac{1}{K} \sum_{n=0}^{K-1} c_n e^{2\pi i f_n k \Delta T}, \quad (5.10)$$

with

$$i = \sqrt{-1}, \quad c_n = a_n + i b_n, \quad \text{and} \quad f_n = \frac{n}{K \Delta T}.$$

This is done with the discrete Fourier transform

$$c_n = \sum_{k=0}^{K-1} \widehat{F}(k\Delta T) e^{-2\pi i f_n k \Delta T}. \quad (5.11)$$

Then, the power spectrum is defined as follows:

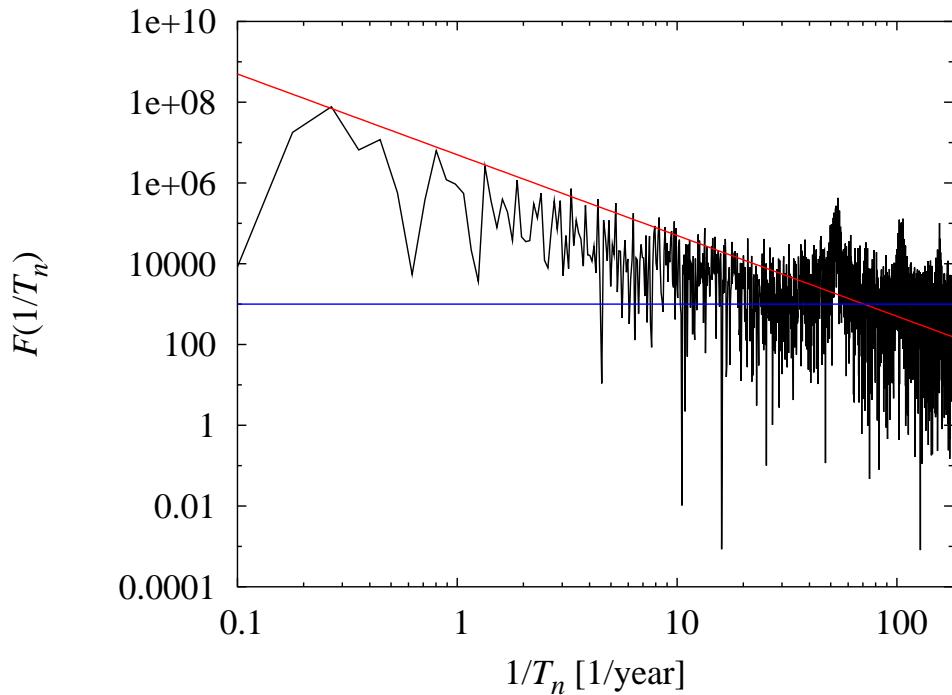
$$F(f_n) := |c_n|^2 = |a_n|^2 + |b_n|^2. \quad (5.12)$$

It is a measure for the intensity of the particular frequencies  $f_n = 1/T_n$  with the oscillation period  $T_n = K\Delta T/n$ . In Fig. 5.6 the power spectrum of the MDT of the cross section 5877-019 in driving direction east can be seen.

First at all the power spectrum declines with a power law

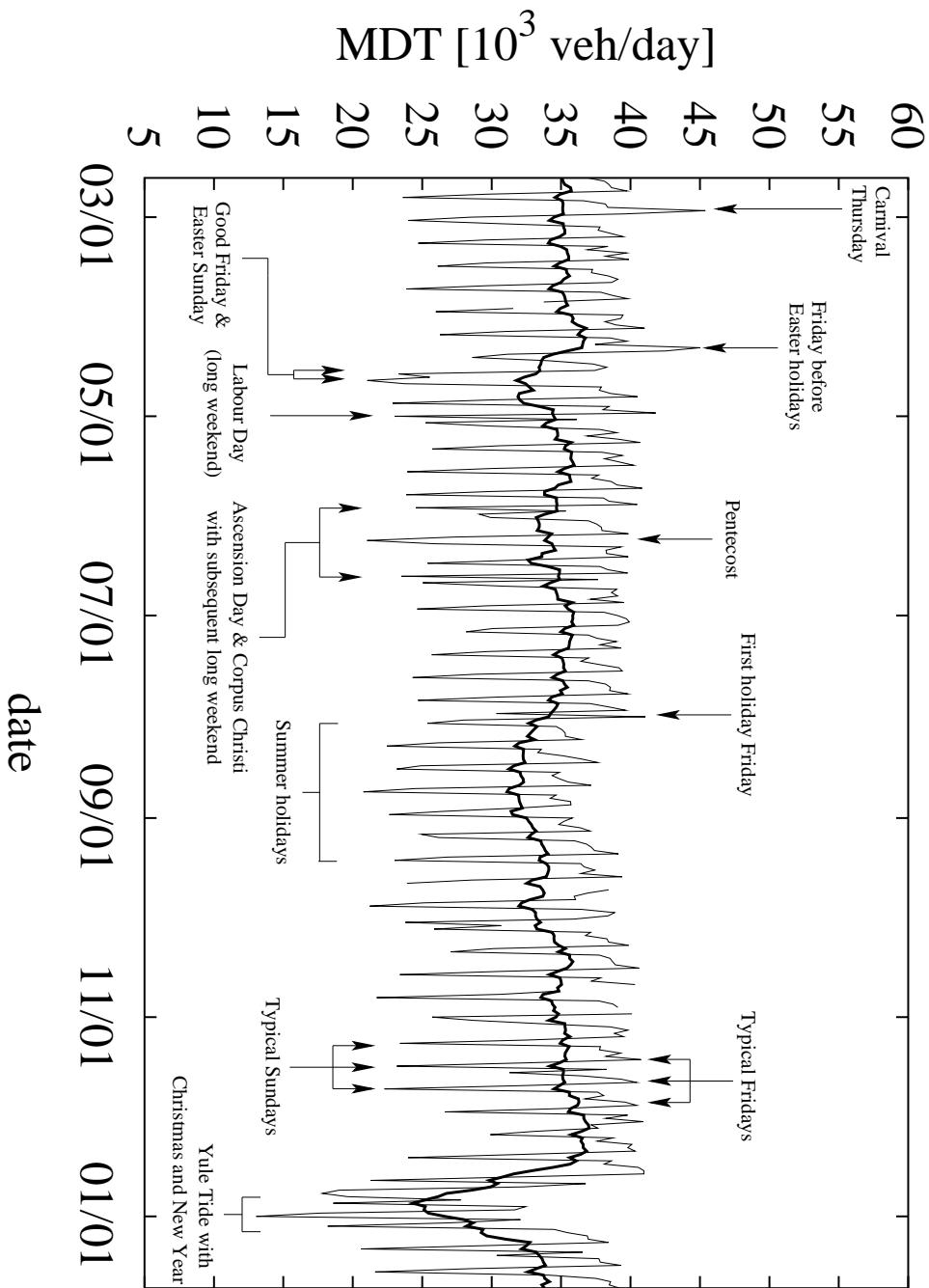
$$F(f_n) = F_0 f_n^{-\epsilon}. \quad (5.13)$$

This can be seen in the double logarithmic scale of the axes and the line with the slope  $-\epsilon$ . Investigations of different cross sections show that  $\epsilon = 2.0 \pm 0.08$ . After the decline it passes into a constant value that indicates gaussian white noise. From this noise clearly distinguish multiple frequencies of  $52 \text{ year}^{-1}$  and indicate weekly recurrent structures.



**Fig. 5.6.** The power spectrum of MDT of the cross section 5877-019 driving direction east. The blue line mark the Gaussian white noise and the red one the power law with the exponent  $\epsilon = 2.0$ . From the noise multiple frequencies of  $52 \text{ year}^{-1}$  clearly distinguish and indicate weekly recurrent structures.

These weekly structures can be seen in more detail using a higher resolution of the MDT during the year as in Fig. 5.7. The traffic demand of typical days can be seen as well as the effects of different causes in consequence of particular days. The traffic demand time series maps an interesting image of the social human behaviour. Knowing the causes and the corresponding effects is an essential part for each further investigation in regard to traffic forecast. In the 15 days moving averages the decreasing traffic because of Easter, summer, and winter holidays can be seen very clearly.



**Fig. 5.7.** Seasonal recurring structures in the MDT of the cross section 5877-019 during the year 2003. Within the typical days the highest traffic volume is measured on Fridays, the least on Sundays. Furthermore the effects of different causes in consequence of particular days can be observed. The measured traffic demand maps an interesting image of the social human behaviour. The central line marks the 15 days moving averages.

To investigate whether there are periodic seasonal fluctuations apart from the daily differences and apart from differences caused by special causes (for instance, school holidays, bridging days or long weekends) the traffic demand time series of comparable weekdays are investigated. Therefore all weekdays are chosen that fulfill the following conditions:

- the MDT is reliably measured according to Sec. 4.2,
- the days are no holidays,
- the days are beyond school holidays,
- no bridging days, days before or after holidays, or long weekends are taken into account.

The results can be seen in Fig. 5.8 (a) and in higher resolution of the year 2003 in Fig. 5.8 (b). The workingdays behave in a very similar manner. This is why the Tuesdays, Wednesdays, and Thursdays (TWT) are drawn with one line. The largest differences can be explained with this classification. Nevertheless a few fluctuations are still left.

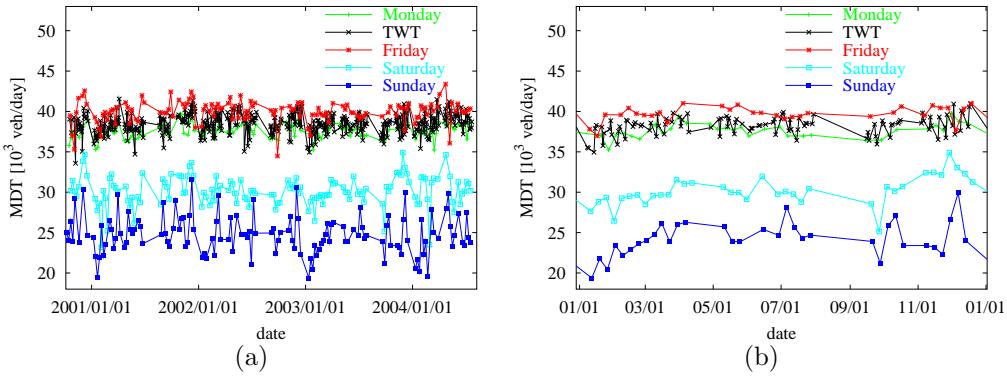
Filtering all the holidays and the special conditions most seasonal fluctuations disappear. The only one left can be seen around the turn of the year, especially at Saturdays and Sundays. Whereas before Christmas there is an increasing traffic flow at Saturdays because of much shopping traffic at the Sundays after the yule tide there is especially little traffic. Unfortunately investigating the particular reasons for the human behaviour would extent the scope of this work and is part of a psychological analysis. Nevertheless for traffic demand forecast the presence of the effects is of enormous importance. Note that these lie all outside the school holidays, so no day between Christmas and New Year is drawn in the graph.

The fluctuations that are still left are definitely higher within the Saturdays and Sundays than within workingdays, so it can be supposed, that there is a strong influence of spare time activities and special events. Those special events are very important to be taken into account for traffic flow forecast. Furthermore especially at weekends the weather conditions may strongly influence the traffic demand. Examples and problems that come with this are discussed later in Sec. 5.4.

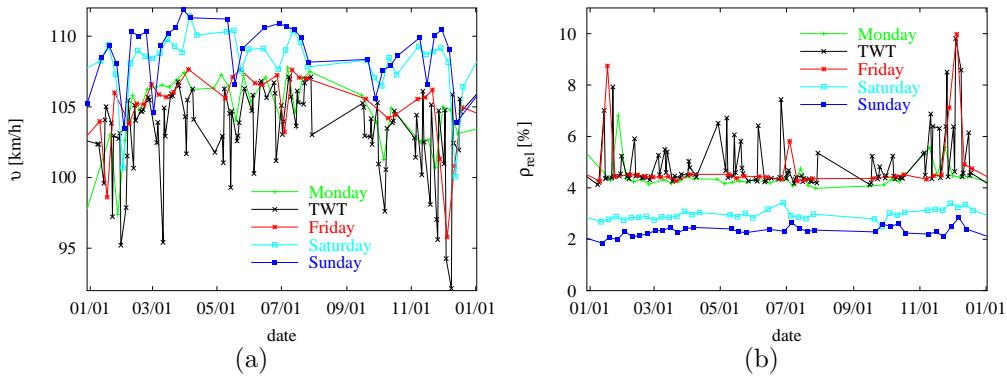
A short look should be taken on the other traffic data. It can clearly be seen in Fig. 5.9 that the MDT, the mean daily velocity, and the mean daily occupancy strongly dependent on each other. Whereby the velocity and the occupancy are mirror images of each other, the fluctuations of both are higher within days with a high traffic demand. This effect can be better seen on smaller timescales and is discussed later in Sec. 5.3.

Also the increasing density and decreasing velocity before the yule tide can be seen. Apart from this no general seasonal fluctuations can be observed.

The high peaks of the density and the low peaks of the velocity indicate the presence of a traffic congestion. So another way to investigate the traffic data is to take a look at the probabilities of a congestion. Therefore the number of days



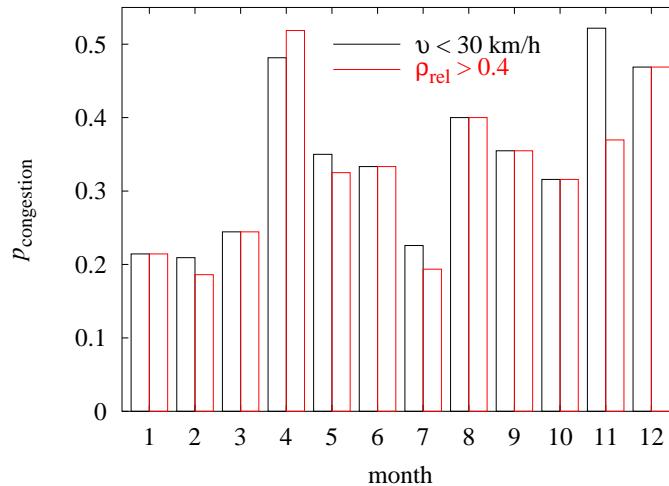
**Fig. 5.8.** (a) Seasonal fluctuations of the MDT of the cross section 5877-019 within particular weekdays. To avoid misinterpretations days with special attributes like holidays are separated out. Classifying the days in regard to the weekdays can help to explain many large changes. Workingdays behave in a very similar manner. Thus, Tuesdays, Wednesdays, and Thursdays are drawn with one line (TWT). (b) The same graph but in higher resolution for the year 2003.



**Fig. 5.9.** (a) The seasonal time series of the mean daily velocity  $v$  and (b) the mean daily occupancy  $\rho_{rel}$  for the cross section 5877-019 driving direction east for the year 2003 are mirror images of each other. On the one hand the classification that is also used for the MDT works very similar. On the other hand in opposite to the traffic demand, the fluctuations are higher within the working days.

condition	$v < 30 \text{ km/h}$	$\rho_{\text{rel}} > 0.4$
Monday	0.152	0.182
TWT	0.352	0.352
Friday	0.194	0.161
Saturday	0.028	0.028
Sunday	0.030	0.030

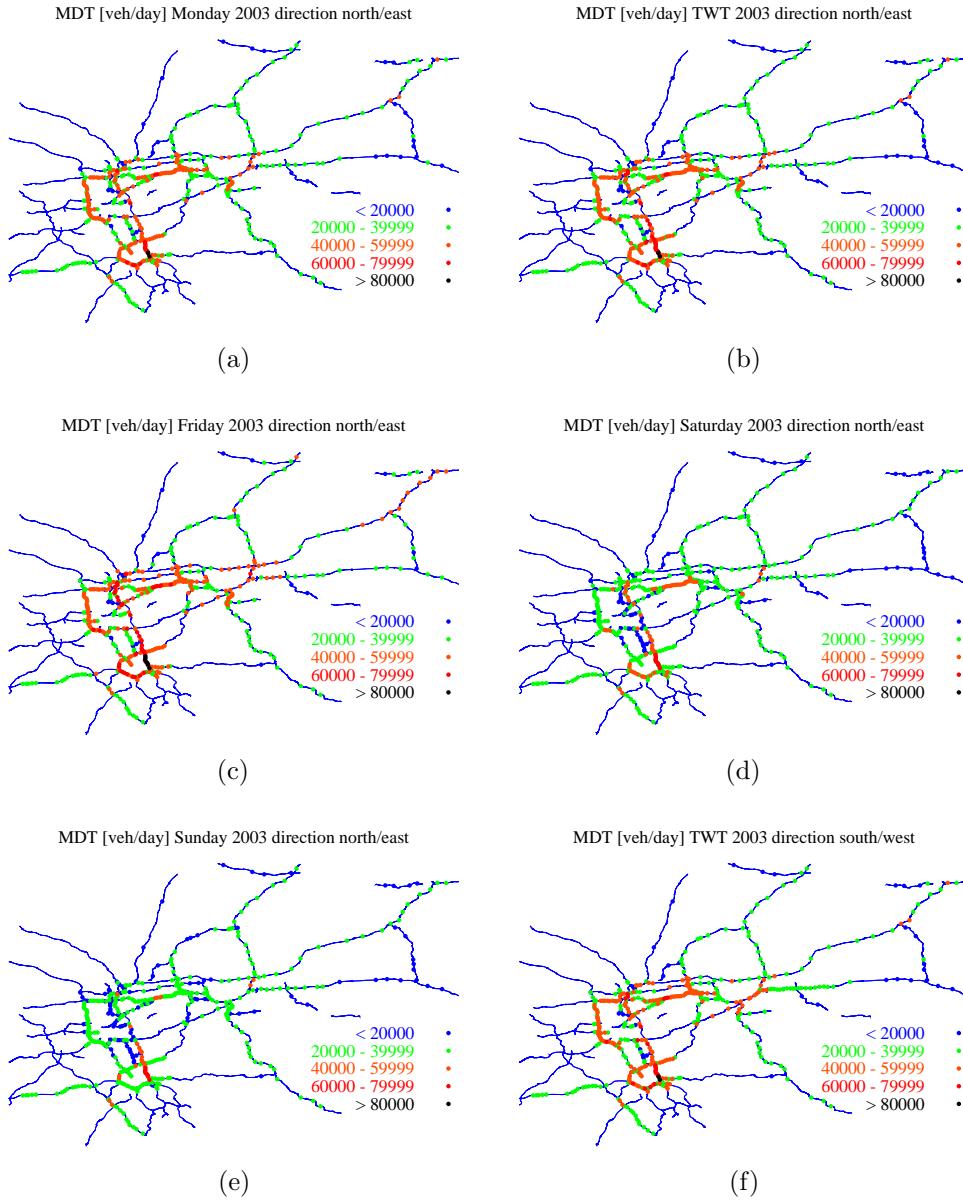
**Table 5.3.** Probabilities for congestions at the cross section 5877-019 driving direction east. To investigate the probabilities for congestions the number of days that fulfill a certain condition is divided by the number of all days of the particular class. As condition is chosen either  $v < 30 \text{ km/h}$  or  $\rho_{\text{rel}} > 0.4$  measured at one minute of the day.



**Fig. 5.10.** Congestion probabilities  $p_{\text{congestion}}$  at the cross section 5877-019 driving direction east of “normal” Tuesdays, Wednesdays, and Thursdays during the different months of the year. As a definition of a congestion hold the two conditions  $v < 30 \text{ km/h}$  and  $\rho_{\text{rel}} > 0.4$  measured at one minute of the day. Whereby the higher jam probabilities during November and December can be expected because of the shopping traffic, the congestion probability during April is also remarkably higher.

with traffic congestion is divided by the complete number of days. As definitions for a day with a congestion hold two different conditions. A day is counted as congested, if in one minute a velocity  $v < 30 \text{ km/h}$  or an occupancy  $\rho_{\text{rel}} > 0.4$  is measured. As can be seen in Table 5.3 the results for these definitions are very similar.

It is not surprising that the probability for a traffic congestion differs strongly among the classes. The days with the highest probability of a congestion are the Tuesdays until the Thursdays. A relative high number of these days are available each month, so it makes sense to use this class of days to investigate the congestion probability in dependence of the month to look for seasonal differences.



**Fig. 5.11.** MDT of the year 2003 for all cross sections in driving direction north or east on (a) Mondays, (b) TWT, (c) Fridays, (d) Saturdays, and (e) Sundays, and (f) in driving direction south or west for TWT. The numbers are given in vehicles per day. Obviously there are some stable general structures. For instance, there is more traffic on working days than on weekends. Also at some locations there is always more traffic than at other.

The results from 2000/10/01 to 2004/07/31 can be seen in Fig. 5.10. The higher congestion probabilities during November and December are expected because of the already mentioned shopping traffic. Apart from this, the high congestion probability during April is remarkable, what may have something to do with unexpected changes of the weather.

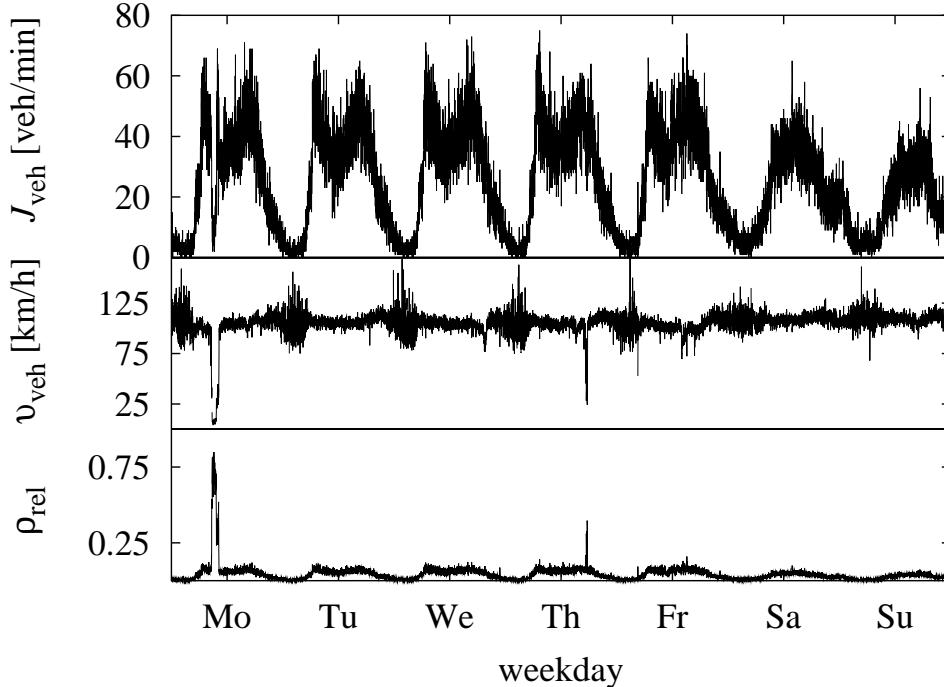
Up to now, all the investigations are made for particular cross sections. Although it could be seen in Table 5.2 that there are correlations between different locations, in general there is a strong dependence of the traffic states on the location. For a complete knowledge based system all the investigations have to be done for the 4,480 loop detectors or the 2,440 cross sections, respectively. This would not only extend the scope of this work. By and by there could appear basic changes in the traffic network like expansions, what leads to a completely different traffic demand. Furthermore the traffic demand can change from time to time due to other external influences. Thus, such investigations have to be repeated frequently. How this can be avoided and how the system can be made adaptive is discussed later. At this point a short look should be taken on all the cross sections. Therefore in Fig. 5.11 the MDT of the year 2003 for all of the cross sections in driving direction north or east can be seen for all the classes of days mentioned above. As can be seen, many results can be straight forwardly transferred. The classification of the days leads to similar results for all the cross sections. The most traffic is measured during working days, the less during the weekend. Furthermore, the regions with much traffic and the regions with little traffic are similar for all the days.

But also in this MDT many differences can be seen and it can be supposed, that a general classification for all cross sections is unsufficient. Furthermore, not all the differences can be explained with recurring structures in the MDT, and not all the recurring structures can be seen in the MDT. Therefore in the next section a look is taken at smaller timescales and at the day to day differences.

### 5.3 Daily Differences

It is not arbitrary that the MDT is used in the previous section to investigate the seasonal differences. Because there is in general little traffic at night, these states of low traffic flow divide the traffic time series naturally in parts and it makes sense to investigate the daily traffic time series. In Fig. 5.12 the differences can be seen for the week from Monday 2004/07/05 to Sunday 2004/07/11 for the traffic flow  $J_{\text{veh}}$ , the velocity  $v_{\text{veh}}$  and the occupancy  $\rho_{\text{rel}}$ . It is also obvious that on this timescale there are similarities during workingdays, whereas the weekends are completely different. Note the very high velocity fluctuations at night. They come from the fact, that each vehicle can drive as fast as its driver wants because of missing vehicle interaction. Because of this effect the velocity is nearly unpredictable or, in other words, a very high forecast error has to be expected. On Monday an outburst of the velocity and the occupancy can be seen indicating a traffic congestion. Obviously also on Thursday there is a traffic

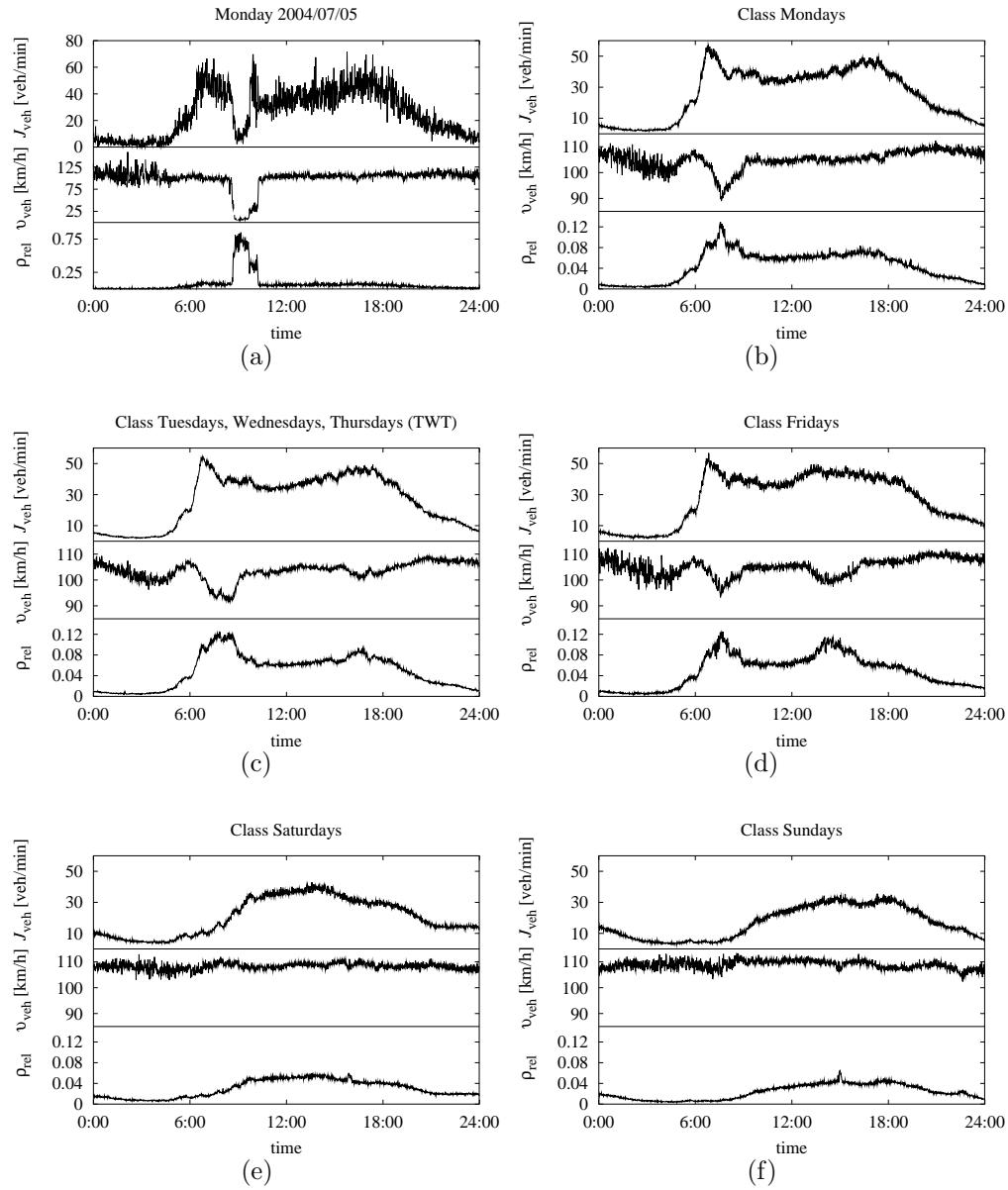
congestion at a time, when at the other weekdays only a small irruption can be supposed.



**Fig. 5.12.** The traffic time series of the traffic flow  $J_{\text{veh}}$ , the velocity  $v_{\text{veh}}$  and the occupancy  $\rho_{\text{rel}}$  for the cross section 5877-019 on the BAB 40 in driving direction east from Monday 2004/07/05 to Sunday 2004/07/11. Because of little traffic demand at night each day can be considered separately. Note that the days on the abscissa mark the noon of the particular day.

In a higher resolution this can be seen in the traffic time series of Monday 2004/07/05 in Fig. 5.13 (a). During the traffic congestion from 8:30 to 10:15 the occupancy increases and the velocity and the traffic flow decreases drastically. Outside the interval of the traffic congestion the traffic flow follows certain patterns that are characteristical for the traffic demand of a Monday. With the velocity three regions can be generally distinguished. One is of course the traffic congestion with a significantly lower velocity. The second is during the day outside the traffic congestion, that means from 5:00 to 8:30 and from 10:15 to 21:00. In this interval, a mean velocity of  $\bar{v} = 103.8$  km/h with a standard deviation of  $\sigma(v) = 5.3$  km/h is observed. The third region is at night, that means between 0:00 and 5:00 and after 21:00. Here, the mean velocity is  $\bar{v} = 106.6$  km/h and the standard deviation  $\sigma(v) = 11.2$  km/h, what is significantly higher.

Because of the scaling the only difference that can be seen clearly in the occupancy is during the traffic congestion. But it is a well known fact, that in free flow the



**Fig. 5.13.** Daily traffic time series of the traffic flow  $J_{\text{veh}}$ , the velocity  $v_{\text{veh}}$  and the occupancy  $\rho_{\text{rel}}$  of the cross section 5877-019 on the BAB 40 driving direction east for (a) the particular Monday 2004/07/05 and for the classified mean time series of the year 2003 for (b) Mondays, (c) Tuesdays until Thursdays, (d) Fridays, (e) Saturdays, and (f) Sundays. Note the different scaling on the ordinate for the particular Monday and the other ones. In the classified traffic time series the fluctuations are much lower as well as the interruptions of the occupancy and the velocity.

density and thus also the occupancy comes linear with the traffic flow. To see this, another scaling must be used. This can be fulfilled when classified traffic time series are observed and the fluctuations are much lower.

In Fig. 5.13 (b) – (f) the classified traffic time series of certain classified days of the year 2003 can be seen. What is precisely meant with a classified mean traffic time series  $\bar{x}(\mathcal{G}_{D_n}, t)$  is explained in the following. The daily traffic time series  $x(d, t)$ , whereby  $x$  can be any kind of traffic data (flow, velocity, or occupancy), at day  $d$  is averaged for a certain group of days  $\mathcal{G}_{D_n}$ . Thereby  $D_n$  indicates the kind of days, for instance, Mondays, Sundays, or TWT. In the following manner this is done for each minute  $t$  separately:

$$\bar{x}(\mathcal{G}_{D_n}, t) = \frac{\sum_{d \in \mathcal{G}_{D_n}} x(d, t)}{\sum_{d \in \mathcal{G}_{D_n}} 1}. \quad (5.14)$$

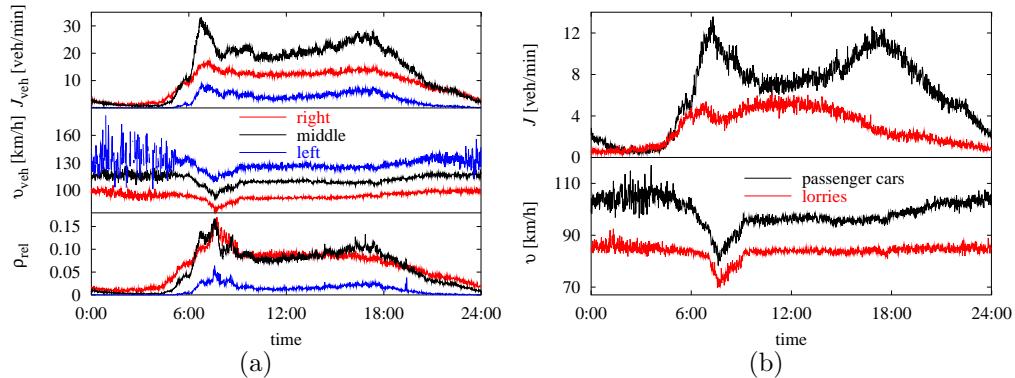
Thus, what is meant with the classified traffic time series is the mean value of each minute separately, within a certain group of days. The classified days used here are the same that are chosen in Sec. 5.2 and consist of the “normal” weekdays of the year 2003. That means the groups are  $\mathcal{G}_{\text{Mo}} \hat{=} \text{Mondays}$ ,  $\mathcal{G}_{\text{TWT}} \hat{=} \text{Tuesdays}$  until Thursdays,  $\mathcal{G}_{\text{Fr}} \hat{=} \text{Fridays}$ ,  $\mathcal{G}_{\text{Sa}} \hat{=} \text{Saturdays}$ , and  $\mathcal{G}_{\text{Su}} \hat{=} \text{Sundays}$ .

As mentioned the fluctuations are much lower, so note the different scaling on the ordinate in comparison to Fig. 5.13 (a). It can be seen that the characteristics of the traffic flow and the occupancy are very similar, but it is important, that the peak in the morning coincides with the irruption of the velocity. They are both a little bit later than the morning peak of the traffic flow.

In opposite to the time series of the particular Monday in Fig. 5.13 (a), no values of the occupancy are remarkable higher than  $\rho_{\text{rel}} = 0.12$  and no values of the velocity are under  $v = 90 \text{ km/h}$ . The peaks are only an indicator for a certain probability of a traffic congestion. With this classification the traffic congestion of the particular Monday cannot be forecast.

Using classified traffic time series for traffic forecast is a common method of many former traffic forecast approaches. Especially the heuristics and knowledge based systems already mentioned in Sec. 3.4 use this. But also the ATHENA model (cp. Sec. 3.4 on page 43) and the Kohonen map (cp. Sec. 3.5.2 on page 56) cluster the traffic time series in certain weekdays. This comes because recurring structures in the human behaviour lead to recurring structures in the traffic demand.

Many of these patterns, their effects and their occurrence are analysed in detail in [Chrobok et al., 2000] for the urban traffic flow of the city of Duisburg. Some of them can also be seen in the traffic flow of Fig. 5.13 (b) – (f), although the data are stemming from the motorway BAB 40. The clear structures like the small morning peak and the broaden afternoon peak on working days can be seen as well as smaller morning peaks on Saturdays, increased evening traffic on Fridays and Saturdays, and in general low traffic on Sundays. The traffic pattern reflects on detail the human life.



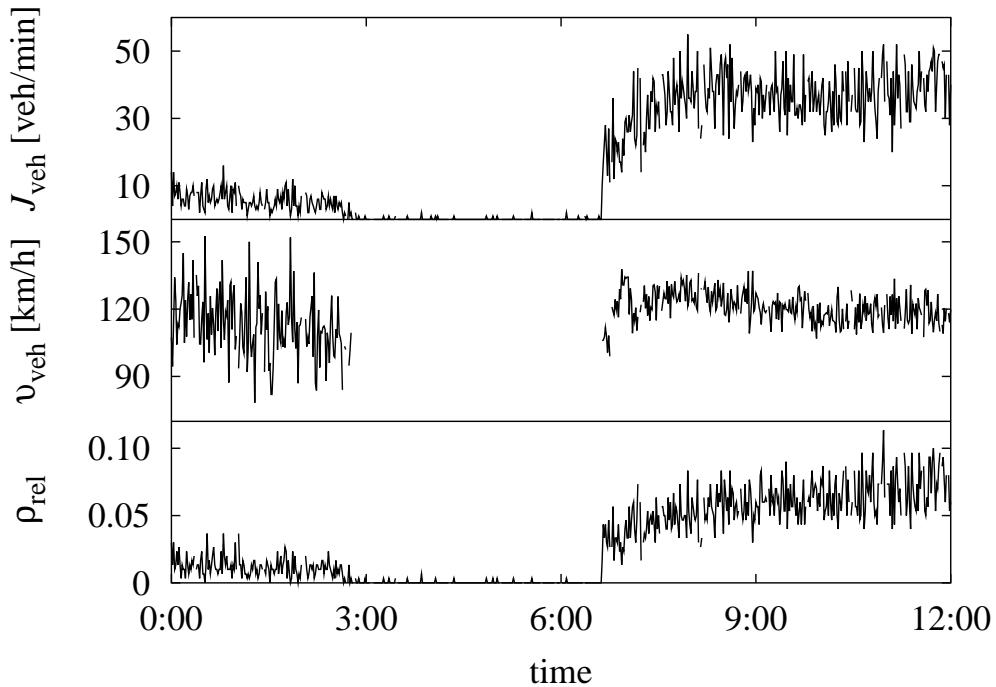
**Fig. 5.14.** (a) Classified traffic time series “normal Mondays” of the year 2003 for cross section 5877-019 driving direction east for the right, middle, and the left lane. (b) The same classification for the right lane of this cross section whereas passenger cars and lorries are distinguished.

In the same manner such classified daily traffic time series can be calculated for any group of classified days  $\mathcal{G}_{D_n}$ . Important is that  $N$  groups  $\mathcal{G}_{D_n}$  with  $n = 1, 2, \dots, N$  are chosen so that the pairwise  $\mathcal{G}_{D_n}$  are disjunctive and all days  $d$  are elements of the set union of all  $\mathcal{G}_{D_n}$ . Then, the real forecast procedure is the particular choice of these groups: the classification. In former works different classifications are used as can be read in Sec. 3.4. There are several reasons why the choice of the classification can differ strongly. In some cases, especially in older works, the classification is the result of considerations that are made without any or with a sparse database. But even when the classification is the result of an intensive data analysis, it depends on the location where the data is stemming from. This problem is already mentioned at the end of Sec. 5.2. And moreover, for the lane resolved information the data has to be analysed in regard to each lane. This is exemplary done in Fig. 5.14 (a). For the differentiation of lorries and passenger cars the particular time series have to be investigated, what is exemplary done in Fig. 5.14 (b). All the analyses have to be done for all the cross sections, what is unfeasible manually.

How to get a suitable classification automatically for all the loop detectors is an essential part of Chapter 6. Another question is the predictability of the irruption of the velocity in Fig. 5.13 (a). This problem falls together with the predictability of special events that influence the traffic state. To answer this question the effects of special events are discussed in the next section.

## 5.4 Special Events

The most obvious special event that influences the traffic state is an accident. An example for an accident at the 2001/05/11 at the cross section 4214-015 on the BAB 2 driving direction east can be seen in Fig. 5.15. At 2:38 a lorry falls over and the motorway is blocked. The traffic flow and the occupancy are nearly zero and the velocity cannot be measured in that interval. This state holds for exactly four hours until the place of accident is vacated.



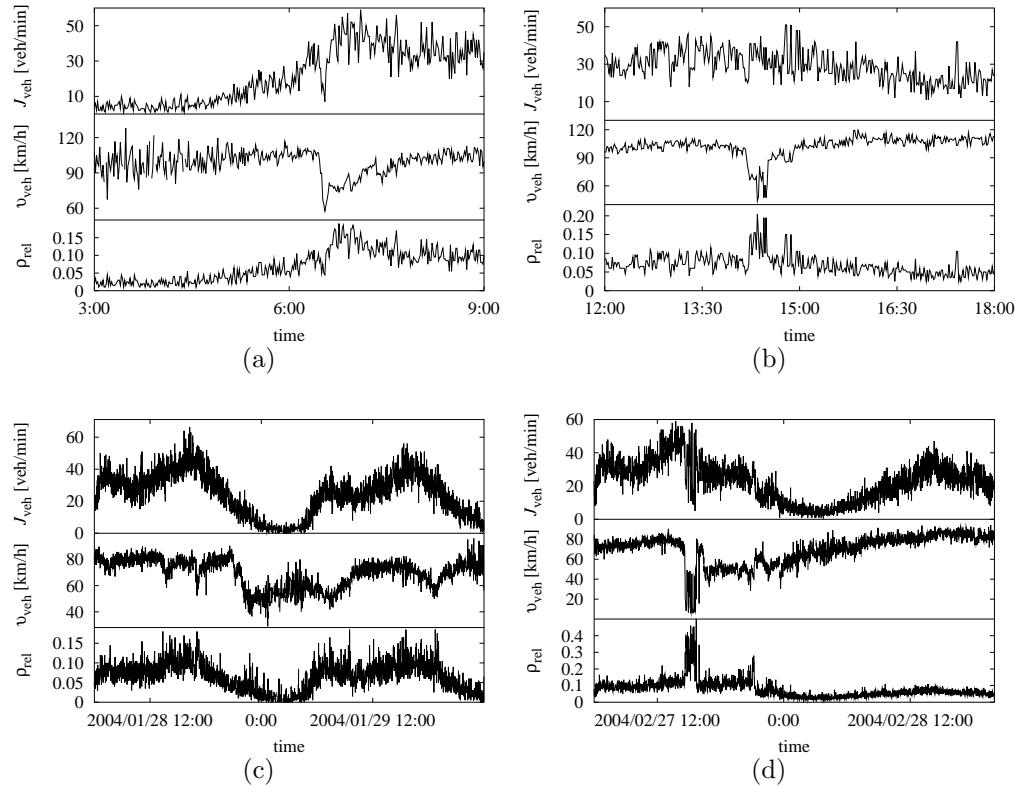
**Fig. 5.15.** The traffic flow, velocity, and occupancy measured during an accident on the 2001/05/11 at the cross section 4214-015 on the BAB 2 driving direction east. At 2:38 a lorry falls over. Flow and occupancy are nearly zero and the velocity is not measured anymore. This state lasts for exactly four hours until the place of accident is vacated

It is quite obvious, that neither the accident itself, nor the effect on the traffic state can be forecast. What can be performed is only a detection of the congestion. Using analyses of former congestions a mean duration of the congestion can be used as an approximation for the particular case. An accident is an example for an event, that is nearly unpredictable.

Another factor that frequently influences the traffic state is the weather. Different moderate weather conditions, like sunshine or rain, influence above all the traffic demand caused by spare time activities. This effect can be mainly measured on weekends and lead to high fluctuations of the mean daily traffic demand within

the Sundays or Saturdays during the year (see Fig. 5.8 in Sec. 5.2).

To influence the commuter traffic, it needs very extreme weather conditions. In Fig. 5.16 four examples of influences of the weather on the traffic state can be seen. The first two examples are measurements at two days with rainstorms at the cross section 5875-010 on the BAB 40 driving direction east. In both cases the velocity decreases, but in a little bit different manner. Whereas on the 2002/06/20 also the traffic flow goes down, on the 2002/08/24 the irruption of the traffic flow can hardly be distinguished from the other fluctuations.



**Fig. 5.16.** Traffic flow  $J_{\text{veh}}$ , velocity  $v_{\text{veh}}$ , and occupancy  $\rho_{\text{rel}}$  measured at days with extreme weather conditions. (a) A rainstorm on the 2002/06/20 6:30 at the cross section 5875-010 on the BAB 40 driving direction east. Flow and velocity show a strong irruption at this point in time. (b) Similar rainstorm on the 2002/08/24 14:11 measured at the same cross section. Whereby the irruption of the traffic flow can hardly be distinguished from the fluctuations, the velocity decreases and the occupancy increases for a while. (c) Snow flurry that starts in the evening of 2004/01/28 influences the velocity of cross section 5885-013 on the BAB 40 driving direction east. (d) Similar snow flurry on the 2004/02/27 starting at about 14:30 at the same cross section.

The effect of even more drastical weather conditions can be seen in Fig. 5.16 (c) and (d). At the cross section 5885-013 on the BAB 40 driving direction east

two cases of extreme snowy conditions are analysed. In both cases the velocity decreases for hours. But at 2004/01/28 the snow starts falling in the evening when there is little traffic demand. The traffic flow and the occupancy are not effected. At 2004/02/27, when the snow starts falling at about 14:30 the traffic demand is significantly higher. This condition leads to a traffic congestion as can be seen in the increasing values of the occupancy.

The four examples of extreme weather conditions have one commonality. In all cases the velocity is influenced what can be interpreted as follows. In opposite to good weather conditions that leads to an increasing traffic demand because of spare time activities, such bad weather conditions normally lead to a wetness, slipperiness, or a reduced visibility. That causes the drivers to reduce their speed. Finally this effect can be measured by a reduced mean velocity. Thus, the value that is influenced by bad weather conditions is the velocity. Only if the traffic flow is high enough this leads to a congestion.

It is also obvious, that the predictability of the effect strongly depends on the ability of a weather forecast. Using the information about the future weather, approximations of the traffic states can be made using a statistical analysis of similar former conditions.

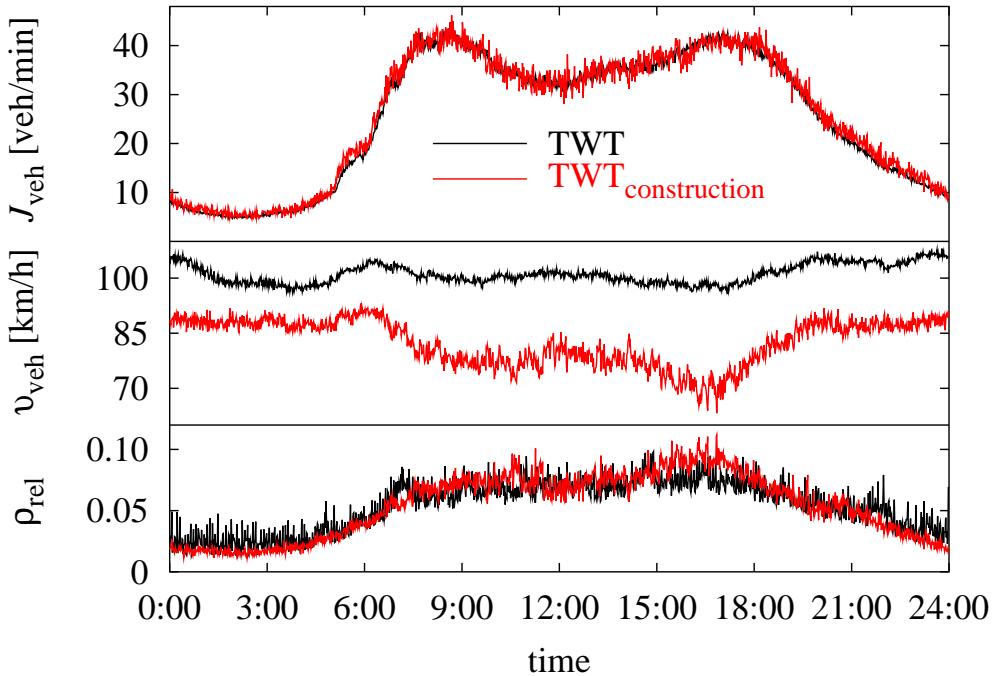
Up to now, all effects that have been discussed in regard to special events can clearly be seen in the traffic data. But to be sure, that the effect is really caused by the event, the particular traffic time series has to be compared with the “normal” one. To be precise, for an analysis of the effect itself, this is mandatory.

As definition of the normal traffic time series can act the classified traffic time series defined by Eq. 5.14. As group of days  $\mathcal{G}_{D_n}$  those days  $d$  have to be chosen, for that with the absence of the event would hold  $d \in \mathcal{G}_{D_n}$ .

With this background the effect of the road works on the BAB 2 between the junctions “Bönen” and “Hamm” can be investigated. Analysing the effects of construction areas it has to be mentioned, that often at the places of the road works the loop detectors are for technical terms out of order. What can be observed by the detectors are the impacts around the construction area. Lane reductions, speed limits, or complex lane guidances lead to slower velocities and can even cause congestions.

The example in Fig. 5.17 shows the classified traffic time series of the class TWT without, and  $TWT_{construction}$  with the presence of the construction area that is a little bit downstream of the cross section 4229-014 on the BAB 2 driving direction east. The reduced mean velocity can clearly be seen. A little bit harder to see is the increased mean occupancy caused by the tailbacks.

How the traffic state is exactly influenced strongly depends on many details of the particular road works as well as on the location of the detectors. Nevertheless it is possible with statistical analyses to increase the predictability of the impacts of different road works under certain conditions. For instance, in [Beckmann and Zackor, 2001, Ober-Sundermeier, 2003] different kinds of construction areas are investigated in detail to develop a unique method to systematically forecast the length and the duration of tailbacks caused by construction areas in dependence



**Fig. 5.17.** The traffic flow  $J_{\text{veh}}$ , velocity  $v_{\text{veh}}$ , and occupancy  $\rho_{\text{rel}}$  on normal Tuesdays, Wednesdays, and Thursdays (TWT) and on the same days when there is a construction area downstream (TWT<sub>construction</sub>). The data are stemming from the cross section 4229-014 on the BAB 2 driving direction east. In the traffic flow there are only a little bit more fluctuations, because there are less days in this class. On the contrary the velocity is significantly lower than on normal days because of a higher frequency of tailbacks.

on the lane reduction and lane guidance.

In opposite to accidents or weather conditions the event “construction area” is exactly known before, at least one day, in most cases even longer before. The task of the forecast is just to estimate the impact on the particular detector. The fact, that the cause that influences the traffic state is known before, also holds for the special events that are discussed in the following.

In general, the traffic demand on motorways is the result of the need for mobility of each single road user. Similarities in the behaviour of the road users lead to observable structures like the commuter traffic. In the example of the commuter traffic the structures are very stable and recurring effects can be measured.

Big events like fairs, sporting events, demonstrations or other processions are examples for events that cause a very similar behaviour of many road users but are not periodically recurring like the commuter traffic. Especially they are not always on the same weekday or at the same time. Like construction areas the event itself is always already known before and the problem is to estimate the

impact.

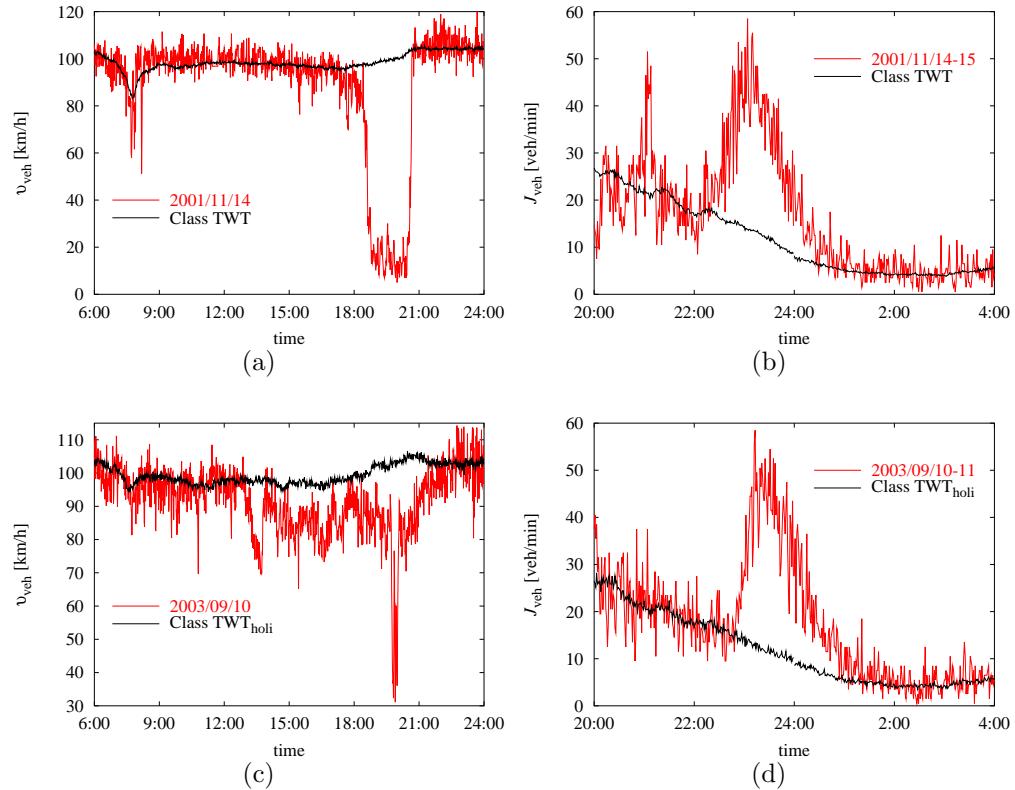
In Fig. 5.18 the example of a sporting event can be seen. Shown is the velocity before and the traffic flow after two football matches of the German national team in Dortmund. The data are stemming from the cross section 5907-027 on the BAB 40 in direction east for the trip to, and in direction west for the trip from the stadium. The match on Wednesday the 2001/11/14 against the Ukraine starts at 20:30, the match against Scotland on Wednesday the 2003/09/10 at 20:45. The particular classified traffic time series are also shown, therefore hold the classes TWT and TWT<sub>holi</sub>, respectively, because the 2003/09/10 lies within the summer holidays.

The holidays might also be the reason for the differences in the velocity irruptions that have been measured. Whereas on the 2003/09/10 in Fig. 5.18 (c) moderate deviations from the classified time series are measured since 13:00 and at 20:00 only a short traffic congestion is observed, on the 2001/11/14 in Fig. 5.18 (a) the velocity fluctuates around the value of the classified time series until 18:30, when there is a traffic congestion that takes two hours. Obviously, when analysing the velocity it is insufficient to observe the effect in regard to the classified time series. On the contrary this can be done with the traffic flow. The similarities in the increased traffic flow after the match in regard to the classified traffic time series are quite remarkable. Because of the difference in the kick off time the rise starts only quarter an hour earlier on the 2001/11/14 in Fig. 5.18 (b) than on the 2003/09/10 in Fig. 5.18 (d).

In the following a closer look is taken on special days in regard to classified traffic time series. Therefore the traffic flow of the day with the special event is divided by the classified traffic flow and the relative traffic flow is observed.

Most of the special events that are discussed up to now only influence the traffic in a particular region. The region might be bigger like in the case of the weather or smaller, like in the case of the road works. But there are also some events apart from holidays or any influence caused by the calendar that influence the overall traffic state. Although in opposite to the recurring structures of the calendar the effects are in most cases much smaller, what is again caused by interesting characteristics of the human behaviour. Recurring events like christmas or New Year's eve lead to strong structures because all humans are familiar with those dates. There are only few extra events that influence the overall traffic state anyway and such events never lead to strong differences. Nevertheless they can be measured.

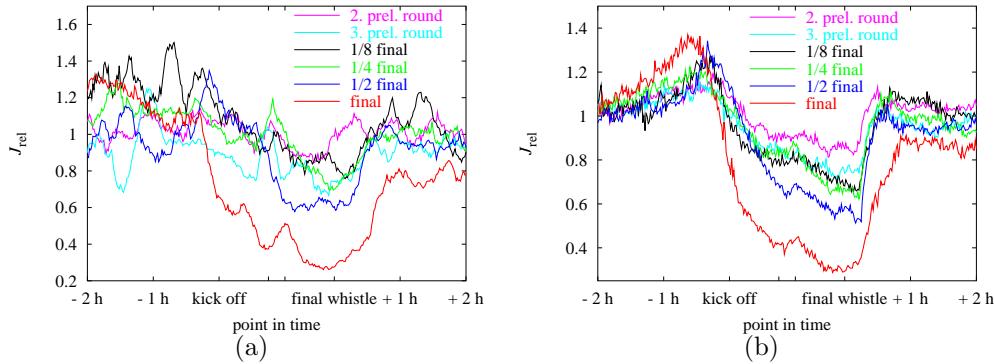
To influence the global traffic state it needs events that are transferred to a broad public, mostly by media. An event that fulfills this condition in Germany is the football world championchip that is a really large media event. In Germany it really influences the traffic state, especially, if the German team is playing and the match is brought forward by television. In Fig. 5.19 (a) the time series of the traffic flow divided by the particular classified traffic time series can be seen for the cross section 5907-027 on the BAB 40 driving direction east for different matches of different importance during the tournament 2002. In Table 5.4 the



**Fig. 5.18.** Sporting events lead to different traffic states on affected motorways that can be observed in the data. Shown are the data of Wednesday the 2001/11/14 when the match Germany against the Ukraine starts at 20:30 ((a) velocity  $v_{\text{veh}}$  on the way to the stadium, (b) traffic flow  $J_{\text{veh}}$  on the way back from the stadium) and of Wednesday the 2003/09/10 when the match Germany against Scotland starts at 20:45 ((c) velocity  $v_{\text{veh}}$  on the way to the stadium, (d) traffic flow  $J_{\text{veh}}$  on the way back from the stadium). The data are stemming from the cross section 5907-027 on the BAB 40 in direction east for the trip to, and in direction west for the trip from the stadium. Furthermore the data of the particular classified traffic time series can be seen. Whereas the similarities in the traffic flow after the matches are remarkable, the velocities show a completely different behaviour that might lie in the fact, that the 2003/09/10 is during the summer holidays.

matches, the kick off times, and the particular used classified day can be seen. The “importance” of every game increases from top to bottom.

Unfortunately the traffic data of the day with the first preliminary round game are completely missing and cannot be shown in Fig. 5.19. Because the kick off is at different times, on the abscissa the data are shifted in regard to the official kick off and final whistle time, respectively.



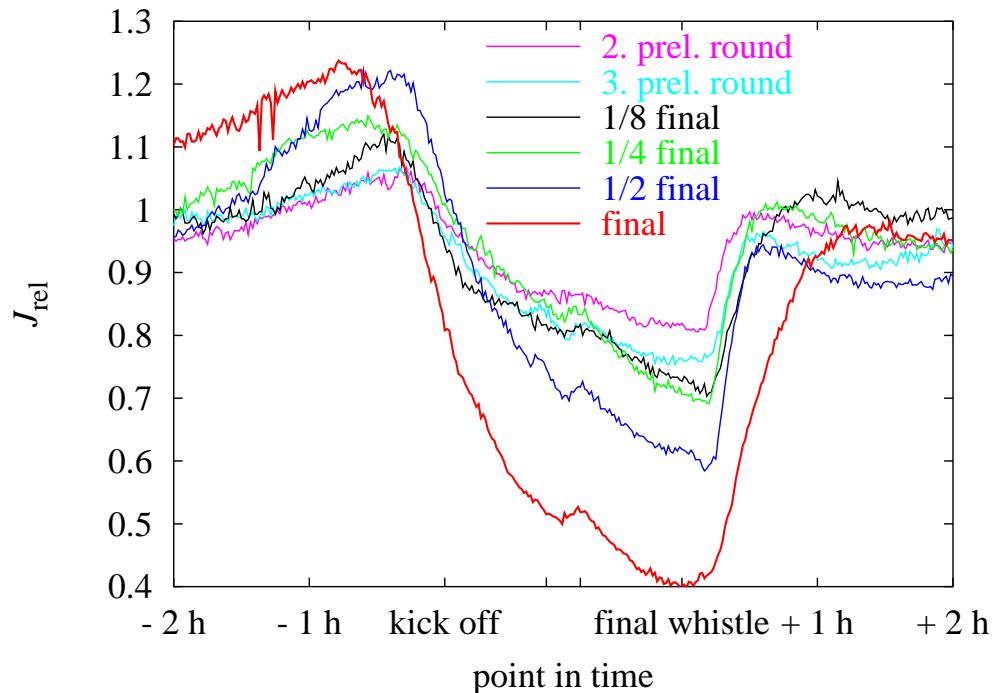
**Fig. 5.19.** Relative traffic flow  $J_{\text{rel}}$  in regard to the particular classified traffic time series during television broadcasts of matches of the German team during the football world championship in 2002. Because the matches are at different times of the day, the data are for a better comparability shifted in regard to the official kick off times. Shown are the data of the second and the third preliminary round match, the eighth, quarter, semi final and the final. Unfortunately the data of the first preliminary round match are missing because of measurement errors. (a) Although an influence can be seen in the relative data of the cross section 5907-027 on the BAB 40 driving direction east a clear pattern is superposed by many fluctuations. (b) The patterns become clearer when the fluctuations disappear because the data of all detectors of the BAB 40 are used. This is why in another step the data of all intact loop detectors are analysed in Fig. 5.20.

Although it can already be seen, that there is a strong influence, the characteristic is superposed by fluctuations. Those fluctuations can be avoided if more detectors are used like in Fig. 5.19 (b), where the data of all intact detectors on the BAB 40 is shown, or in Fig. 5.20, where even all intact detectors of the whole network are used. In this picture the pattern caused by the events can be observed without any disturbance because of noise and the pattern can be interpreted.

First at all, what can be seen in Fig. 5.20, is that the television broadcasts of the six games lead to similar patterns of different intensities. Interesting is, that with an increased importance of the game, the effect also increases. That leads again to the insight that the traffic demand maps a clear image of the social human behaviour. What all curves have in common with each other is the increased traffic flow before the game. Then, a decreasing traffic flow is observed during the match with a local minimum followed by a local maximum around the half time break. Note, that the half time is marked at the abscissa with two tics.

opponent	round	date	time	weekday	class
Saudi Arabia	1. prel. round game	2002/06/01	13:30	Saturday	Saturdays
Ireland	2. prel. round game	2002/06/05	13:30	Wednesday	TWT
Cameroon	3. prel. round game	2002/06/11	13:30	Tuesday	TWT
Paraguay	1/8 final	2002/06/15	8:30	Saturday	Saturdays
USA	1/4 final	2002/06/21	13:30	Friday	Fridays
South Korea	1/2 final	2002/06/25	13:30	Tuesday	TWT
Brazil	final	2002/06/30	13:00	Sunday	Sundays

**Table 5.4.** Important dates for the observed traffic time series during television broadcasts of matches during the Football World Championship 2002. Note that the kick off times are the local times in Germany. The importance of the matches increases from top to bottom.



**Fig. 5.20.** Relative traffic flow  $J_{\text{rel}}$  in regard to the particular classified traffic flow as in Fig. 5.19 but for all intact loop detectors (about 3,200) of the whole network at the particular days. The six TV broadcasts of the football matches lead to different intensities of similar patterns. In all cases there is an increased traffic flow before the kick off. Then, a decreasing curve is observed during the match with a local minimum followed by a local maximum around the half time break (the two ticks without any description mark the end of the first, and the start of the second half time). A strong global minimum can be seen just before the end of the match. Note, that on the ordinate the official times are marked and the real final whistle is in most cases a little bit later. Interesting is that an increasing importance of the game leads to an increased pattern.

Finally, a strong global minimum is observed just at the end of the match. Only a fraction of about 0.4 of the classified traffic flow are measured at the day with the cup final.

The reason for the similarities in the patterns might lie in the fact, that the human is a creature of habit. The human behaviour not only leads to strong recurring structures in the day to day traffic time series. Finally, each event that leads to an increased or decreased need for mobility, can be mapped like a fingerprint in the traffic data. Actually this fact offers the opportunity to forecast future traffic states.

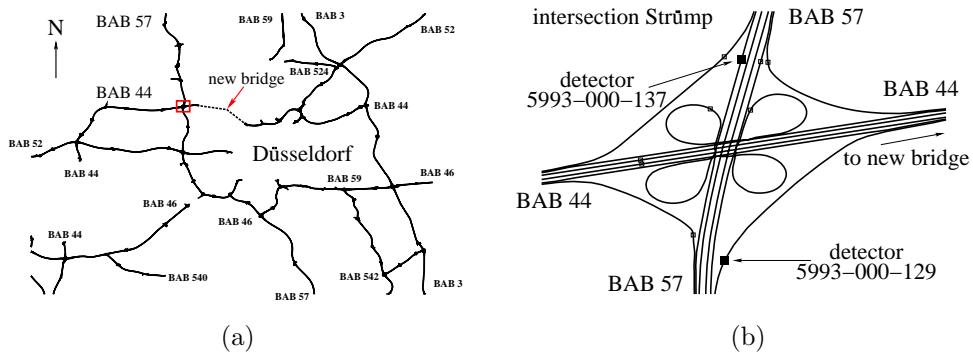
Of course there are many more special events that cause differences in the traffic data and it is a crucial question how to handle them when the data should be forecast. In practice, there is especially a need for future estimations of the impact of such mass events like football matches. Special events must be considered in every traffic forecast application. Because they are in general known before the problem is to estimate their impact. Some approaches are discussed at the end of Sec. 6.2. Indeed the biggest problem in operation is, to transfer the information about events systematically to the particular traffic forecast application as it is discussed in Chapter 7.

Up to know special events have been discussed in that sense, that they have an impact for a certain period. In most cases this is even the fact. But the example of the long term construction area in Fig. 5.17 shows, that those time intervals might be very long. External influences can change the traffic state for weeks, month, years, or even basically. In practice this happens often because of changes in the network, like in the example with the long term construction area. Basic changes of the traffic demand happen because of extensions of the network. An example of this is given in the following.

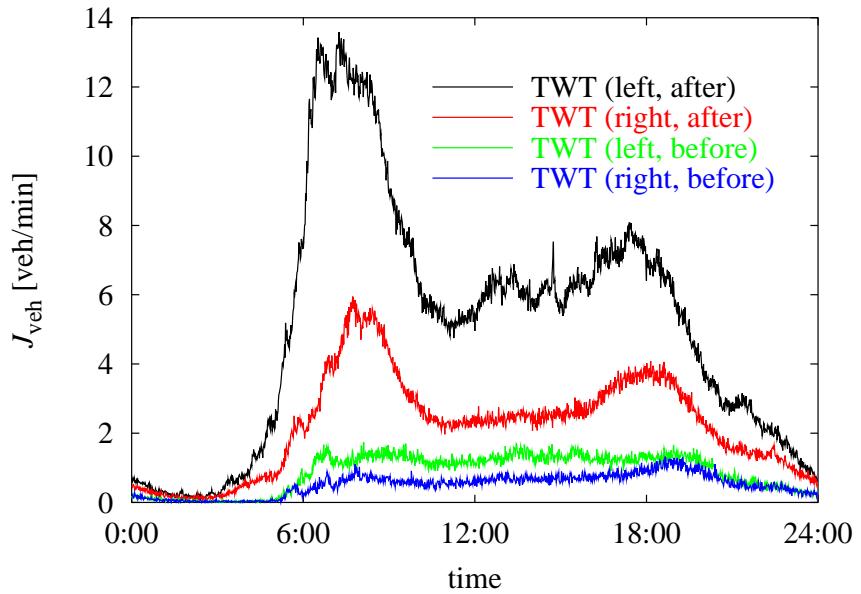
At the 2002/05/31 a basic change in the network has been fulfilled. A new bridge across the Rhine river offers from that moment the opportunity to reach the inner city of Düsseldorf much faster using the motorway BAB 44. The situation of the motorways can be seen in Fig. 5.21. Before the bridge has been opened vehicles could also turn in this direction but the motorway ends at a junction before the Rhine river.

These vehicles, that turn off from the motorway BAB 57 in direction to the junction or the new bridge, respectively, can be measured. The detector 5993-000-137 is located on the parallel lane from the north downstream of the turning lane to the right but, what is important, upstream of the turning lane to the left and the turning lanes from the left and the right from the BAB 44. So, what is measured at this point is exactly the traffic that turns to the left in direction to the bridge. The detector 5993-000-129 is located just on the right turning lane from the BAB 57 to the BAB 44 in direction to the bridge.

In Fig. 5.22 can be seen, how the classified traffic flow of the detectors changes. Shown is the traffic flow of the class TWT up to the 2002/05/31 and of the same class thenceforward to the 2004/07/31 for both, the detector 5993-000-129 (right) and 5993-000-137 (left). It can be clearly seen that the classified traffic flow of



**Fig. 5.21.** (a) At the 2002/05/31 a new bridge has been opened that offers the opportunity to reach the inner city of Düsseldorf faster via the motorway BAB 44. All drivers who want to use the bridge have to pass the intersection of the BAB 57 and the BAB 44 that is called “Strümp”. (b) An enlargement of the rectangle area in (a) that marks the intersection Strümp. The large squares mark the detectors that measure the vehicles that turn off in direction to the new bridge left or right, respectively.



**Fig. 5.22.** The opening of the bridge has a large impact on the classified traffic flow of turning lanes to the bridge. The flow of the class TWT of the detectors 5993-000-129 (right) and 5993-000-137 (left) clearly increases after opening the bridge.

both detectors clearly increases.

This example shows, that the classified traffic time series are not constant, apart from changes because of the secular trend. On the one hand their stability offers the possibility for forecasts. On the other hand it has to be continuously checked, whether they are still suitable and adjustments have to be made from time to time. Thereby it is very important to distinguish special events with a finite duration from basic changes.

The analyses made up to now are basically structured from long time scales to short ones. The former works discussed in Chapter 3 have shown, that long term analyses of traffic data also provide suitable forecast results for long forecast horizons. For instance, the stability of classified traffic time series provide useful forecasts from one day ahead up to a horizon of months. Nevertheless often the so called one-step-ahead forecast is of basic interest, what lies in the fact that many traffic control applications need fast reactions to the current traffic state. This is why in the next section effects on very small timescales are analysed: the fluctuations or random noise.

## 5.5 Random Noise

In every set of measured data appear random fluctuations. When talking about noise in a data set we have to distinguish two different kinds of noise: *measurement noise*, and *dynamical noise*.

Measurement noise refers to corruption of observations by errors which are independent of the dynamics. The dynamics satisfy  $\vec{x}_{n+1} = F(\vec{x}_n)$ , but the measured scalars are  $s_n = s(\vec{x}_n) + \eta_n$ , where  $s(\vec{x})$  is a smooth function that maps all essential scalars of the dynamics to the measured value and  $\eta_n$  are random numbers. Apart from the errors in measurement discussed in Sec. 4.2 the remaining measurement noise has an order of magnitude that can be seen in Table 4.2.

Dynamical noise, in contrast, is a feedback process following  $\vec{x}_{n+1} = F(\vec{x}_n + \vec{\eta}_n)$ , where the system is perturbed by a small random amount  $\vec{\eta}_n$ . In traffic data, these perturbations are due to the randomness in the human behaviour on short timescales. In case of no interactions between the vehicles the drivers could drive as fast as they want, for instance, at night, what leads to a large variance in the velocity. If there exist interactions, large fluctuations appear because of the perception and reaction of the driver-vehicle entity. Even most recent spatial temporal microscopic traffic models, like proposed in [Kerner et al., 2002, Knospe et al., 2000, Lee et al., 2004], can only describe some basic features of the driver behaviour in a statistical manner, but are not able to forecast the concrete behaviour of the particular driver, what is even probably impossible. As can be seen in the raw data, for instance, in Fig. 5.13 (a), especially the fluctuations of the traffic demand are of a large order of magnitude, much larger than those of the measurement noise, what is negligible. This leads to a broad discussion with many theoretical considerations about the uncertainty in forecasting traffic data

(see, for instance, [Davis and Yang, 2001, Lan, 2001, Yang and Davis, 2002]). When talking about noise the characteristics is of great importance. Therefore the power spectrum (cp. Sec. 5.2 on page 92) provides useful information. In Fig. 5.23 the power spectrum for two cross sections of the daily traffic time series of a typical Monday with and without traffic congestion can be seen.

The behaviour of the noise is quite similar to that of the MDT in Sec. 5.2. Two regions can be clearly distinguished. First the descent of the amplitude follows a power law with an exponent that is nearly  $\epsilon = 2.0$ . This descent is more distinctive in case of a congestion. In a traffic time series without congestion this effect is only faintly observable. Note that because of the double logarithmic scale this holds for only a small frequency range. For higher frequencies and for a large range hold  $\epsilon = 0$  what is typical for Gaussian white noise and a strong indicator for randomness and unpredictability.

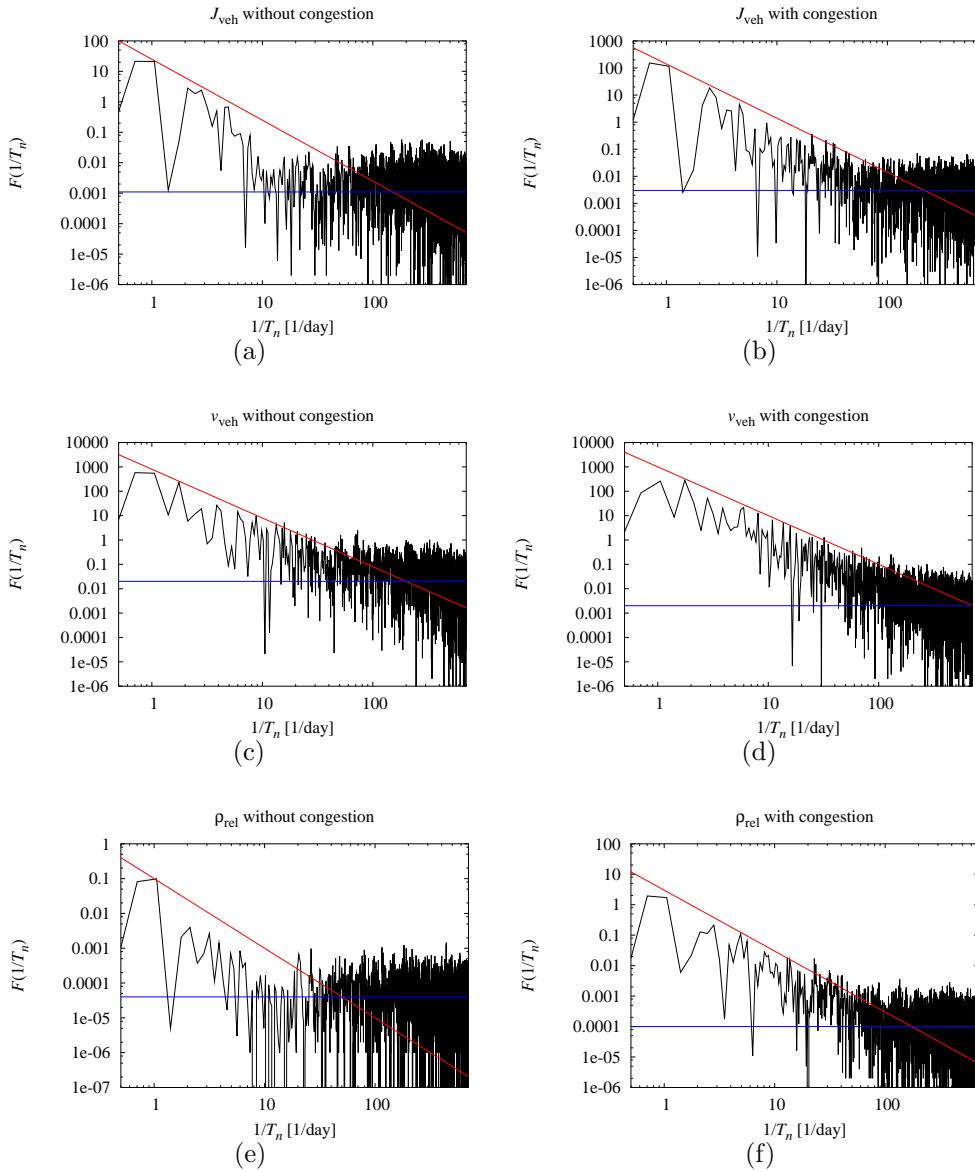
These results are in total agreement with the observations in [Helbing, 1997], but disagree with many theoretical considerations, that suppose for  $\epsilon$  values of 1.0 [Choi and Lee, 1995, Nagel and Herrmann, 1993, Nagel and Paczuski, 1995, Zhang and Hu, 1995], 1.4 [Musha and Higuchi, 1978], or 1.8 [Yukawa and Kikuchi, 1996], that try to find an easy model describing the traffic on short timescales. Furthermore all such models act on a spatial temporal base and do not try to really forecast the traffic state.

Because the peaks of white noise cannot be forecast the mean value is of basic interest for forecasts. The question is then, how to extract a suitable mean value. In the following a few basic methods are discussed.

One easy method (AVG) is to divide the daily traffic time series  $x(t_m)$  in  $M/L$  intervals  $I$  of a length  $L$  and use the mean value:

$$\text{AVG}_L(x, \tau_n) = 1/L \sum_{t_m \in I} x(t_m). \quad (5.15)$$

Thereby the time steps  $t_m$  with  $m = 1, \dots, M$  of the raw data change to the averaged time steps  $\tau_n$  with  $n = 1, \dots, M/L$ . For instance, for a daily traffic time series of minute aggregated data it holds  $L = 1$  min. This method is often used in both, scientific research and traffic management applications. The choice of  $L$  depends strongly on the particular objective. Whereas for raw data can also hold  $L = 20$  s [Dia, 2001],  $L = 30$  s [Ahmed and Cook, 1979, Ahmed, 1983], or sometimes even single vehicle are available [Knospe et al., 2002b], the proposals for further analyses range from  $L = 30$  s [Persaud and Hall, 1989],  $L = 1$  min [Ahmed and Cook, 1983, Davis and Nihan, 1991, Lee and Choi, 1998],  $L = 5$  min [Chen et al., 1997, Dougherty and Cobbett, 1997, Faouzi, 1996, Iwasaki and Saito, 1999, Nair et al., 2001, Stephanedes et al., 1981, Sun et al., 2003],  $L = 9$  min [Wild, 1997],  $L = 15$  min [Kim, 1994, Krause, 1988, Smith, 1995],  $L = 30$  min [Danech-Pajouh and Aron, 1991, Kirby et al., 1997, van der Voort et al., 1996, Williams, 2001],  $L = 60$  min [Lambole et al., 1997, Leutzbach and Siegner, 1974, Siegner and Schmitt, 1980, van Iseghem and Danech-Pajouh, 1999, Zackor et al., 1996], or anyway the MDT is observed as in Sec. 5.2 with  $L = 1440$  min.



**Fig. 5.23.** Power spectrums of the daily traffic time series without (left) and with a traffic congestion (right) for traffic flow  $J_{\text{veh}}$  ((a) and (b)), velocity  $v_{\text{veh}}$  ((c) and (d)), and occupancy  $\rho_{\text{rel}}$  ((e) and (f)). The time series without congestion are stemming from Monday 2004/04/19 at the cross section 5972-050 of the BAB 44 driving direction east. The time series with congestion are stemming from Monday 2003/06/30 at the cross section 4347-007 of the BAB 3 driving direction north. The blue line mark the Gaussian white noise and the red one the power law with the exponent  $\epsilon = 2.0$ . This descent is more distinctive in case of a congestion.

In many of the works a model for a one-step-ahead forecast is investigated and the interval is equal to the raw data. In most cases the interval is even equal to the interval for that the forecast has to be calculated. An exception is the work of Siegener and Schmitt [1980], where an interpolation between hourly aggregated traffic data is used for forecasts on smaller timescales.

A disadvantage of this method is, that the number of data values reduces from  $M$  to  $M/L$ , what is often undesired. To avoid this, the moving average method (MOV) can be used. It calculates for each time step  $t_m$  the average of the interval  $[t_m - N; t_m + N]$  with the length  $2N + 1$ :

$$\text{MOV}_{2N+1}(x, t_m) = \frac{x(t_m - N) + \dots + x(t_m + N)}{2N + 1}. \quad (5.16)$$

The algorithm is the same as in Eq. 3.13 where it is used for forecast. The only difference is that for forecast at a certain time step  $t_f$  when the forecast is calculated only data values  $t_m < t_f$  can be used. In the same manner the exponential smoothing method proposed in Eq. 3.18 can be used to eliminate the noise with the following modification:

$$\text{EXP}_\alpha(x, t_m) = \alpha \sum_{n=0}^{\infty} \left[ \frac{x(t_m + n) + x(t_m - n)}{2} \right] [1 - \alpha]^n. \quad (5.17)$$

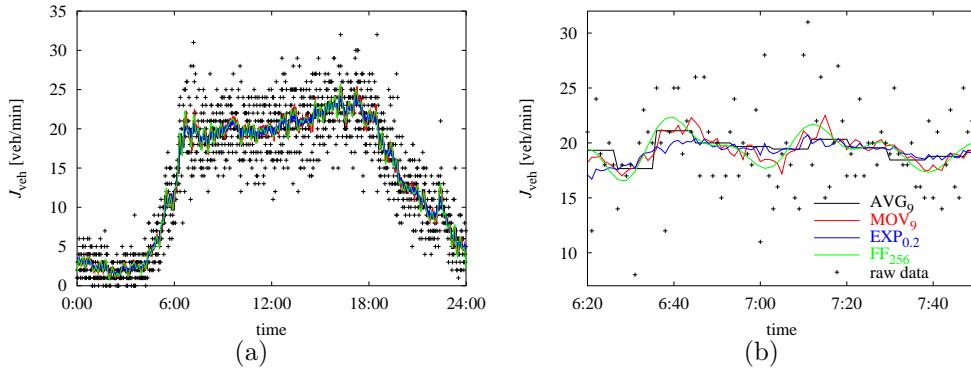
Thereby the parameter  $\alpha$  affects the influence of the value  $x(t_m)$  on its smoothed value  $\text{EXP}_\alpha(x, t_m)$ . For  $\alpha = 1$  the smoothed curve is identical to the raw data. Note that precisely Eq. 5.17 holds for infinite time series. Numerical approximations have to go from  $n = 0, \dots, N$  but only make sense for  $N \gg 1$ . In any case the term  $[1 - \alpha]^{N+1} x(t_N)$  should be added to avoid a systematical under approximation of  $\text{EXP}_\alpha(x, t_m)$  (consider, that in case of a daily traffic time series of minute aggregated data it holds  $N < M/2$ ).

A large set of methods to eliminate noise uses a Fourier filter. Therefore the Fourier transform is calculated using Eq. 5.11. Then on the Fourier coefficients  $c_n$  different mathematical algorithms can be applied. An easy method is, for instance, to set coefficients of higher order to zero and use only the first  $P$ . Then, in the backtransformation (Eq. 5.10) the new coefficients are used, that is, in the example the first  $P$ :

$$\text{FF}_P(x, t_m) = \frac{1}{K} \sum_{n=0}^P c_n e^{2\pi i f_n k \Delta T}. \quad (5.18)$$

In Fig. 5.24 the results of eliminating noise are shown for all of the four methods with parameters that can be seen in table Table 5.5. Because the noise is larger for a single lane than for a cross section, the investigation is done for the traffic flow of detector number 5874-023-001 on the right lane of BAB 40 driving direction east of Monday the 2002/10/28. The differences of the methods are clearly

smaller than the order of magnitude of the fluctuations and can only be seen in Fig. 5.24 (b) where a higher resolution is used. In a similar manner the variation of the parameters only lead to marginal changings. This is why for the following investigations 9 min moving averages  $MOV_9$  are used what is arbitrary. For completeness it has to be mentioned that especially using Fourier transform there are many different kinds of filtering algorithms. For technical issues please refer to the standard literature, for instance, [Hamming, 1983].



**Fig. 5.24.** Different methods eliminating random noise. The traffic data are stemming from the loop detector 5874-023-001 on the right lane of BAB 40 driving direction east of Monday the 2002/10/28. The different methods are explained in the text, the example parameters can be seen in Table 5.5. (a) A difference between the different algorithms is hardly to see because of the magnitude of the fluctuations. (b) Only with a high resolution the differences of the four different algorithms of Table 5.5 can be seen.

Algorithm	Equation	Parameter	Value
$AVG_L$	Eq. 5.15	$L$	9
$MOV_{2N+1}$	Eq. 5.16	$2N + 1$	9
$EXP_\alpha$	Eq. 5.17	$\alpha$	0.2
$FF_P$	Eq. 5.18	$P$	256

**Table 5.5.** Averaging algorithms and example parameters used in Fig. 5.24.

To quantify the magnitude of the noise the probability functions are investigated. Therefore, for 320 days of the class TWT from 2001/01/09 to 2004/07/20 of the detector 5874-023-001 on the right lane of BAB 40 driving direction east, the difference  $\Delta J_{veh}$  of the raw data  $J_{veh}$  and the 9 min moving averages  $MOV_9(J)$  is calculated:

$$\Delta J_{veh} = J_{veh} - MOV_9(J_{veh}). \quad (5.19)$$

After that, the values are rounded up to the first decimal place and the number

$N(\Delta_{J_{\text{veh}}})$  of same values  $\Delta_{J_{\text{veh}}}$  is divided by the total number of measurements  $N$  to get the probability distribution:

$$P(\Delta_{J_{\text{veh}}}) = \frac{N(\Delta_{J_{\text{veh}}})}{N}. \quad (5.20)$$

The result can be seen in Fig. 5.25. The values follow in good approximation a Gaussian distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right), \quad (5.21)$$

with a standard deviation  $\sigma$  around a mean value  $\mu$ . The outliers at the edges indicate transitions from or to traffic congestions, and could not be interpreted as fluctuations. To evaluate the predictability both, the standard deviation as well as the outliers are of large interest, especially for one-step-ahead forecasts.

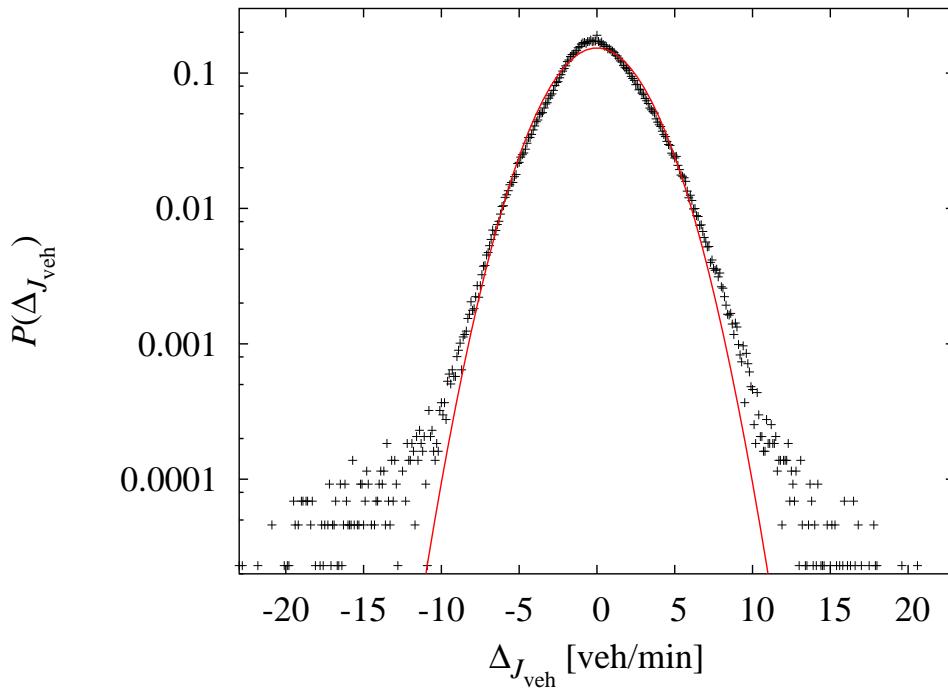
The same investigations are done for the other traffic data and the results can be seen in Fig. 5.26. Whereas the distribution of the traffic flow of lorries  $J_{\text{lor}}$  in Fig. 5.26 (a) and the occupancy  $\rho_{\text{rel}}$  in Fig. 5.26 (b) only show a little bit different characteristic of the outlierers, the transitions from free flow to congestion and vice versa influence drastically the distribution of the velocities of passenger cars  $v_{\text{pc}}$  in Fig. 5.26 (c) and that of lorries  $v_{\text{lor}}$  in Fig. 5.26 (d). To analyse the data in the right way, the approximations of the velocities are done with a superposition of three Gaussian distributed functions with positive factors that sum to one:

$$P(\Delta_v) = [1 - \alpha - \beta] P(0, \sigma) + \alpha P(\mu_-, \sigma_-) + \beta P(\mu_+, \sigma_+), \quad (5.22)$$

with  $0 < \alpha < 1$ ,  $0 < \beta < 1$ , and  $\alpha + \beta < 1$ . Thereby  $P(\mu_-, \sigma_-)$  is the estimated distribution of the systematical values under the mean (black in the graph), and  $P(\mu_+, \sigma_+)$  that of systematical values over the mean (blue). For completeness it has to be mentioned, that for the velocity of passenger cars hold  $\alpha = 0.05$  and  $\beta = 0.15$ , for that of lorries  $\alpha = 0.15$  and  $\beta = 0.23$ . Note, that in case of the fluctuations under the mean value the data swings around the analytical function because of the same round off errors mentioned in Sec. 4.2.5.

What is of basic interest here is of course the order of magnitude of the fluctuations that can be approximated by the central distributions. The standard deviations as well as the variances of the random noise that are approximated with this method can be seen in Table 5.6 for all the raw data.

All the methods of eliminating noise discussed up to now smooth the curve over adjacent data values. With the smoothing parameters in Table 5.5 only the number of values are chosen that are taken into account or their weight, respectively. But with the choice of these parameters comes always the problem, that on the one hand the fluctuations do not disappear if too few values are taken into account, on the other hand also some essential peaks are smoothed if the interval is chosen too large.

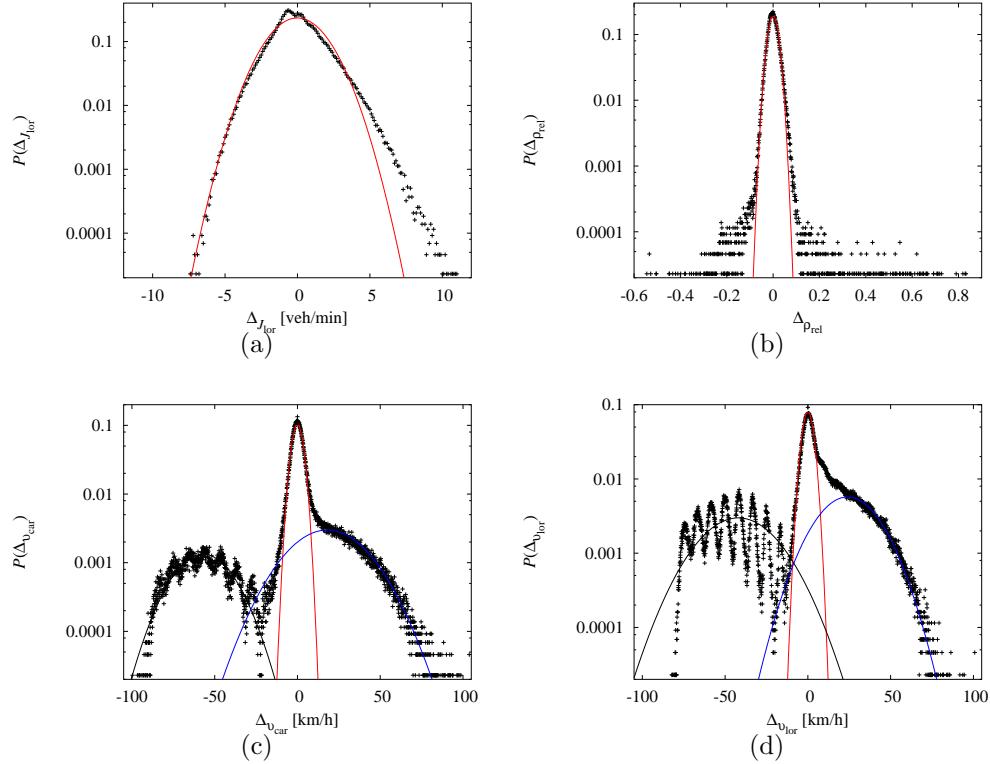


**Fig. 5.25.** The probability distribution  $P(\Delta J_{\text{veh}})$  of the differences  $\Delta J_{\text{veh}}$  of raw data of traffic flow  $J_{\text{veh}}$  and the 9 min moving averages  $\text{MOV}_9(J_{\text{veh}})$  during 320 days of measurements of the detector 5874-023-001. Note the semilogarithmic plot. Apart from outliers at the edges the points follow in good approximation a Gaussian distribution (red) with zero mean  $\mu$  and a standard deviation of  $\sigma = 2.6 \text{ veh/min}$ .

Another method to investigate fluctuations and that does not use adjacent data values is the use of the classified traffic time series with Eq. 5.14. Thereby the data are averaged over the same time of day within the days of the same class. In this case, the noise within the classification is considered.

In Fig. 5.27 (a) an example of the class TWT can be seen. In comparison to the 9 min moving averages it can be clearly seen, that the curve is even smoother. This strongly depends on the particular classification and the number of days used for that. The standard deviation  $\sigma$  is now calculated for each minute separately. To get an impression of the fluctuations within the particular class, the raw data of a single day is not enough. Therefore in Fig. 5.27 (b) also curves added by  $\pm\sigma$  (red) as well as  $\pm 2\sigma$  (blue) can be seen. It is quite clear, that also within a certain class the fluctuations are very high.

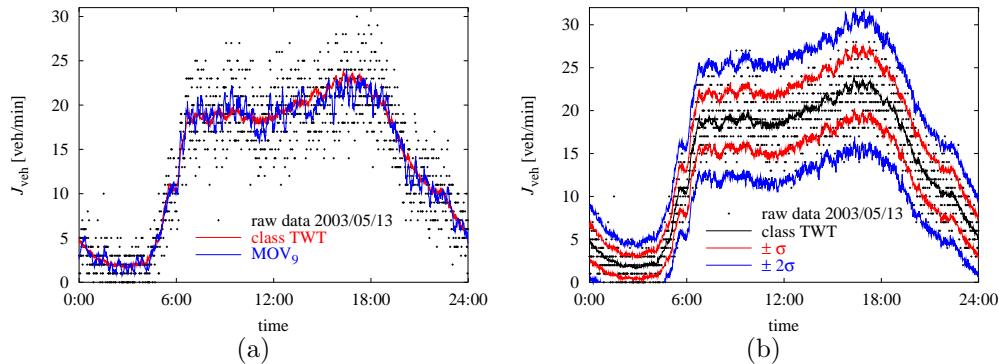
If the traffic data is considered as a superposition of different long and short term effects like done in [Zackor et al., 1996] or in [van Iseghem and Danech-Pajouh, 1999], the question comes up, how the fluctuations influence the time series in detail. The crucial point is thereby, whether the superposition is additive (ho-



**Fig. 5.26.** The same investigations like Fig. 5.25 for the other raw data of detector 5874-023-001 for (a) the flow of lorries  $J_{\text{lor}}$ , (b) the occupancy  $\rho_{\text{rel}}$ . The raw data are the points, the red curve mark an estimated Gaussian distribution. For the velocities of passenger cars  $v_{\text{pc}}$  (c) and lorries  $v_{\text{lor}}$  (d) the outliers that indicate systematical fluctuations under or over the mean are separately approximated with the black curve for under and the blue curve for over the mean.

Data	$\sigma$	$\sigma^2$
$J_{\text{veh}}$	2.6 veh/min	$6.76 \text{ [veh/min]}^2$
$J_{\text{lor}}$	1.7 veh/min	$2.89 \text{ [veh/min]}^2$
$v_{\text{pc}}$	3 km/h	$9 \text{ [km/h]}^2$
$v_{\text{lor}}$	3 km/h	$9 \text{ [km/h]}^2$
$\rho_{\text{rel}}$	0.02	0.0004

**Table 5.6.** Standard deviation  $\sigma$  and the variance  $\sigma^2$  of random noise for the different minute aggregated raw data. The values are approximated with the Gaussian distributions that can be seen in Fig. 5.25 and Fig. 5.26.



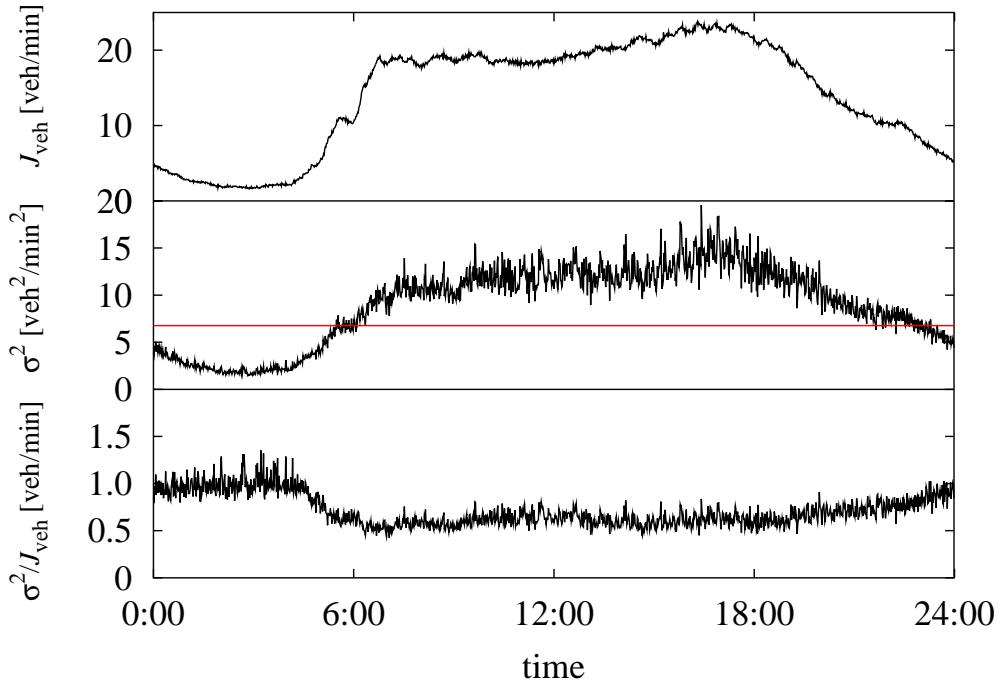
**Fig. 5.27.** (a) The raw data of Tuesday 2003/05/13 of the loop detector 5874-023-001 on the right lane of the BAB 40 driving direction east in comparison to the classified traffic time series of the class TWT (red) and the 9 min moving averages  $MOV_9$  (blue). Whereas the moving averages smooth the curve over adjacent values, with the classified traffic time series the fluctuations are considered within the particular class, in this case TWT. (b) A basic characteristic of the fluctuations is their standard deviation  $\sigma$  that is now calculated for each minute separately within the class TWT. In addition to the raw data and the classified data the lines for  $\pm\sigma$  (red) and  $\pm 2\sigma$  (blue) are shown.

moscedasticity) or multiplicative (heteroscedasticity). In detail, homoscedasticity means, that the variance  $\sigma^2$  of the fluctuations is independent of the mean value. The complement, heteroscedasticity, means, that for different mean values different variances are measured.

To answer this question, a closer look is taken on the variances  $\sigma^2$  within the classified traffic time series. Therefore the loop detector 5874-023-001 on the right lane of the BAB 40 driving direction east is chosen.

In Fig. 5.28 three time series can be seen. On the top is the classified traffic time series of the flow  $J_{\text{veh}}$  for the class TWT (320 days from 2001/01/09 to 2004/07/20). In the central are shown the variances  $\sigma^2$  of the traffic flow values within this class for each minute separately. To compare them with the order of magnitude of the variance of the distribution  $P(\Delta J_{\text{veh}})$  (see Table 5.6), this is marked as the red line. It is quite clear, that the order of magnitude is quite similar, although, if the variance is calculated for each minute separately, it can be seen, that it follows nearly the characteristics of the mean daily traffic flow and is in any case heteroscedastic. Even more, with an increasing traffic flow, also the variance increases.

This motivates to consider the fraction  $\sigma^2/J_{\text{veh}}$  of the variance  $\sigma^2$  in regard to the classified flow  $J_{\text{veh}}$ , what can be seen at the bottom. The values lie between 0.4 and 1.4, and although the characteristics of the daily traffic time series disappears, it seems, that the values are a little bit higher at night than during the day. Further investigations show, that this is quite typical for the classified traffic flow.



**Fig. 5.28.** The variance of classified traffic flow within the days of the classification. On the top is the classified traffic flow  $J_{\text{veh}}$  of the class TWT for the detector 5874-023-001 on the right lane of the BAB 40 driving direction east. In the central is the variance  $\sigma^2$  for all traffic flow values of the class TWT for each minute separately. At the bottom is the variance divided by the classified traffic flow  $\sigma^2/J_{\text{veh}}$ . All values are calculated for each minute separately. The red line mark the variance of the distribution  $P(\Delta_{J_{\text{veh}}})$  (see Table 5.6).

Depending on the particular measurement location and choice of the classification sometimes also at other times the variance can reach values up to one, two, or even three times of the value of the classified flow.

To interpret this result it has to be considered what is done in detail. First, the classified traffic flow is calculated with Eq. 5.14, so it depends on the class  $D_n$  and is calculated for each time step  $t$  separately:  $J_{\text{veh}} = J_{\text{veh}}(\mathcal{G}_{D_n}, t)$ . To be precise, what can be seen in Fig. 5.28 is the mean value of the traffic flow of all days that are elements of the group  $\mathcal{G}_{D_n}$  of the class  $D_n$ . In the particular case are this all Tuesdays, Wednesdays, and Thursdays:  $D_n = \text{TWT}$ .

In the same manner the variance depends on the group  $\mathcal{G}_{D_n}$  and the time  $t$ :  $\sigma^2 = \sigma(\mathcal{G}_{D_n}, t)^2$ . It is a measure for the fluctuations within the particular class  $D_n$  and so for the stability of the classified traffic time series. The higher the variance, the higher are deviations within the data of the classified days. If the classified traffic time series is used for forecast the variance is a crucial information

for the quality that can be expected for the forecast.

Obviously the deviations of the classified traffic flow are a little bit higher at night. For the rest of the day hold in good approximation

$$\frac{\sigma(\mathcal{G}_{D_n}, t)^2}{J_{\text{veh}}(\mathcal{G}_{D_n}, t)} \approx \text{const.} \quad (5.23)$$

Investigations with other loop detectors confirm Eq. 5.23 for several periods of the day. Nevertheless, sometimes the fraction is increased. For this, there could be several reasons.

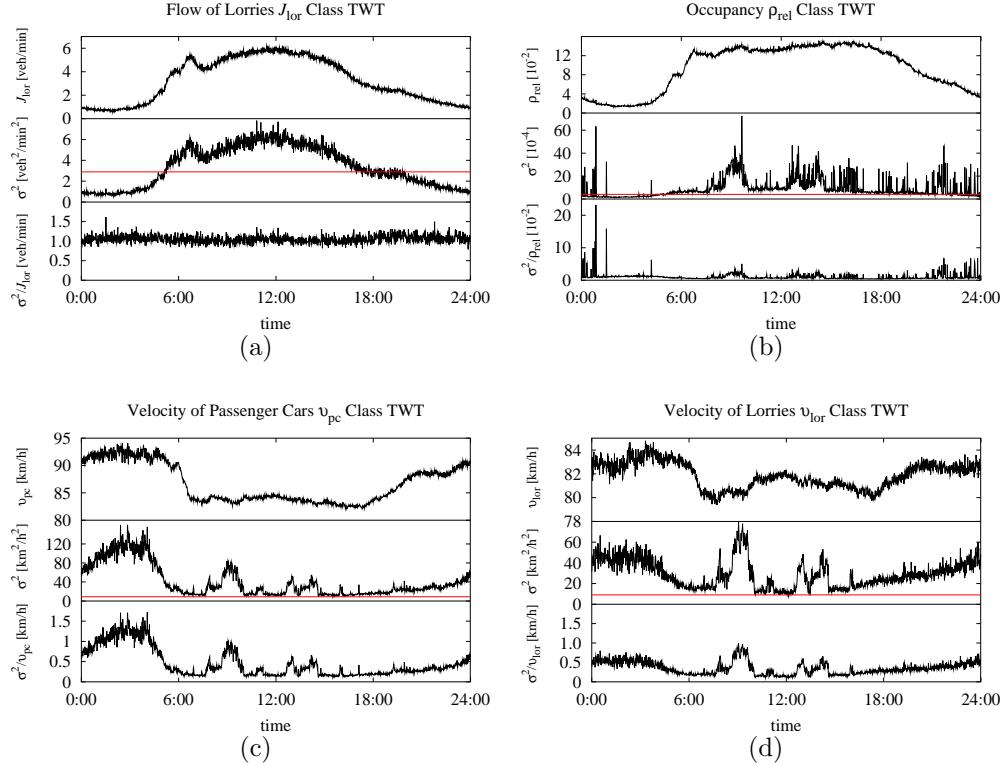
One possible reason is, that the classification might be insufficient for the particular detector. The classification that is used up to now distinguishes the different weekdays. But is this in general a suitable classification? To answer this question, in Sec. 6.1 an algorithm is developed, that classifies daily traffic time series automatically.

Another reason is that there are days with special events that cannot be considered within the classification. Whether the influence of special events can be forecast or not depends strongly on the kind of the particular event. This question is discussed in Sec. 5.6.2.

Furthermore, it can be sometimes basically difficult or even impossible to estimate the traffic data at certain times. For instance, fluctuations at night, when there is no interaction between the vehicles, are nearly unpredictable.

Up to now the investigations are only done for the classified traffic flow  $J_{\text{veh}}(\mathcal{G}_{D_n}, t)$ . Especially the variance  $\sigma_x(\mathcal{G}_{D_n}, t)^2$  in regard to the variance of the distribution  $P(\Delta_x)$  (see Table 5.6), and the fraction  $\sigma_x(\mathcal{G}_{D_n}, t)^2/x(\mathcal{G}_{D_n}, t)$  offer useful information about the quality of using classifications of any kind of traffic data  $x$  for forecasts. Thus, the same investigations are done for the other available traffic data of the detector 5874-023-001 on the right lane of the BAB 40 driving direction east. For that also the 320 days of the class TWT from 2001/01/09 to 2004/07/20 are used.

For the traffic flow of lorries  $J_{\text{lor}}(\mathcal{G}_{D_n}, t)$  that can be seen in Fig. 5.29 (a) holds Eq. 5.23 quite exactly. Furthermore  $\sigma_{J_{\text{lor}}}(\mathcal{G}_{D_n}, t)^2$  is of the same order of magnitude as the variance of  $P(\Delta_{J_{\text{lor}}})$  (red line). This is completely different in the occupancy data  $\rho_{\text{rel}}(\mathcal{G}_{D_n}, t)$  in Fig. 5.29 (b). In the fraction  $\sigma_{\rho_{\text{rel}}}(\mathcal{G}_{D_n}, t)^2/\rho_{\text{rel}}(\mathcal{G}_{D_n}, t)$  the highest outliers of all the traffic data are observed. Furthermore the variance  $\sigma_{\rho_{\text{rel}}}(\mathcal{G}_{D_n}, t)^2$  is much higher than the variance of  $P(\Delta_{\rho_{\text{rel}}})$ . This also holds for both the velocity of passenger cars  $v_{\text{pc}}(\mathcal{G}_{D_n}, t)$  in Fig. 5.29 (c) and that of lorries  $v_{\text{lor}}(\mathcal{G}_{D_n}, t)$  in Fig. 5.29 (d). For both velocities the variance is also significantly higher at night. The reason therefore is already mentioned: because of the missing vehicle to vehicle interaction the velocity is nearly random. Furthermore, during the day the curve is very unstable at typical times with a high congestion probability. But for the other periods the magnitude of  $\sigma_v(\mathcal{G}_{D_n}, t)^2/v(\mathcal{G}_{D_n}, t)$  is smaller than one for both velocities and thus lower than those of the traffic flow.



**Fig. 5.29.** The same graphs like in Fig. 5.28 for the other raw data obtained by the loop detectors: a) traffic flow of lorries  $J_{\text{lor}}(\mathcal{G}_{D_n}, t)$ , b) occupancy  $\rho_{\text{rel}}(\mathcal{G}_{D_n}, t)$ , c) velocity of passenger cars  $v_{\text{pc}}(\mathcal{G}_{D_n}, t)$ , and d) velocity of lorries  $v_{\text{lor}}(\mathcal{G}_{D_n}, t)$ . The variances of the distribution  $P(\Delta_x)$  for each particular traffic data  $x$  from Table 5.6 can be seen as red lines. Whereas the fraction  $\sigma^2/J_{\text{lor}}$  of the flow of lorries is very stable, that of the velocities ( $\sigma^2/v_{\text{pc}}$  and  $\sigma^2/v_{\text{lor}}$ ) is in general lower. Only at night and at typical congestion times it is increased. The variance  $\sigma^2$  of the classified occupancy shows many large outliers what is an indicator for large forecast errors, at least if the forecast is done with classification.

## 5.6 Interpretation

At this point the basic results of the empirical data analysis are discussed in connection with the results of the former works in Chapter 3. The objective is to understand the physical background that is necessary to formulate a new traffic forecast method that can be used for applications in general. Therefore it is important to analyse which data can be forecast and which method is appropriate for which forecast horizon.

### 5.6.1 Data to Forecast

In the forecast methods mentioned in Chapter 3 and in all analyses up to now following data is estimated:

- flow  $J$  (even distinguished in that of all vehicles  $J_{\text{veh}}$ , passenger cars  $J_{\text{pc}}$ , or lorries  $J_{\text{lor}}$ ),
- demand  $Q$ ,
- density  $\rho$ ,
- occupancy  $\rho_{\text{rel}}$ ,
- velocity  $v$  (even distinguished in that of all vehicles  $v_{\text{veh}}$ , passenger cars  $v_{\text{pc}}$ , or lorries  $v_{\text{lor}}$ ), and
- travel time  $T$ .

Which data is chosen of course depends on which data is available for analysis, and which data one is interested in. In most cases like in this work flow  $J$ , occupancy  $\rho_{\text{rel}}$ , and velocity  $v$  are directly measurable using inductive loop detectors. To measure the density  $\rho$  directly is very difficult. In most cases  $\rho$  is estimated assuming homogenous traffic conditions from  $J$  and  $v$  using Eq. 2.4. In this case it makes more sense to forecast  $J$  or  $v$  directly, and if one is interested in  $\rho$  it can be calculated thereafter.

This holds also for travel times  $T$ . Although in the meanwhile there are several methods of directly measuring  $T$  as used in [Matsumura et al., 1998, Sun et al., 2003], even in newer works  $T$  is estimated from other measured raw data [Bovy and Thijs, 2000, van Lint, 2004], in most cases this is  $v$ .

Both,  $v$  and  $T$  are completely unsuitable for forecast in case of free flow. What is meant with free flow here, is that there is no interaction between vehicles, like often observed at night. Both data depend completely on the desired  $v$ , which can strongly differ and lead to high deviations in forecasting it, as shown in the variance of the classified time series.

At this point it must be known, that there is a strong dependence among the traffic data. How the data depend on each other in detail has much to do with the

dynamics of the traffic and is part of many former investigations and discussions. For instance, whereas Kerner and Rehborn [1996] propose a *three phase theory* of traffic, Helbing et al. [1999] distinguish five different kinds of congestions. The interested reader is referred to the extensive literature, especially to the references in [Ceder, 1999, Fukui et al., 2003, Helbing et al., 2000, ITS International, 2002, 2003, Taylor, 2002].

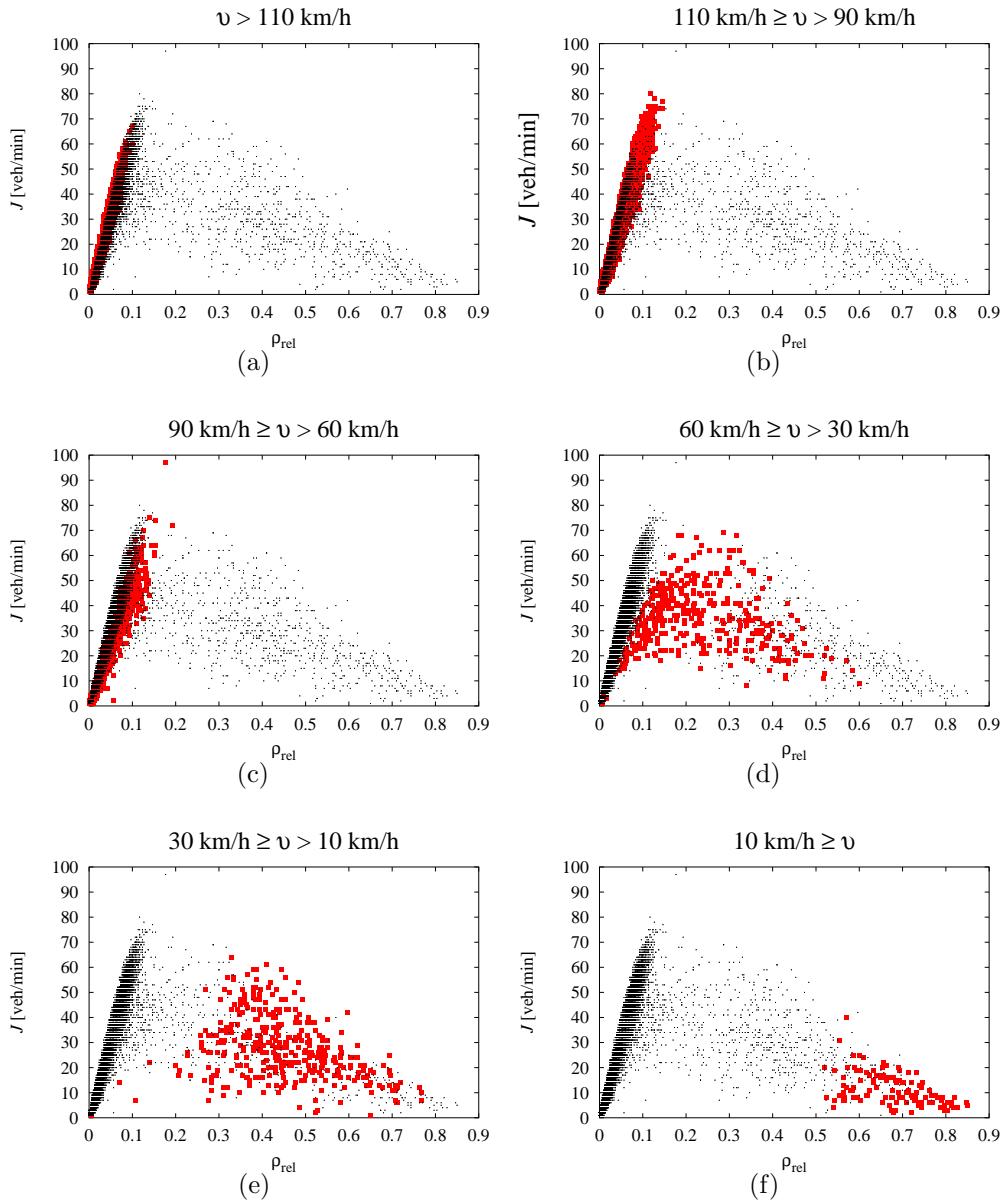
Apart from this discussion it is an undisputable fact, that free flow traffic and congested traffic can be clearly distinguished. The free flow regime is clearly characterised by high velocities. Each vehicle has no or nearly no interaction with other vehicles and can drive as fast or nearly as fast as the driver wants. This can be seen in an easy way in the fundamental diagrams in Fig. 5.30. The regions for velocities  $v > 90$  km/h cover a large region of the possible traffic flow  $J$ , whereas only little changes of the occupancy  $\rho_{\text{rel}}$  and the velocity  $v$  are observed. Thus, in free flow the traffic data that describes the traffic state most exactly is the traffic flow  $J$ .

This changes drastically in congested traffic. Congested traffic could appear, when the traffic flow  $J$  crosses a certain maximum limit. During a congestion  $J$  decreases, the occupancy  $\rho_{\text{rel}}$  (as well as the density  $\rho$ ) increases and the velocity  $v$  decreases. This can be seen in Fig. 5.30 (d) – (f).

How this happens in detail is investigated with spatial temporal traffic models and is part of the discussions mentioned above. Many of the traffic models that are proposed in this extensive literature could make estimations of the spatial temporal extension and moving of a traffic congestion. In opposite to this, the formation of a traffic congestion can only be phenomenologically described. The exact time of a formation of a traffic congestion can only be statistically described and has large variances in regard to the fluctuations that are observed in the velocity and the occupancy.

The traffic data that are most feasible for forecasts are the traffic demand  $Q$  or the traffic flow  $J$ , respectively. But whereas the traffic flow  $J$  is clearly defined as a measureable observable in Sec. 2.1, up to now it is not clearly defined what is meant with traffic demand. This lies in the fact, that both data are very similar and in many applications the traffic demand and the traffic flow are only two expressions for the same thing.

This is not the general case. There is a basic difference between the traffic demand and the traffic flow. The traffic flow is an observable that is clearly defined by Eq. 2.1 and can be easily measured by the inductive loop detectors. The traffic demand is a measurement for the need for mobility on a certain section. It is the number of vehicles, whose drivers try to pass the cross section. The traffic demand can also be defined as the traffic flow on motorways with infinite capacity. That means motorways, where all drivers could always drive as fast as they can. In general, that means for the most time  $t$  of the day, the traffic demand  $Q(t)$  is very similar to the traffic flow  $J(t)$ . But to be precise, three cases have to be distinguished. In the region where there is no vehicle interaction holds



**Fig. 5.30.** Local fundamental diagrams of the cross section 5877-019 driving direction east. Shown is the traffic flow  $J$  aggregated over the three lanes of the cross section on the ordinate and the occupancy  $\rho_{\text{rel}}$  averaged over the three lanes on the abscissa. The small points mark all measured data. The red squares show the data for the particular velocities  $v$ .

$$J(t) = Q(t). \quad (5.24)$$

In this case the traffic demand is exactly the traffic flow. This happens mostly at night.

During the most time of the day there is a lot of vehicle interaction, but the vehicles could drive nearly as fast as they want and

$$J(t) \approx Q(t). \quad (5.25)$$

The third case is a traffic congestion. The capacity of the road could not cope with the traffic demand:

$$J(t) < Q(t). \quad (5.26)$$

In case of a congestion, vehicles cannot pass the road in the desired time and the traffic flow is reduced by a negative addend  $\Delta_{\text{jam}}(t)$ :

$$J(t) = Q(t) + \Delta_{\text{jam}}(t). \quad (5.27)$$

These vehicles have to pass it a little bit later. Then the traffic flow is the sum of the demand of this time and the flow  $\Delta_{\text{delay}}(t)$  that come from delayed vehicles:

$$J(t) = Q(t) + \Delta_{\text{delay}}(t). \quad (5.28)$$

If a conservation of vehicles is assumed it holds

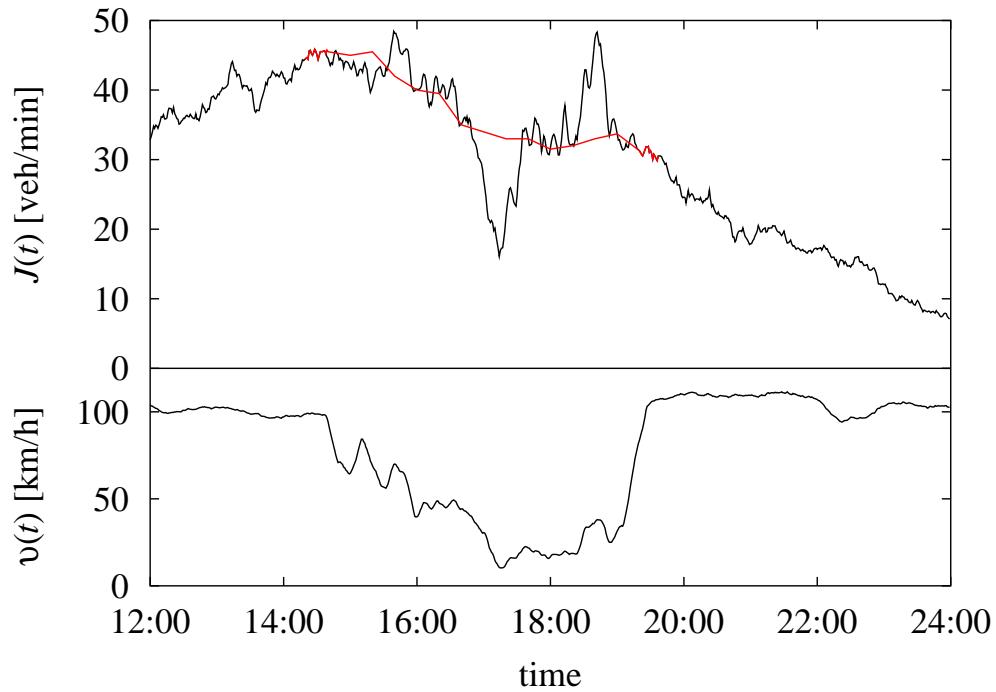
$$\int_{t_{\text{jam}}}^{t_{\text{end}}} [\Delta_{\text{delay}}(t) + \Delta_{\text{jam}}(t)] dt = 0. \quad (5.29)$$

Thereby  $t_{\text{jam}}$  is the time the congestion begins and  $t_{\text{end}}$  the virtual time when the last delayed vehicle passes the cross section because of this congestion.

The traffic flow can always be measured in opposite to the traffic demand. Thus, the traffic demand is the data to be estimated. A schematical sketch of such an estimation can be seen in Fig. 5.31. Some of these considerations are also already used in [Krause, 1988], but it comes out that the estimation of the traffic demand is very difficult. Especially in practise Eq. 5.29 does not hold exactly because of underestimations of the measured  $J(t)$  in case of a traffic congestion. The problem is already discussed in detail in Sec. 4.2.6.

The reason, why the traffic flow  $J$  shows that stable structures is, that all daily or seasonal, recurrent or special, predictable or unpredictable causes first lead to a certain traffic demand  $Q$ . Then, all the measureable traffic data are only the observables of the traffic states, that are caused by  $Q$ . Because  $J$  is very similar to it, the most feasible forecast results are obtained for  $J$ .

Nevertheless, if anyone wants to forecast anything in traffic, the first objective must be to forecast the traffic demand. Unfortunately the traffic demand is not measurable in an easy way. But the traffic flow is a reliable approximation for it.



**Fig. 5.31.** 21 min moving averages of measured traffic flow  $J(t)$  and velocity  $v(t)$  of the cross section 5877-019 at Thursday 2004/07/15. For the time beyond the traffic congestion, that is indicated by the decreased velocity, it holds  $J(t) \approx Q(t)$ . For the time with congestion the demand  $Q(t)$  is schematically estimated with the red line. In case of the traffic congestion vehicles could not pass the cross section in the desired time and  $J(t) < Q(t)$ . Because they have to pass it later, then  $J(t) > Q(t)$ .

Very important is the reason, why there are recurring structures in the traffic demand. Of course a vehicle is neither present or absent on a network because it is for instance 9:00. Rather, these temporal factors are only descriptive variables that represent the effects of the temporal and location specific distribution of activities that generate traffic demand. A basic method must be found, how recurring structures can be found and how they could be used for forecasts. To structure the problem, in the next section some considerations about the different methods in regard to different forecast horizons are done.

### 5.6.2 Methods Depending on the Forecast Horizon

In Chapter 3 several former works forecasting different traffic data are presented. With the empirical results obtained in this work it is now quite comprehensible how they work, and which method is most feasible for forecasts. It can be seen in the following, that this depends strongly on the forecast horizon.

Especially the ARIMA and the parametric regression methods presented in Sec. 3.1.3 work quite well for short term forecasts, especially for the one-step-ahead forecast. How this works is quite easy to understand with the results of Sec. 5.5. The most ARIMA methods used for traffic forecasts are of low order and are basically used to eliminate the noise. This can also be understood if the similarities of Eq. 5.16 and Eq. 3.13, or Eq. 5.17 and Eq. 3.18 are considered, respectively. Therefore the most recent measured values are used depending on the particular choice of the order of the model and the parameters. Physically, this method is completely equal to the naive predictor. The traffic state is estimated to keep as it is, perhaps with a positive or negative trend. The particular choice of the parameters may depend on the particular investigated traffic data, but is no basically different approach. Note, that investigations in [Chrobok et al., 2004] show, that a trend model fast leads to large over- or underapproximations.

For short forecast horizons there is no better method to forecast traffic data than these low order parametric regression techniques. Especially the investigations in Sec. 3.6 have proved, that it does not depend on the complexity of the algorithm how feasible the forecast results for one-step-ahead forecasts are.

The only way to improve these results is to use adjacent data, for instance, with a spatial temporal traffic model. An investigation of this is not part of this work, but in the forecast application presented in Chapter 7 a microscopic traffic simulation model is used to enhance the short term forecast.

For longer forecast horizons classified historical traffic time series show feasible results. Also the empirical results of the fluctuations within a certain classification show, that a classification using similar conditions make sense. But to improve the forecast results, also other external influences than the calendary differences have to be considered. The circumstances that influence the traffic must be as similar as they could. On the other hand not for every day of the year a classification is needed. The results that could be seen in Fig. 5.8 and Fig. 5.9 show, that only a part of the mean daily traffic (MDT) differences can be explained with the changings according to the weekdays. The question is, how an adequate classification according to the circumstances that crucially influence the traffic can be determined.

A way to do this is the application of a clustering algorithm as done in the ATHENA method (cp. Sec. 3.4 on page 43). After the clustering each cluster forms a class  $D_n$  and with the data of all days of the group  $\mathcal{G}_{D_n}$  the classified traffic time series can be calculated and can be used for forecast. This procedure solves the problem of automatically finding a suitable classification. The problem left is to assign different attributes to the days. These attributes must be also those that come because of the calendar like “day of the week” or “holiday” as well as those that come because of special events like “football”. The attributes must be known and assigned from the first. Otherwise it is not possible to assign later the day whose data should be forecast to the adequate cluster.

The classification according to attributes is nothing else than a way to put external knowledge in the forecast algorithm. The analysis of the special events show,

that any kind of knowledge can enhance the forecast result. Thus, those attributes should be all causes that influence the traffic. But as discussed in Sec. 5.4, not all of them can be forecast. Thereby, the predictability of the event or the cause itself has to be distinguished from its effect on the traffic state or the accuracy with which it can be estimated, respectively.

In regard to this, in Table 5.7 the events of Sec. 5.4 are qualitatively considered. If the cause is not known before as in case of accidents, it does not make sense to assume the attribute “accident” to the time series that should be forecast. In the most cases, for instance, in case of road works or big events, the cause is well known before and even in case of weather conditions there is a certain possibility to feasibly forecast them hours before.

event	predictability cause	effect
accident	unpredictable	unpredictable
weather	depends on weather forecast	partially predictable
road works	known before	partially predictable
big events	known before	partially predictable
network changings	known before	nearly unpredictable

**Table 5.7.** Different special events discussed in Sec. 5.4 in regard to the predictability of the cause and the predictability of the effect. For an adequate forecast cause and effect must be predictable.

The predictability of the effect with known cause depends on the stability with which the same cause is leading to the same effect. In case of accidents there are a variety of different possible circumstances. It is often even difficult to make statements about for how long an incident influences the traffic hours after it happened. Similar conditions hold for all unique events, especially changings of the network. In this case it is impossible to apply any model based on cause and effect.

For all recurring events with stable effects the impact can be approximated. If the event is always at the same time, a cluster including this event can be considered. If the event is at different times like the television broadcasts of the football matches, the netto effect must be filtered.

To consider circumstances that are not known but that also influence the traffic state, nonparametric regression techniques in form of a state space model can help (see Sec. 3.2). The benefit of a state space model is, that no considerations about any events or even calendar differences are necessary. To forecast the traffic states only the nearest neighbours in phase space have to be determined. The problem with this method is, that the effects of special events, that are as

well known before as their effects on traffic states in past, could not be considered until the first changings in the measurements appear.

To understand this clearly the case of the television broadcast of the football matches in Fig. 5.20 is considered. What a state space model does is, that states in the past are compared with the current state. From the manner, how a similar state in the past evolves, the future state is estimated. Apart from the fact how the states are defined and which distance in phase space is used in detail, we could consider the classified traffic time series as an aggregation of similar states in the past and deviations from the fraction  $J_{\text{rel}} = 1$  as distances. Now, the evolution of the past is taken for the future approximation. But this means in most cases exactly those states for that also hold  $J_{\text{rel}} \approx 1$ . Not until the first clear deviations of the traffic flow are measured, such as two or one hour before the kick off time, this effect can be forecast in general. With larger forecast horizons, this event is even completely unpredictable with the state space model.

This is based on the fact, that the state space model works only on the level of the measured traffic data and does not use any external knowledge about the reasons for that. But the measured traffic data and all the observables are only the symptoms of the activities that lead to different traffic conditions. Thus, solely using a state space model with measured traffic data can never reach the accuracy of a system where the causes that lead to the traffic are considered.

Nevertheless, a state space model is useful when the causes that lead to a particular traffic state are not known and this is mostly the case. The events that can be seen in Table 5.7 are only a small amount of everything that influences the traffic. All the particular reasons of all the single drivers are not known, but they influence the traffic. In this case a state space model works quite well and offers useful information.

Note, that especially for a single cross section the forecast using a state space model is even more difficult because of the fluctuations. Nevertheless all considerations have to be done in regard to the particular location, that means for each cross section separately. If any lane resolved information is needed, this must even be done for each loop detector separately. Thus, before using a state space model the data have to be smoothed.

Up to now it could be seen, that different forecast algorithms mentioned in Chapter 3 work for different forecast horizons. The information of causes that influence the traffic are the base for feasible forecasts and they can be used in form of classifications. For all causes that are not known most recently measured traffic data can be used in form of a state space model. For short term forecasts the most recently measured traffic data prepared with low order parametric regression models provide feasible results.

These answers lead directly to some problems that have to be considered in the next chapter to develop a new general basic traffic forecast algorithm. The first problem is how a suitable classification can be found and how the system can be fed with information. Furthermore this must be done automatically for each loop detector separately. This is the largest part of the next chapter.

Thereafter, the forecast data must be calculated. In doing so basic changes of the network have to be considered as well as events that are not part of the classification. Because parametric regressions of most recently measured data show the best results for short term horizons, it has to be investigated, for how long this is the case. Finally the most recently measured data is considered according to a state space model.

Before this is done, a short look should be taken on the neural networks, that are not considered in this work for further investigations. The neural networks mentioned in Sec. 3.3 are a mathematical formulation of processes that are expected to happen in the human brain. Thereby the basic idea is the opportunity of self learning. If neural networks are considered as sufficiently sophisticated the whole process of knowledge based systems could be treated by them automatically. But the state of the art is far away from that. The neural networks used today are not sufficiently comparable with the human brain. Those that are used for traffic forecasts consist only of a few layers with some neurons. The former works show, that they are neither able to enhance the performance of even the naive predictor on short timescales [Dougherty and Cobbett, 1997, Lee and Choi, 1998], and even the ability to recognise recurring structures is limited to distinguish the seven days of the week [van der Voort et al., 1996]. Many authors conclude that the results are unsatisfying, but that they expect the neural network to be superior to other methods sometimes. But the enhancement of neural networks is not part of this work.

# Chapter 6

# New Forecast Algorithm

“Savoir pour prévoir, afin de pouvoir.”

**Auguste Comte**  
(1798 – 1857), French philosopher

In this chapter a new traffic forecast procedure is introduced. Therefore, some open questions are investigated that come from the former works and the empirical results. In the following section a method is proposed that automatically calculates the classification for each loop detector separately. Thereafter it is analysed how the forecast data can be feasibly calculated with this classification, especially in regard to automatic adaptivity.

Then it is considered how most recently measured data can be considered in the forecast algorithm. Therefore the forecast horizons are quantified, for which parametric regression methods of the most recently measured data outperform the classified traffic flow time series. Furthermore another approach to use most recently measured data is proposed that uses the method of nearest neighbours in phase space. Finally, the results of long term forecast, short term forecast, and combinations are shown for all the available traffic values.

## 6.1 Classification with Cluster Analysis

The idea of a classification is, that the causes that influence the traffic are considered in that sense, that for a certain time interval of a traffic time series under certain conditions the traffic data can be estimated from past time series under similar conditions. At night there is in general very little traffic and it makes often sense to choose the 24 hours of a day as time interval.

The conditions can be assigned to the days in form of attributes. Thereby it is essential to distinguish attributes with a large influence on traffic from those without one. This distinction is important. If the classification goes too far and

too many attributes per day are considered, days exist with unique attributes and the classification cannot be used for forecasts.

Whereby long term forecasts with classified traffic time series are quite feasible, in most works that are mentioned in Sec. 3.4 the classification is done manually. Thereby, the individual classification differs and it depends, where the data are exactly stemming from. As discussed there is also a strong spatial dependence and if a lane resolved information is needed the classification must be chosen even according to this.

This leads directly to the question, how the process of classification can be done automatically. About this only little investigations are done. Most of the works deal with the problem of automatic pattern recognition [Bajwa et al., 2003, Whiteley and Davis, 1993]. In some works cluster algorithms are applied, but the attributes of the clusters are analysed manually. Within the ATHENA and the KARIMA approach (see Sec. 3.4 on page 43 and Sec. 3.5.2 on page 56, respectively) the properties of the day of the week changings are arranged with the  $k$ -means cluster algorithm or a self organizing neural network, respectively. In both cases the clusters are interpreted manually. In newer works like in [Nowotny et al., 2003] or in [Chung, 2003] other cluster algorithms are used and also holidays and weather conditions are considered as attributes.

Which cluster algorithm works best is often an essential question. All algorithms can basically be distinguished in partitioning clustering and hierarchical clustering. Partitioning clustering partitioned objects in a particular number of clusters. In hierarchical clustering this is not done, but a series of partitions takes place resulting in a *dendrogram*, which is a two dimensional diagram that illustrates the fusions or divisions made at each successive stage of analysis. That may run from a single cluster containing all objects to a certain number of clusters each containing a single object.

The question, which cluster algorithm is most feasible according to daily traffic data is considered in [Kleinhans, 2002]. Because a dendrogram of such a large number of objects like thousands of daily traffic time series is difficult to interpret, the author states that hierarchical cluster algorithms are not useful for that purpose in general. Finally, a  $k$ -medoid cluster algorithm is for several reasons found to be most feasible for that purpose. It also outperforms the  $k$ -means method that is often used, for instance, for the ATHENA method. Thus, in the following the  $k$ -medoid cluster algorithm is used. First, a short introduction is given. The interested reader is even referred to [Kaufman and Rousseeuw, 1990].

### 6.1.1 *K*-Medoid Cluster Algorithm

What is needed first for every cluster algorithm is the definition of an object  $\vec{s}_i$  and a distance  $d_{i,j}$  to compare two objects  $\vec{s}_i$  and  $\vec{s}_j$ . For that the definition of a state space vector proposed in Sec. 3.2.1 can be used as an object and the  $p$ -norm (Eq. 3.37) of the difference of two vectors as the distance. Kleinhans [2002] states that for daily traffic time series as state space vectors the Manhattan

metric ( $p = 1$ ) works better than the Euclidean distance ( $p = 2$ ), because the squaring leads to an exorbitant influence of outliers. Then, the algorithm acts in two phases.

In the first phase, an initial partition with  $k$  clusters is calculated. Therefore all objects  $\vec{s}_i$  are divided in so called *normal objects*  $\vec{n}_j$  and *representatives*  $\vec{r}_l$ . Thereby, each normal object belongs to exactly one of all representatives. A representative with all its normal objects forms a cluster.

As the first representative the object is chosen for that the sum of the distances is minimum:

$$\vec{r}_1 = \vec{s}_i, \quad (6.1)$$

with

$$i = \arg \min_i \left( \sum_{j \neq i} d(\vec{s}_i, \vec{s}_j) \right). \quad (6.2)$$

Thereby the operator  $\arg$  means the argument of the operator  $\min$ , in this case  $i$ . To obtain a new representative  $\vec{r}_g$ , the normal object  $\vec{n}_c$  acts as candidate. Then for each other normal object  $\vec{n}_j$  the difference between the distance  $D_j$  to its current representative and the distance to the candidate  $d(\vec{n}_j, \vec{n}_c)$  is calculated and used as the contribution  $C_{jc}$  if it is positive:

$$C_{jc} = \max(D_j - d(\vec{n}_j, \vec{n}_c), 0). \quad (6.3)$$

The candidate  $\vec{n}_{c_0}$  for that holds

$$c_0 = \arg \max_c \left( \sum_j C_{jc} \right) \quad (6.4)$$

is chosen as a new representative  $\vec{r}_g = \vec{n}_{c_0}$ , that acts then as a representative for all normal objects  $\vec{n}_{j_g}$ , for that it is the nearest representative

$$\vec{r}_g = \arg \min_{\vec{r}_l} (d(\vec{n}_{j_g}, \vec{r}_l)). \quad (6.5)$$

Up to  $k$  ( $l = 1, 2, \dots, k$ ) representatives candidates are chosen and new representatives are calculated.

In the second phase, an attempt is made to improve the set of representatives. This means, that the sum of the distances

$$T_{j_g g} = \sum_{j_g} C_{j_g g} \quad (6.6)$$

of the normal objects  $\vec{n}_{j_g}$  to their representatives  $\vec{r}_g$  must be minimum.

Therefore, for each pair consisting of a representative  $\vec{r}_l$  and a normal object  $\vec{n}_c$ , the part  $P(C_{jl}, C_{jc})$  for the normal object  $\vec{n}_j$  is calculated if  $\vec{n}_c$  would be a representative and  $\vec{r}_l$  a normal object. Here  $P(C_{jl}, C_{jc})$  is the part of this change in the overall distance of the normal objects to their representatives. It is calculated according to the following conditions:

- if  $d(\vec{n}_j, \vec{r}_l) > D_j$  and  $d(\vec{n}_j, \vec{n}_c) > D_j$ , then  $P(C_{jl}, C_{jc}) = 0$ .
- if  $d(\vec{n}_j, \vec{r}_l) = D_j$ , two cases must be distinguished:
  - if  $d(\vec{n}_j, \vec{n}_c) < E_j$ , then  $P(C_{jl}, C_{jc}) = d(\vec{n}_j, \vec{n}_c) - d(\vec{n}_j, \vec{r}_l)$ ,
  - else, if  $d(\vec{n}_j, \vec{n}_c) \geq E_j$ , then  $P(C_{jl}, C_{jc}) = E_j - D_j$ .

Thereby  $E_j = d(\vec{n}_j, \vec{r}_s)$  is the distance from  $\vec{n}_j$  to that representative  $\vec{r}_s$ , so that  $E_j$  is the second smallest distance among all the distances from  $\vec{n}_j$  to the representatives.

- if  $d(\vec{n}_j, \vec{r}_l) > D_j$  and  $d(\vec{n}_j, \vec{n}_c) < D_j$ , then  $P(C_{jl}, C_{jc}) = d(\vec{n}_j, \vec{n}_c) - D_j$ .

Finally the sum of all contributions of the change of  $\vec{n}_c$  and  $\vec{r}_l$  is calculated as

$$T_{lc} = \sum_j P(C_{jl}, C_{jc}), \quad (6.7)$$

and this pair  $(\vec{n}_c, \vec{r}_l)$  is chosen for that  $T_{lc}$  is minimum. If this minimum negative, the exchange is executed and the second phase starts from the beginning, otherwise the algorithm ends.

Finally, the representatives  $\vec{r}_l$  together with all the normal objects  $\vec{n}_{jl}$  that belong to them, act as one cluster.

### 6.1.2 Clustering of Days According to Attributes

The first objective is to calculate classified daily traffic time series. Thus, as objects or state space vectors  $\vec{s}_i$  act in general the daily traffic time series. The results of investigations about the variances within a certain classification around the mean value in Sec. 5.5 have shown, that also in cases of a suitable classification the values of the velocity and the occupancy are very unstable. The analyses of Kleinhans [2002] have proofed, that this is also the case if multivariate objects are used for the cluster algorithm.

Thus, univariate state space vectors are used here. Furthermore to avoid misinterpretations because of fluctuations, the 24 hourly aggregated values  $\mu_j$  of the traffic flow are used:

$$\vec{s}_i = (\mu_1, \mu_2, \dots, \mu_{24}). \quad (6.8)$$

With this the  $k$ -medoid algorithm can be applied to the traffic flow of different days of a particular cross section.

In [Danech-Pajouh and Aron, 1991] and [van der Voort et al., 1996] the clustering is only applied to group the days according to the day of the week changes. The days that belong to each cluster are interpreted manually. Kleinhans [2002] uses different numbers of  $k$  and visualises the clusters in the calendar with different colours. Nevertheless the properties of the clustered days are interpreted manually.

Thereby it is an indisputable fact, that the attributes of the days have to be entered manually. Especially for feasible forecasts, it is essential to know the crucial attributes of a day before. Thus, each attribute that is supposed to influence the traffic state and to improve the forecast, must be assigned to each particular day. Then, when the cluster algorithm is applied to all the days, the clusters must be interpreted according to the attributes. Therefore, in [Danech-Pajouh and Aron, 1991] the authors use a predisposed number of  $k = 4$  groups and the case that each weekday has its own characteristics is excluded a priori.

To avoid this, if  $m$  attributes are considered the days have to be clustered in at least  $m$  groups. If also misinterpretations of undetected errors in measurements as well as special events that are not assigned to a particular day have to be taken into account, the number of groups have to be chosen as at least  $k = m + 1$ . This rule is called in the following the  $m + 1$ -rule and means the minimum number of  $k = m + 1$  groups if an analysis in regard to  $m$  attributes is done. How large  $k$  has to be finally chosen is discussed later.

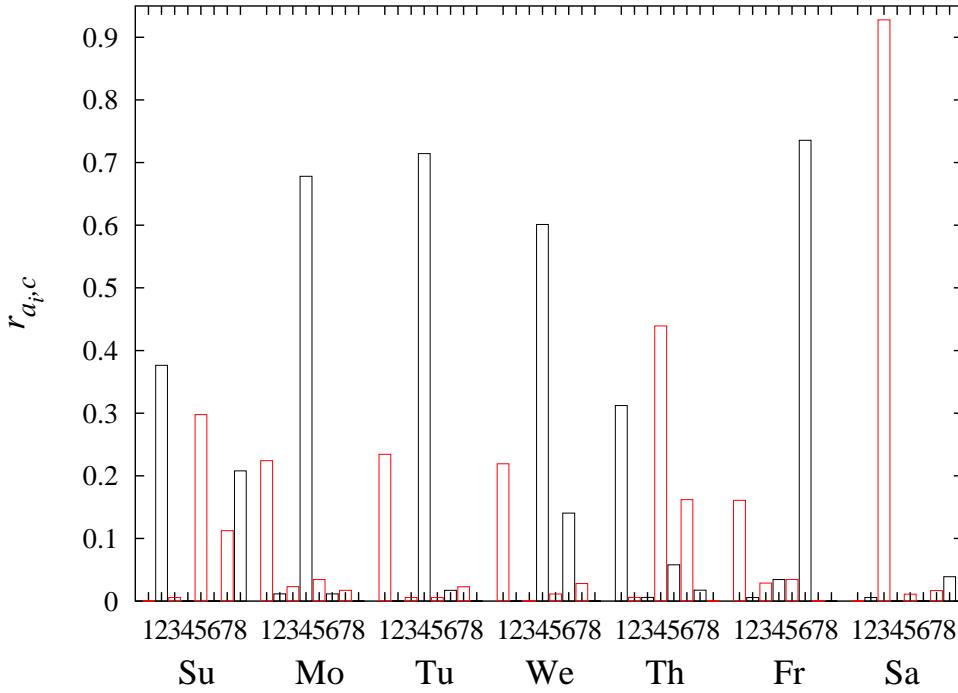
Then, the cluster algorithm is applied and the ratio  $r_{a_i,c}$  with that days with the attribute  $a_i$  belong to cluster  $c$  can be easily calculated by dividing the number of days  $N_{a_i,c}$  with attribute  $a_i$  that belong to cluster  $c$  by the total number of days  $N_{a_i}$  with attribute  $a_i$ :

$$r_{a_i,c} = \frac{N_{a_i,c}}{N_{a_i}}. \quad (6.9)$$

An example for such ratios in case of the seven weekdays as attributes can be seen in the histogram in Fig. 6.1. For convenience the following abbreviations in regard to the day of week attributes are used:

- Monday  $\hat{=}$   $a_1 = \text{Mo}$ ,
- Tuesday  $\hat{=}$   $a_2 = \text{Tu}$ ,
- Wednesday  $\hat{=}$   $a_3 = \text{We}$ ,
- Thursday  $\hat{=}$   $a_4 = \text{Th}$ ,
- Friday  $\hat{=}$   $a_5 = \text{Fr}$ ,
- Saturday  $\hat{=}$   $a_6 = \text{Sa}$ ,
- Sunday  $\hat{=}$   $a_7 = \text{Su}$ .

As database hold in this example all days with more than 1,200 minutes feasible traffic flow data according to Sec. 4.2 of the days from 2001/01/01 to 2004/07/31 ( $N_{\text{days}} = 1307$ ) of the loop detector 5877-019-001 on the right lane of the BAB 40 driving direction east. The number  $N_{\text{days}} = 1307$  of the days when data is available is essential, because if  $k = N_{\text{days}}$ , each day forms a separate cluster. This means that each day is unique and forecasting using classification is not possible. If classification should be used it must be ensured that  $k \ll N_{\text{days}}$  which



**Fig. 6.1.** Histogram of the seven weekdays according to 8 clusters after a  $k$ -medoid cluster algorithm is applied to the traffic flow data of loop detector 5877-019-001. Shown is the ratio  $r_{a_i, c}$  with that the day of week represented by the attribute  $a_i$  belongs to one of the eight clusters  $c$ . The number of clusters is  $k = 8$  according to the  $m + 1$ -rule, what is a lower bound if  $m$  attributes are investigated. Whereas the days from Monday to Thursday mainly belong to the groups  $c = 1$  and  $c = 4$ , Fridays belong to  $c = 6$ , Saturdays to  $c = 3$ , and Sundays to  $c = 2$ ,  $c = 5$ ,  $c = 7$ , and  $c = 8$ .

can be interpreted as an upper bound in opposite to the  $m + 1$ -rule, what is a lower bound.

The question is, how the result can be interpreted automatically. Therefore it is proposed to form a  $k$ -dimensional vector of belongings  $\tilde{b}_{a_i}$  with the ratio of belongings  $r_{a_i, c}$  of the different clusters  $c = 1, 2, \dots, k$ . For the example above the belonging vectors can be seen in Eq. 6.10 whereby the ratios are rounded to two digits.

$$\begin{aligned}
\vec{b}_{\text{Su}} &= (0.00, 0.38, 0.01, 0.00, 0.30, 0.00, 0.11, 0.21), \\
\vec{b}_{\text{Mo}} &= (0.22, 0.01, 0.02, 0.68, 0.03, 0.01, 0.02, 0.00), \\
\vec{b}_{\text{Tu}} &= (0.23, 0.00, 0.01, 0.71, 0.01, 0.02, 0.02, 0.00), \\
\vec{b}_{\text{We}} &= (0.22, 0.00, 0.00, 0.60, 0.01, 0.14, 0.03, 0.00), \\
\vec{b}_{\text{Th}} &= (0.31, 0.01, 0.01, 0.44, 0.06, 0.16, 0.08, 0.00), \\
\vec{b}_{\text{Fr}} &= (0.16, 0.01, 0.03, 0.03, 0.03, 0.74, 0.00, 0.00), \\
\vec{b}_{\text{Sa}} &= (0.00, 0.01, 0.93, 0.00, 0.01, 0.00, 0.02, 0.04).
\end{aligned} \tag{6.10}$$

Now the belonging vectors can be compared using distance measures  $d(a_{i_1}, a_{i_2}) = \|\vec{b}_{a_{i_1}} - \vec{b}_{a_{i_2}}\|$  in regard to a certain norm  $\|\cdot\|$ . Which norm finally is used influences the result only in marginal cases. But for comparability of the distances especially in regard to different dimensions  $k$  of the vectors that are needed later, it has to be considered that for all the belonging vectors hold  $\|\vec{b}_a\|_1 = 1$  apart from their dimension  $k$ . Thereby  $\|\cdot\|_1$  denotes the 1-norm or Manhattan metric. This lies in the fact, that the vectors consist of the ratios of belongings  $r_{a_i, c}$  that sum to one independently of the dimension  $k$ . If another  $p$ -norm should be used, all vectors have to be normalized in regard to that norm  $\vec{b}_{\text{norm}} = \frac{\vec{b}}{\|\vec{b}\|_p}$  before calculating the distance.

For instance, consider the simple case of two very crucial attributes that arrange the time series in  $k = 2$  clusters. The belonging vectors are

$$\vec{b}_1 = (1, 0) \tag{6.11}$$

and

$$\vec{b}_2 = (0, 1), \tag{6.12}$$

the distance according to the Manhattan metric

$$d_{\text{Manhattan}, 1, 2} = \|\vec{b}_1 - \vec{b}_2\|_1 = 2, \tag{6.13}$$

and according to the 2-norm

$$d_{\text{Euclidean}, 1, 2} = \|\vec{b}_1 - \vec{b}_2\|_2 = \sqrt{2}. \tag{6.14}$$

If now a cluster algorithm with, for instance,  $k = 4$  is applied, the time series may be arranged because of additional attributes dividing the two clusters of the 2-medoid algorithm in equal parts resulting in

$$\vec{b}_3 = (0.5, 0.5, 0, 0) \tag{6.15}$$

and

$$\vec{b}_4 = (0, 0, 0.5, 0.5). \tag{6.16}$$

Whereby the Manhattan metric stays the same

$$d_{\text{Manhattan}, 3, 4} = \|\vec{b}_3 - \vec{b}_4\|_1 = 2, \tag{6.17}$$

the Euclidean distance is now

$$d_{\text{Euclidean},3,4} = \|\vec{b}_3 - \vec{b}_4\|_2 = 1. \quad (6.18)$$

Thus, if the Euclidean distance should be used, the vectors first have to be normalized. In the example the 2 dimensional vectors do not change because the norms of  $\vec{b}_1$  and  $\vec{b}_2$  are 1. But the 4 dimensional vectors are now

$$\vec{b}_{3,\text{norm}} = (\sqrt{0.5}, \sqrt{0.5}, 0, 0) \quad (6.19)$$

and

$$\vec{b}_{4,\text{norm}} = (0, 0, \sqrt{0.5}, \sqrt{0.5}), \quad (6.20)$$

and the distance is, as in the 2 dimensional case,  $d_{\text{Euclidean},3,4,\text{norm}} = \sqrt{2}$ .

For the belonging vectors in Eq. 6.10 the distances  $d_{\text{Manhattan}}$  in regard to the Manhattan metric can be seen in the upper right half of Table 6.1. The values on the diagonal are always zero, because the distance of a vector to itself is zero for every norm. In the lower left half of the table the Euclidean distances  $d_{\text{Euclidean},\text{norm}}$  of the normalized vectors can be seen.

$a_i$	Su	Mo	Tu	We	Th	Fr	Sa
Su	0	1.862	1.932	1.921	1.827	1.908	1.844
Mo	1.383	0	0.115	0.280	0.524	1.460	1.887
Tu	1.407	0.051	0	0.268	0.561	1.552	1.944
We	1.401	0.208	0.196	0	0.345	1.306	1.944
Th	1.363	0.403	0.407	0.282	0	1.193	1.922
Fr	1.392	1.320	1.317	1.164	1.061	0	1.910
Sa	1.385	1.391	1.408	1.414	1.406	1.387	0

**Table 6.1.** The distances of the belonging vectors  $\vec{b}_{a_i}$  in Eq. 6.10 in regard to the Manhattan metric  $d_{\text{Manhattan}}$  in the upper right half of the table and to the Euclidean distance  $d_{\text{Euclidean},\text{norm}}$  in the lower left half. Note, that it is important to normalize the vectors of Eq. 6.10 in regard to the particular  $p$ -norm, if the distances should be comparable for different dimensions of the vectors. The distances offer useful information which days behave in a similar manner and which weekdays are completely different. As easily can be seen, the distances of the days Monday to Thursday are significantly smaller than those of the other days, what is in excellent agreement with the results in Sec. 5.3.

It can be seen, that the distances among the working days Monday to Thursday are significantly smaller than those of the other days, what is in excellent agreement with the results in Sec. 5.3. Whereas the quantitative values differ, the qualitative result becomes quite clear. As mentioned above, the choice of the norm is not essential for the basic result. For following investigations the Euclidean distance is used.

To assign the attributes to a certain group, the distances are compared with a certain limit  $d_{\text{lim}}$ , and if for the distance  $d(\vec{b}_{a_{i_1}}, \vec{b}_{a_{i_2}})$  of the belonging vectors with

the attributes  $a_i$  and  $a_j$  hold

$$d(\vec{b}_{a_{i_1}}, \vec{b}_{a_{i_2}}) < d_{\lim}, \quad (6.21)$$

the attributes are joined to a class  $D_n = (a_{i_1}, a_{i_1})$ .

Normally this goes straight forward as can be seen by the values of Table 6.1. To handle some conflicts, the following easy clustering algorithm is applied. Starting from the smallest distance  $d(\vec{b}_{a_{i_1}}, \vec{b}_{a_{i_2}})$  for that hold  $d(\vec{b}_{a_{i_1}}, \vec{b}_{a_{i_2}}) < d_{\lim}$ , the particular two attributes  $a_{i_1}$  and  $a_{i_2}$  are joined. Then, the next smallest distance  $d(\vec{b}_{a_{i_3}}, \vec{b}_{a_{i_4}})$  between the attributes  $a_{i_3}$  and  $a_{i_4}$  is considered. In the case, that  $d(a_{i_3}, a_{i_4}) < d_{\lim}$  and the element  $a_{i_3}$  is already in a class  $D_n = (a_{i_3}, \dots, a_{i_x})$ , then the element  $a_{i_4}$  joins this group, only if the distance limit is not exceeded even for all other elements in this group. Otherwise, the algorithm proceeds with the next smallest distance, up to the case that the next smallest distance exceeds the limit  $d(\vec{b}_{a_{i_x}}, \vec{b}_{a_{i_y}}) \geq d_{\lim}$ .

Because this kind of classification needs two cluster algorithms, the whole process of assigning particular attributes to the days, using the  $k$ -medoid cluster algorithm, forming the belonging vectors, and interpreting them to get the final classifications is called in the following *double cluster assignment* (DCA). The results of the DCA are always a classification of certain classes  $D_n$ , that come from daily traffic time series and attributes  $a_i$  that are assigned to the days before.

$d_{\lim}$	$N_{D_n}$	$D_n$
0.0	7	Su, Mo, Tu, We, Th, Fr, Sa
0.06	6	Su, (Mo,Tu), We, Th, Fr, Sa
0.21	5	Su, (Mo,Tu,We), Th, Fr, Sa
0.41	4	Su, (Mo,Tu,We,Th), Fr, Sa
1.32	3	Su, (Mo,Tu,We,Th,Fr), Sa
1.39	2	(Su,Sa), (Mo,Tu,We,Th,Fr)
1.42	1	(Su,Mo,Tu,We,Th,Fr,Sa)

**Table 6.2.** Different choices of the distance limit  $d_{\lim}$  influence the result of the weekday classification  $D_n$ . For this example the Euclidean distances in Table 6.1 are used, so the result is easily to reconstruct. Depending on  $d_{\lim}$  days form a group of crucial attributes or classes  $D_n$ . Chosen are those limits  $d_{\lim}$  just when the classification changes rounded to two digits.

With the parameter  $d_{\lim}$  it can be controlled how many deviations in the DCA are accepted. If  $d_{\lim}$  is chosen too large, there is no distinction between the attributes. If  $d_{\lim}$  is chosen too small, each attribute may be detected as a crucial one. Actually, this is in case of the weekday classification no problem. But in the following many other attributes are investigated. If each would form a separate group, the result is that each day is unique and thus unpredictable. In Table 6.2 the classification for the example in Table 6.1 can be seen for those values of  $d_{\lim}$ , for that the classification changes.

According to the results of Sec. 5.3 and especially that in Fig. 5.13, the weekdays Monday until Thursday are very similar for the loop detector 5877-019-001 or the cross section 5877-019, respectively. Thus, a distance limit of  $0.41 \leq d_{\text{lim}} \leq 1.32$  should be reasonable. For the following examples  $d_{\text{lim}} = 0.6$  is chosen. The classifications differ, because of the large dependency in space. Therefore, for the following observations, 10 randomly chosen loop detectors apart from the detector 5877-019-001 are investigated. The topological data can be seen in Table 6.3, and the results of the classification with the DCA in Table 6.4.

loop detector	motorway BAB	direction	$N_1$	lane
4135-015-033	1	south	2	right
4214-015-002	2	east	3	middle
4215-006-035	2	west	3	middle
4349-025-001	3	north	3	right
4488-096-033	4	west	2	right
5877-019-001	40	east	3	right
5877-019-003	40	east	3	left
6211-000-033	52	west	2	left
6250-032-001	57	north	2	left
6289-007-034	59	south	2	left
6256-009-002	61	north	2	left

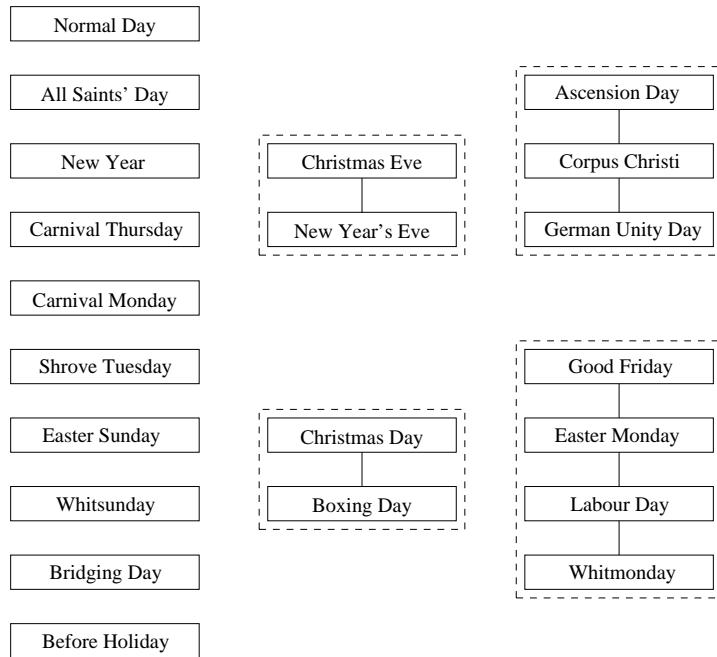
**Table 6.3.** Topological data of several randomly chosen exemplary loop detectors. The following investigations are done for all these detectors to get an impression about the different spatial dependencies. Shown is the particular motorway on that the detector is placed, the driving direction, the number of lanes  $N_1$  at the particular place of the particular motorway in the particular direction, and the lane, where the detector is placed.

Thus, for the weekdays an automatic classification algorithm is developed. This can be applied for arbitrary loop detectors. With the distance limit  $d_{\text{lim}}$  the tolerance can be adjusted. But only for daily differences according to the weekdays, the classification would be unnecessary. In this case, just the seven weekdays can be taken as suitable classification for all loop detectors. But as shown in Chapter 5, there are more differences in traffic time series than those that come because of weekdays. Therefore even other attributes should be considered, for instance, special days.

To investigate this, as example 20 days of a year are marked with special attributes. In most cases such attributes are holidays. But also attributes according to special regional habits make sense. In the Rhine region such a habit is, for instance, Carnival, with the days “Carnival Thursday”, “Carnival Monday”, or the “Shrove Tuesday”. Another crucial attribute is a bridging day, that means a day between a holiday and weekend. Also days before holidays behave in a different manner as, for instance, shown in [Chrobok, 2000]. Together with the normal days there are 21 different days so according to the  $m + 1$ -rule a 22-medoid algo-

loop detector	$N_{D_n}$	$D_n$
4135-015-033	4	Su, (Mo,Tu,We,Th), Fr, Sa
4214-015-002	5	Su, Mo, (Tu,We,Th), Fr, Sa
4215-006-035	5	Su, Mo, (Tu,We,Th), Fr, Sa
4349-025-001	4	Su, (Mo,Tu,We,Th), Fr, Sa
4488-096-033	4	Su, (Mo,Tu,We,Th), Fr, Sa
5877-019-001	4	Su, (Mo,Tu,We,Th), Fr, Sa
5877-019-003	5	Su, Mo, (Tu,We,Th), Fr, Sa
6211-000-033	4	Su, (Mo,Tu,We,Th), Fr, Sa
6250-032-001	5	Su, Mo, (Tu,We,Th), Fr, Sa
6289-007-034	4	Su, (Mo,Tu,We,Th), Fr, Sa
6256-009-002	5	Su, (Mo,Tu), (We,Th), Fr, Sa

**Table 6.4.** Classifications of the weekdays for the exemplary loop detectors of Table 6.3. As for the results of Table 6.2 the DCA is used with  $k = 8$  clusters and a distance limit for the belonging vectors of  $d_{\text{lim}} = 0.6$  whereby the Euclidean distance of the normalized belonging vectors are used.



**Fig. 6.2.** Clustering of traffic data of the loop detector 5877-019-001 shows that some special days behave in a similar manner. These are connected with a line and are framed with a dashed rectangle. The kind the days are grouped is comprehensible. It is quite remarkable that the algorithm proposes this kind of interpretation automatically without any other information or manually interpretation. Applied is the DCA with a 22-medoid cluster algorithm because of the  $m + 1$ -rule (with the normal days 21 different attributes are investigated).

rithm is applied for the DCA. If again a distance limit of  $d_{\lim} = 0.6$  according to the Euclidean distance is applied, the clustered attributes for the loop detector 5877-019-001 lead to interesting results that can be seen in Fig. 6.2.

The results are quite comprehensible. The similar characteristics of Christmas Eve and New Year's Eve is automatically detected. Both days are for many people half-day working days. In the same manner the similarity of Christmas Day and Boxing Day is detected. Note, that Christmas Day and Boxing Day are in Germany very similar holidays (first and second Christmas Day). Ascension Day and Corpus Christi are in one group. Both days are holidays at a Thursday and this situation always leads to a long weekend. A similar situation holds for the holidays of Easter Monday and Whitmonday.

The manner the special days are grouped in detail is not surprising. But note, that the results differ for each particular detector location. Also important for such an application is the fact, that each day appears only one time a year, so only three or four days have been available for this analysis. The results in Fig. 6.2 show, that the algorithm also works with other attributes than the day of week. In a similar manner different attributes can be observed with the DCA, whether they influence the traffic state or not. In Table 6.5 the monthly classification can be seen for all the loop detectors that are also investigated in the example above in regard to weekdays. The attributes are denoted from 1  $\hat{=}$  January to 12  $\hat{=}$  December. Then, according to the  $m + 1$ -rule a 13-medoid cluster algorithm is applied, for each month the belonging vectors are calculated and normalized according to the Euclidean distance and compared with a distance limit of  $d_{\lim} = 0.6$ .

loop detector	$N_{D_n}$	$D_n$
4135-015-033	7	(1,2), (3,4,5,9), (6,10), 7, 8, 11, 12
4214-015-002	3	(1,2), (3,11,12), (4,5,6,7,8,9,10)
4215-006-035	4	(1,2), (3,4,5,6,11), (7,8,9,10), 12
4349-025-001	6	1, (2,11), (3,10), (4,5,6,7,9), 8, 12
4488-096-033	7	1, 2, (3,4), (5,9,10), (6,7), 8, (11,12)
5877-019-001	6	1, 2, (3,11), (4,5,6,9,10), (7,8), 12
5877-019-003	7	(1,2), 3, (5,6), (4,7,9), 8, (10,11), 12
6211-000-033	4	(1,2,3,4,5), (6,7,9,10), 8, (11,12)
6250-032-001	4	(1,2), (3,10,11), (4,5,6,7,8,9), 12
6289-007-034	6	(1,4,10), (2,3,5,9), (6,7), 8, 11, 12
6256-009-002	7	1, (2,3,4,6), 5, (7,8), 9, 10, (11,12)

**Table 6.5.** Classifications  $D_n$  of the months using the DCA for several exemplary loop detectors. The attributes are denoted from 1  $\hat{=}$  January to 12  $\hat{=}$  December.

As shown in Sec. 5.2 many seasonal differences come because of school holidays and the effects differ depending on the motorway. This dependency can also be seen in the monthly analysis in Table 6.5. But school holidays can also directly be analysed. Therefore the attributes of the school holidays can be assigned to

the days. In North Rhine-Westphalia there are four different school holidays in the period the analysis is made: *Easter*, *Summer*, *Autumn*, and *Winter*. Together with the possibility of no school holidays and according to the  $m+1$ -rule, the time series are clustered with the DCA in  $k = 6$  groups and interpreted as mentioned above. The results can be seen in Table 6.6.

loop detector	$N_{D_n}$	$D_n$
4135-015-033	2	(Noholidays,Easter,Summer,Autumn), Winter
4214-015-002	2	(Noholidays,Easter,Summer,Autumn), Winter
4215-006-035	4	(Noholidays,Summer), Easter, Autumn, Winter
4349-025-001	3	(Noholidays,Autumn), (Easter,Summer), Winter
4488-096-033	3	(Noholidays,Autumn,Easter), Summer, Winter
5877-019-001	3	Noholidays, (Easter,Summer,Autumn), Winter
5877-019-003	3	(Noholidays,Autumn), (Easter,Summer), Winter
6211-000-033	3	(Noholidays,Easter,Summer), Autumn, Winter
6250-032-001	2	(Noholidays,Easter,Summer,Autumn), Winter
6289-007-034	3	Noholidays, (Easter,Summer,Autumn), Winter
6256-009-002	2	(Noholidays,Easter,Summer,Autumn), Winter

**Table 6.6.** Similar Effects of different school holidays on the traffic flow time series of several loop detectors. In all cases there is a basic difference between days with and without holidays in Winter.

Now a closer look should be taken at the rule of the parameters that have to be chosen: the distance limit  $d_{\lim}$  and the number of clusters  $k$ . Therefore it must be understood what exactly happens during both cluster algorithms. The traffic time series are grouped in  $k$  clusters. In case of  $k = 2$  in each cluster are traffic time series that are most possible different. These are in most cases that of working days and of weekends. Consider, that this is consequential the case. Then all time series of working days are in one cluster resulting in the normalized belonging vectors  $\vec{b}_{\text{working}} = (1, 0)$  and all time series of weekends in the other one resulting in  $\vec{b}_{\text{weekend}} = (0, 1)$ . So, the Euclidean distance between, for instance, Mondays and Sundays, would be  $\|\vec{b}_{\text{working}} - \vec{b}_{\text{weekend}}\|_2 = \sqrt{2}$  and that between Sundays and Saturdays  $\|\vec{b}_{\text{weekend}} - \vec{b}_{\text{weekend}}\|_2 = 0$ .

But because of holidays, that can also be on weekdays which are normally working days, this is not always the case. At Good Friday, for instance, there is little traffic demand and this day is without a doubt in the cluster of the weekend. Because of this, it holds in general  $\|\vec{b}_{\text{working}} - \vec{b}_{\text{weekend}}\|_2 < \sqrt{2}$ . This is the only reason for the choice of the distance limit  $d_{\lim} < \sqrt{2}$ .

With the number of clusters  $k$  it can be adjusted, how many dimensions the belonging vector has. In opposite to the distance limit, with the parameter  $k$  it can be basically adjusted, how many different time series are distinguished. Consider the example above and the case that now all time series are grouped in  $k = 3$  clusters. If now the working days stay the same, but the Sundays and

the Saturdays are grouped in separate clusters resulting in the belonging vectors  $\vec{b}_{\text{Sa}} = (0, 1, 0)$  and  $\vec{b}_{\text{Su}} = (0, 0, 1)$  the Euclidean distance increases directly from zero to  $\sqrt{2}$ . With the right choice of the parameter  $k$  also classifications of special events can be considered as it is shown in the next example.

As discussed in Sec. 5.4 an often considered case is that of sporting events where there is a lot of additional traffic supposed at certain sections. Now, for days when such events are expected an own classified traffic time series can be calculated. But if this is done for all events in a large scale network this leads directly to the case of unique days. So what can be done?

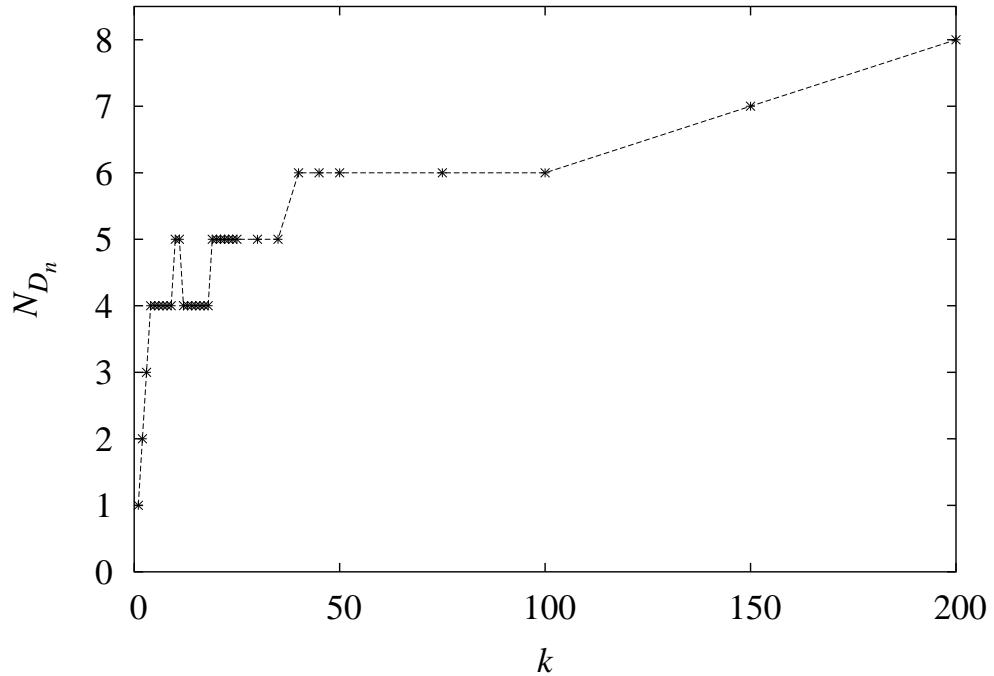
Consider the case of a football match at a certain city at Saturday at 15:30, what is in Germany a usual kick off time. In the state of North Rhine-Westphalia there are at 2004/07/31 six football clubs in the first devision and each particular match only influences a certain region. For the next example the loop detectors 5877-019-001 on the right lane of the BAB 40 driving direction east and 5924-013-001 on the right lane of the BAB 42 driving direction east are considered. The first one is also known from the classification above and is far away from any football stadium. The second one is several hundred meters upstream from the junction *Gelsenkirchen Schalke*, which is the off-ramp that leads directly to a football stadium. Now, the days are marked with the weekday attributes, but the Saturdays are divided in such with (Fb) and without (Sa) a football match. The classification for several values of  $k$  for both loop detectors can be seen in Table 6.7.

$k$	5877-019-001	5924-013-001
8	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)
9	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)
10	Su, (Mo,Tu,We), Th, Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)
11	Su, (Mo,Tu,We), Th, Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)
12	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, Sa, Fb
13	Su, (Mo,Tu,We,Th), Fr, (Sa,Fb)	Su, (Mo,Tu,We,Th), Fr, Sa, Fb

**Table 6.7.** Classifications  $D_n$  of the weekdays with different numbers of clusters  $k$ . With an adequate number of clusters, also special events can be classified. The Saturdays are distinguished in days without (Sa) and with (Fb) a football match. Whereby the loop detector 5877-019-001, which is far away from the stadium, is not affected, the traffic time series of loop detector 5924-013-001 form different clusters of Saturdays for  $k > 11$ .

So the cluster algorithm can also help to distinguish special events. Nevertheless the adequate choice of  $k$  is a very complex problem. As can be seen in Table 6.7 from  $k = 11$  to  $k = 12$  in case of loop detector 5877-019-001, increasing of  $k$  can also lead to a decreasing number of classes. This is in general not the case.

In Fig. 6.3 the classification according to weekdays and the Saturday with football can be seen for the loop detector 5877-019-001 for several values of  $k$ . If the number of clusters  $k$  is exactly the number of investigated daily traffic time series



**Fig. 6.3.** Dependency of the number of classes  $N_{D_n}$  according to 8 attributes (the 7 weekdays and Saturdays with Football) depending on the number of clusters  $k$ . The limited case is that each daily traffic time series form a separate cluster, what means that the classification does not make sense. This can be seen in the distinction in normal Saturdays and Saturdays with Football, although this loop detector is placed where there is no influence in regard to traffic because of football.

$k = N_{\text{days}}$ , each time series form a separate cluster, and an interpretation of the clusters and so forecasting using a cause and effect strategy is impossible. This can be seen in the number of classes that is  $N_{D_n} = 8$ , although the place with the detector is not influenced by the football match.

Thus, it should always hold  $k \ll N_{\text{days}}$ . In our example hold  $N_{\text{days}} = 1307$  so  $N_{\text{days}}$  is of an order of magnitude  $N_{\text{days}} \approx 10^3$ . Thus, for  $k$  should not exceed an order of  $k \approx 10^1$  and the example  $k = 22$  of the special days is already a high number for the choice of  $k$ .

After an adequate number of clusters  $k$  is chosen, the question comes up straight forward, how the attributes should be chosen and how the different classifications can be connected. For this, a few considerations have to be made about the attributes that make sense.

The classification algorithm should use the principle of cause and effect as already discussed and should not be depending on special conditions that means on the particular social environment that lead to the traffic demand. And in fact, the

method can be applied on different national or religious holidays, different school holidays, or different seasonal fluctuations. Furthermore, even with a different calendar the algorithm works in the same manner. That is necessary, because the attributes can also be certain regional or even local events that lead to different traffic demand.

Nevertheless, this algorithm has always to be fed with information from outside and some basic rules have to be fulfilled. One important task is to avoid random correlations in the interpretation of the attributes. Therefore, the  $k$ -medoid cluster algorithm should be applied on attributes in a special manner. It is important, that each day is assigned to one attribute. Furthermore there should be no day that is assigned to two attributes. It can be said, that the attributes should be pairwise disjunctive and all days are elements of the set union of all attributes. This basic rule for the attributes is in the analysis above already considered. The seven weekdays are disjunctive as well as the months, and the school holidays. Also the special days are disjunctive for the time, the data is investigated here (for completeness, it has to be mentioned that in the year 2008 Easter is so early, that the Ascension Day is on the first of May which is also the Labour Day in Germany). To be precise, for the culture in Germany one has to distinguish the special days in days that are fix in regard to the calendar (for instance, New Year, Labour Day, or Christmas Day) and those that are fix in regard to Easter (for instance, Carnival Monday, Easter Monday, or Corpus Christi). Another case that can happen is that the bridging day is also a day before a holiday.

day $d$	$\vec{\mathcal{C}}(d)$
2004/02/23	[Mo, Noholidays, CM]
2004/04/09	[Fr, Easter, (GF,EM,LD,WM)]
2004/05/19	[(Tu,We), Noholidays, BH]
2004/05/21	[Fr, Noholidays, BD]
2004/08/04	[(Tu,We), Summer, ND]
2004/10/03	[Su, Noholidays, [AD,CC,GU]]

**Table 6.8.** A few examples of classification vectors  $\vec{\mathcal{C}}(d)$  for several days  $d$  when the traffic time series of the loop detector 5877-019-001 are classified in  $N_{\mathcal{A}} = 3$  attribute groups  $\mathcal{A}_1 \cong$  Weekdays,  $\mathcal{A}_2 \cong$  SchoolHolidays, and  $\mathcal{A}_3 \cong$  SpecialDays. According to the special days abbreviations are used for Carnival Monday (CM), Good Friday (GF), Easter Monday (EM), Labour Day (LD), Whitmonday (WM), Before Holiday (BH), Bridging Day (BD), Normal Day (ND), Ascension Day (AD), Corpus Christi (CC), and German Unity Day (GU).

Knowing this, the complete classification algorithm can be defined. All attributes  $a_i$  that are considered to influence the traffic have to be arranged in  $N_{\mathcal{A}}$  attribute groups  $\mathcal{A}_j$  ( $j = 1, 2, \dots, N_{\mathcal{A}}$ ). Each group  $\mathcal{A}_j$  has to consist of  $m_j$  disjunctive elements  $a_{j,l} \in \mathcal{A}_j$  ( $l = 1, 2, \dots, m_j$ ) in that sense, that no day  $d$  can be assigned to two attributes of this group. Furthermore all  $N_{\text{days}}$  considered days  $d$  can be assigned to one attribute of this group.

loop detector	$N_{D_n}$	$D_n$
4135-015-033	5	Su, Mo, (Tu,We,Th), Fr, Sa
4214-015-002	6	Su, Mo, Tu, (We,Th), Fr, Sa
4215-006-035	6	Su, Mo, (Tu,We), Fr, Sa
4349-025-001	6	Su, Mo, (Tu,We), Th, Fr, Sa
4488-096-033	5	Su, Mo, (Tu,We,Th), Fr, Sa
5877-019-001	5	Su, (Mo,Th), (Tu,We), Fr, Sa
5877-019-003	5	Su, (Mo,Tu), (We,Th), Fr, Sa
6211-000-033	5	Su, Mo, (Tu,We,Th), Fr, Sa
6250-032-001	6	Su, Mo, Tu, (We,Th), Fr, Sa
6289-007-034	5	Su, (Mo,Tu,We), Th, Fr, Sa
6256-009-002	5	Su, (Mo,Tu), (We,Th), Fr, Sa

**Table 6.9.** Classifications of the weekdays for the exemplary loop detectors of Table 6.3 using the DCA with  $k = 22$  as well as a distance limit of  $d_{\text{lim}} = 0.6$  in regard to the Euclidean distance between belonging vectors, that are normalized in regard to the Euclidean distance.

loop detector	$N_{D_n}$	$D_n$
4135-015-033	5	Noholidays, Easter, Summer, Autumn, Winter
4214-015-002	4	Noholidays, (Easter,Summer), Autumn, Winter
4215-006-035	5	Noholidays, Easter, Summer, Autumn, Winter
4349-025-001	4	Noholidays, (Easter,Summer), Autumn, Winter
4488-096-033	4	(Noholidays,Autumn), Easter, Summer, Winter
5877-019-001	5	Noholidays, Easter, Summer, Autumn, Winter
5877-019-003	5	Noholidays, Easter, Summer, Autumn, Winter
6211-000-033	5	Noholidays, Easter, Summer, Autumn, Winter
6250-032-001	3	(Noholidays,Easter,Summer), Autumn, Winter
6289-007-034	5	Noholidays, Easter, Summer, Autumn, Winter
6256-009-002	4	(Noholidays,Summer), Easter, Autumn, Winter

**Table 6.10.** Classifications of the school holidays for the exemplary loop detectors of Table 6.3 using the DCA with  $k = 22$  as well as a distance limit of  $d_{\text{lim}} = 0.6$  in regard to the Euclidean distance between belonging vectors, that are normalized in regard to the Euclidean distance.

As already mentioned, examples for such groups are the attributes discussed above, for instance, the seven weekdays from  $a_{j_1,1} = \text{Su}$  to  $a_{j_1,7} = \text{Sa}$ , the holidays from  $a_{j_2,1} = \text{Noholidays}$  to  $a_{j_2,5} = \text{Winter}$ , or even the special days from  $a_{j_3,1} = \text{NormalDay}$  to  $a_{j_3,21} = \text{GermanUnityDay}$ .

Then, the DCA is applied knowing the boundary conditions that an increasing  $k$  in the  $k$ -medoid cluster algorithm leads to a refinement, whereas it should hold  $k \ll N_{\text{days}}$ . For each group  $\mathcal{A}_j$  the elements  $a_{j,l}$  are clustered with the DCA as mentioned above to form  $N_{\mathcal{C}_{j,D_{j,n}}}$  groups  $\mathcal{C}_{j,D_{j,n}}$  of crucial attributes or classes  $D_{j,n}$ . In the same manner, the groups  $\mathcal{G}_{j,D_{j,n}}$  can be defined consisting of the days  $d \in \mathcal{G}_{j,D_{j,n}}$  that are assigned to the particular class  $D_{j,n}$ . Now, each of the  $N_{\text{days}}$  days  $d$  can be assigned to its  $N_{\mathcal{C}_{j,D_{j,n}}} = N_{\mathcal{A}}$  dimensional classification vector  $\vec{\mathcal{C}}(d)$  consisting of the classes  $D_{j,n}$ . Finally, two days  $d_1$  and  $d_2$  with the same classification vector  $\vec{\mathcal{C}}(d_1) = \vec{\mathcal{C}}(d_2)$  can be assigned to the same group  $\mathcal{G}_{\vec{\mathcal{C}}(d_1)}$ .

In Table 6.8 are a few examples for several days  $d$  of the loop detector 5877-019-001 with their classification vectors  $\vec{\mathcal{C}}(d)$  when the traffic time series are grouped in  $k = 22$  clusters. The result of the DCA according to weekdays are the five classes Su, (Mo,Th), (Tu,We), Fr, and Sa. The school holidays are distinguished in a separate class each, and the special days according to Fig. 6.2.

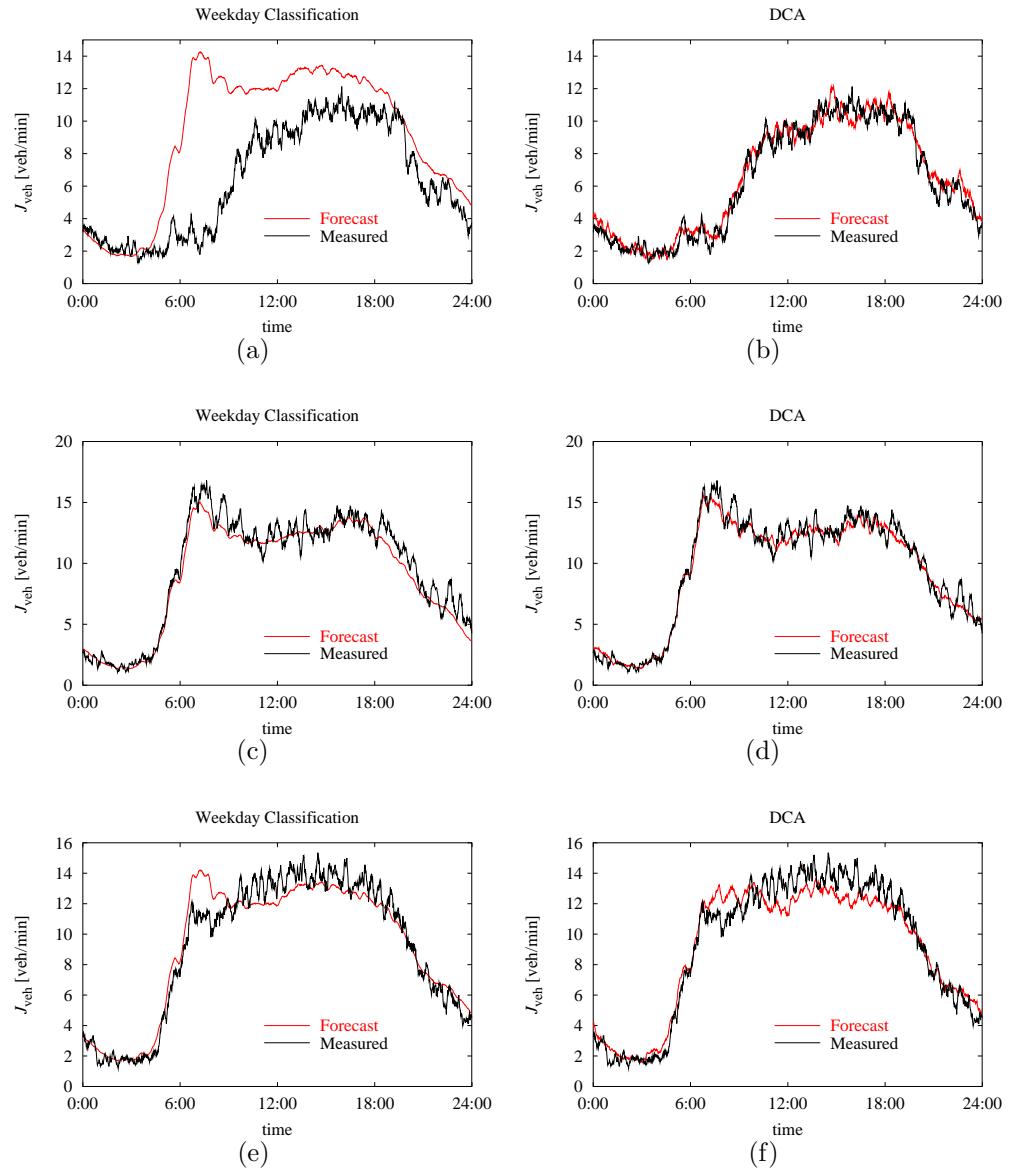
With the classification, the knowledge about the causes that have an effect on the traffic is used. After the classification the question comes up, how the forecast can be calculated from the classified days. This is considered in the following section. Because a classification is needed for all further calculations and for traceability, the result of the DCA is calculated in the following for all the loop detectors of Table 6.3. Therefore, a  $k = 22$ -medoid cluster algorithm is used as well as a distance limit of  $d_{\text{lim}} = 0.6$  in regard to the Euclidean distance between belonging vectors, that are normalized in regard to the Euclidean distance.

The results in regard to weekdays can be seen in Table 6.9 and in regard to school holidays in Table 6.10.

## 6.2 Calculating the Forecast

After an appropriate classification of the traffic time series is calculated using any kind of knowledge, the question is, how it can be used to calculate the forecast data. This problem is close to the question, how many data of a day with the same classification vector are available. Although it can be avoided, that each day is assigned to a separate cluster, it maybe the case, that formulating the classification vector for the day  $d_0$  leads to the case, that the vector  $\vec{\mathcal{C}}(d_0)$  does not exist for former days  $d_{\text{former}}$  when traffic data are available.

In this case a general hierarchical disposition of the attribute groups helps to achieve several days that can be considered for the forecast. For instance, in case of the 2004/01/01 what has the classification vector  $\vec{\mathcal{C}}(2004/01/01) = [(\text{Mo,Th}), \text{Winter}, \text{New Year}]$  it may be feasible to just look for other New Year's.



**Fig. 6.4.** Calculating the forecast traffic time series of traffic flow  $J_{\text{veh}}$  with Eq. 5.14 with a classification according to the weekdays (left) and according to the DCA classification algorithm proposed in Sec. 6.1 (right). The crucial attribute “Good Friday” is automatically separated and lead to a large difference of the forecast traffic time series at the 2004/04/09 (b) in regard to a normal classification according to weekdays in (a). But also the small differences of the attribute “Before Holiday” lead to an enhancement in (d) in regard to (c) at the 2004/05/19, as well as the attribute “Bridging Day” (f) in regard to (e) at the 2004/05/21. For a better overview not the raw data but the 15 min moving averages are shown.

The general case is, that  $N_{\text{class}}$  days of a certain classification vector are available. One possibility to achieve the forecast time series is to use the already in Chapter 5 proposed Eq. 5.14. Using this algorithm already the improvements of the enhanced classifications can be clearly seen in the 15 min moving averages of the forecast and measured traffic flow  $J_{\text{veh}}$  of selected days of the loop detector 5877-019-001 in Fig. 6.4.

Most improvements are obtained, because the cluster algorithm separates the special days from the normal weekdays. But to use the algorithm in operation it must be shown, that the results do not get worse in periods, when other classifications, like that according to the weekdays, work quite well. Thus, for all days from 2004/08/01 to 2004/09/30, when there are no special days, the minute aggregated traffic flow is forecast for several loop detectors and the performance is tested in comparison to a classification according to weekdays. In Table 6.11 the forecast results for a weekday classification are tested with several quality measures. The data are the arithmetic mean of the particular quality measure of the number of days  $N_{\text{days}}$ , when appropriate data are available. In Table 6.12 this is done with the DCA algorithm. It can clearly be seen, that the new algorithm provides even for these days improvements for most of the loop detectors.

loop detector	$\bar{e}_{\text{MAE}}$ [veh/min]	$\bar{e}_{\text{RMSE}}$ [veh/min]	$\bar{e}_{\text{RMSEP}}$	$\bar{e}_{\text{ME}}$ [veh/min]	$N_{\text{days}}$
4135-015-033	2.906	3.985	0.269	0.689	53
4214-015-002	2.594	3.507	0.296	0.654	34
4215-006-035	2.732	3.990	0.552	0.665	35
4349-025-001	2.757	3.647	0.288	1.292	61
4488-096-033	2.334	3.034	0.295	0.804	35
5877-019-001	1.828	2.348	0.273	0.451	58
5877-019-003	1.374	2.047	0.792	0.325	44
6211-000-033	2.101	3.157	0.535	0.953	61
6250-032-001	1.906	2.469	0.254	0.362	61
6289-007-034	3.933	5.515	0.420	0.796	4
6256-009-002	3.244	4.650	0.441	0.894	61

**Table 6.11.** Forecast results with a classification according to the seven weekdays. The quality measure data are the arithmetic mean calculated over  $N_{\text{days}}$  days for: the mean absolute error  $\bar{e}_{\text{MAE}}$ , the root mean square error  $\bar{e}_{\text{RMSE}}$ , the root mean square error proportional  $\bar{e}_{\text{RMSEP}}$ , and the mean error  $\bar{e}_{\text{ME}}$ . Chosen are those days, when appropriate data are available from 2004/08/01 to 2004/09/30. The measures are calculated with the measured data and the classified traffic flow of the particular weekday calculated with Eq. 5.14.

In some cases the cluster classification does not outperform the weekday classification for all quality measures. This can be, because random fluctuations have a larger influence on the more detailed classification of the DCA. It calculates the forecast with time series from a smaller number of days  $\sum_{d \in \mathcal{G}_{\mathcal{C}}} 1$  with that

loop detector	$\overline{e}_{\text{MAE}}$ [veh/min]	$\overline{e}_{\text{RMSE}}$ [veh/min]	$\overline{e}_{\text{RMSEP}}$	$\overline{e}_{\text{ME}}$ [veh/min]	$N_{\text{days}}$
4135-015-033	2.880	3.968	0.260	0.499	53
4214-015-002	2.586	3.465	0.282	0.468	34
4215-006-035	2.748	3.999	0.538	0.604	35
4349-025-001	2.691	3.565	0.278	1.119	61
4488-096-033	2.316	3.003	0.284	0.555	35
5877-019-001	1.813	2.326	0.274	0.342	58
5877-019-003	1.333	2.004	0.824	0.182	44
6211-000-033	2.023	3.026	0.532	0.737	61
6250-032-001	1.899	2.458	0.251	0.317	61
6289-007-034	3.996	5.595	0.430	0.719	4
6256-009-002	3.267	4.678	0.438	0.969	61

**Table 6.12.** Forecast results with a classification according to the DCA. The quality measure data are the arithmetic mean calculated over  $N_{\text{days}}$  days for: the mean absolute error  $\overline{e}_{\text{MAE}}$ , the root mean square error  $\overline{e}_{\text{RMSE}}$ , the root mean square error proportional  $\overline{e}_{\text{RMSEP}}$ , and the mean error  $\overline{e}_{\text{ME}}$ . Chosen are those days, when appropriate data are available from 2004/08/01 to 2004/09/30. The measures are calculated with measured and classified traffic flow, whereby Eq. 5.14 is used and the classification is obtained with the DCA. Even for the period from 2004/08/01 to 2004/09/30, when there are no special days, the cluster classification provides an improvement for most of the loop detectors.

loop detector	$\overline{e}_{\text{MAE}}$ [veh/min]	$\overline{e}_{\text{RMSE}}$ [veh/min]	$\overline{e}_{\text{MRE}}$	$\overline{e}_{\text{ME}}$ [veh/min]	$N_{\text{days}}$
4135-015-033	2.846	3.925	0.258	0.498	53
4214-015-002	2.547	3.417	0.278	0.469	34
4215-006-035	2.712	3.944	0.531	0.604	35
4349-025-001	2.664	3.532	0.275	1.118	61
4488-096-033	2.278	2.955	0.279	0.554	35
5877-019-001	1.797	2.305	0.271	0.342	58
5877-019-003	1.324	1.988	0.817	0.182	44
6211-000-033	1.999	2.982	0.523	0.737	61
6250-032-001	1.893	2.450	0.250	0.317	61
6289-007-034	3.980	5.553	0.427	0.719	4
6256-009-002	3.257	4.663	0.436	0.968	61

**Table 6.13.** Forecast results of the CMS forecast method. An improvement of the forecast is obtained, if the classified traffic time series is smoothed, in this case with a 15 minute moving average. Apart from this, the classified traffic flow is obtained in the same manner like in Table 6.12.

the classified traffic data is calculated in Eq. 5.14. As already discussed, these fluctuations are very large in comparison to the mean and unpredictable.

To cope with this problem the classified traffic time series  $\bar{x}_k(\mathcal{G}_{\vec{C}}, t)$  should be smoothed over the time steps  $t$  before used for forecasting. This smoothing is the more important the smaller the difference of adjacent time steps (in our case one minute) and the lower the number of days  $\sum_{d \in \mathcal{G}_{\vec{C}}} 1$  in the particular group  $\mathcal{G}_{\vec{C}}$ . The

smoothing can be done with, for instance, an ARIMA model.

As discussed, which model and which parameters work best strongly depends on the particular kind of data. At this point it is only important, that the fluctuations from minute to minute are avoided. Thus, in Table 6.13 the results can be seen, if the classified traffic flow  $\bar{J}_{\text{veh}}(\mathcal{G}_{\vec{C}}, t)$  is smoothed with a 15 minute moving average over adjacent minutes. This combination of the classification with the DCA, calculating the forecast with Eq. 5.14 and smoothing the time series over adjacent minutes with Eq. 3.13 is called in the following *classified mean smoothed* (CMS) forecast method.

With this smoothing, the weekday classification is outperformed by the DCA classification for nearly all quality measures, except for two loop detectors. The results for the detector 6289-007-034 have to be handled with care, because feasible data is only available for four days. The problem with the loop detector 6256-009-002 is, that there is a basic change in the traffic data. How this problem can be solved is discussed in the following.

To consider basic changes in traffic demand as discussed in Sec. 5.4 on page 112 the model must be adaptive. Therefore Eq. 5.14 is very inertial, because every new data set is weighted in the same manner like the old ones. This can be changed in different manners with parametric regression methods, as already proposed by Janko [1994].

Then, as index  $i$  hold the chronological numbering of the days  $d_i$  that belong to the particular group  $d_i \in \mathcal{G}_{\vec{C}}$ . With that, in general all models that are introduced in Sec. 3.1 can be used. But as already stated in that section, often low order ARIMA models produce feasible results. When, for instance, a moving average approach over the last  $N_{\text{MOV}}$  days of the particular class should be used, Eq. 5.14 can be changed in regard to Eq. 3.13 to:

$$\bar{x}(\mathcal{G}_{\vec{C}}, N_{\text{MOV}}, t) = \frac{\sum_{i=N_{\mathcal{G}_{\vec{C}}}-N_{\text{MOV}}+1}^{N_{\mathcal{G}_{\vec{C}}}} x(d_i, t)}{N_{\text{MOV}}}. \quad (6.22)$$

Here  $N_{\mathcal{G}_{\vec{C}}} = \sum_{d \in \mathcal{G}_{\vec{C}}} 1$  is the number of days  $d_i$  ( $i = 1, 2, \dots, N_{\mathcal{G}_{\vec{C}}}$ ).

For each time step  $t$  the moving average of the traffic data  $\bar{x}(\mathcal{G}_{\vec{C}}, N_{\text{MOV}}, t)$  is calculated separately over the last  $N_{\text{MOV}}$  days. It is obvious that  $N_{\text{MOV}} \leq N_{\mathcal{G}_{\vec{C}}}$  must always hold.

In a similar manner, an exponential smoothing model can be applied:

$$\begin{aligned}\bar{x}(\mathcal{G}_{\vec{C}}, \alpha, 1, t) &= x(d_1, t), && \text{for the first day } d_1, \\ \bar{x}(\mathcal{G}_{\vec{C}}, \alpha, i, t) &= \alpha x(d_i, t) + (1 - \alpha) \bar{x}(\mathcal{G}_{\vec{C}}, \alpha, i - 1, t), && \text{for each other.}\end{aligned}\quad (6.23)$$

Note that for such adaptive procedures it is even more important to smooth the resulting classified time series  $\bar{x}(t)$  over adjacent minutes, because the averaging over a large number of days is missing. The adaptivity of the forecast is at the expense of the smoothing of the fluctuations. Thus, when the methods are applied in the following investigations, the resulting classified time series  $\bar{x}(t)$  are smoothed with 15 minute moving averages. Furthermore, the combination of the classification with the DCA, calculating the forecast with Eq. 6.23, and smoothing the time series over adjacent minutes with Eq. 3.13 is called in the following *classified adaptive smoothed* (CAS) forecast method.

To show, how both adaptive procedure work, in the following the example of the loop detector 5993-000-137, which is placed on the turning lane to the left from the motorway BAB 57 to the motorway BAB 44 in direction to the new bridge opened at the 2002/05/31 as discussed in Sec. 5.4 on page 112. The DCA algorithm divides the weekdays in four classes, the school holidays in four classes, and all special days in exactly ten classes. In the following example the traffic time series for all days  $d_i$  with the classification vectors

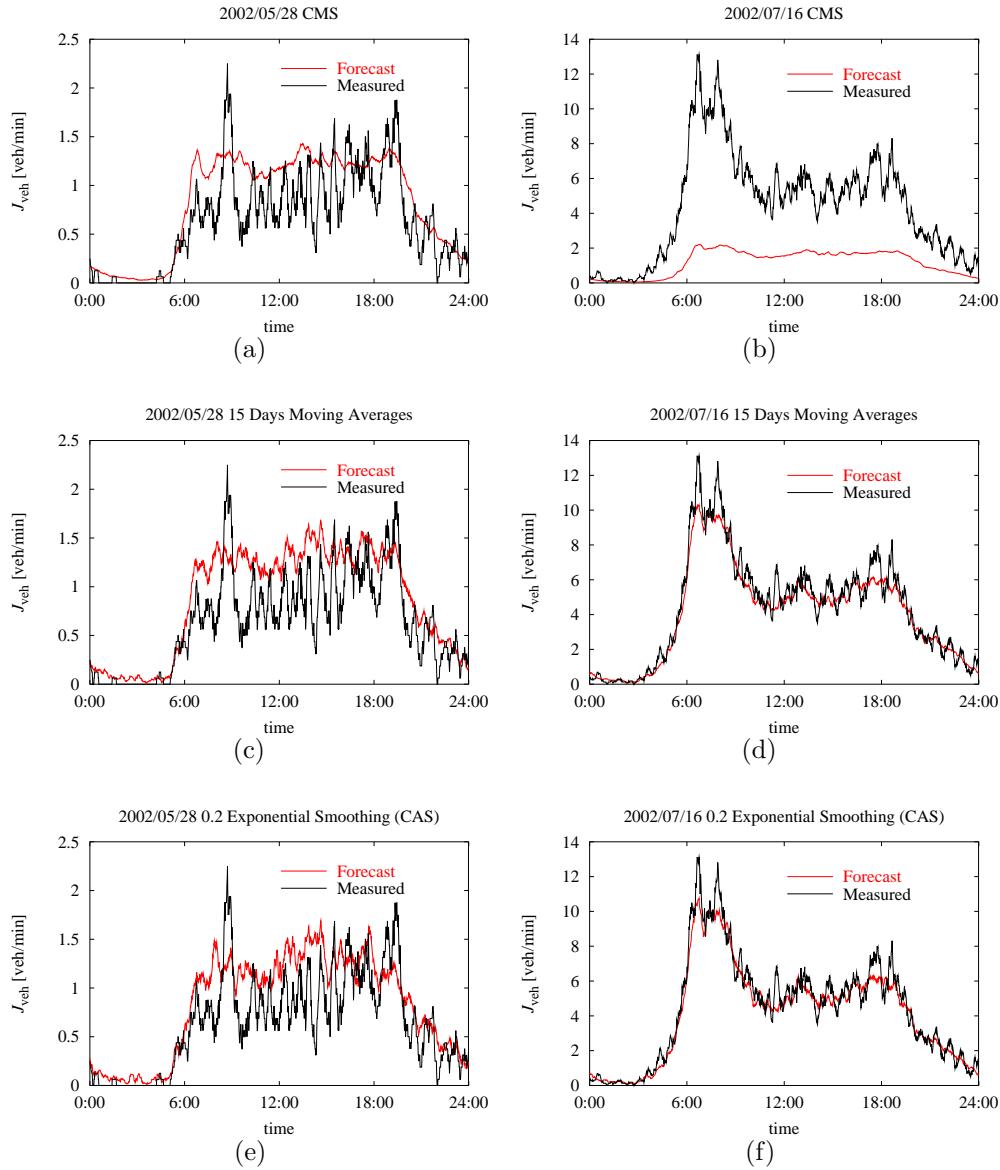
$$\vec{C}(d_i) = [(Mo, Tu, We, Th), (No\text{holidays}, Easter), \text{Normal Day}] \quad (6.24)$$

are considered. That means all normal weekdays from Monday to Thursday that lie outside school holidays or within Easter holidays.

The improvements of the adaptive procedures in regard to the CMS can be clearly seen in Fig. 6.5. On the left there are shown the data of Tuesday 2002/05/28, that means before the bridge is opened, on the right the data of Tuesday 2002/07/16 after the opening. Shown are the measured traffic flow  $J_{\text{veh}}$  in comparison to the forecast traffic flow time series  $\bar{J}_{\text{veh}}(\mathcal{G}_{\vec{C}})$  according to the classification vector of Eq. 6.24. The classified time series is then calculated as CMS (a) and (b), as the 15 days moving average (Eq. 6.22,  $N_{\text{MOV}} = 15$ , thereafter the resulting time series is smoothed with 15 min moving averages over adjacent 15 minutes) in (c) and (d), and as the CAS in (e) and (f).

It can clearly be seen, that the CMS in Fig. 6.5 (b) is outperformed by the adaptive procedures in Fig. 6.5 (d) and (f). Note, that as in Fig. 6.4 for a better overview also for the measured data not the raw data are shown, but the 15 min moving averages.

The general forecast results of the CAS with  $\alpha = 0.2$  can be seen in Table 6.14 for several loop detectors. Especially for the loop detector 6256-009-002 an improvement is obtained as mentioned above. Note, that in some cases the forecast results are negatively influenced by the adaptivity, but apart from loop detector 4215-006-035 the results are even better than with the weekday classification in



**Fig. 6.5.** The adaptive forecasts clearly outperform the classified mean (CMS) in case of basic changes in the traffic data. This has been the case at the loop detector 5993-000-137 because of a new bridge (cp. Sec. 5.4 on page 112). It can be seen the measured and the forecast traffic flow  $J_{\text{veh}}$ . On the left is the data for Tuesday 2002/05/28, before the bridge is opened. On the right is the same for Tuesday 2002/07/16 after the opening. The forecast time series is calculated as CMS (a) and (b), as the 15 days moving average (Eq. 6.22,  $N_{\text{MOV}} = 15$ , thereafter the resulting time series is smoothed with 15 min moving averages over adjacent 15 minutes) in (c) and (d), and as the CAS in (e) and (f). Note, that as in Fig. 6.4 for a better overview also for the measured data not the raw data are shown, but the 15 min moving averages over the time of the day.

loop detector	$\bar{e}_{\text{MAE}}$ [veh/min]	$\bar{e}_{\text{RMSE}}$ [veh/min]	$\bar{e}_{\text{RMSEP}}$	$\bar{e}_{\text{ME}}$ [veh/min]	$N_{\text{days}}$
4135-015-033	2.901	3.947	0.262	0.469	53
4214-015-002	2.536	3.390	0.268	0.401	34
4215-006-035	2.767	4.019	0.532	0.818	35
4349-025-001	2.570	3.405	0.253	0.552	61
4488-096-033	2.288	2.966	0.277	0.538	35
5877-019-001	1.824	2.343	0.273	0.442	58
5877-019-003	1.325	1.995	0.839	0.205	44
6211-000-033	1.902	2.880	0.550	0.396	61
6250-032-001	1.901	2.453	0.245	0.204	61
6289-007-034	4.317	6.114	0.473	0.484	4
6256-009-002	3.165	4.522	0.441	0.511	61

**Table 6.14.** Forecast results for the traffic flow for the loop detectors of Table 6.3 for the days from 2004/08/01 to 2004/09/30 with the CAS with  $\alpha = 0.2$ . An adaptive procedure offers the opportunity to react on basic changes of the traffic data. Although in some cases like the loop detector 4215-006-035 the adaptivity has a negative influence in regard to the CMS in Table 6.13, the results are even better than that of the weekday classification in Table 6.11.

Table 6.11. Furthermore, if the algorithm works in operation, it is also recommended to repeat the cluster algorithm with the attribute interpretation from time to time.

With the classification of traffic time series, many kind of knowledge can be used for forecasting. For most effects observed and analysed in Chapter 5 it is possible to find solutions to estimate future traffic states. Nevertheless with the DCA only those attributes can be used, that can be assigned to the whole day.

Consider the television broadcast of the football matches (Table 5.4). An attribute could be “television broadcast of a football match starting at 13:30”, but not “television broadcast of a football match in general”, because different kick off times lead to effects at different times. In case of different times of the cause and so the effect, the classification must be done without this event. Then, the only way is to filter the net effect out. For the example of Fig. 5.20 the following basic model is assumed:

$$\bar{x}_{\text{Forecast}}(d_p, t) = \frac{\sum_{d_n \in \mathcal{G}_{\text{Event}}} \frac{x(d_n, t + \Delta t_n)}{\bar{x}(\mathcal{G}_{\vec{C}(d_n)}, t + \Delta t_n)}}{\sum_{d_n \in \mathcal{G}_{\text{Event}}} 1} \bar{x}(\mathcal{G}_{\vec{C}(d_p)}, t). \quad (6.25)$$

The forecast traffic time series  $\bar{x}_{\text{Forecast}}(d_p, t)$  for the time  $t$  of the particular day  $d_p$  that is influenced by the special event is the product of the mean influence of the event and its classified traffic time series  $\bar{x}(\mathcal{G}_{\vec{C}(d_p)}, t)$  without this event. The mean influence of the event is calculated by the mean of the fraction of the measured traffic data  $x(d_n, t + \Delta t_n)$  at days  $d_n \in \mathcal{G}_{\text{Event}}$  with this event and the

classified traffic time series  $\bar{x}(\mathcal{G}_{\vec{C}(d_n)}, t + \Delta t_n)$  that would be that of the days  $d_n$  without this event.  $\Delta t_n = t_n - t_p$  is the particular time difference, with that the events start at day  $d_n$  at  $t_n$  and at day  $d_p$  at  $t_p$ . The sum goes over all days  $d_n$  that are in the group  $\mathcal{G}_{\text{Event}}$  with all days with the particular event and is finally divided by the number of days  $\sum_{d_n \in \mathcal{G}_{\text{Event}}} 1$  with the particular event.

Another approach is, to assume the different traffic as a cumulative one:

$$\bar{x}_{\text{Forecast}}(d_p, t) = \bar{x}(\mathcal{G}_{\vec{C}(d_p)}, t) + \frac{\sum_{d_n \in \mathcal{G}_{\text{Event}}} x(d_n, t + \Delta t_n) - \bar{x}(\mathcal{G}_{\vec{C}(d_n)}, t + \Delta t_n)}{\sum_{d_n \in \mathcal{G}_{\text{Event}}} 1}. \quad (6.26)$$

Whereas this is a feasible approach for the traffic flow and even the occupancy or density, the different velocities in Fig. 5.18 have shown that the forecast is not straight forward for all data. Furthermore in case of such organized special events many control strategies are applied to avoid congestions. People are told to avoid the particular area or to use public transport. Sometimes the infrastructure is temporary changed or even streets may be blocked. To consider all this, much information is needed that is not available for this work.

At this point it has to be mentioned, that traffic event management is a basic problem. Even nowadays the cumulative traffic of different events lead to large congestions. This effect is sometimes amplified by construction areas that are set up just to the wrong time. To discuss all this would extent the scope of this work, but traffic forecast considering event management offers enough interesting stuff for future research. What is basically missing is an event database (see also Sec. 8.2).

Up to now, the forecast model calculates classified traffic time series. It is a powerful tool for large forecast horizons. Nevertheless as discussed most recently measured traffic data offer valuable information for very short term forecasts. How most recent measured data can be considered is discussed in the next section.

### 6.3 Considering Recently Measured Data

A case that is often considered in regard to short term forecasts is the so called one-step-ahead forecast. This means precisely, the forecast of the next traffic data value with the knowledge of all values before. In the case of minute aggregated traffic data that are one minute forecasts. As discussed low order parametric regression models provide feasible results. For the set of loop detectors in Table 6.3 the averaged mean absolute error  $\overline{\text{MAE}}$  for the days with feasible traffic data for the period 2004/08/01 to 2004/09/30 can be seen in Table 6.15 for the following forecast models:

- naive forecasting  $\hat{x}_t$  (Eq. 3.12),
- exponential smoothing  $s_t$  with  $\alpha = 0.2$  (Eq. 3.18),

- double smoothing average  $s_{\text{double},t}(\tau)$  with  $\alpha = 0.6$  (Eq. 3.21),
- single smoothing with linear trend  $s_{\text{trend},t}(\tau)$  with  $\alpha = 0.05$  (Eq. 3.22),
- moving average  $\mu_t$  with  $N = 15$  (Eq. 3.13), and
- double moving average  $\mu_{\text{double},t}(\tau)$  with  $N = 35$  (Eq. 3.15).

The parameters are chosen, after the results are compared for different parameters. But note, that the optimal choice of the parameters strongly depends on the particular data set. For all trend models hold  $\tau = 1$  because one-step-ahead forecasts are calculated.

loop detector	$\hat{x}_t$	$s_t$	$s_{\text{double},t}(\tau)$	$s_{\text{trend},t}(\tau)$	$\mu_t$	$\mu_{\text{double},t}(\tau)$	$N_{\text{days}}$
4135-015-033	3.286	2.712	3.517	2.768	2.776	2.845	53
4214-015-002	3.290	2.545	3.416	2.517	2.519	2.557	34
4215-006-035	3.182	2.513	3.328	2.529	2.499	2.560	35
4349-025-001	3.318	2.528	3.434	2.519	2.506	2.551	61
4488-096-033	2.577	1.997	2.683	2.015	1.969	2.005	35
5877-019-001	2.284	1.800	2.397	1.787	1.786	1.817	58
5877-019-003	1.632	1.333	1.716	1.332	1.324	1.355	44
6211-000-033	2.447	1.895	2.532	1.884	1.886	1.911	61
6250-032-001	2.517	1.922	2.615	1.893	1.893	1.919	61
6289-007-034	3.937	3.110	4.096	3.198	3.139	3.303	4
6256-009-002	3.997	3.098	4.139	3.091	3.071	3.138	61

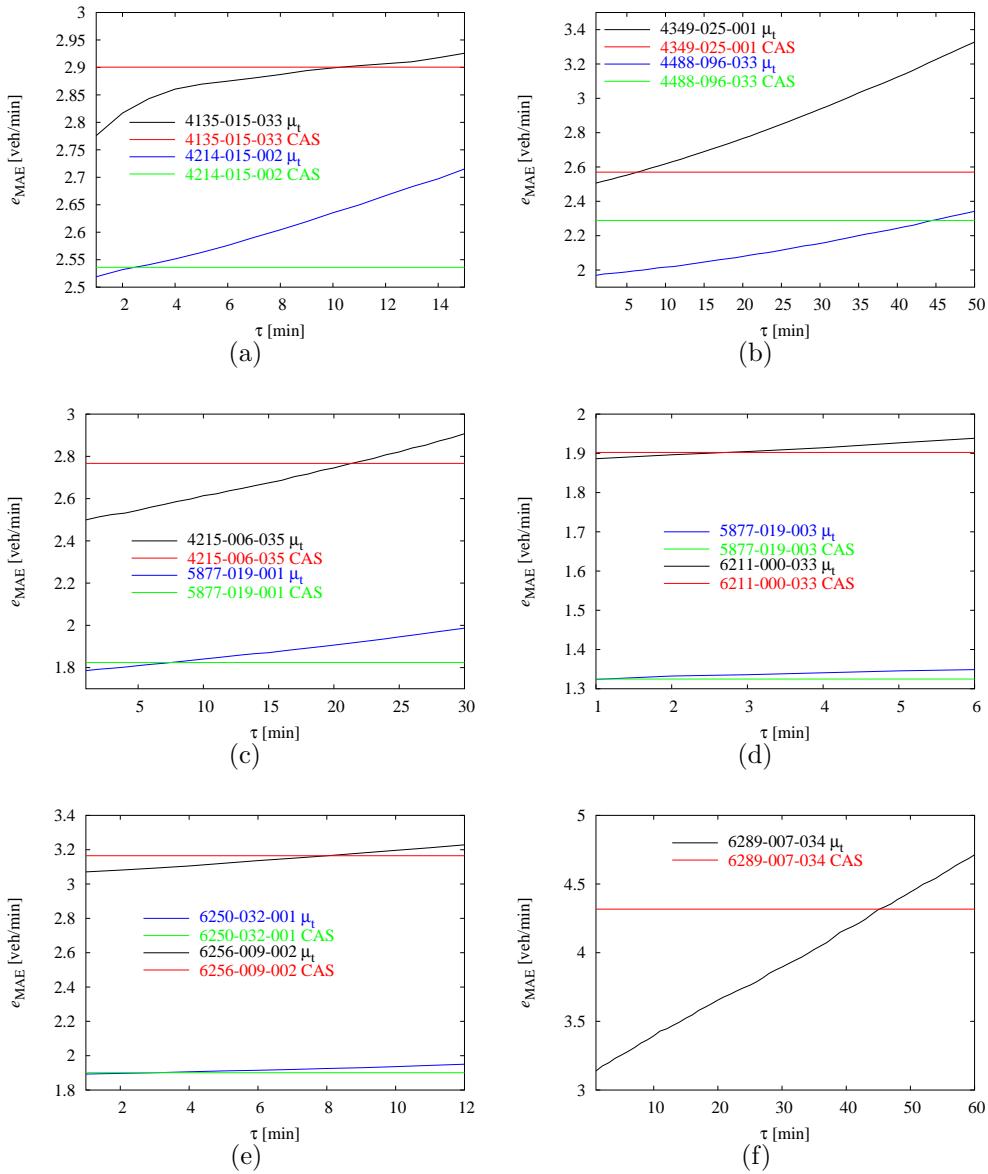
**Table 6.15.** The mean absolute error  $\overline{e_{\text{MAE}}}$  in veh/min averaged over  $N_{\text{days}}$  from 2004/08/01 to 2004/09/30 with feasible traffic data for the loop detectors of Table 6.3 and several one-step-ahead forecast models.

Especially the exponential smoothing and the moving average forecast model outperform the forecasts with the classified traffic time series of Table 6.11 to Table 6.14. It can also clearly be seen, that all trend models are outperformed by the constant ones, what is in excellent agreement with the urban traffic data investigated in [Chrobok et al., 2004].

Feasible methods to combine the current with historical data already exist and are proposed in Sec. 3.5.1. At this point a more fundamental question should be investigated: For how long does it make sense to consider the most recently measured data?

To investigate this question, the short term forecast method must be considered with an increasing forecast horizon  $\tau$  and compared with the results of the adaptive classified traffic time series. For the following investigation the 15 minute moving average model  $\mu_t$  is used, which show the best results for several loop detectors in Table 6.15.

In Fig. 6.6 the mean absolute error  $e_{\text{MAE}}$  can be seen for traffic flow forecasts with different forecast horizons  $\tau$  for all loop detectors of Table 6.3. As classified



**Fig. 6.6.** Mean absolute error  $e_{MAE}$  of the loop detectors 4135-015-033 and 4214-015-002 (a), 4349-025-001 and 4488-096-033 (b), 4215-006-035 and 5877-019-001 (c), 5877-019-003 and 6211-000-033 (d), 6250-032-001 and 6256-009-002 (e), and 6289-007-034 (f) for different forecast horizons  $\tau$  calculated with 15 min moving averages  $\mu_t$  and the classified traffic time series calculated with the CAS (with  $\alpha = 0.2$  in Eq. 6.23). The short term method outperforms the historical data from only the one-step-ahead forecast (5877-019-003) up to 45 min (6289-007-034). On average the short term forecast outperforms the historical data for 13.5 min.

traffic data act the CAS (with  $\alpha = 0.2$  in Eq. 6.23), that means the  $e_{MAE}$  is stemming from Table 6.14. Because this is a classified forecast, it is constant for all horizons  $\tau$ . To be precise, this  $e_{MAE}$  is the mean value for horizons up to about two months (from 2004/08/01 to 2004/09/31).

The short term method outperforms the historical data from only the one-step-ahead forecast (5877-019-003 in Fig. 6.6 (d)) up to 45 min (6289-007-034 in Fig. 6.6 (f)). On average the short term forecast outperforms the historical data for 13.5 min.

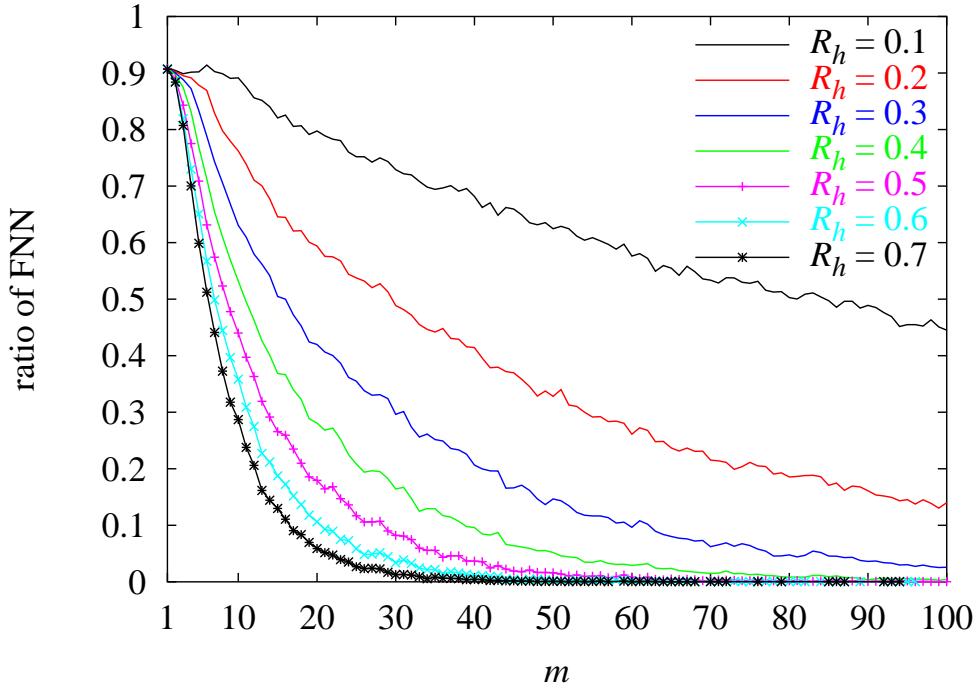
For which forecast horizons  $\tau$  the most recently measured data can outperform the historical traffic time series is depending on the particular traffic time series and especially what happens at this day. All unpredictable or only hardly predictable events lead to an increased horizon  $\tau$ , for that the ARIMA models outperform any knowledge based system, because the knowledge does not help. In case of the accident (cp. Sec. 5.4) the 15 min moving averages outperform the adaptive historical time series for 82 min for the loop detector 4214-015-002.

For other special events like extreme weather conditions, but also for many recurring structures for that the reasons are unknown, it would be reasonable, if the structures are identified in the historical data base and used for the forecast. An approach to consider this are the nonparametric regression techniques proposed in Sec. 3.2.1. In the following an approach is proposed combinig phase space embedding techniques with knowledge based classified traffic time series.

As described in Sec. 3.2.1, first at all, delay reconstruction vectors  $\vec{s}_n$  of the whole time series have to be defined. Therefore, just the time difference in number of samples  $\nu\Delta t$  as to be chosen. Because minute aggregated data is used, the time difference is given as  $\Delta t = 1$  min, and because there is no reason, why valuable information should be neglected  $\nu$  is chosen as  $\nu = 1$ .

In using phase space embedding methods it is an important fact [Kantz and Schreiber, 1999], that the data is free of noise, or at least, that the noise is reduced. This also has been comfirmed by the results of Shin et al. [1999] (see also Sec. 3.2.2 on page 29). Thus, according to the results in Sec. 5.5, 9 min moving averages instead of the raw data are used for the following investigations. Whenever past data is investigated, the data are smoothed with Eq. 5.16. But note, that if the algorithm is applied for forecasts, the 9 min moving averages of the current state space vector can only be calculated from the data that is already measured, that means, Eq. 3.13 has to be used.

With this, to define the state space vectors  $\vec{s}_{t_0}$  at time step  $t_0$ , only a suitable minimum embedding dimension has to be found. If this would be done, it would be straight forward to search for the NNs and use the FNN method in the original sense. But as explained, the phase space embedding technique should be combined with the knowledge based classification. To do this, not the whole phase space is searched for similar states. For a particular day  $d_0$  at the time of day  $t_0$  only state space vectors at days  $d_n \in \mathcal{G}_{\vec{C}(d_0)}$  that belong to the group of days  $\mathcal{G}_{\vec{C}(d_0)}$  with the same classification vector  $\vec{C}(d_0)$  like the day  $d_0$  are considered.



**Fig. 6.7.** Ratio of false nearest neighbours (FNN) for different embedding dimensions  $m$  and different thresholds  $R_h$ . To obtain a small fraction of FNN the embedding dimension has to be chosen as large as possible.

The *knowledge based embedding vectors* of the smoothed traffic time series  $x_{d_n,t}$  can then be described as

$$\vec{s}_{d_n,t_0} = (x_{d_n,t_0-m}, x_{d_n,t_0-[m-1]}, \dots, x_{d_n,t_0-1}, x_{d_n,t_0}). \quad (6.27)$$

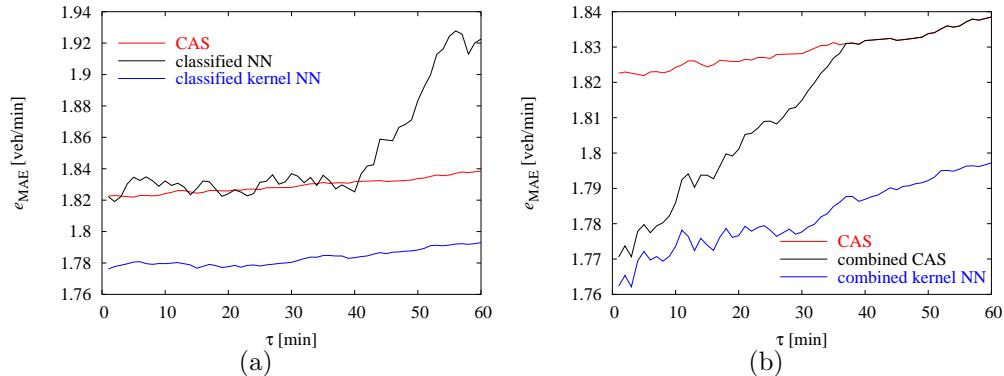
Now, the question of a suitable embedding dimension  $m$  has to be investigated. This is done with the method of false nearest neighbours (FNN) as described in Sec. 3.2.1. Note, the NNs are only searched among certain classified days  $d_n \in \mathcal{G}_{\vec{C}(d_0)}$  at the time step  $t_0$ .

To compare the state space vectors, in the following the Euclidean distance is used as norm.

In Fig. 6.7 the ratio of FNN for different choices of the embedding dimension  $m$  and different choices of thresholds  $R_h$ , for which  $R_i > R_h$  holds (see Eq. 3.36), can be seen. For this investigation the data of the days  $d_n$  from 2001/01/01 to 2004/09/31 with the classification vector  $\vec{C}(d_n) = [(\text{Mo,Th}), \text{Noholidays}, \text{Normal Day}]$  of loop detector 5877-019-001 are used. As time steps  $t_0$  are chosen every 60 min of a day ( $t_0 = 0, 60, \dots, 1380$ ).

The results show, that the embedding dimension should be chosen as large as possible to obtain a fraction of FNN that is as small as possible. The maximum

number of  $m$  depends on the time step  $t_0$  for that the state space vector is calculated. The dimension  $m$  can just be chosen as  $m = t_0$ , that is, the whole day up to  $t_0$ . This is done in the following, when the phase space vectors are used for forecast.



**Fig. 6.8.** The mean absolute error  $e_{\text{MAE}}$  for different forecast horizons  $\tau$  using different forecast methods that consider most recently measured data. (a) If only the interpolated value  $x_{d_n, t_0 + \tau}$  of the NN is used (classified NN, black), the forecast performance decreases in regard to the CAS (red). This problem can be solved, if the interpolated values of all possible state space vectors are weighted using a kernel function (classified kernel NN). (b) The forecast performance can be enhanced, if the combinations according to Eq. 3.95 are used. As current data  $J_{\text{cur}}(t_0)$  act the 15 min moving averages of the most recently measured data and as historical data time series  $J_{\text{hist}}(t)$  the CAS (black) and the classified kernel NN (blue).

If only the interpolated value  $x_{d_n, t_0 + \tau}$  of the classified NN is used, the forecast results do not necessarily improve, as can be seen in Fig. 6.8 (a). Shown is the mean absolute error  $e_{\text{MAE}}$  for forecasts of the traffic flow of the period from 2004/08/01 to 2004/09/30 for the loop detector 5877-019-001 for three different forecast methods for forecast horizons up to  $\tau = 60$  min.

Apart from the already mentioned CAS that can be seen as nearly a constant value, the result of the forecast with the classified NN interpolated value  $x_{d_n, t_0 + \tau}$  according to Eq. 6.27 is shown, which does not improve the result. The reason is that using only one value for forecast is very unstable and exactly a problem, that is already known as the *bias/variance dilemma* [German et al., 1992].

This is why a third method is proposed here that is similar to the kernel estimation (see Sec. 3.2.2 on page 26). According to Eq. 3.42 all iterpolated values  $x_{d_n, t_0 + \tau}$  of all possible phase space vectors are weighted with a factor  $\theta_j$  that depends according to Eq. 3.43 on a kernel function  $K$  and a bandwidth parameter  $w_n$ . The only difference here to the work of [Faouzi, 1996] is the different bandwidth  $w_n = \gamma_m m$ , that is chosen here because of the varying embedding dimension  $m$ . As a trade-off between bias and estimator variance  $\gamma_m$  is chosen as  $\gamma_m =$

0.04 veh/min for the state space vectors that consist of traffic flow. As can be seen in Fig. 6.8 (a) the method of weighting the NN with a kernel function leads to a clear improvement of the forecast.

For the calculations of the next section also state space vectors are used, that consist of other traffic data. For those, that consist of velocity  $\gamma_m = 1$  km/h and for those that consist of occupancy  $\gamma_m = 0.1$  is chosen. Note, that the smaller the values of  $\gamma_m$  the more similar is the result to the nearest neighbour model, and the higher the value the more is the result similar to the mean value of all interpolated values.

Finally it is compared, how a combination of these methods with the combination algorithm of Eq. 3.95 that has been proposed by Wild [1997] behave on the mean absolute error  $e_{\text{MAE}}$ . Therefore the 15 min moving averages of the most recently measured data are used as  $J_{\text{cur}}(t_0)$  and the time series of the CAS and the classified kernel NN as  $J_{\text{hist}}(t)$ . The parameters  $t_{\text{thmax}} = 37$  min and  $\eta = 0.57$  are chosen according to [Chrobok et al., 2004]. The result can be seen in Fig. 6.8 (b) for forecast horizons  $\tau$  up to 60 min.

## 6.4 Forecast Results

The differences and especially the difficulty to forecast the velocity and the occupancy have been already discussed in Sec. 5.6.1. But for applications, like the particular one that is mentioned in the next chapter, also estimations for other traffic data is needed. For these cases, the dependence of the data among each other offer the opportunity to apply the forecast algorithm also to the other traffic data. Thus, at least, the forecast results are presented for all the available traffic data. To get a broad overview about the forecast performance, the results are presented as distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$ .

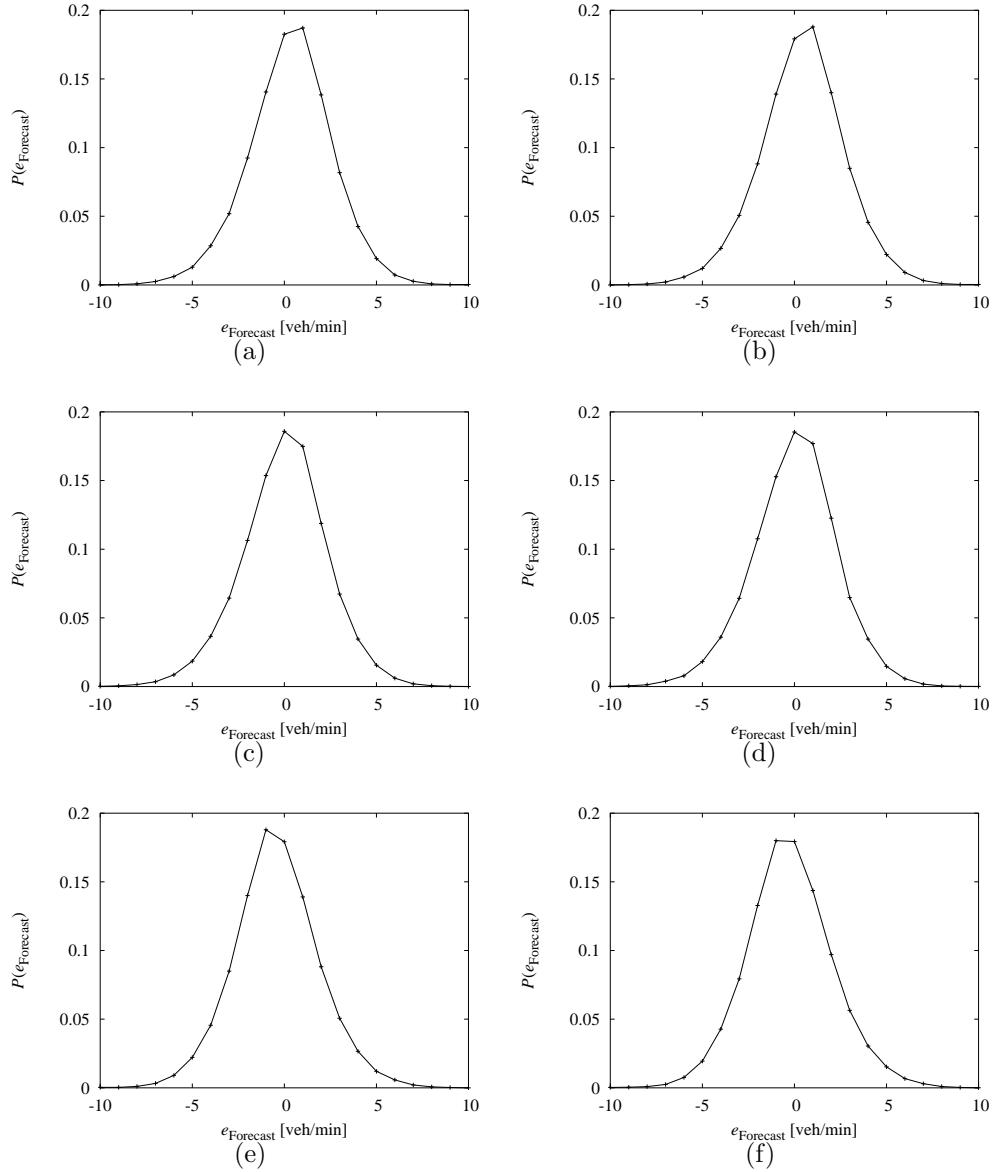
This can be seen in the following figures for all the essential traffic data measured by the loop detectors, that is, the flow  $J_{\text{veh}}$  (Fig. 6.9), the flow of lorries  $J_{\text{lor}}$  (Fig. 6.10), the velocity of passenger cars  $v_{\text{pc}}$  (Fig. 6.11), the velocity of lorries  $v_{\text{lor}}$  (Fig. 6.12), and the occupancy  $\rho_{\text{rel}}$  (Fig. 6.13).

Shown are always the results for forecasts of traffic data of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. As classification act the three dimensional classification vector with the attributes of the weekdays, that of school holidays, and that of the special days. Note, that the classification is always done according to the traffic flow  $J_{\text{veh}}$ .

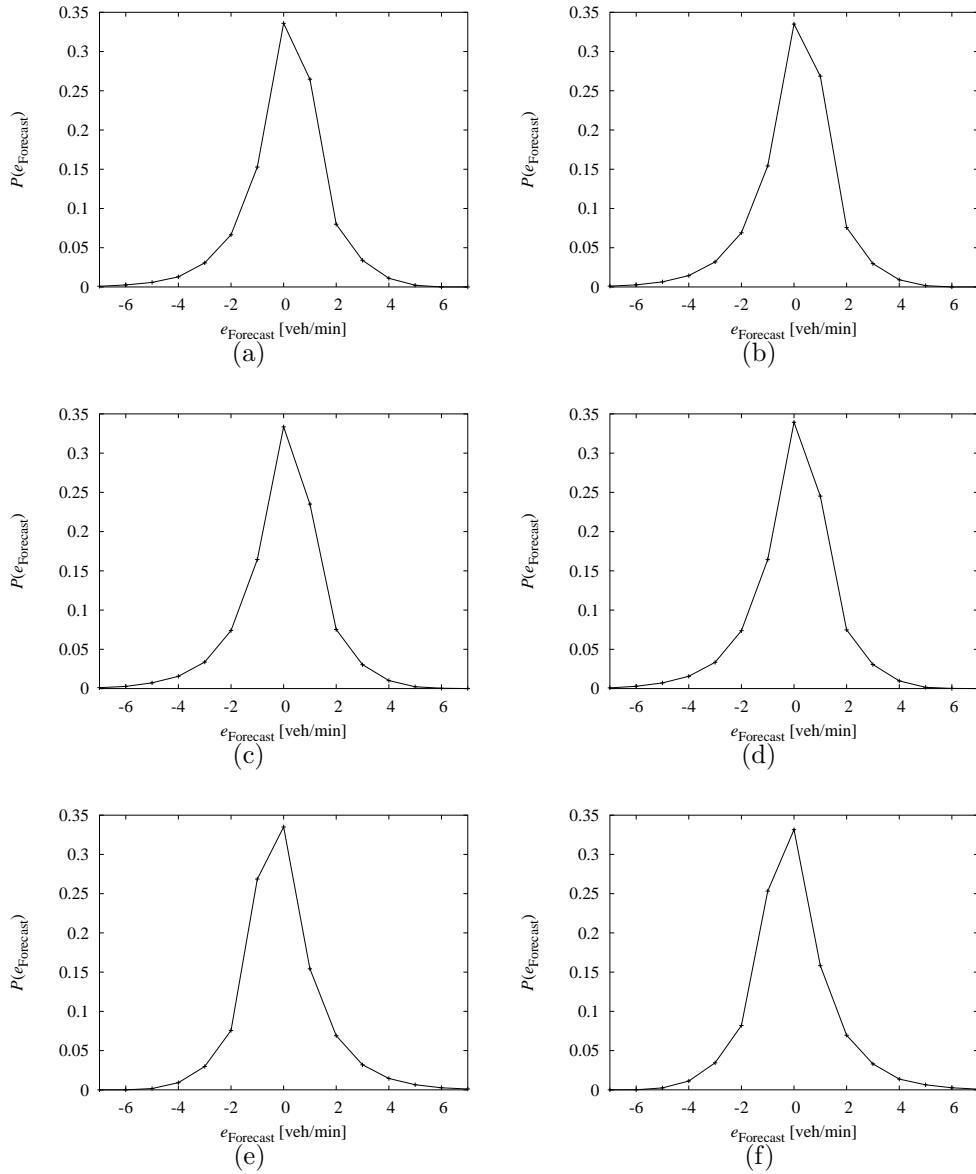
In the figures can be seen:

- forecasts of classified traffic time series for arbitrary forecast horizons  $\tau$ :
  - (a) CMS,
  - (b) CAS,
- one-step-ahead forecasts ( $\tau = 1$  min) of parametric regression methods:

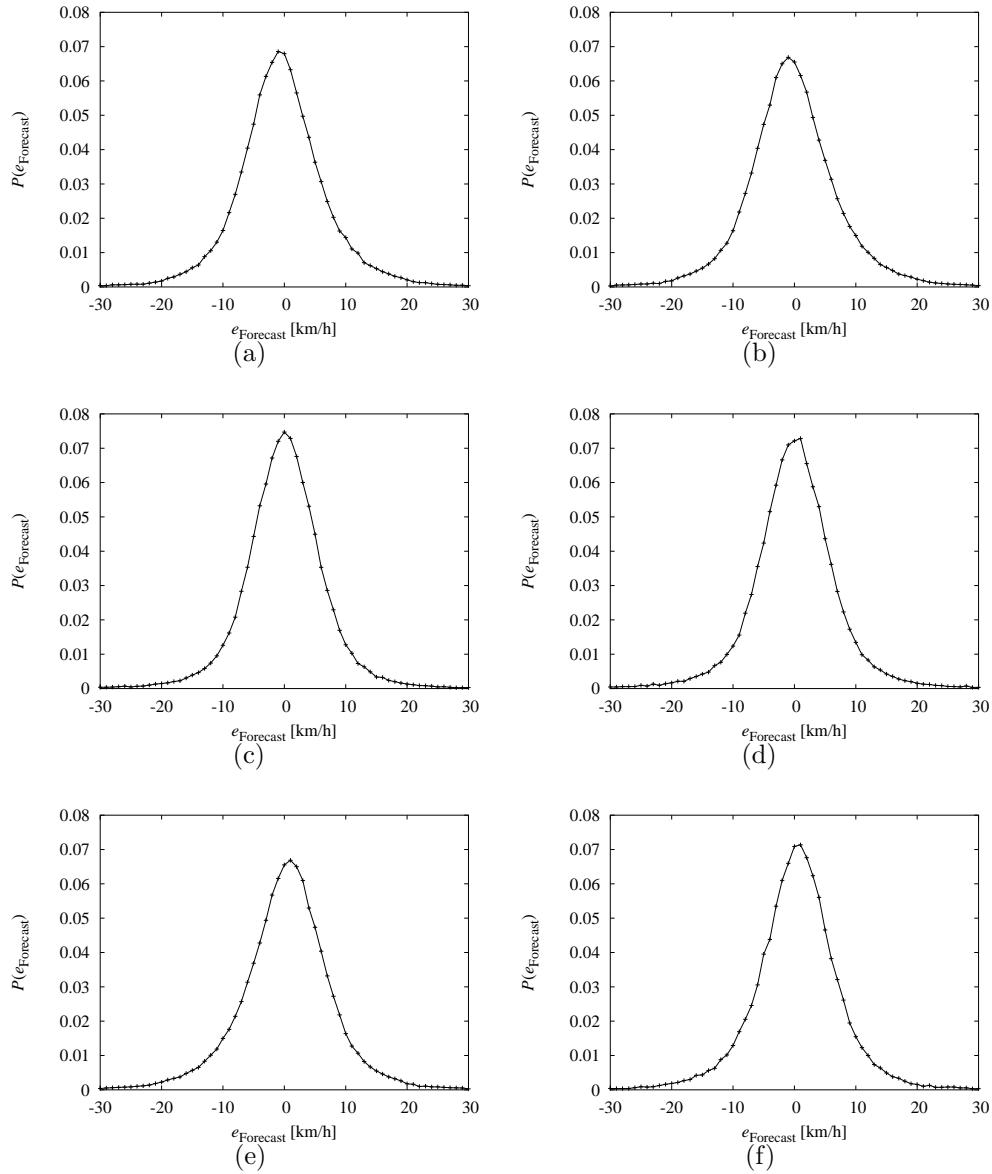
- (c) exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ),
- (d) moving averages (Eq. 3.13 with  $N = 15$ ),
- forecasts for a forecast horizon of  $\tau = 30$  min with combinations according to Eq. 3.95 with  $t_{\text{thmax}} = 37$  min and  $\eta = 0.57$ . As  $J_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $J_{\text{hist}}(t)$ :
  - (e) CAS,
  - (f) classified kernel NN.



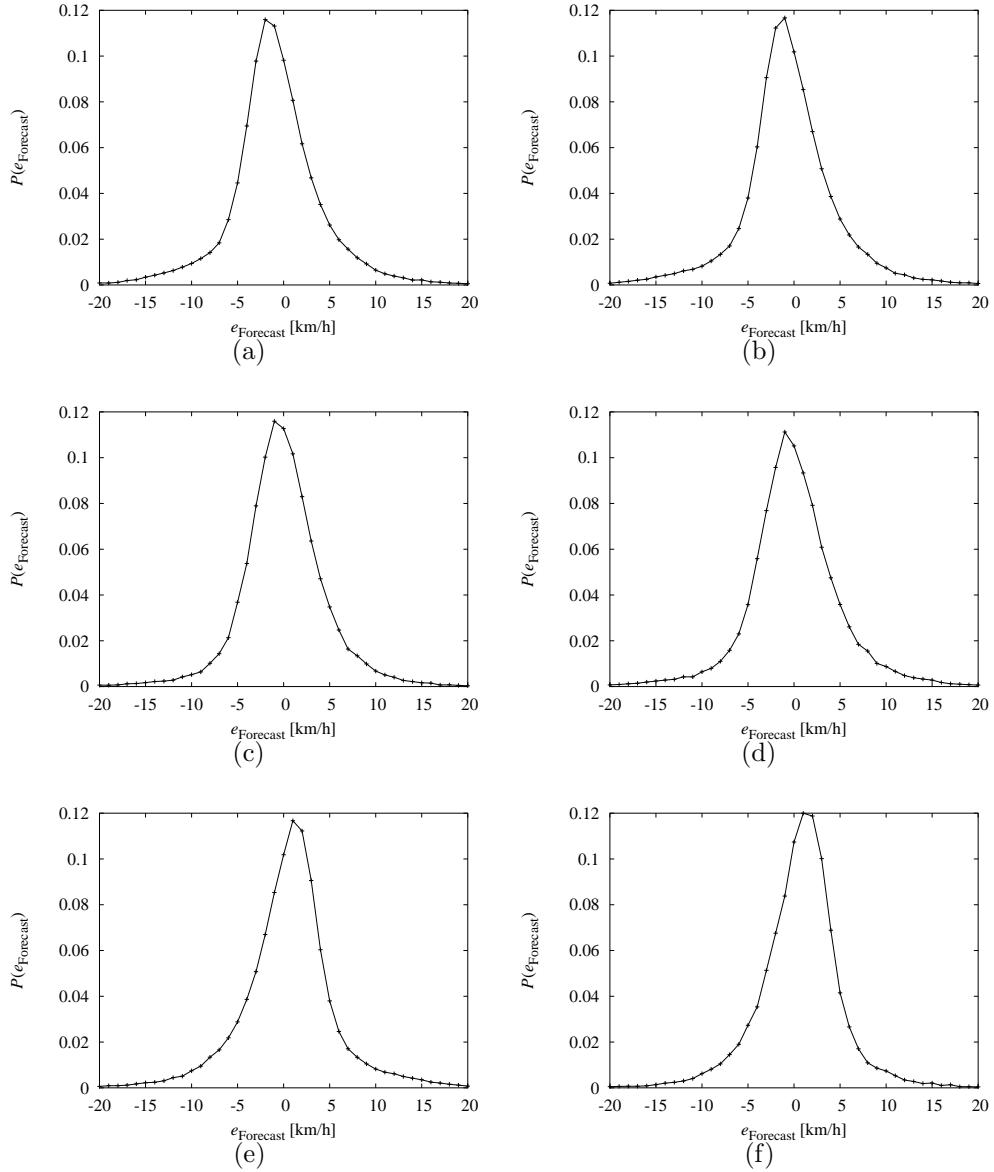
**Fig. 6.9.** Forecast results of forecasting the traffic flow  $J_{\text{veh}}$  of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. Shown are the distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$  rounded to integer values for the CMS (a) and the CAS (b) for arbitrary forecast horizons  $\tau$ . The one-step-ahead forecasts calculated with (c) the exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ) and (d) the moving averages (Eq. 3.13 with  $N = 15$ ). Furthermore, for a forecast horizon of  $\tau = 30$  min two combinations according to Eq. 3.95 are calculated, whereby as  $J_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $J_{\text{hist}}(t)$  the (e) CAS and the (f) classified kernel NN.



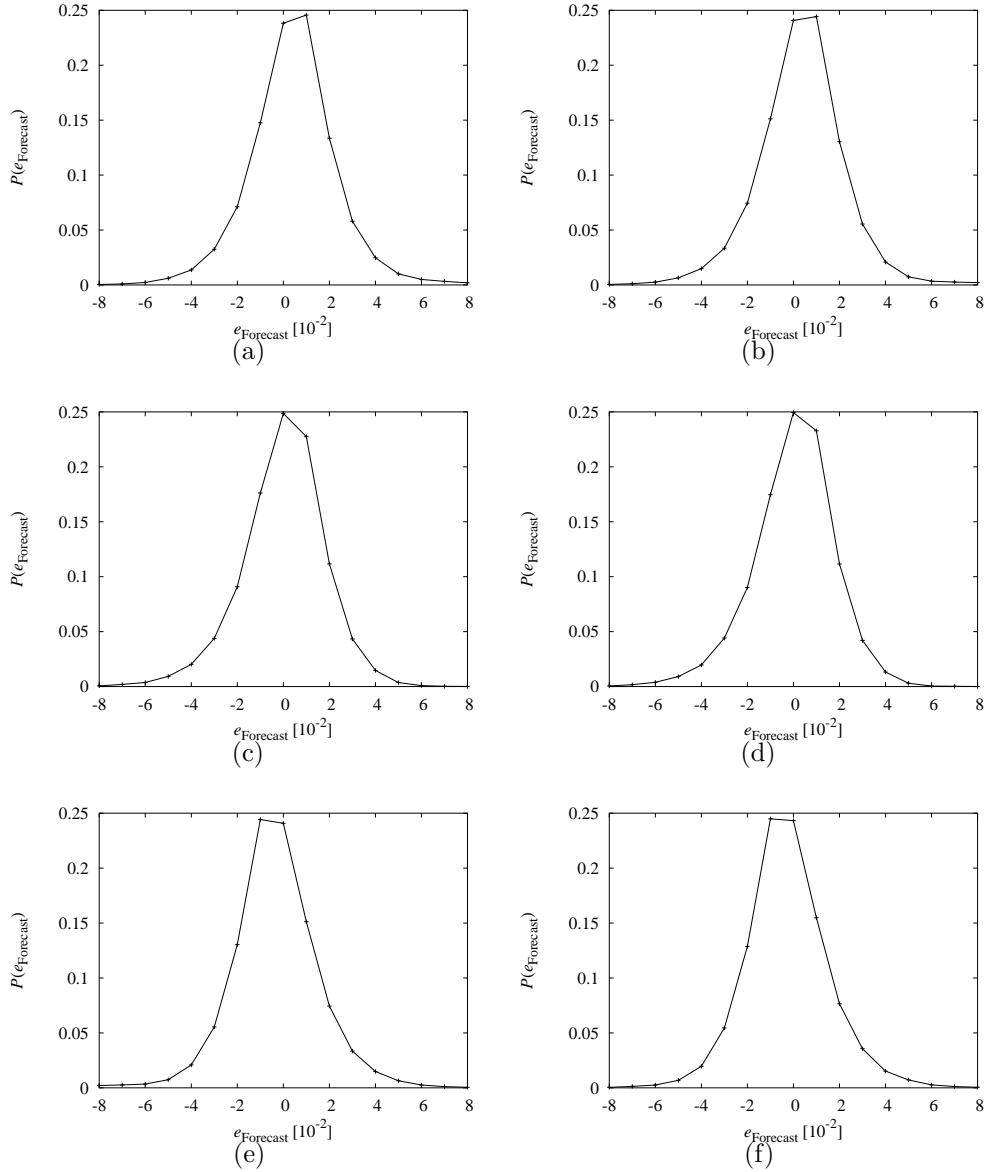
**Fig. 6.10.** Forecast results of forecasting the traffic flow of lorries  $J_{\text{lor}}$  of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. Shown are the distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$  rounded to integer values for the CMS (a) and the CAS (b) for arbitrary forecast horizons  $\tau$ . The one-step-ahead forecasts calculated with (c) the exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ) and (d) the moving averages (Eq. 3.13 with  $N = 15$ ). Furthermore, for a forecast horizon of  $\tau = 30$  min two combinations according to Eq. 3.95 are calculated, whereby as  $J_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $J_{\text{hist}}(t)$  the (e) CAS and the (f) classified kernel NN.



**Fig. 6.11.** Forecast results of forecasting the velocity of passenger cars  $v_{pc}$  of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. Shown are the distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$  rounded to integer values for the CMS (a) and the CAS (b) for arbitrary forecast horizons  $\tau$ . The one-step-ahead forecasts calculated with (c) the exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ) and (d) the moving averages (Eq. 3.13 with  $N = 15$ ). Furthermore, for a forecast horizon of  $\tau = 30$  min two combinations according to Eq. 3.95 are calculated, whereby as  $v_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $v_{\text{hist}}(t)$  the (e) CAS and the (f) classified kernel NN.



**Fig. 6.12.** Forecast results of forecasting the velocity of lorries  $v_{\text{lor}}$  of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. Shown are the distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$  rounded to integer values for the CMS (a) and the CAS (b) for arbitrary forecast horizons  $\tau$ . The one-step-ahead forecasts calculated with (c) the exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ) and (d) the moving averages (Eq. 3.13 with  $N = 15$ ). Furthermore, for a forecast horizon of  $\tau = 30$  min two combinations according to Eq. 3.95 are calculated, whereby as  $v_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $v_{\text{hist}}(t)$  the (e) CAS and the (f) classified kernel NN.



**Fig. 6.13.** Forecast results of forecasting the occupancy  $\rho_{\text{rel}}$  of the loop detector 5877-019-001 for the period from 2004/08/01 to 2004/09/30. Shown are the distributions  $P(e_{\text{Forecast}})$  of the forecast error  $e_{\text{Forecast}}$  rounded to integer values for the CMS (a) and the CAS (b) for arbitrary forecast horizons  $\tau$ . The one-step-ahead forecasts calculated with (c) the exponential smoothing method (Eq. 3.18 with  $\alpha = 0.2$ ) and (d) the moving averages (Eq. 3.13 with  $N = 15$ ). Furthermore, for a forecast horizon of  $\tau = 30$  min two combinations according to Eq. 3.95 are calculated, whereby as  $\rho_{\text{cur}}(t_0)$  act the 15 minute moving averages of the most recently measured data and as  $\rho_{\text{hist}}(t)$  the (e) CAS and the (f) classified kernel NN.

# Chapter 7

# Application

“Das Neue dringt herein mit Macht.”

**Johann Christoph Friedrich von Schiller**  
(1759 – 1805), German poet and dramatist

This chapter describes an application that uses the results of this work: the traffic information system OLSIM (**OnLine Traffic SIMulation**). One essential part of the system are forecast algorithms that are used to forecast traffic states 30 and 60 min ahead. These forecast algorithms are applications of the scientific results of this work. The whole system is even much more complex and at this point only the framework of the system can be shortly described. The interested reader is referred to the extensive literature, such as [Chrobok et al., 2004,a,b, 2002, 2003, 2001, Hafstein et al., 2003, Marinossen et al., 2002, Pottmeier et al., 2004, 2003].

## 7.1 Traffic Information System OLSIM: Motivation and Background

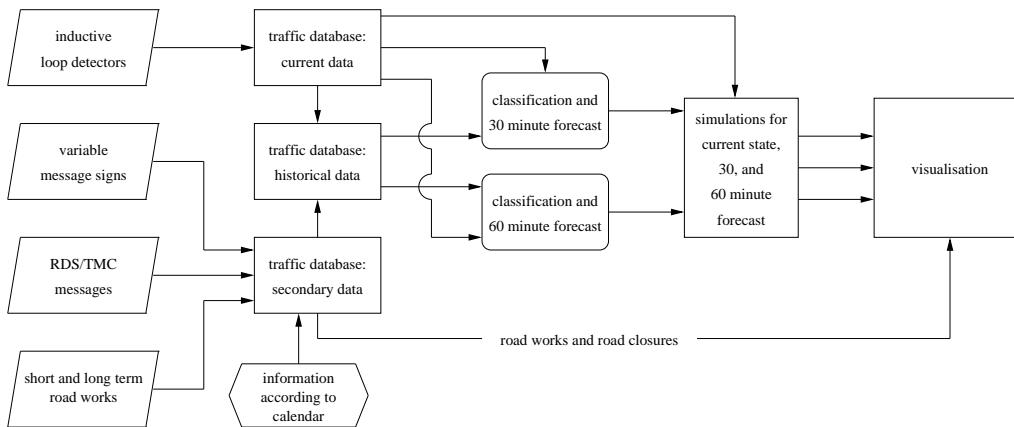
The traffic information system OLSIM is established, to give the Internet user the opportunity to get information about the motorway traffic state in North Rhine-Westphalia (NRW). Apart from the current traffic state, a 30 and a 60 minute forecast is available. The approach to generate the current traffic state in the whole motorway network is to use the locally measured traffic data of the loop detectors discussed in this work as the input into a spatial temporal traffic simulation. Using simulations is nowadays the state of the art in every traffic information or management system. With a simulation traffic states are also estimated in regions that are not covered by measurements. Furthermore, it directly provides estimates for other valuable quantities like trip travel times. To provide information about future traffic states, forecast traffic data have to be filled into the simulation instead of the current data. The basis of the calculation is

the procedure proposed in Chapter 6. But for technical reasons, not all potentials can be used for application.

In the following the principle of the whole information system is introduced. Thereafter, the concrete forecast algorithm used by the system is explained. Then, the spatial temporal traffic model is outlined and a short description is given, how the model is applied to the real network. At last, the application is introduced that visualises the information.

## 7.2 Principle

The intention in developing the traffic information system OLSIM is to offer the opportunity to inform the road user fast and efficiently about the current and the predictive traffic state. Therefore all available information has to be collected and prepared in a manner that is useful for the user. The general setup of the traffic information system OLSIM is depicted in Fig. 7.1.



**Fig. 7.1.** The architecture of the traffic information system OLSIM. Different input data are available that are depicted by the parallelograms on the left. After storing them in different databases, the data are prepared for the 30 and 60 minute forecast with algorithms that result from this work and are described on detail in Sec. 7.3. Finally, the current and the two forecast traffic data sets are filled in one simulation each that calculate the network wide traffic state from the point data. This state can finally be visualised.

In addition to the loop detector data discussed in this work in detail, other secondary data is available. The data of the control states of about 1,800 variable message signs that are located across the network are provided every minute. The location and the duration of road works are available as well. The information of short term construction areas are sent daily, those of permanent construction areas every week.

Another data source are the so called RDS/TMC-messages (Radio Data System

/ Traffic Message Channel). These messages are information provided by the traffic warning service and include all kind of warnings concerning the current traffic, for instance, traffic jams, accidents, road closures, and reroutings. These data are sent to the OLSIM system immediately when they are generated.

Apart from that, all the information according to the calendar like the attributes weekday, month, holiday, school holidays, and so on are available. As discussed, they are, of great importance. All these data are stored in different databases. The most important one is that of the current data of the loop detectors. All the other data are stored in a secondary database. Both data bases offer the basic information for the classification.

Then, the data is fed into one of three simulations. Whereas for the online simulation that maps the current traffic state only the raw data is needed, the 30 and the 60 minute forecast simulations need the data of the particular forecast procedures. Then, the results of all the three simulations are visualised. Furthermore some more informations of the secondary database like road works or road closures are given to the user too.

### 7.3 Applied Forecast Model

As discussed in detail, first a classification to calculate historical data is needed. For technical terms, it is not possible, to apply the DCA for each particular loop detector separately. Thus, a classification for all the loop detectors is needed. As a trade-off, for all loop detectors three dimensional classification vectors are used, that consist of the classes according to the weekdays  $D_{1,n}$ , that of the holidays  $D_{2,n}$ , and that of special days  $D_{3,n}$ . The particular classes can be seen in Table 7.1.

$D_{1,n}$	Su, Mo, Tu, We, Th, Fr, Sa
$D_{2,n}$	Noholidays, (Easter, Summer, Autumn), Winter
$D_{3,n}$	Normal Day, Bridging Day, Before Holiday, Holiday, (All other days)

**Table 7.1.** Classes  $D_{j,n}$  used by the traffic information system OLSIM. For technical terms it is not possible to apply the DCA for each particular loop detector.

Then, the forecast data are calculated for the traffic flow of all vehicles  $J_{\text{veh}}$ , the flow of lorries  $J_{\text{lor}}$ , the velocity of passenger cars  $v_{\text{pc}}$ , the velocity of lorries  $v_{\text{lor}}$ , and the occupancy  $\rho_{\text{rel}}$ .

As 60 minute forecast directly act the 15 minute moving averages of the classified adaptive forecast using Eq. 6.23 with  $\alpha = 0.2$ . For the 30 minute forecast this time series is combined with the current traffic data using Eq. 3.95, whereby as current data act the 15 min moving averages of the most recently measured data, and the parameters are chosen as  $t_{\text{thmax}} = 37$  min and  $\eta = 0.57$ .

## 7.4 Spatial Temporal Traffic Model

As mentioned, to obtain a network wide traffic state, a spatial temporal traffic model is needed to generate the 2 dimensional line information from the point information measured by the loop detectors. Because the data is fed into the simulator and processed by it every minute it has to be at least as fast as real time. So, the simulation model has to fulfill certain requirements in regard to accuracy and efficiency. A class of models that fulfill these requirements also in large-scale networks are the so called *cellular automata models* (see, for instance, [Esser and Schreckenberg, 1997, Nagel et al., 2000, Rickert and Wagner, 1996, Schreckenberg and Wolf, 1998]).

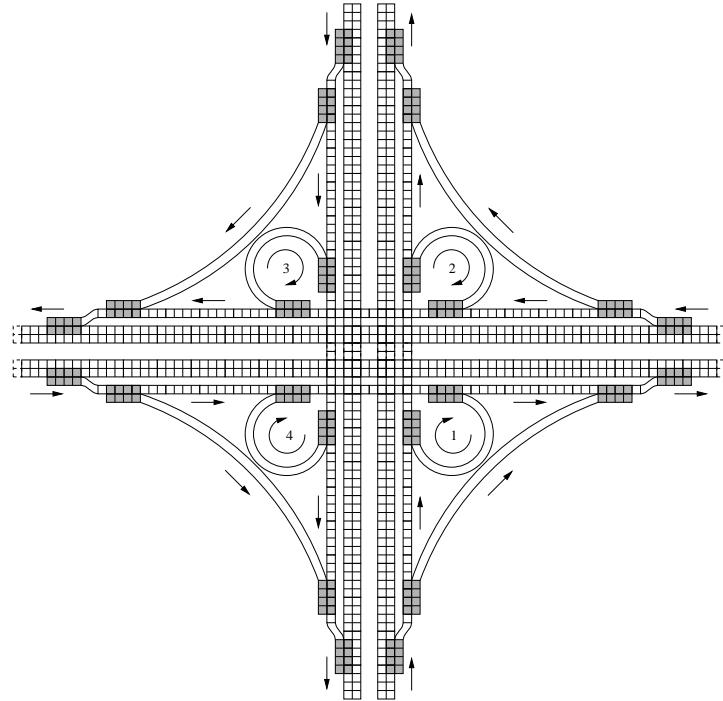
The model that is used by the system OLSIM is based on the first cellular automaton model for traffic flow, that has been able to reproduce some characteristics of real traffic [Nagel and Schreckenberg, 1992]. After several refinements using velocity dependent randomisation [Barlovic et al., 1998] or anticipation and the human reaction on brake lights (comparable to the dependency on velocity and acceleration) [Knospe et al., 2000], the model becomes more and more realistic. One basic problem left is, that all those models are developed under laboratory conditions. That means, that in most cases models are designed for traffic on only one lane. An approach to enhance the model above for two lanes is done in [Knospe et al., 2002a]. But the network consists of many other elements like on-, off-ramps, intersections, lane closures, speed limits, and so on. Thus, other improvements have to be made until the model has been able to simulate the motorway network of NRW faster than real time [Wahle et al., 2002]. The model that is finally used by the traffic information system OLSIM is described in [Hafstein et al., 2004]. How the topology of all the elements above is implemented, is described in the following.

## 7.5 Network Topology

An important point in the design of a simulator is the representation of the road network. Therefore a digitised image of the network is needed. The problem with geographic information systems is that there is no unique format for every kind of application [Grande et al., 2004]. Existing geographic information systems could not fulfill the requirements that are needed for a microscopic simulator of this detailing. Thus, a new geographic information system called **Olsim Track Data Format (OTDF)** has been designed.

The approach is based on the work of [Froese, 1998] and divides the network into links. The main links connect the junctions and highway intersections representing the carriageway. Each junction and intersection consist of another link, like on- and off-ramps or turning lanes. Then, several attributes are assigned to each link. The attributes are, for instance, the length, the number of lanes, or a possible speed limit. Each link together with all attributes is called *track*. Each track is divided into the cells that are needed for the cellular automaton traffic model.

The tracks are joined using elements that are called *exits*. Each single exit represents a junction of two tracks. It comprises all the important information about the junction that is needed by the simulation model, especially where exactly a vehicle can enter or leave a track. With tracks and links together it is possible to build a digital image of the topology of each traffic network. An example of an intersection can be seen in Fig. 7.2.



**Fig. 7.2.** The complex structure of an intersection in OTDF. The vehicles are simulated on the *tracks*. The small cells indicate the discretisation in the microscopic simulation model and stay in this case for 20 cells of 1.5 m length. The tracks are connected with the exits (grey regions). Those exits comprise all the important information where exactly a vehicle can enter or leave a track. With the tracks and the exits it is possible to map a digital image of any complex traffic network (picture from [Hafstein et al., 2004]).

In a complex network there are always more than one routes to get from an origin to the destination. Moreover, from a physical point of view, loops lead to the fact, that the number of ways is actually infinite. Looking at Fig. 7.2 this can be understood in an easy way. A vehicle can use the junctions 1 to 4 driving in a circle arbitrarily often before it continues in a different direction. This is avoided in OTDF with so called *connections*. The connections comprehend all necessary routing information at the junctions and intersections. With the connections OTDF becomes a fully routing capable data format.

With tracks, exits, and connections the digital image of the topology of the motorway network of NRW is built. Table 7.2 shows some design parameters of the

network that hold for the 2004/10/31.

Another crucial information concerns the positions of the installed loop detectors. They also have to be included in the digital map of the network. The positions in the simulation are called *checkpoints*, and at these checkpoints the simulation is adapted to the measured traffic data of the loop detectors. Therefore, algorithms have to be found, to incorporate the real world measurements into the simulation. First methods for this procedure are given for urban traffic in [Esser and Schreckenberg, 1997]. However, these suffer from a major drawback, since they destroy the dynamics of the system and are only sufficient for urban traffic, where the dynamics are governed by intersections, mainly traffic lights.

area <sup>a</sup>	34,083 km <sup>2</sup>
inhabitants <sup>a</sup>	≈ 18,075,000
overall motorway length <sup>a</sup>	2,173 km
on- and off-ramps	876
intersections	73
number of tracks	3,814
lane resoluted track length	12,549.562 km
cells in simulator	8,366,375
loop detectors	4,480

**Table 7.2.** Design parameters of the North Rhine-Westphalian motorway network that is digitised in OTDF for the traffic information system OLSIM. The data are the status quo of the 2004/10/31.

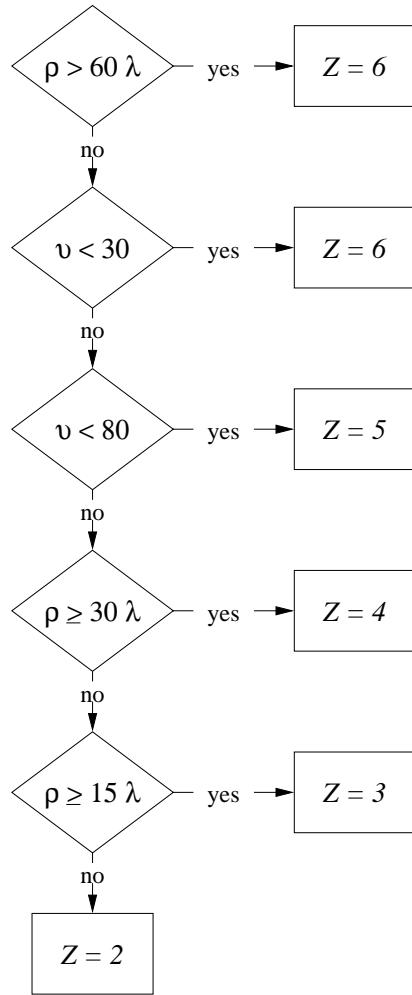
<sup>a</sup>The data are stemming from the state office for data processing and statistics NRW.

Thus, several methods like *flow tuning*, *density tuning*, and *tuning of the mean gap* are investigated that follow the idea to add or remove vehicles *adiabatically*, that is, without disturbing the system [Froese, 1998, Kaumann, 2000, Kaumann et al., 2000]. The method that is finally used by the system is an enhancement of the tuning of the mean gap. Before vehicles are added or removed from the simulation, the first step is to try to move vehicles behind the checkpoint in front of it and vice versa [Hafstein et al., 2004]. This method is preferred to pure insert/removal strategies because these can completely fail due to positive feedback if a non existing congestion is simulated. Note, that the algorithms how to adapt microscopic traffic simulations with real world measurements are still part of present research.

## 7.6 Visualisation

There are several user interfaces that could be possible to transfer the information about the network wide traffic state to the public. Because the traffic information system OLSIM is designed for a broad public, a web-based visualisation has been chosen.

On the website a map of NRW is drawn, where the motorways are coloured according to the simulated traffic state on the particular track. Therefore a traffic state  $Z$  is generated from the mean velocity  $v$  and the density  $\rho$  on each track. This is done as in [Klukas-Illen and Weiss, 2001] whereby for the traffic information system OLSIM a few enhancements are made. In Fig. 7.3 can be seen, how the state  $Z$  is calculated from the mean velocity  $v$  in km/h and the density  $\rho$  in veh/km for each track. The state  $Z = 1$  is missing, because it only indicates, that no traffic state is available.

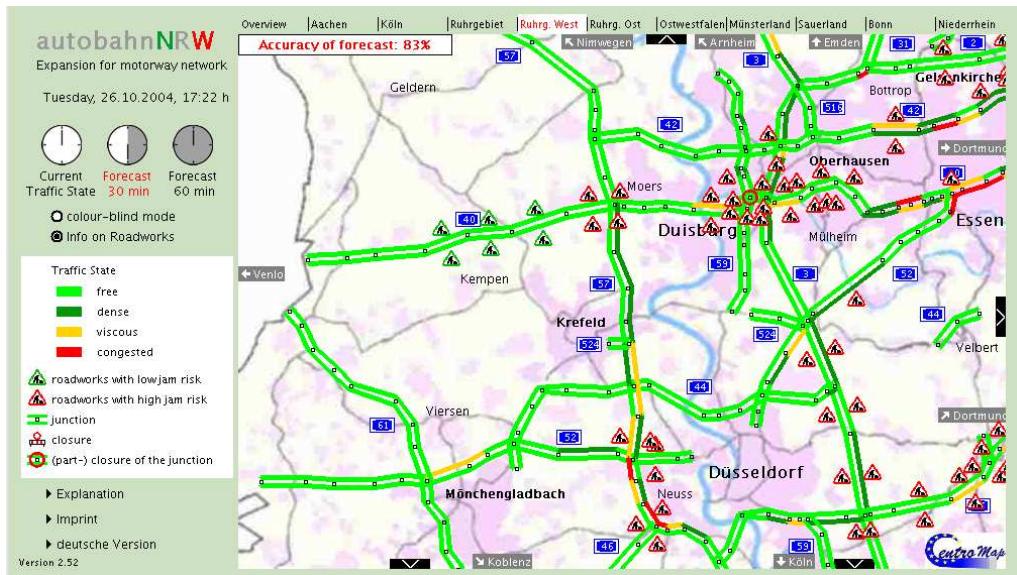


**Fig. 7.3.** Calculation of the traffic state  $Z$  with the density  $\rho$  in veh/km and the velocity  $v$  in km/h. If there are two lanes on the motorway hold  $\lambda = 2$ , otherwise  $\lambda = 3/2$ . The state  $Z = 1$  is not shown in the diagram because it only indicates, that no state is available.

Because psychological investigations have shown, that most humans can only distinguish four traffic states maximum, many even only three [Kochs et al.,

2004], the states  $Z = 2$  and  $Z = 3$  can both be seen in the map as light green for free flow. The other ones are dark green ( $Z = 4$ ) for dense flow, yellow ( $Z = 5$ ) for stop and go traffic, and red ( $Z = 6$ ) for a traffic jam.

On the website the user can choose whether he wants to see the current traffic state, that of the 30, or that of the 60 minute forecast. Furthermore he can select different maps in a way whether he wants an overview or one of several detailed maps. For completeness it has also to be mentioned, that the map offers some more information than only the traffic states like construction areas, road closures, or many kinds of detailed text information. A screenshot of this can be seen in Fig. 7.4. At the 2004/10/31 the website is available at the address: <http://www.autobahn.nrw.de>.



**Fig. 7.4.** Screenshot of the visualisation of the 30 minute traffic forecast of NRW at the 2004/10/26. At this time the website is available at the address: <http://www.autobahn.nrw.de>.

# Chapter 8

## Conclusions and Outlook

”What we call Progress is the exchange of one nuisance for another nuisance.”

**Oliver Wendell Holmes**  
(1809 – 1894), American physician and essayist

In this chapter first a short summary of this work is given. The main conclusions are shortly described. Nevertheless, in each scientific work not all questions can be answered. Moreover, answering questions may lead to new, sometimes more questions. Thus, a short outlook is given, that may be the basis for interesting future works.

### 8.1 Conclusions

The objective of this work has been the development and application of an advanced traffic forecast algorithm with a physical background. The state of the art report of current methods and applications of forecasting traffic has shown, that the methods can be basically classified in four groups: parametric regression, nonparametric regression, neural networks, and knowledge based systems. A comparison of the different methods is difficult because in the works highly different kinds of traffic data are used.

Thus, the data of 4,480 inductive loop detectors of the motorway network of NRW are empirically analysed. Because there are many sources for possible measurement errors, several methods are proposed with that misinterpretations can be avoided. The basic traffic data that are measured with the loop detectors are the traffic flow, the velocity, and the occupancy. Thereby, for flow and velocity a distinction in lorries and passenger cars is possible.

In statistically analysing these data in detail it becomes clear, that days with similar properties show also similar characteristics in the traffic data. Many of

the properties are dominated by the calendar. But also special events that possibly depend on particular regions influence measurably the traffic state.

Many of the recurring structures can be forecast for arbitrary forecast horizons with heuristics or knowledge based systems. Thereby the causes are identified, that have an impact on traffic and their effect is measured. This method has the benefit, that the knowledge about future causes like sport events or road works can be used for the forecast. Any traffic forecast system that should forecast certain events must be fed with information about them.

In practise this can be done with a classification of the days according to several attributes that are implied by certain causes, for instance, the days of the week. To assure, that there is a limited number of attributes, the classification must be done taking the effect of each particular cause into account. The effect highly depends on the particular traffic value. Some events influence the traffic flow, some the velocity.

Sometimes even cognitive abilities are required to identify different patterns. This is the reason why neural networks are often proposed in former works, but unfortunately at present common neural networks are not able to cope with this problem automatically.

For horizons of minutes up to hours the most recently measured data offer useful information for forecast. Especially for the one-step-ahead forecast the main problem is to smooth the fluctuations which can be performed by parametric regression techniques, so called ARIMA methods. In most cases models of low order like exponential smoothing or moving averages provide feasible results. The physical background is, that the traffic state is expected to stay as it is.

Another method to use most recently measured data is the nonparametric regression technique, whereby the traffic state is analysed with a state space model. This leads to a few forecast enhancements, because not all recurring events that influence the traffic are known from the first. But with a state space model the traffic data cannot be forecast until the first impacts of the event are measured. Using these results a new traffic forecast model is developed. To classify the daily traffic time series a double cluster assignment (DCA) method is proposed. Thereby the days are assigned to several attributes, that are supposed to influence traffic states. Then, the DCA calculates automatically which attribute influences each particular loop detector with an automatic cluster interpretation.

When the classification is done, the forecast traffic data time series can be calculated. This can be straight forwardly done with just the mean value of the classified days for each time of day. Because the number of days used in each classification can differ strongly, it is always important to smooth the classified traffic time series, for instance, with moving averages. To consider also basic changes in the network, not just the mean value of the classified days should be used for forecast. Instead of this, the classified traffic time series should be calculated with parametric regression techniques. The enhancement of the forecast is shown with 15 days moving averages and 0.2 exponential smoothing, what is also used for the further investigations.

To use the most recently measured data, it is investigated, for which forecast horizons the short term parametric regression methods outperform the classified traffic data. The results show clearly, that normally this is the case for one hour maximum. With this result, a combination of the short term and the long term forecast is straight forward.

Additionally an approach to combine the most recently measured data with the historical information is investigated. It is shown, that the forecast with classified traffic time series can be enhanced, when among all days of the same class the nearest neighbours (NNs) in phase space are used and the forecast is calculated with a kernel function that weights the NNs according to their distance in phase space with a Gaussian distributed kernel function.

Finally the forecast results are shown for two classified traffic time series working for arbitrary forecast horizons, two parametric regression methods for one-step-ahead forecasts, and two combinations for 30 minute horizons.

Thereafter an application of the new traffic forecast algorithm is presented. The traffic information system OLSIM offers among other information a 30 and a 60 minute traffic forecast. To obtain a network wide traffic state, a spatial temporal traffic model is used to transform the point information of the loop detectors into a line information. Therefore an advanced cellular automaton traffic model is used.

## 8.2 Outlook

Unfortunately for technical reasons not all algorithms that are proposed in this work can be used in the application. Thus, in the future the application has to be enhanced applying the complete forecast algorithm, especially applying the double cluster assignment (DCA) for each loop detector separately.

Furthermore an event database that has to be established and kept up to date is basically missing. Note that in a traffic network like that of NRW there are many regional events that influence locally the traffic. Taking them into account using the new forecast technique can enhance the result.

The DCA proposed in this work is used for the automatic recognition of causes and their impacts on traffic. In future it may be, that other systems can be evolved with cognitive abilities that outperform this method. As candidates often neural networks are discussed but there have to be many improvements as the results of former works in Sec. 3.3 have shown.

An opportunity for interesting future research offers the visualisation or the user interface, respectively. Currently, a web based visualisation is displayed. The question is, whether the visualisation in the Internet is in future the optimal way to reach the user. A problem at the moment is, that most travellers have got an Internet connection at home or at work, but not in their vehicle. This leads to the effect, that the system is used to plan the route *pre-trip*, that means before the trip is started. With so called *Advanced Traveller Information Systems* (ATIS)

[Adler and Blue, 1998, Barfield and Dingus, 1998], it is possible to inform the traveller *en-route*, so that he can respond to this information. The information can be brought to the traveller with, for instance, onboard systems, mobil phones, or variable message signs.

This leads directly to the question whether another kind of information can be offered to the driver. With the information about the traffic state, the driver has no information how long it takes to pass a certain section. The intention at the moment is to enhance the traffic information system OLSIM in that sense, that also information about current and forecast travel times is provided. The travel times follow directly from the travel times of virtual vehicles in the microscopic traffic simulation.

Once travel times are calculated, the question can also be investigated, how the route with the shortest travel time can be determined and a *dynamic* route guidance system can be developed. Such a dynamic route guidance systems calculates the route with the shortest travel time considering not only the static infrastructure but also the dynamic traffic states.

With the enhancements of traffic forecast and information systems another very interesting problem comes up that is different from forecasts in other fields, for instance, weather. The weather does not care about the forecast and cannot be influenced therefore giving no response. In the case of stock market forecasts, if many investors trust the forecaster, there can be a positive response. But forecasting traffic can lead to a negative response falsifying the prognosis [Arnott et al., 1991].

Consider the case that there are more than one possible route to get from an origin to the destination. If a congestion is forecast on one of those routes, say route *A*, and the information is brought to the travellers, many of them may decide to avoid that route *A*. If enough travellers do so, the forecast congestion may stay away or appear on other routes, say *B* or *C*. This phenomenon is called the *self-destructing forecast*.

Whether a self-destructing forecast appears or not, first at all depends strongly on how many travellers get the information. If only a small part of them use the information, those travellers have got an advantage over the others. In traffic the minority has got benefits, the majority stays in a congestion. But in cases of public systems, like the traffic information system OLSIM, the information is generally obtainable for all travellers. The question is, at which time a certain user limit is passed so that the traffic state is measurably influenced by the respond.

To estimate this influence, the driver reaction must be forecast. However driver behaviour is difficult to understand and at the moment there are only theoretical investigations into driver behaviour.

An approach of simulating the driver behaviour with a strategic and a tactical layer is done in [Wahle et al., 2000]. Further theoretical analyses show [Selten et al., 2004], that even in the simple case of a two route scenario three differ-

ent kinds of drivers can be distinguished <sup>1</sup>. First feedbacks of the OLSIM users let expect, that in fact there are even many more strategies, how the travellers respond to information. But detailed investigations are required in the future.

Consider the case, far in future, that nearly all road users can obtain reliable traffic forecasts immediately and also their reactions can be estimated. Let  $\vec{f}$  be the traffic forecast vector consisting of the forecast traffic states of all sections in the network. This leads to the reaction of the drivers that can be described with a *driver reaction function*  $\vec{R}(\vec{f})$ . This function has to be applied on the forecast vector to obtain the traffic state vector  $\vec{s}_1$  that is the result of the forecast and the reaction of the drivers:

$$\vec{s}_1 = \vec{R}(\vec{f}). \quad (8.1)$$

So, forecasting the traffic state  $\vec{f}$  generates the traffic state  $\vec{s}_1$ .

Equation 8.1 leads to several interesting open questions. A driver who knows  $\vec{s}_1$  has got benefits and a traffic information system that forecasts  $\vec{s}_1$  from the first outperforms others. But if  $\vec{s}_1$  can be directly forecast this leads with the reaction of the drivers to the traffic state  $\vec{s}_2$  that can be easily calculated using Eq. 8.1 as

$$\vec{s}_2 = \vec{R}(\vec{s}_1) = \vec{R}(\vec{R}(\vec{f})) = \vec{R}^2(\vec{f}), \quad (8.2)$$

and the question comes up, why not directly forecasting  $\vec{s}_2$  and so on.

This procedure can be iterated up to  $n$  times:

$$\vec{s}_n = \vec{R}^n(\vec{f}), \quad (8.3)$$

whereby  $n$  can mean arbitrary often. The question is how the states  $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n$  behave, especially under what conditions the iteration can converge.

But searching for a state  $\vec{s}_n$  for that finally hold  $\vec{R}^n(\vec{f}) \approx \vec{s}_n$  is based on the assumption that the forecaster wants to provide a forecast that is as accurate as possible.

But, moreover the driver reaction can also be used to control traffic. To do this the forecaster can perhaps generate a certain traffic state  $\vec{o} = \vec{R}(\vec{f}_{\text{gen}})$  with his forecast  $\vec{f}_{\text{gen}}$ . This can be, for instance, the optimal state in regard to the least number of congestions or the smallest cumulative sum of travel times. But in this case the forecaster has to consider that the reaction function can change if the reality deviates too much from the forecasts. In this manner, the reaction function also depends on the difference  $\vec{o} - \vec{f}_{\text{gen}}$ .

Apart from the driver reaction there is at least another question: Can information about individual traffic be extended to public transport? At the present even in sophisticated regions of the world the timetable is the only available information about public transport. If it would be possible to obtain and analyse data about latenesses of public transport systems, forecast methods can be developed that are not restricted to individual traffic. The future vision is an advanced traffic management and control system for intermodal transport.

---

<sup>1</sup>At this point the author wants to give special thanks for helpful discussions and suggestions to Reinhard Selten, Nobel-Prize winner in economics of the year 1994.



# Abbreviations

ABE	Absolute Error
AR	Auto Regressive
ARE	Absolute Relative Error
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
ATIS	Advanced Traveller Information System
CA	Cellular Automaton
CAS	Classified Adaptive Smoothed
CMS	Classified Mean Smoothed
DCA	Double Cluster Assignment
EC	Equality Coefficient
FE	Forecast Error
FNN	False Nearest Neighbour
MA	Moving Average (Eq. 3.2)
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MAXE	Maximum Error
MDT	Mean Daily Traffic
ME	Mean Error
MLP	Multi Layer Perceptron
MOV	Moving Averages (Eq. 3.13 or Eq. 5.16)
MRE	Mean Relative Error
MSE	Mean Square Error
N/A	Not Available
NN	Nearest Neighbour
NRW	North Rhine-Westphalia
OLSIM	Online Traffic Simulation
OTDF	Olsim Track Data Format
PE	Percent Error
RDS/TMC	Radio Data System / Traffic Message Channel
RE	Relative Error
RMSE	Root Mean Squared Error
RMSEP	Root Mean Squared Error Proportional
RRMSE	Relative Root Mean Squared Error
TLRN	Time Lagged Recurrent Network
TWT	Tuesdays, Wednesdays, and Thursdays



# Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich der Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Professor Dr. Michael Schreckenberg betreut worden.

Duisburg, den 15. Februar 2005

---

Roland Chrobok

## Teilpublikationen

- Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg (2000).  
Three categories of traffic data: Historical, current, and predictive. In  
E. Schnieder and U. Becker (Eds.), *Proc. of the 9th IFAC Symposium Control in Transp. Systems*, Braunschweig, pp. 250 – 255. Int. Fed. of Automatic Control.
- Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg (2004).  
Different methods of traffic forecast based on real data. *Euro. J. Op. Res.* 155, 558–568.
- Chrobok, R., A. Pottmeier, S. Hafstein, and M. Schreckenberg (2004b).  
Traffic forecast in large scale freeway networks. *Int. J. of Bifurcation and Chaos* 14(6), 1995–2005.



# Curriculum Vitae

Roland Chrobok

10.04.1975	Geboren in Mülheim an der Ruhr
1994	Latinum, Graecum, Abitur
1994 – 1995	Wehrdienst, sechswöchiges Praktikum an der technischen Schule des Heeres und der Fachschule des Heeres für Technik, Abzeichen für Leistungen im Truppendienst in Gold
1995 – 2000	Studium der Physik und Mathematik an der Gerhard-Mercator-Universität Duisburg und der Ruhr Universität Bochum
10.09.1997	Vordiplom Physik mit der Note "sehr gut"
14.08.2000	Diplom der Physik an der Gerhard-Mercator-Universität Duisburg mit der Arbeit "Statistische Analyse von Zählschleifendaten als Methode zur Verkehrsprognose" und der Note "sehr gut"
seit 01.09.2000	Wissenschaftlicher Mitarbeiter am Lehrstuhl "Physik von Transport und Verkehr" der Universität Duisburg-Essen
03.03.2001	Förderpreis "Logistik – Güterverkehr – Infrastukturkosten" der Interessengemeinschaft der deutschen Logistikunternehmer "Club km92"
01.07.2001	Siemens Förderpreis Straßenverkehrstechnik "Verkehrsprognose"
12.07.2001	1. Platz BMW Scientific Award
20.11.2003	Heinz-Billing-Preis der Max-Planck-Gesellschaft



# Acknowledgements

First I thank my adviser, Michael Schreckenberg, for the opportunity to work in the interesting field of traffic, his invaluable support and trust in my work throughout my thesis. He always encouraged me to follow my own ideas.

I am especially grateful to the current and former colleagues at the group of Physics of Transport and Traffic. My deepest gratitude goes to Sigurður Hafstein, Florian Mazur, and Andreas Pottmeier, the cooperators in the successful development of the traffic information system OLSIM.

For the network support I thank Torsten Huiszinga, Andreas Pottmeier, and Andreas Kefel.

For the fruitful cooperation within the different projects my gratitude goes to Robert Barlović, Thomas Grunewald, Sigurður Hafstein, Oliver Kaumann, Andreas Kefel, Hubert Klüpfel, Florian Mazur, Andreas Pottmeier, Joachim Wahle, Daniel Weber, and Matthias Woltering.

I am especially grateful to Birgit Dahm for helping me with the administrative work.

For sharing insights I thank Frank Königstein, Sabine Knoche, Wolfgang Knospe, Tobias Kretz, Tim Meyer-König, Anja Ober-Sundermeier, Nathalie Waldau, Michael Wiedemann, and Marko Wölki.

I would like to thank the Ministry of Transport, Energy and Spatial Planning of North Rhine-Westphalia for financial support and for the availability of the traffic data. I am very grateful to Georg Stüben and Udo Ziegler for the fruitful work and for the possibility of validating the results of this work with such a broad public.

For reading the manuscript I am thankful to Conny Chrobok, Sigurður Hafstein, Torsten Huiszinga, Florian Mazur, Andreas Pottmeier, and Jeremy Stanton.

Last but not least I would like to thank my parents Margret and Heinz Chrobok and my sister Conny Chrobok for their loving support during the thesis.



# Bibliography

- Abdulhai, B., H. Porwal, and W. Recker (1999). Short term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks. See *Transp. Research Board* [1999].
- Adler, J. and V. Blue (1998). Toward the design of intelligent traveler information systems. *Transp. Res. C* 6, 157–172.
- Ahmed, M. S. and A. R. Cook (1979). Analysis of freeway traffic time-series data by using box-jenkins techniques. *Transp. Res. Record* 722, 1–9.
- Ahmed, S. A. (1983). Stochastic processes in freeway traffic. *Traffic Engineering and Control* 24, 306–310.
- Ahmed, S. A. and A. R. Cook (1981). Point process models for freeway incident detection. See Hurdle et al. [1981], pp. 20–30.
- Ahmed, S. A. and A. R. Cook (1983). Application of time series analysis techniques to freeway incident detection. *Transp. Res. Record* 841, 19–21.
- Akaike, H. (1973). Information theory and extension of the maximum likelihood principle. In *Proc. of the 2nd Int. Symp. on Information Theory*, Budapest, pp. 267–281.
- Akaike, H. (1974). A new look at the statistical model identification. In *IEEE Transactions on Automatic Control*, Volume 19, pp. 716–723.
- Anderson, T. W. (1971). *Statistical Analysis of Time Series*. New York: John Wiley and Sons.
- Arnott, R., A. de Palma, and R. Lindsey (1991). Does providing information to drivers reduce traffic congestion? *Transp. Res. A* 25(5), 309–318.
- Bajwa, S., E. Chung, and M. Kuwahara (2003). Sensitivity analysis of short term travel time prediction model's parameters. See ITS International [2003].
- Barfield, W. and T. Dingus (1998). *Human Factors in Intelligent Transportation Systems*. Mahwah: Lawrence Erlbaum Associates Inc.

- Barlovic, R., L. Santen, A. Schadschneider, and M. Schreckenberg (1998). Metastable states in cellular automata for traffic flow. *Eur. Phys. J. B* 5, 793–800.
- Beckmann, A. and H. Zackor (2001). Untersuchung und Eichung von Verfahren zur aktuellen Abschätzung von Staudauer und Staulängen infolge von Tages- und Dauerbaustellen auf Autobahnen. In *Forschung Straßenbau und Straßenverkehrstechnik*, Volume 808. Bonn-Bad Godesberg: Bundesministerium für Verkehr.
- Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- Bovy, P. H. and R. Thijs (2000). *Estimators of Travel Time for Road Networks, New Developments, Evaluation Results and Applications*. Delft: Delft University Press.
- Box, G. E. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Brockwell, P. and R. Davis (1996). *Introduction to Time Series and Forecasting*. New York: Springer.
- Brown, R. G. and P. Y. Hwang (1992). *Introduction to Random Signals and Applied Kalman Filtering*. New York: John Wiley and Sons.
- Buckley, D. (Ed.) (1974). *Proc. of the 6th Int. Symp. on Transp. and Traffic Theory*. Sydney: A.H. & A.W. Reed Pty Ltd.
- Bundesministerium für Verkehr (2000). *Technische Lieferbedingungen für Streckenstationen*. Dortmund: Verkehrsblattverlag.
- Bundesministerium für Verkehr (2002). *Verkehr in Zahlen 2002/2003*. Berlin: Bundesministerium für Verkehr, Bau- und Wohnungswesen.
- Bundesministerium für Verkehr (2003). *Grundlagen für die Zukunft der Mobilität in Deutschland - Bundesverkehrswegeplan 2003*. Berlin: Bundesministerium für Verkehr, Bau- und Wohnungswesen.
- Ceder, A. (Ed.) (1999). *Proc. of the 14th Int. Symp. on Transp. and Traffic Theory*. Amsterdam: Pergamon.
- Chadenas, C. and H. Kirschfink (1999). *Ganglinien NRW: Validierung*. Aachen: Heusch/Boesefeld GmbH.
- Chen, H., M. S. Dougherty, and H. R. Kirby (1997). An investigation of detector spacing and forecasting performance using neural networks. See ITS International [1997].

- Choi, M. and H. Lee (1995). Traffic flow and 1/f fluctuations. *Phys. Rev. E* 52(6), 5979–5984.
- Chrobok, R. (2000). Statistische Analyse von Zählschleifendaten als Methode zur Verkehrsprognose. Diplomarbeit, University Duisburg-Essen, Duisburg.
- Chrobok, R., S. Hafstein, A. Pottmeier, and M. Schreckenberg (2004). OLSIM: Future Traffic Information. In M. H. Hamza (Ed.), *Proc. of the IASTED Int. Conference: Applied Simulation and Modelling*, Rhodes, CD-ROM.
- Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg (2000). Three categories of traffic data: Historical, current, and predictive. See Schnieder and Becker [2000], pp. 250–255.
- Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg (2004). Different methods of traffic forecast based on real data. *Euro. J. Op. Res.* 155, 558–568.
- Chrobok, R., A. Pottmeier, S. Hafstein, and M. Schreckenberg (2004a). OLSIM: A new generation of traffic information systems. In V. Macho and V. Kremer (Eds.), *Forschung und wissenschaftliches Rechnen, GDWG-Bericht Nr. 63*, pp. 11–25. Max-Planck-Gesellschaft.
- Chrobok, R., A. Pottmeier, S. Hafstein, and M. Schreckenberg (2004b). Traffic forecast in large scale freeway networks. *Int. J. of Bifurcation and Chaos* 14(6), 1995–2005.
- Chrobok, R., A. Pottmeier, S. Marinossen, and M. Schreckenberg (2002). On-line simulation and traffic forecast: Applications and results. See Hamza [2002], pp. 113–118.
- Chrobok, R., A. Pottmeier, J. Wahle, and M. Schreckenberg (2003). Traffic forecast using a combination of on-line simulation and traffic data. See Fukui et al. [2003], pp. 345–350.
- Chrobok, R., J. Wahle, and M. Schreckenberg (2001). Traffic forecast using simulations of large scale networks. See Stone et al. [2001], pp. 434–439.
- Chung, E. (2003). Classification of traffic pattern. See ITS International [2003].
- Clarke, G. M. and D. Cooke (1978). *A Basic Course in Statistics*. London: Edward Arnold.
- Czuka, F.-J., G. Folkerts, and H. Kirschfink (2000). *Verkehrslage-Prognose NRW*. Aachen: Heusch/Boesefeldt GmbH.
- Danech-Pajouh, M. and M. Aron (1991). ATHENA: A method for short term inter urban motorway traffic forecasting. *Recherche Transports Sécurité - English issue 6*, 11–16.

- Daubechies, I. (1988). Orthonormal basis of compactly supported wavelets. *Communications on Pure and Applied Mathematics* 41, 909–996.
- Davis, G. A. and N. L. Nihan (1991). Nonparametric regression and short term freeway traffic forecasting. *J. of Transp. Engineering* 117(2), 178–188.
- Davis, G. A., N. L. Nihan, M. M. Hamed, and L. N. Jacobson (1990). Adaptive forecasting of freeway traffic congestion. *Transp. Res. Record* 1287, 29–33.
- Davis, G. A. and S. Yang (2001). Accounting for uncertainty in estimates of total traffic volume: An empirical Bayes approach. *J. of Transp. and Statistics* 4(1), 27–37.
- Deheuvels, P. (1977). Estimation non-paramétrique de la densité par histogrammes généralisés. In *Publications de l’Institut statistique*, Volume XXII, pp. 1–23. Université de Paris.
- Dell’Orco, M. and M. Ottomanelli (Eds.) (2002). *Proc. of the 9th Meeting of the EURO Working Group on Transp.: Intermodality, Sustainability and Intelligent Transp. Systems*, Bari. Politecnico di Bari.
- Deutsche Shell AG (1999). *Mehr Autos - weniger Emissionen*. Hamburg: Abteilung Information und Presse.
- Dia, H. (2001). An object-oriented neural network approach to short term traffic forecasting. *Euro. J. Op. Res.* 131, 253–261.
- Dougherty, M. S. (1995). A review of neural networks applied to transport. *Transp. Res. C* 3(4), 247–260.
- Dougherty, M. S. and M. R. Cobbett (1997). Short term inter-urban traffic forecasts using neural networks. *Int. J. of Forecasting* 13(1), 21–31.
- Dougherty, M. S. and M. Joint (1992). A behavioural model of driver route choice using neural networks. In S. Ritchie and C. Hendrickson (Eds.), *Proc. of the Int. Conference on Artificial Intelligence Applications in Transp. Engineering*, San Buenaventura.
- Esser, J. and M. Schreckenberg (1997). Microscopic simulation of urban traffic based on cellular automata. *Int. J. of Mod. Phys. C* 8(5), 1025–1036.
- Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.
- Faouzi, N.-E. E. (1996). Nonparametric traffic flow prediction using kernel estimator. See Lesort [1996], pp. 41–54.
- Faouzi, N.-E. E. (1999). Combining predictive schemes in short term traffic forecasting. See Ceder [1999], pp. 471–487.

- Ferrari, P. (1988a). The control of motorway reliability. *Transp. Res. A* 25, 419–427.
- Ferrari, P. (1988b). The reliability of the motorway transport system. *Transp. Res. B* 22, 291–310.
- Ferrari, P. (1989). The effect of driver behaviour on motorway reliability. *Transp. Res. B* 23, 139–150.
- Ferrer, J. and J. Barceló (1993). AIMSUN2: Advanced Interactive Microscopic Simulator for Urban and non-urban Networks. Technical report, Departamento de Estadística e Investigación Operativa, Facultad de Informática, Universitat Politècnica de Catalunya.
- Florio, L. and L. Mussone (1996). Neural-network models for classification and forecasting of freeway traffic flow stability. *Control Engineering Practice* 4(2), 153–164.
- Froese, K. (1998). Simulation von Autobahnverkehr auf der Basis aktueller Zähldaten. Diplomarbeit, University Duisburg-Essen, Duisburg.
- Fukui, M., Y. Sugiyama, M. Schreckenberg, and D. Wolf (Eds.) (2003). *Traffic and Granular Flow '01*, Tokyo. Springer.
- Gartner, N. and N. Wilson (Eds.) (1987). *Proc. of the 10th Int. Symp. on Transp. and Traffic Theory*, New York. Elsevier.
- German, S., E. Bienenstock, and R. Doursat (1992). Neural network and the bias-variance dilemma. *Neural Computation* 4, 1–58.
- Gottinger, H. W. (1980). *Elements of Statistical Analysis*. Berlin: de Gruyter.
- Grande, A., B. Kraus, and D. Wiegand (2004). Report - Geografische Informationssysteme. In *c't*, Volume 10, pp. 84–89. Hannover: Heise Zeitschriften Verlag.
- Grossberg, S. (1976). Adaptive pattern recognition and universal recording: (1) parallel development and coding of neural feature detectors. *Biological Cybernetics* 23, 121–134.
- Grunewald, T. (2001). Analyse und Bewertung von Verkehrsleitsystemen am Beispiel von Verkehrsbeeinflussungsanlagen. Diplomarbeit, University Duisburg-Essen, Duisburg.
- Hafstein, S., R. Chrobok, A. Pottmeier, M. Schreckenberg, and F. Mazur (2004). A high-resolution cellular automata traffic simulation model with application in a freeway traffic information system. *Computer-Aided Civil and Infrastructure Engineering* 19(6), 338–350.

- Hafstein, S., R. Chrobok, A. Pottmeier, J. Wahle, and M. Schreckenberg (2003). Cellular automaton modeling of the Autobahn traffic in North Rhine-Westphalia. In I. Troch and F. Breitenecker (Eds.), *Proc. of the 4th MATHMOD, 4th IMCAS Symp. on Mathematical Modelling*, Vienna, pp. 1322–1331.
- Hamming, R. (1983). *Digital Filters*. Englewood Cliffs, N.J.: Prentice-Hall.
- Hamza, M. H. (Ed.) (2002). *Proc. of the 6th IASTED Int. Conference of Internet and Multimedia Systems and Applications*, Hawaii. ACTA Press.
- Harvey, A. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- Haykin, S. S. (1990). *Kalman Filtering and Neural Networks*. New York: John Wiley and Sons.
- Hecht-Nielsen, R. (1990). *Neurocomputing*. Massachusetts: Addison-Wesley.
- Helbing, D. (1997). *Verkehrs dynamik*. Berlin: Springer. (in German).
- Helbing, D., A. Hennecke, and M. Treiber (1999). Phase diagram of traffic states in the presence of inhomogeneities. *Phys. Rev. Lett.* 82(21), 4360–4363.
- Helbing, D., H. Herrmann, M. Schreckenberg, and D. Wolf (Eds.) (2000). *Traffic and Granular Flow '99*, Heidelberg. Springer.
- Hoops, M., R. Kates, and H. Keller (2000). Bewertung von Verfahren zur Erkennung von Störungen im Verkehrsablauf in Theorie, Praxis und Simulation. In *Forschung Straßenbau und Straßenverkehrstechnik*, Volume 797. Bonn: Bundesministerium für Verkehr.
- Hounsell, N. B. and S. Ishtiaq (1997). Journey time forecasting for dynamic route guidance systems in incident conditions. *Int. J. of Forecasting* 13(1), 33–42.
- Hurdle, V., E. Hauer, and G. Steuart (Eds.) (1981). *Proc. of the 8th Int. Symp. on Transp. and Traffic Theory*, Toronto. University of Toronto Press.
- Innamaa, S. (2001). Short term prediction of highway travel time using MLP neural networks. See ITS International [2001].
- ITS International (Ed.) (1997). *Proc. of the 4th World Congress on Intelligent Transport Systems*, Berlin. ITS World Congress, CD-ROM.
- ITS International (Ed.) (1998). *Proc. of the 5th World Congress on Intelligent Transport Systems*, Seoul. ITS World Congress, CD-ROM.
- ITS International (Ed.) (1999). *Proc. of the 6th World Congress on Intelligent Transport Systems*, Toronto. ITS World Congress, CD-ROM.

- ITS International (Ed.) (2001). *Proc. of the 8th World Congress on Intelligent Transport Systems*, Sydney. ITS World Congress, CD-ROM.
- ITS International (Ed.) (2002). *Proc. of the 9th World Congress on Intelligent Transport Systems*, Chicago. ITS World Congress, CD-ROM.
- ITS International (Ed.) (2003). *Proc. of the 10th World Congress on Intelligent Transport Systems*, Madrid. ITS World Congress, CD-ROM.
- Iwasaki, M. and K. Saito (1999). Short term prediction of speed fluctuations on a motorway using historical patterns. See ITS International [1999].
- Jacobs, O. (1993). *Introduction to Control Theory*. Oxford: Oxford University Press.
- Janko, J. (1994). *Probleme der Reisezeitprognose in einem Leitsystem für den Straßenverkehr*. Dissertation, TU Berlin, Berlin.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. of Basic Engineering, Transactions ASME, Series D* 82, 35–45.
- Kantz, H. and T. Schreiber (1999). *Nonlinear Time Series Analysis*. Cambridge: Cambridge University Press.
- Kaufman, L. and P. Rousseeuw (1990). *Finding Groups in Data - An Introduction to Cluster Analysis*. New York: John Wiley and Sons.
- Kaumann, O. (2000). Online-Simulation von Autobahnverkehr: Kopplung von Verkehrsdaten und Simulation. Diplomarbeit, University Duisburg-Essen, Duisburg.
- Kaumann, O., K. Froese, R. Chrobok, J. Wahle, L. Neubert, and M. Schreckenberg (2000). On-line simulation of the freeway network of North Rhine-Westphalia. See Helbing et al. [2000], pp. 351–356.
- Kennel, M. B., R. Brown, and H. D. I. Abarbanel (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45(6), 3403–3411.
- Kerner, B. and H. Rehborn (1996). Experimental properties of complexity in traffic flow. *Phys. Rev. E* 53(5), R4275–R4278.
- Kerner, B. S., S. L. Klenov, and D. E. Wolf (2002). Cellular automata approach to three-phase traffic theory. *J. Phys. A* 35, 9971–10013.
- Kim, C. (1994). *Development and Evaluation of Traffic Prediction Systems*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg.

- Kirby, H. R., S. M. Watson, and M. S. Dougherty (1997). Should we use neural networks or statistical models for short term mororway traffic forecasting? *Int. J. of Forecasting* 13(1), 43–50.
- Kleinhans, M. (2002). Klassifikation von Verkehrsdaten mit Clusterverfahren. Diplomarbeit, University of Cologne, Cologne.
- Klukas-Illen, M. and R. E. Weiss (2001). *Verkehrsbeeinflussung auf BAB in Nordrhein-Westfalen - Lastenheft Unterzentrale*. Aachen: Heusch/Boesfeld GmbH.
- Knospe, W. (2002). *Synchronized traffic: Microscopic modeling and empirical observations*. Dissertation, University Duisburg-Essen, Duisburg.
- Knospe, W., L. Santen, A. Schadschneider, and M. Schreckenberg (2000). Towards a realistic microscopic description of highway traffic. *J. Phys. A* 33, L477–L485.
- Knospe, W., L. Santen, A. Schadschneider, and M. Schreckenberg (2002a). A realistic two-lane model for highway traffic. *J. Phys. A* 35, 3369–3388.
- Knospe, W., L. Santen, A. Schadschneider, and M. Schreckenberg (2002b). Single-vehicle data of highway traffic: microscopic description of traffic phases. *Phys. Rev. E* 65, 056133.
- Kochs, A., N. Meerkamp, and B. Steinhauer (2004). Wie definiert ein Autofahrer Stau? - Nutzung und Akzeptanz von Verkehrslageinformationen im Internet. *Straßenverkehrstechnik* 48(5), 221–226.
- Kohonen, T. (1995). *Self Organising Maps*. Berlin: Springer.
- Koshi, M., M. Iwasaki, and I. Ohkura (1981). Some findings and an overview on vehicular flow characteristics. See Hurdle et al. [1981], pp. 403–426.
- Krause, S. (1988). *Ein Selbstregulierendes Prognoseverfahren zur Verkehrsbeeinflussung*. Dissertation, Rheinisch-Westfälische Technische Hochsule Aachen, Aachen.
- Kühne, R. (1984). Macroscopic freeway model for dense traffic – Stop-start waves and incident detection. See Volmuller and Hamerslag [1984], pp. 21–42.
- Kühne, R. (1987). Freeway speed distribution and acceleration noise – Calculations from a stochastic continuum theory and comparison with measurements. See Gartner and Wilson [1987], pp. 119–137.
- Kühne, R. (1991). Freeway control using a dynamic traffic flow model and vehicle reidentification techniques. *Transp. Res. Record* 1320, 251–259.

- Kurkjian, A., S. Gershwin, P. Houpt, A. Willsky, and E. Chow (1980). Estimation of roadway traffic density on freeways using presence detector data. *Transp. Science* 14(3), 232–261.
- Laffont, S., G. Regniet, and G. Schmidt (1996). *Stau-Prognose für Baustellen auf Autobahnen in Bayern*. Aachen: Heusch/Boesefeld GmbH.
- Lambole, C., J. Satucci, and M. Danech-Pajouh (1997). 24 or 48 Hour Advance Traffic Forecast in Urban and Periurban Environments: The Example of Paris. See ITS International [1997].
- Lan, C.-J. (2001). A recursive traffic flow predictor based on dynamic generalized linear model framework. See Stone et al. [2001], pp. 410–415.
- Ledoux, C. (1997). An urban traffic flow model integrating neural networks. *Transp. Res. C* 5(5), 287–300.
- Lee, H. K., R. Barlovic, M. Schreckenberg, and D. Kim (2004). Mechanical restriction versus human overreaction triggering congested traffic states. *Phys. Rev. Lett.* 92(23), 8702.
- Lee, S., D. Kim, J. Kim, and B. Cho (1998). Comparison of models for predicting short term travel speeds. See ITS International [1998].
- Lee, Y.-I. and C. Y. Choi (1998). Development of a link travel time prediction algorithm for urban expressway. See ITS International [1998].
- Leerkamp, B. (1999). Erhebungs- und Hochrechnungsverfahren des Kfz-Verkehrs für kommunale Planungsaufgaben. *Straßenverkehrstechnik* 12, 612.
- Lesort, J. (Ed.) (1996). *Proc. of the 13th Int. Symp. on Transp. and Traffic Theory*, Amsterdam. Pergamon.
- Leutzbach, W. (1998). *Introduction to the Theory of Traffic Flow*. Berlin: Springer.
- Leutzbach, W. and W. Siegener (1974). *Untersuchung über die Möglichkeit, aus der Analyse von Meßreihen von Dauerzählstellen eine Prognose für zukünftige Verkehrsbelastungen zu entwickeln*. Karlsruhe: Forschungsauftrag 3707 des Bundesministeriums für Verkehr.
- Ma, S., G. He, and S. Wang (2001). A traffic flow forecast supported system based multi-agent. See Stone et al. [2001], pp. 620–624.
- Marinossen, S., R. Chrobok, A. Pottmeier, J. Wahle, and M. Schreckenberg (2002). Simulation framework for the Autobahn traffic in North Rhine-Westphalia. In S. Bandini, B. Chopard, and M. Tomassini (Eds.), *Cellular Automata, Proc. of the 5th Int. Conference on Cellular Automata for Research and Industry*, Geneve, pp. 315–324.

- Matsumura, S., H. Yamashita, S. Iwaki, and H. Sugimura (1998). Experimental verification of travel-time prediction method. See ITS International [1998].
- Maybeck, P. S. (1979). *Stochastic Models, Estimation, and Control*, Volume 1. New York: Academic Press.
- McFadden, J., W.-T. Yang, and S. Durrans (2001). Application of artificial neural networks to predict speeds on two-lane rural highways. See Transp. Research Board [2001].
- McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. L. Cam and J. Neyman (Eds.), *Proc. of 5th Berkeley Symp. on Mathematical Statistics and Probability*, Volume 1, Berkeley, pp. 281–297. University of California.
- Meißner, F. (1998). *Die Prädiktion der Verkehrsentwicklung in großen Straßen- netzen und ihre Realisierung in einem Rechnerverbund*. Dissertation, Technische Universität Hamburg-Harburg, Hamburg.
- Minsky, M. and S. Papert (1969). *Perceptrons*. Cambridge: MIT Press.
- Murat, Y. (1999). Prediction of Traffic Volumes in Bosphorus Bridge using Artificial Neural Networks. In M. Pursula (Ed.), *Proc. of the 11th Mini-Euro Conference on Artificial Intelligence in Transportation Systems and Science*, Helsinki, pp. XLII. Helsinki University of Technology.
- Musha, T. and H. Higuchi (1978). Traffic current fluctuation and the Burgers equation. *Japanese J. of Applied Physics* 17(5), 811–816.
- Nagel, K., J. Esser, and M. Rickert (2000). Large-scale traffic simulations for transport planning. In D. Stauffer (Ed.), *Annual Review of Computational Physics*, Volume VII, pp. 151–202. Singapore: World Scientific.
- Nagel, K. and H. Herrmann (1993). Deterministic models for traffic jams. *Physica A* 199, 254–269.
- Nagel, K. and M. Paczuski (1995). Emergent traffic jams. *Phys. Rev. E* 51, 2909–2918.
- Nagel, K. and M. Schreckenberg (1992). A cellular automaton model for freeway traffic. *J. Physique I* 2(199), 2221–2229.
- Nair, A. S., J.-C. Liu, L. Rilett, and S. Gupta (2001). Non-linear analysis of traffic flow. See Stone et al. [2001], pp. 683–687.
- Neubert, L. (2000). *Statistische Analyse von Verkehrsdaten und die Modellierung von Verkehrsfluß mittels zellulärer Automaten*. Dissertation, University Duisburg-Essen, Duisburg.

- Nowotny, B., J. Asamer, K. Din, and R. Karim (2003). Classification of traffic data time series by cluster analysis, artificial neural networks and ANOVA. See ITS International [2003].
- Ober-Sundermeier, A. (2003). *Entwicklung eines Verfahrens zur Stauprognoze an Engpässen auf Autobahnen unter besonderer Berücksichtigung von Arbeitsstellen*. Dissertation, University of Kassel, Kassel.
- Otokita, T. and K. Hashiba (1998). Travel time prediction based of pattern extraction from database. See ITS International [1998].
- Park, D. and L. Rilett (1999). Forecasting freeway link travel times with a multilayer feedforward neural network. *Computer-Aided Civil and Infrastructure Engineering* 14, 357–367.
- Persaud, B. and F. Hall (1989). Catastrophe theory and patterns in 30-second freeway traffic data – implications for incident detection. *Transp. Res. A* 23(2), 103–113.
- Pottmeier, A., R. Chrobok, S. Hafstein, F. Mazur, and M. Schreckenberg (2004). OLSIM: Up-to-date traffic information on the web. In *Proc. of the third IASTED Int. Conference Communication, Internet, and Information Technology*, US Virgin Islands, pp. 571–576.
- Pottmeier, A., S. Hafstein, R. Chrobok, J. Wahle, and M. Schreckenberg (2003). The traffic state of the Autobahn network of North Rhine-Westphalia: An online traffic simulation. See ITS International [2003].
- Principe, J., W. Lefebvre, and N. Euliano (2000). *Neural Systems: Fundamentals Through Simulation*. New York: John Wiley and Sons.
- Qiao, F., H. Yang, and W. Lam (2001). Intelligent simulation and prediction of traffic flow dispersion. *Transp. Res. B* 35(9), 843–863.
- Reed, R. and R. Marks (1999). *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. Cambridge: MIT Press.
- Rice, J. and E. van Zwet (2001). A simple and effective method for predicting travel times on freeways. See Stone et al. [2001], pp. 227–232.
- Rickert, M. and P. Wagner (1996). Parallel real-time implementation of large-scale, route-plan-driven traffic simulation. *Int. J. of Mod. Phys. C* 7(2), 133–153.
- Rilett, L. and D. Park (2001). Direct forecasting of freeway corridor travel times using spectral basis neural networks. *Transp. Res. Record* 1752, 140–147.
- Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

- Roozemond, D. A. (1997). Forecasting travel times based on actuated and historic data. In L. Sucharov and G. Bidini (Eds.), *Proc. of the 3rd Int. Conference on Urban Transport and the Environment*, Acquasparta, pp. 215–224.
- Sachs, L. (1984). *Applied Statistics*. New York: Springer.
- Saito, T., K. Yasui, T. Ukyo, and J. Nakano (1997). Study on traffic demand prediction methods for intelligent traffic signal control. See ITS International [1997].
- Schnieder, E. and U. Becker (Eds.) (2000). *Proc. of the 9th IFAC Symp. Control in Transp. Systems*, Braunschweig. Int. Fed. of Automatic Control.
- Schreckenberg, M. and D. Wolf (Eds.) (1998). *Traffic and Granular Flow '97*, Singapore. Springer.
- Selten, R., M. Schreckenberg, T. Chmura, T. Pitz, S. Kube, S. Hafstein, R. Chrobok, A. Pottmeier, and J. Wahle (2004). Experimental investigation of day-to-day route-choice behavior and network simulations of Autobahn traffic in North Rhine-Westphalia. In *Human Behaviour and Traffic Networks*, pp. 1–24. Springer.
- Shin, S., J. Choi, and S. Yoo (1999). Randomness of de-noised time series data for travel time forecasting. See ITS International [1999].
- Siegener, W. and W. Schmitt (1980). Prognose von Verkehrsmengen durch aktuelle Fortschreibung von Langzeiterwartungswerten. In *Forschung Straßenbau und Straßenverkehrstechnik*, Volume 316. Bonn-Bad Godesberg: Bundesministerium für Verkehr.
- Smith, B. L. (1995). *Forecasting freeway traffic flow for intelligent transportation systems application*. PhD thesis, University of Virginia, Charlottesville.
- Smith, B. L., B. M. Williams, and R. K. Oswald (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. C* 10(7–8), 1133–1152.
- Sommer, B. (2000). *9. Koordinierte Bevölkerungsvorausberechnung des Statistischen Bundesamtes*. Wiesbaden: Statistisches Bundesamt.
- Stephanedes, Y. J., P. G. Michalopoulos, and R. A. Plum (1981). Improved estimation of traffic flow for real-time control. *Transp. Res. Record* 795, 28–39.
- Stone, B., P. Conroy, and A. Broggi (Eds.) (2001). *Proc. of the 4th Int. IEEE Conference on Intelligent Transp. Systems*, Oakland. Institute of Electrical and Electronics Engineers.
- Sun, H., H. X. Liu, H. Xiao, and B. Ran (2003). Short term traffic forecasting using the local linear regression model. *Transp. Res. Record* 1836, 143–150.

- Taylor, M. (Ed.) (2002). *Proc. of the 15th Int. Symp. on Transp. and Traffic Theory*, Oxford. Pergamon.
- Terrell, G. (1999). *Mathematical Statistics: A Unified Introduction*. New York: Springer.
- Thomopoulos, N. (1980). *Applied Forecasting Methods*. Englewood Cliffs: Prentice-Hall.
- Transp. Research Board (Ed.) (1999). *Proc. of the 78th Annual Meeting of the Transp. Research Board*, Washington D.C. National Academies Press, CD-ROM.
- Transp. Research Board (Ed.) (2001). *Proc. of the 80th Annual Meeting of the Transp. Research Board*, Washington D.C. Mira Digital Publishing, CD-ROM.
- Trigg, D. and A. Leach (1967). Exponential smoothing with an adaptive response rate. *Operations Research Quarterly* 25(3), 53–59.
- van Arem, B., M. van der Vlist, M. Muste, and S. Smulders (1997). Travel time estimation in the GERDIEN project. *Int. J. of Forecasting* 13(1), 73–85.
- van der Voort, M., M. Dougherty, and S. Watson (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transp. Res. C* 4(5), 307–318.
- van Iseghem, S. and M. Danech-Pajouh (1999). Forecasting traffic one or two days in advance - an intermodal approach. *Recherche Transports Sécurité* 65, 79–97.
- van Lint, J. (2004). *Reliable Travel Time Prediction for Freeways*. PhD thesis, Delft University of Technology, Delft.
- Vemuri, A., M. Poycarpou, and P. Pant (1998). Short term forecasting of traffic delays in highway construction zones using on-line approximators. *Mathematical and Computer Modeling* 27(9–11), 311–322.
- Volmuller, I. and R. Hamerslag (Eds.) (1984). *Proc. of the 9th Int. Symp. on Transp. and Traffic Theory*, Utrecht. VNU Science Press.
- Wahle, J., A. Bazzan, F. Klügl, and M. Schreckenberg (2000). Anticipatory traffic forecast using multi-agent techniques. See Helbing et al. [2000], pp. 87–92.
- Wahle, J., R. Chrobok, A. Pottmeier, and M. Schreckenberg (2002). A microscopic simulator for freeway traffic. *Network and Spatial Economics* 2, 371–386.
- Weiss-Electronic (2001). *Hinweise zur Schleifenverlegung*. Trier: Weiss-Electronic.

- Whiteley, J. R. and J. Davis (1993). Qualitative interpretation of sensor patterns. *Int. J. of Forecasting* 8(2), 54–63.
- Whittaker, J., S. Garside, and K. Lindveld (1997). Tracking and predicting a network traffic process. *Int. J. of Forecasting* 13(1), 51–61.
- Wild, D. (1996). *Die Prognose von Verkehrsstärken anhand klassifizierter Granulien*. Dissertation, Technische Hochschule Karlsruhe, Karlsruhe.
- Wild, D. (1997). Short term forecasting based on a transformation and classification of traffic volume time series. *Int. J. of Forecasting* 13(1), 63–72.
- Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: An evaluation of ARIMAX modeling. See Transp. Research Board [2001].
- Williams, B. M. and B. Smith (1999). Traffic condition forecasting for ITS operations. See ITS International [1999].
- Wolf, D. E., M. Schreckenberg, and A. B. (eds.) (1996). *Traffic and Granular Flow*. Singapore: World Scientific.
- Wright, C. (1974). Estimating daily traffic totals from incomplete data. See Buckley [1974], pp. 387–404.
- Yang, S. and G. A. Davis (2002). Bayesian estimation of classified mean daily traffic. *Transp. Res. A* 36, 365–382.
- Yin, H., S. Wong, J. Xu, and C. Wong (2002). Urban traffic flow prediction using a fuzzy-neural approach. *Transp. Res. C* 10(2), 85–98.
- You, J. and T. Kim (2000). Development and evaluation of a hybrid travel time forecasting model. *Transp. Res. C* 8, 231–256.
- Yukawa, S. and M. Kikuchi (1996). Density fluctuations in traffic flow. *J. Phys. Soc. Japan* 65(4), 916–919.
- Yun, S., S. Namkoong, J. Rho, S. Shin, and J. Choi (1998). A performance evaluation of neural network models in traffic volume forecasting. *Mathematical and Computer Modeling* 27(9–11), 293–310.
- Zackor, H. (1976). Untersuchung von Steuerungsmodellen zur Verkehrsstromführung mit Hilfe von Wechselwegweisern - Teil III: Entwicklung von Steuerungsmodellen. In *Forschung Straßenbau und Straßenverkehrstechnik*, Volume 199. Bonn-Bad Godesberg: Bundesministerium für Verkehr.
- Zackor, H., T. Röhr, A. Lindenbach, W. Balz, and H. Frik (1996). Entwicklung von Verfahren zur großräumigen Prognose der Verkehrsentwicklung und Folgerungen für den Datenaustausch von Verkehrsrechnerzentralen. In *Forschung*

- Straßenbau und Straßenverkehrstechnik*, Volume 263. Bonn-Bad Godesberg: Bundesministerium für Verkehr.
- Zarchan, P. and H. Musoff (2000). *Fundamentals of Kalman Filtering: A Practical Approach*. American Institute of Aeronautics and Astronautics.
- Zhang, X. and G. Hu (1995). 1/f noise in a two-lane highway traffic model. *Phys. Rev. E* 52(5), 4664–4668.

