



Lehrstuhl für Verkehrstechnik  
Technische Universität München  
Univ.-Prof. Dr.-Ing. Fritz Busch

Schriftenreihe Heft 6

Svetlana Vukanovic

Intelligent link control framework with empirical  
objective function: INCA

München 2009

Die Deutsche Bibliothek – CIP Einheitsaufnahme

Vukanovic, Svetlana:

Intelligent link control framework with  
empirical objective function: INCA

Hrsg.: Fritz Busch, München, 2009

Schriftenreihe des Lehrstuhls für Verkehrstechnik  
der Technischen Universität München, Heft 6

Zugl.: München, Techn. Univ., Diss., 2006

ISBN 978-3-937631-05-9

Copyright ©

Lehrstuhl für Verkehrstechnik der Technischen Universität München 2009

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Druck: TypeSet GmbH, Ismaning

ISBN 978-3-937631-05-9

ISSN 1612-9431

Lehrstuhl für Verkehrstechnik · Institut für Verkehrswesen

Technische Universität München · 80333 München

Telefon: 089 / 289 – 22438 · Telefax: 089 / 289 – 22333 · E-Mail: [info@vt.bv.tum.de](mailto:info@vt.bv.tum.de)

Internet: [www.vt.bv.tum.de](http://www.vt.bv.tum.de)

Bei der vorliegenden Veröffentlichung handelt es sich um die Dissertation von  
Frau Dr.-Ing. Svetlana Vukanovic

Vorsitzender: Univ.-Prof. Dr.-Ing. Günther Leykauf, TU München

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Fritz Busch, TU München
2. Univ.-Prof. Dr.-Ing. Dr. Dusan Teodorovic, University of  
Belgrade and Virginia Tech State University,  
Washington
3. Univ.-Prof. i.R. Dr./UCB Hartmut Keller, TU München



## Acknowledgments

I am deeply indebted to everyone who has supported me in the past five years. I wish to thank my advisor, Professor Fritz Busch, for being supportive and understanding and always providing fruitful discussions. Sincere appreciation is extended to Professor Hartmut Keller for giving me the chance to start this endeavour in the first place as well as for his moral, mental, personal and scientific support. I wish to thank Professor Dusan Teodorovic for serving as my co-advisor. He has been of enormous help in providing feedback and commentary on my work and showing me how to approach a problem in an innovative way. Special thanks goes to Professor Markos Papageorgiou for providing feedback and commentary on my work and giving inspiring and useful advice.

Many thanks to all those colleagues and friends who offered their ideas, thoughts and humour which helped me grow and produce the results I needed. This goes to all the colleagues at the former Fachgebiet Verkehrstechnik und Verkehrsplanung with whom I have had the privilege to work at the beginning of my stay in Germany and all the colleagues at Transver for offering me the pleasant and inspiring place during my research. Particularly special thanks go to Dr. Ronald Kates for his enthusiasm, showing me how wonderful research can be, as well as for the rich discussions and suggestions which helped shape my thesis work; Samuel Denaes for asking the right questions and offering personal and professional encouragement; Klaus Bogenberger for helping me to get used to the new surroundings and supporting me in my scientific and personal development; and my great friends and colleagues Andrea David and Esin Bozyazi for giving me the example and being great companions during all this time.

I owe a big word of thanks to Bavarian (motorway) authorities (OBB, ABDS) and people at traffic control centre Freimann (VRZ), for unselfishly offering needed information and support in their interpretation.

Finally, many thanks, hugs and kisses go out to my friends and family, who were there for me when I needed them. And as they know I needed them for much more than a piece of paper. Hoping not to do injustice to all the others I especially thank Marija, Vlada, Olga (your friendship made me being a richer person); Svetlana (for being as you are); Katarina, Borko, Dejan, Srđan, Vladimir (without you life would not be half as adventurous); Filip (your genuinely straightforward way of reasoning offers a completely new view to the world around us); Dimitrije (your love of life is an inspiration), Sandra, Milka, Maja, Ivan (you helped me to understand people better); and cousins Jovan, Iva and Dušanka (growing up with you was pleasure).

None of this would have been possible without the extraordinary love, support and encouragement of my parents Smiljan and Gordana (technique does after all run in the family), and my sister Ksenija (you are a wonderful being).



## Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Context.....	2
1.2	Research objectives and scope.....	6
1.3	Pursued approach.....	8
1.4	Outline .....	10
<b>2</b>	<b>Link control systems.....</b>	<b>11</b>
2.1	Traffic flow problem .....	12
2.1.1	Efficiency .....	12
2.1.2	Safety.....	15
2.2	Components of link control system.....	18
2.2.1	Data collection systems .....	20
2.2.2	Communication component (variable message signs) .....	21
2.2.3	Importance of human factor .....	22
2.2.4	Warning and harmonisation strategies .....	23
2.3	Brief history and overview of world practice.....	26
2.3.1	History .....	26
2.3.2	Practice .....	27
2.4	Link control models: State-of-the-art.....	28
2.4.1	“Heuristic” approaches .....	29
2.4.2	Optimal control approaches.....	34
2.4.3	Existing operational measure of performances .....	35
2.4.3.1	Efficiency: total time spent in the system.....	35
2.4.3.2	Safety: Detection versus false alarm rate .....	36
2.4.3.3	Integrated performance measures.....	37
2.5	Effects of link control systems .....	38
2.5.1	Traffic safety .....	38
2.5.2	Efficiency .....	40

<b>2.6 Conclusion .....</b>	<b>41</b>
<b>3 Data driven model for link control on motorways using an empirical objective function: INCA .....</b>	<b>45</b>
<b>3.1 Data driven approaches.....</b>	<b>48</b>
3.1.1 Model bias, variance and Parsimony principal .....	49
3.1.2 Optimising model parameters and model complexity .....	50
3.1.3 Data driven methods in the context of link control.....	51
3.1.4 Common components of data-driven approaches.....	55
<b>3.2 INCA general link control framework.....</b>	<b>55</b>
<b>3.3 Objective function .....</b>	<b>58</b>
3.3.1 Definition of an incident .....	58
3.3.2 Analysis of typical link control situations .....	60
3.3.3 Approach for estimation of link control effects .....	63
3.3.4 Estimation of the accident time densities.....	64
3.3.5 Reconstruction of the traffic situation at the section.....	66
3.3.6 Modelling of link control effects .....	67
3.3.6.1 Calculation of time costs.....	68
3.3.6.2 Warning.....	69
3.3.6.3 Harmonisation.....	70
3.3.7 Calculation of the objective function – summary .....	71
3.3.8 Objective function parameters .....	72
<b>3.4 Data management .....</b>	<b>73</b>
3.4.1 Data checking.....	73
3.4.2 Plausibility checks.....	74
3.4.3 Data replacement (completion).....	75
<b>3.5 Knowledge base.....</b>	<b>77</b>
3.5.1 Utilised information .....	78
3.5.2 Determining traffic state .....	82
<b>3.6 Control model.....</b>	<b>84</b>
3.6.1 Hierarchical control approach.....	86



---

3.6.2	Local “segment-based” control .....	87
3.6.2.1	Usage of traffic state identification .....	88
3.6.2.2	Model for warning .....	89
3.6.2.3	Model for harmonisation decision .....	92
3.6.2.4	Control termination rule .....	95
3.6.3	Global “supervisor” layer .....	96
<b>3.7</b>	<b>Optimisation.....</b>	<b>98</b>
3.7.1	Parameters and constraints .....	100
3.7.2	Parameter optimisation using Downhill simplex method .....	101
3.7.3	Estimation of the generalisation error using Bootstrap .....	103
3.7.3.1	Aggregation of the results from Bootstrap samples .....	104
3.7.4	Reduction of the degrees of freedom using Ridge regression .....	106
3.7.5	Adjustment of the objective function .....	107
3.7.6	Optimisation process .....	108
3.7.7	Generating the ideal control .....	111
<b>3.8</b>	<b>Assessment of control performances .....</b>	<b>113</b>
3.8.1	Assessment objectives and indicators .....	113
3.8.2	Assessment analysis .....	114
3.8.3	Assessment indicators .....	115
3.8.4	Reference cases.....	116
<b>3.9</b>	<b>Conclusion .....</b>	<b>116</b>
<b>4</b>	<b>Technical implementation.....</b>	<b>118</b>
<b>4.1</b>	<b>INCA control software (INCAS).....</b>	<b>119</b>
<b>4.2</b>	<b>INCAS GUI.....</b>	<b>121</b>
<b>4.3</b>	<b>On-line operation in Munich motorway control centre.....</b>	<b>123</b>
<b>5</b>	<b>Optimisation and evaluation results .....</b>	<b>124</b>
<b>5.1</b>	<b>Test site .....</b>	<b>125</b>
<b>5.2</b>	<b>Training and test data characteristics .....</b>	<b>126</b>
5.2.1	Data quality .....	126
5.2.2	Number of speed drops.....	128

5.2.3	Accidents and work zones.....	129
5.2.4	Optimisation and evaluation – data and approach used.....	130
<b>5.3</b>	<b>Objective function evaluation .....</b>	<b>131</b>
5.3.1	Introduction.....	131
5.3.2	Evaluation of the ideal control.....	131
5.3.3	Empirical examples.....	134
5.3.4	Summary .....	135
<b>5.4</b>	<b>Control model optimisation.....</b>	<b>136</b>
5.4.1	Optimisation approach and settings .....	136
5.4.2	Choice of the harmonisation model .....	137
5.4.3	Choice of warning model.....	138
5.4.4	Influence of the parameters of the objective function.....	141
5.4.5	Influence of the Bootstrap.....	143
5.4.6	Influence of Ridge regression .....	145
5.4.7	Convergence.....	147
5.4.8	Overview of the optimised (INCA) and current (MARZ) parameters.....	148
<b>5.5</b>	<b>Evaluation of the INCA on-line performances.....</b>	<b>149</b>
5.5.1	Applied optimisation approach and derived control models .....	149
5.5.2	On-line results .....	151
5.5.3	Empirical examples.....	156
<b>5.6</b>	<b>Extendibility and flexibility.....</b>	<b>160</b>
5.6.1	Changes in the warning model.....	160
5.6.2	Performance assessment .....	161
<b>5.7</b>	<b>Summary.....</b>	<b>162</b>
<b>6</b>	<b>Conclusion and future research.....</b>	<b>164</b>
<b>6.1</b>	<b>Summary and conclusion .....</b>	<b>164</b>
<b>6.2</b>	<b>Future research .....</b>	<b>168</b>
6.2.1	Improved knowledge about the link control effects.....	168
6.2.2	Extension and refinement of certain model aspects.....	168
6.2.3	Additional sources of information .....	170

---

6.2.4 On-line close-loop tests .....	170
6.2.5 Application of INCA framework in other areas .....	171
<b>Nomenclature .....</b>	<b>172</b>
<b>References .....</b>	<b>175</b>
<b>A. Appendix.....</b>	<b>186</b>
A.1. Automatic incident detection/estimation/prediction algorithms .....	186
A.2. DR-FAR .....	191
A.3. Downhill Simplex.....	194
A.4. Accidents, road works and weather conditions .....	196

## Zusammenfassung

In Deutschland werden bis 2007 1.270 km des Autobahnnetzes mit Streckenbeeinflussungsanlagen (SBA) ausgestattet (BMVBF, 2002A). Das Hauptpotential von SBA liegt in der dynamischen Anzeige von Geschwindigkeitsbeschränkungen auf Streckenabschnitten, womit unter Angabe von Umfeld- und Verkehrsbedingungen (z.B. Nebel oder Stau), die Unfallwahrscheinlichkeiten reduziert und die Kapazität erhöht wird.

SBA sollten für die Auswahl der Steuerungsentscheidung und die Erfüllung ihrer Zielsetzung verschiedene Indikatoren und Algorithmen benutzen. Vorhandene Modelle verwenden in der Regel strikte Wenn-Dann-Regeln, die bei empfindlichen Algorithmen häufig zu restriktive Steuerungen aktivieren. Außerdem werden Kapazitätsbetrachtungen und Sicherheitsüberlegungen meist getrennt voneinander betrachtet. Das ist hauptsächlich auf einen fehlenden Indikator zurückzuführen, der beide SBA-Effekte kombiniert. Zur Lösung dieser Problematik soll in der vorliegenden Arbeit eine Methode entwickelt werden, die ausreichend flexibel für Optimierungsalgorithmen ist und für die Interpretation von veränderlicher Zuverlässigkeit und Priorität der vorhandenen Information (Rohdaten, Indikatoren und Algorithmen) geeignet ist.

In der Arbeit wird das umfassende, praktisch anwendbare, adaptive Steuerungssystem INCA entwickelt und vorgestellt, das auf einem datengetriebenen (data-driven) Ansatz aufbaut. Die entscheidenden Komponenten des Systems sind die Kosten-Nutzen basierte Zielfunktion, das flexible Informationsfusionsmodell und der Optimierungsprozess.

Die Kosten der Schaltung einer Geschwindigkeitsbeschränkung werden für die Zielfunktion in Relation zu denen ohne Schaltung geschätzt, indem die Zeitverluste der Fahrer berechnet werden, die sich an die Beschränkung halten. Wenn Situationen mit und ohne Steuerung berücksichtigt werden, beschreibt das Modell auch relative Nutzen, die sich aufgrund einer Reduktion der Unfallwahrscheinlichkeit infolge einer rechtzeitigen Warnung oder Verkehrsharmonisierung ergeben. Die Zielfunktion hat verschiedene konfigurierbare Parameter, die an die jeweiligen Nutzeranforderungen angepasst werden können.

Das datengetriebene Modell kombiniert die verfügbaren Informationen auf flexible Art und Weise, was die veränderliche Zuverlässigkeit und die verschiedenen Randbedingungen berücksichtigt. Die Hauptherausforderung bei der Entwicklung war, ein Modell zu finden, das verständlich und steuerbar ist, aber trotzdem die hohen Anforderungen an die Genauigkeit erfüllt. Die Steuerungsentscheidung wird in zwei Stufen herbeigeführt: zunächst für jede einzelne Kontrollstation (AQ) und danach koordiniert für den gesamten Streckenabschnitt. Jede Kontrollstation ist mit einer „intelligenten“ lokalen Steuerungseinheit ausgerüstet, die für die Auswahl der besten Steuerungsentscheidung in der jeweiligen Situation für den folgenden Streckenabschnitt zuständig ist. Die Entscheidung für das Ziel „Harmonisierung“ wird mit dem modifizierten Belastungsalgorithmus nach dem MARZ (BAST, 1999) herbeigeführt. Die

Gefahrenwarnung erfolgt auf der Basis eines Logit-basierten Regressionsmodells, das das Gewicht der vorhandenen Information unter definierten Verkehrszuständen bestimmt. Andere Steuerungsalgorithmen (z.B. Wetterwarnung) sind im nächsten Schritt enthalten. Zusätzlich werden auf globaler Ebene verschiedene globale Regeln verwendet die zu einer Koordination der Steuerungsentscheidungen beitragen.

Das Ziel des Optimierungsprozesses ist, den bestmöglichen Parameter aus einem minimalen Datenumfang zu ermitteln und damit die Robustheit und Übertragbarkeit des Modells sicher zu stellen. Um das zu erreichen, werden ineffektive oder instabile Parameter mittels der „Ridge-Regression“ ausgeschlossen. Die Verteilung der Parameter und somit die Varianz des Modells mit vom Umfang begrenzten Trainingsdaten wird mittels der Bootstrap Methode ermittelt. Die Suche der optimalen Parameter wird mittels der Downhill Simplex Methode durchgeführt. Weiterer Teil des Optimierungsprozesses ist ein Genetischer Algorithmus.

Um INCA zu testen und zu bewerten, wurde das vorgeschlagene Steuerungssystem in der Verkehrsrechnerzentrale Freimann bei München installiert. Als Teststrecke wurde ein 38km langer Autobahnabschnitt der BAB A8 von Salzburg in Fahrtrichtung München mit 20 AQ ausgewählt. Über zwei Monate wurden Daten gesammelt und für die Kalibrierung verwendet. INCA wurde in einem offenen Regelkreis über weitere zwei Monate getestet.

Die generierten „idealen“ Schaltungen wurden bewertet und konnten zeigen, dass die Zielfunktion Schaltungen hinsichtlich des gewünschten Verhaltens angemessen „belohnt“ oder „bestraft“. Des Weiteren wird gezeigt, dass INCA die Wichtigkeit der neuen Indikatoren erkannte und die Gewichte angemessen zuweist. Die Schaltungen, die On-Line erzeugt wurden, werden mit der Zielfunktion sowie der Detektions- und Fehlalarmrate bewertet und mit Schaltungen nach aktuellen MARZ-Algorithmen sowie mit „idealen“ Schaltungen verglichen. Nach den Detektions- und Fehlalarmraten führte INCA zu signifikant besseren Schaltentscheidungen als die anderen getesteten Algorithmen, vor allem bei den Schaltungen von 80 km/h und 40 km/h, wohingegen bei Schaltungen von 60 km/h weiteres Verbesserungspotential für INCA ausgemacht wurde. Die Gesamtverbesserung im Vergleich zum bestehenden System betrug 40%.

Daneben wird nachgewiesen, dass INCA auf beliebigen Autobahnen mit Verkehrserfassungssystemen anwendbar ist, wenn eine historische Datenbasis zur Verfügung steht. Im Gegensatz zu existierenden Ansätzen bietet INCA einen systematischen Rahmen für die Steuerung von SBA. Dabei kann durch die vorgestellte Optimierungsmethode und Zielfunktion für jeden Autobahnabschnitt ein allgemeines Steuerungsmodell abgeleitet werden. Daneben bietet INCA eine Übersicht über die Wichtigkeit der jeweiligen Algorithmen und deren Zuverlässigkeit. Des Weiteren können mit INCA durch die ständige Beobachtung der Steuerungsqualität signifikante strukturelle Änderungen in Verkehrsnachfragemustern oder schwerwiegende technische Probleme in kurzer Zeit erkannt werden. Die modulare Struktur des Systems erlaubt Erweiterungen, Umkonfigurationen und die Entwicklung weiterer unabhängiger Komponenten.



# 1 Introduction

Mobility distinguishes the modern world from the world that existed centuries ago. People travel hundreds of kilometres on a daily basis to get to work, goods are efficiently transported across large distances, and information needs less than a second to be available in another part of the globe. These changes impose new challenges and requirements upon all areas of human existence, and thus upon traffic engineering as well.

Increased mobility of people, together with growing motorization, can lead to demands that are much greater than the capacity of the existing road network and higher exposure of drivers to potentially unsafe situations. When too many vehicles attempt to use a transportation infrastructure with a limited capacity – congestion occurs. Today, congestion on motorways is a common phenomenon in all industrialized countries and leads to reduced traffic safety, increased delays, air pollution, and drivers' frustrations. Many reports (FHWA, 1998; ETSC, 2001; SHELL, 2000) estimate that this trend will continue to grow. Construction of new roads is not only extremely costly, but may also have negative side effects on local social and economical surroundings. Therefore, better utilisation of the existing network and safer driving experience have become major concerns of the traffic researchers and practitioners. Motorway traffic management systems have been proposed and implemented all over the world as a strategy for improving performance of the existing road network. The main idea behind management strategies is to avoid/reduce congestion (which reduces the available capacity) by protecting (and slightly increasing) the available capacity. Typical strategies employed are link control systems (LCS), ramp metering, and guidance systems.

Developments in traffic management and control have always depended on advances in information technology. Simply, algorithms and models can work only with available information and "hardware" capacity. Consequently, the first controls (e.g. first ramp metering in early 60s (MAY, 1964) consisted of fix-time pre-defined programs. With advances in technology the first simple time-responsive algorithms and models emerged. Today, more sophisticated and advanced models that significantly improve motorway traffic performance are available, but the problems with which these new models have to cope have become different. Firstly, the initial complex non-linear problem of traffic modelling has grown in its *complexity* as the network and control dissemination tools expanded. Secondly, the major stated *goal* at the beginning was the improvement in traffic performances, but today safety as well as environmental impacts play an important role when choosing and applying traffic management strategy. Not only are these goals sometimes conflicting, but assessing the control ability to achieve them and expressing this ability as a unique and especially an operational measure is not an easy task. Thirdly, the paradigm has changed from "not having enough data" to having "too much data". This new *available data* is different in nature, frequency and reliability. Distinguishing which data is reliable and relevant for the problem

under consideration (and when) becomes a new challenge (this is not only the case in traffic management and control but in our everyday life as well). Finally, algorithms must satisfy rigorous criteria in order to be applied in practice. They have to be robust and reliable (in order to trigger high compliance of drivers in a long term), but should also be reasonably easy to operate and maintain.

In the past few decades different methods have been investigated and proposed for controlling large non-linear systems, systematic integration of available data and producing the best decisions from them. These methods are mostly coming from the domains of artificial intelligence (AI) and advanced control theory. Today, the largest potential is observed in combining knowledge from these two areas – combining the model architecture from the control theory with the learning mechanism from AI.

In this chapter the context and background of traffic management systems and link control system in general will be further explained, as well as motivation why this problem is relevant to be tackled. Subsequently, the main objectives of the dissertation will be presented, the scope of the dissertation will be narrowed, and the research approach will be explained. This thesis particularly concentrates on the methods and models for control of link control systems (with focus on variable speed limits) on motorways. The final part of this introduction then gives a brief outline of the subjects covered in each chapter of the thesis.

## **1.1 Context**

Accidents<sup>1</sup> have great effect on the safety of responders and on the mobility of the travelling public. It is estimated that, in addition to delay costs, direct economic loss due to accidents and fatalities amounts to \$200/€160 billion per year in the USA/EU (FHWA, 2004A; ETSC, 2001). Secondary accidents make up 14-18% of all accidents and are estimated to cause 18% fatalities on motorways. Although today roads are much safer place than before, there is still great potential for further improvement. The best illustration for this is the latest resolution of the European Commission to initiate and finance new projects with the ambitious aim to decrease the number of accidents on the European roads by 50% by 2010 (ETSC, 2001). Another problem with whom the modern world is faced is decreased efficiency due to disturbed operating conditions at motorways. There are several factors that can influence the operating conditions of a motorway. These factors include traffic conditions (e.g. traffic composition), recurring and non-recurring events, geometric and weather conditions.

Different strategies can be implemented in order to optimise motorway capacity usage and increase traffic safety, e.g. link control systems, ramp metering, etc. Generally, these strategies vary in terms of the degree of traffic control they provide, implementation costs,

---

<sup>1</sup> Safety professionals prefer the word “crash” to “accident” because the latter suggests that occurrence is due to pure fate and cannot be influenced by human decision. However, the word “accident” is wider accepted among all traffic professionals, as well as among general public, and will be used in this thesis.



circumstances under which they should be applied, technological requirements, and institutional considerations (e.g. jurisdiction, privacy and liability concerns).

Link control systems<sup>2</sup> consist of a succession of Variable Message Signs (VMS), with displays separated by a distance of 1 to 3 km along a motorway. The signs can display speed limits as well as other restrictive, advisory or informative information, conveyed by pictograms or short phrases. Each link control is complemented by detection equipment to measure traffic and sometimes weather data in the vicinity. Link control systems are designed to improve utilization of the existing motorway network through effective information and control mechanisms based on real-time *incident* and *congestion* detection, and recognition of situations that could lead to traffic breakdown and/or unsafe driving experience.

In Germany, the first link control system was installed in 1965, on a 30 km section of the BAB A8 corridor between Holzkirchen and Munich (ZACKOR, 1972). This first LCS installation was manually controlled, based on images from video cameras. In 1970, the Federal Ministry of Transportation (BMV) introduced first “Framework for Link Control Systems” for LCS improvement and implementation (BUSCH, 1971). Considering future development, BMV’s “Program for Link Control Systems” was introduced in 1980 through construction of 111 variable message signs, which have been regularly upgraded since. In 1997, “Merkblatt für Verkehrsrechnerzentralen und Unterzentralen” (MARZ) (BAST, 1999), was published. MARZ includes the description of algorithms and LCS control procedures, as they *should* be implemented in the existing link control systems.

LCS installations are also implemented in other industrialised countries such as the Netherlands, Austria, Spain, Switzerland, Denmark, Japan, Australia, the USA, etc. (TRB, 2000A; MIDDELHAM, 2002; DG-TREN, 2002). However, specific LCS implementation and goals vary from country to country, and even within a same country. Probably the most specific practice is in Germany. In Germany there are no general speed limits on motorways and therefore LCS has been considered as a legal means to reduce speeds when necessary, but also allowing people to drive faster when this does not increase risk for users.

The main goal of link control systems is to increase safety. Another often stated goal is better usage of available motorway capacity. Typical strategies or tactics for fulfilling these goals are (FGSV, 1992; VUKANOVIC ET AL., 2003A): harmonization<sup>3</sup> of traffic flow, warning the road users, and prohibiting truck passing. The harmonization approach uses speed limits that generally correspond to or exceed the speed at the critical density of the fundamental diagram. For fulfilling these goals, every sign can display different information (chosen from a set of predetermined available control actions – e.g. speed limit signs in increments of 10 km/h or 20 km/h).

---

<sup>2</sup> Please note that description given here is in accordance with the German standards and that each country has its own installation and implementation specification.

<sup>3</sup> In literature (especially Dutch) the term “homogenization” is also often used

Ideally, a link control system model should be able to recognize and appropriately influence traffic in all its states (free flow, congestion, incidents, etc.) and under all traffic conditions (VUKANOVIC ET AL., 2005). However, traffic situations and their characteristics can be very different. For example, from traffic flow theory it is known that in free-flow conditions state information travels in the same direction as traffic does, while in congested condition state information may also flow in the opposite direction (LIGHTHILL AND WHITHAM, 1955). Moreover, incidents can influence and change traffic characteristics in various ways. If an incident or accident happens in free-flow condition and closes one lane, demand would still be lower than the reduced available capacity and no (or minor) changes would be observed in traffic data. On the other hand, in dense traffic the same event would result in propagation of traffic in the upstream direction. Studies show (SIEBER, 2003) that regardless of how advanced the algorithm is it will always have better performances under certain specific circumstances, while under other conditions its performances would decrease: for example, there are special algorithms that deal only with accidents that happen in free-flow state and perform very well under such conditions, but produce numerous false alarms under dense traffic, and vice versa. Additionally, the accuracy of the algorithm is directly linked to the accuracy and dependability of the data collection systems (MAHMASSANI ET AL., 1999). LCS usually use loop detector data for their operation, which is known to produce sporadic errors. Reduction of these errors and possible inclusion of other available data (e.g. floating car data) will improve overall system performances. Hence, algorithm reliability is not an absolute value, but should be considered as a function of location specific traffic (and data) characteristics and situation, and integrated as such into the system.

Appropriate reaction in all such situations, and consequently utilisation of all available information, is of crucial importance for system reliability and thus also for its implementation in a real-time environment. Information integration and data fusion methods can improve performances of LCS beyond what any individual component or algorithm could achieve alone (MAHMASSANI ET AL., 1999; SIEBER, 2003). Combining sensor data as well as the indicator given by different algorithms using the same data, would exploit strengths and weaknesses of different sensors and algorithms under different conditions.

One of the distinct features of LCS is its influence on long motorway corridors, with a succession of variable message signs. For this reason controls can not be observed as independent actions, but should be considered in a broader (coordinated) context (VUKANOVIC ET AL., 2003A; HEGYI ET AL., 2003). In addition, for safety reasons it is often required that the driver not encounter a decrease in the speed limit larger than a pre-specified amount that is usually 10-20 km/h. Finally, a mechanism for prevention of too frequent changes in control signals should be incorporated into control model as well.

Control actions should be determined taking future traffic conditions and control effects into consideration. Future traffic conditions may be directly estimated by applying prediction models, or indirectly by developing a model (e.g. data driven) that would be able to learn

from the previous relations between present and following traffic states and thus automatically include this knowledge into the decision-making process.

Basically, two different approaches to link control may be found in literature: optimal control and heuristic. Optimal control approach uses the traffic flow model to estimate the influences of different speed limits and to make the final decision. Heuristic approaches focus on the combination of available information acquired through various incident detections, traffic state estimation and prediction algorithms, as well as from weather detection algorithms. Among these, automatic incident detection algorithms play the most important role in the existing systems (HOOPS ET AL., 2000), as they are responsible for triggering warning controls and are a major contributor to increased safety (the main LCS goal). Each of these approaches has its positive and negative aspects which are discussed in detail in chapter 2.4. In practice, only “heuristic” approaches are implemented.

All the above mentioned approaches lack a comprehensive measure of system performance, which would be used in control and/or control model optimisation. Proposed systems usually deal either with traffic flow or with safety performances, but they do not optimise the system to satisfy both of these goals. Traffic flow performance is usually measured by travel time spent in the system, while safety performances are usually measured by detection rate versus false alarm rate (DR-FAR). It must be stated that there are approaches attempting to achieve this, usually by introducing some additional, more or less heuristic, constraint upon the system (one of the optimal control approaches) (HEGYI ET AL., 2003). Failure to achieve this is not surprising since on the one hand knowledge regarding LCS impact is still quite limited (even though there have recently been some new studies on the subject), while on the other - combining travel time information with safety gains is not a straightforward process.

As demonstrated by implemented link control systems (OBB, 2000; DG-TREN, 2002), well-designed and maintained LCS represents an efficient motorway traffic management component in increasing traffic safety<sup>4</sup> and reducing motorway congestion, even with modest compliance reported in recent studies. The evidence strongly suggests that there is considerable additional potential to improve traffic safety by improving the performance of incident detection algorithms and traffic management strategies. Improvement of detection (and control in general) will increase the potential for effective and timely warning, while reducing false alarms should improve currently very modest compliance and thus the impact of traffic management strategies as a whole.

*Improvement of motorway performance has motivated this research, which strove to develop a new, theoretically founded but also practically applicable, link control framework that should overcome shortcomings of the existing approaches and could easily be installed into the existing traffic management centres and handled by traffic operators*

---

<sup>4</sup> In Germany reduction in accident rates of up to 35% have been reported. Studies in the Netherlands and UK report similar, although slightly lower figures.

## 1.2 Research objectives and scope

The **main objective** of this dissertation is to develop a comprehensive link control framework in order to overcome limitations of the existing approaches. The objectives of this thesis, roughly in order of appearance, include:

- Analysis of the current traffic problem on motorways, and identifying situations where link control could improve motorway operation
- Review of literature on existing link control systems;
- Development of a new link control model that could easily be applied in practice;
- Inclusion of an optimisation procedure, able to determine the most appropriate model structure and parameters, for the purpose of achieving the best possible balance between system robustness and accuracy.
- Development of a new one-dimensional objective function that could be used in system learning/optimisation, monitoring, modification and improvement;
- Calibration, testing and evaluation of the new model.

In order to overcome the previously described shortcomings of existing approaches, the **main properties** of the new link control system should be (see Figure 1-1):

- Traffic responsiveness: the determination of the variable speed limit in real time according to the current traffic status;
- Coordination: coordination of variable message panels along the motorway;
- Objective validity: validity based on the sufficiently accurate objective function and not “heuristic” rules
- Adaptiveness and/or prediction: the adaptation to temporarily changing traffic conditions and/or making control decisions based on predicted future states;
- Robustness: the developed model should be able to cope with different traffic conditions (free flow, congestion, synchronized flow, incidents, etc.), as well as with corrupted or missing data<sup>5</sup>.
- Extendibility (open system architecture): the integration of additional algorithms and/or indicators into the control system, or integration of multiple different management systems into a single integrated control system;
- Generality and transferability: no site-specific installation, and implementation on a different motorway with reasonable calibration and tuning efforts;

---

<sup>5</sup> It should be noted that robustness is a quality that is difficult to assess in absolute sense, as there are probably situations in which the model eventually fails (e.g. floods, earthquakes, etc).

- Operability: the implementation of the algorithm on a typical German motorway and its traffic management centres, without extensive costs, calibration, detection and computing efforts.

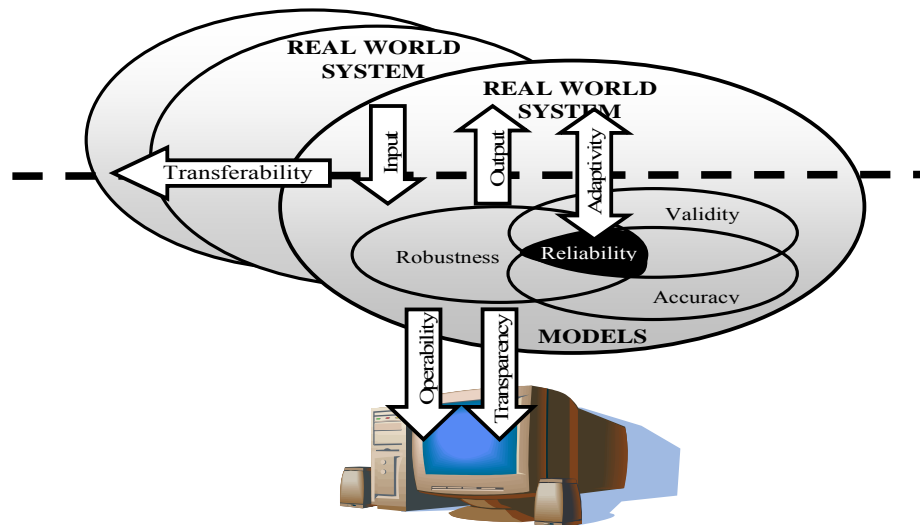


Figure 1-1: Requirements for successful traffic control (modified figure from VAN LINT, 2004)

**The focus of this dissertation** is to develop a model that is *general* and not location-specific, at least in terms of mathematical structure and input-output relations. For example, such a model should be applicable on different motorway routes, with different geometrical properties (detector and sign spacing, location of on- and off-ramps, number of lanes). General models fulfil other criteria important for successful control systems like, for example, transparency and transferability. This is due to the fact that model reliability also depends on reliability of the real world (VAN LINT, 2004), and particularly on the people and systems operating these models. The vast majority of models are far too complex and require huge efforts to operate and/or interpret them. A model that requires specific design for every location is not likely to be deployed on a larger scale. Furthermore, the traffic control model must be tractable in order to be understandable and accepted by the user. Therefore, although technology has advanced in last years, widely used models are usually of a simple structure.

Finally, **this dissertation focuses on** the development of a new one-dimensional objective function to be used in system calibration, adaptation and monitoring. The objective function should capture all major effects of link control systems, e.g. safety gains due to warning, safety gains due to harmonisation, but also travel time losses due to displayed control. Link control systems are influencing long motorways corridors and one control station influences adjoining traffic too. Therefore, the objective function should be able to estimate the spatial effects of the control (at least on the section between two control stations).

**The scope** of the research effort will be narrowed however. Link control systems have different goals and wide range of possible controls that could be displayed. This research will focus on control by variable speed limits, whereas other controls (e.g. weather controls, manual interventions) will be taken into consideration, but not investigated in detail.

Furthermore, this model should consider the possibility of its coordination with other motorway control measures (e.g. ramp control), which will be reflected in its modular structure.

*The main contribution of the research presented in this dissertation is the design, development, application, and evaluation of a new link control framework, which includes one-dimensional measure of different control effects, a flexible control model, and (model and parameter) optimisation procedure.*

### **1.3 Pursued approach**

It is generally accepted that traffic control problem is a complex, nonlinear, spatiotemporal problem. Traffic flow theory provides invaluable insight into the nature of the complexities and constraints it poses on models and methods that tackle it (KELLER, 2002A; DAGANZO, 1997; PAPAGEORGIOU, 1991). The obvious solution approach would then be to exploit traffic flow simulation model to estimate the effects of variable speed limits and select the best strategy. However, as shown in chapter 2.4.2, control of link control systems based on information exclusively from the traffic flow model, has several potentially critical pitfalls (it requires huge quantities of data, estimation or/and prognosis of boundary conditions, still limited knowledge about the effects of the link control systems, etc.) and could lead to the most accurate, but not necessarily the most robust and reliable, and therefore not the “best” model for practical implementation.

As such, we argue that a data-driven approach to derive the best controls from data directly represents a more suitable choice, given that sufficient data is available. The second compelling argument in favour of data driven approaches is that feedback processes are automatically dealt with as well, since these are also “present” in the data due to user response to the system. It simply does not matter for a model whether a specific traffic condition is the result of a feedback process following its installation, or not, as long as it is familiar with the resulting data. A carefully designed data-driven model could have a very flexible and extendable structure. If supported by appropriate mechanisms, a data-driven model could automatically decide if the new available information has any relevance and what is its importance.

Since the focus is on data-driven models, this thesis deals with link control model on motorways, provided that sufficient data regarding the particular motorway is available. This implies that some data collection system is installed and that a sufficiently large database of input (detection data and optionally ambient and external conditions) and output data (variable speed limits per sign per control period) is compiled. Historically, data collection systems consist of local detection (inductive loops, radar detectors) resulting in local aggregated characteristics (flows, local mean speeds) of traffic stream. Hence, both the new model and the objective function should be able to work with this aggregated, macroscopic data.

Link control generally pursues both “local” and “global” objectives. Local control is based on the information from the local detector(s), whereas global control can take into account predefined global objectives and/or information from detectors along the road (DENAES ET AL. 2004). In this thesis a “hierarchical” approach has been chosen: first the optimal segment control is determined and then it is modified under consideration of global constraints and dynamic, if required. “Segment” control refers to the structure of the local modules that incorporate information from local and neighbouring detector stations in order to make the best local decision. As such, this approach is neither local nor global. Moreover, one of the major features of the proposed solutions is their flexible model structure, where also information from further detection stations could be included, if necessary. Special “coordination” rules have been incorporated into the model in order to guarantee that controls at subsequent VMS are in accordance with each other.

LCS should utilize various indicators, algorithms, etc., in order to carry out a LCS control decision and achieve its objectives. For this reason, there is a need for a method which is sufficiently flexible to allow optimisation, and which is particularly capable of representing variable reliability and priority of the available information (raw data, indicators and algorithms) (VUKANOVIC ET AL., 2005). Different approaches can be used in order to combine available data. Here, bearing in mind other constraints imposed upon system (e.g. satisfying practical requirements, tractability, etc.) a (logit-based) regression model has been selected. To allow different “weighting” schemes under different traffic conditions, classification of the traffic states is introduced into the system and one such regression model is assigned to each specified traffic state.

A regression model supported with a logit function is an appealing solution, especially for practical applications, since it offers the possibility for information “weighting”, it is easy to incorporate additional mechanisms for eliminating model bias and variance, and it is flexible and tractable. It is also suitable for modelling processes where discrete decisions should be made, as is the case with link control systems. It deals with uncertainty of available information by assigning different “weights” to each of them. Here, the difference between harmonization and warning strategies is maintained by representing them with two independent components. Of these two controls, the more restrictive one is chosen as the final output from the “segment-based” speed limit model. This is especially convenient for increasing optimisation performance and system tractability. The “weights” that are assigned to each algorithm and the model decision points, are then the subject of an optimisation process carried out using the downhill-simplex model. Ridge regression has been incorporated into the model in order to eliminate/reduce correlation between the available information which is known to negatively influence performance of regression-based models (VUKANOVIC ET AL., 2004). The (re-sampling) Bootstrap method is integrated into the optimisation procedure, in order to estimate parameter (and thus model) variance, with limitedly available learning data, and allowing for the stable reduction of model complexity (elimination of irrelevant or unstable information from the model).

The complex LCS objectives as described above are modelled and represented by a cost-benefit based objective function (VUKANOVIC ET AL., 2003A; KATES ET AL., 2002). As far as travel time is concerned, we may estimate the “cost” of a speed limit compared to no speed limit by calculating the delay that would result from drivers’ obeying the control. Although it is known that compliance is *not* 100 %, the losses of effectiveness is modelled as equal to the time that *would have been* lost if all drivers had complied. This aspect will be important in avoiding false alarms. If situations with and without control are considered, this model also describes relative “benefits” (i.e., reduction of losses) associated with a reduction in the probability of an accident due to a timely warning or traffic harmonization. In the case of a warning, the benefit is attributable to removing the “surprise” that would occur in the case of a sharp speed drop without the warning. The formulation thus includes a statistical model for the relationship between accidents and sharp speed drops, which constitute an acute hazard. In the case of harmonization, in order to estimate a variable accident prevention rate, a model based on the results of KATES ET AL., 1999 and STEINHOFF, 2002 is incorporated into the objective function.

For a typical LCS, the data supply permits testing and re-calibration several times per year. Additionally, by continuously monitoring performance, significant structural changes in traffic patterns or severe technical problems can be detected within a relatively short time.

Field (empirical) data has been used for model calibration. Although the influence of existing control has been incorporated in this data it was assumed that it still represents the best information available. The model has been calibrated with different settings of the initial parameters in order to investigate their influences on the derived parameters and final model performance. The new model has been installed at Munich traffic control centre and evaluation of the system was carried out during a two-month on-line open-loop test at BAB A8 Ost (Salzburg direction Munich). Model outputs have been compared to ideal and existing control (reference case) with the mean of DR-FAR curves and the new objective function. Flexibility and ability of the proposed framework to automatically recognize more or less reliable information have been tested by comparing the model performance prior to and following the inclusion of additional indicators.

## 1.4 Outline

The remainder of this dissertation is organized as follows. *Chapter 2* gives the background and an overview of the *link control systems*. Traffic flow problems and the potential of link control systems to reduce them are described first. Next, an overview of link control systems and their major components is given, followed by the relevant state of the art, an overview of the practice in link control systems, and LCS effects reported up to the present. The chapter is concluded with the needs and possible approaches for development of a comprehensive link control framework. *Chapter 3* presents the *methodology used in this thesis*. It provides the background of the data driven approaches and the conceptual framework of a new system. It



describes the basic model hypothesis and gives a description of the proposed system architecture. Each component of the framework is subsequently described in detail. *Chapter 4* describes the *technical implementation* of the proposed framework: the developed software and its installation at the Munich traffic control centre. *Chapter 5* describes the characteristics of available test and training data, validation of the objective function, calibration approach, model flexibility, and results of the empirical-based *evaluation of the performances* of here proposed link control model. The developed model is tested with empirical (on- and off-line) data. The performance of the new model is evaluated using the objective function and DR-FAR curves, and compared to performances of the currently standard model in Germany (which was in operation during the test period) and ideal control. *Chapter 6* concludes this dissertation by highlighting the main contributions of this thesis and indicating *directions for future extension* of some of the presented ideas.

## 2 Link control systems

Link control systems (LCS) with variable message signs (VMS) have been widely implemented as strategic instruments for increasing traffic *safety* and *efficiency*. Link control systems consist of succession of VMS, with displays at a distance of 1 to 3 km along a motorway<sup>6</sup>, and complemented by detection equipment to measure traffic and weather data in the vicinity. Traffic controls such as dynamic speed limits, overtaking prohibitions, headway distance (RÄMÄ AND KULMALA, 2000), lane closure or opening as well as traffic and environmental related information can be displayed at the each VMS location (control station). Link control systems can pursue their goals through several (in literature frequently differently classified) strategies (FGSV, 1992):

1. *Harmonisation* of traffic flow, especially in situations where traffic volume is approaching a critical capacity value through variable speed limits
2. Increasing traffic safety in special situations such as fog, rain, ice, through various information and enforcement controls
3. Increasing traffic safety through warning about congestion, accident, work zones, with variable speed limits and lane closure
4. Freeing the left and, when existing, middle lane for auto drivers when truck percentage is high, through prohibiting truck passing
5. Increasing traffic safety near on- and off-ramps through the equalization of the speeds of the mainline and entrance lanes.

One of the major LCS potentials lies in dynamic speed limits which provide the possibility for timely *warning* of drivers about danger ahead so as to timely adjust their behaviour and *harmonisation* of traffic flow by high densities to prevent or postpone disturbances. In order

---

<sup>6</sup> Detector and VMS spacing varies from country to country and even within the same country. The description given here is in accordance with German standards.

to be able to fulfil these goals such critical situations should be timely identified and the action that would maximally increase safety and efficiency should be automatically selected. Hence, LCS usually provide a range of intervention options and pursue **multiple objectives** where its successful implementation depends on permanent interplay of these strategies and their effects, which are not easy to assess.

The aim of this chapter is to identify the background of link control systems and the rationale for the development of new control algorithm. To demonstrate all the positive effects of a well designed link control systems motorway safety and congestion problem will be described first. A more detailed description of automatic link control problem, as well as a brief history and state of the art and practice will be given next through the introduction of a general automatic link control framework and its major components. In the last sections of this chapter the effects of link control systems and the needs of development of a new system will be discussed and an ideal link control strategy will be presented.

## 2.1 Traffic flow problem

### 2.1.1 Efficiency

Today motorway congestion is our reality and a major cause of reduced efficiency. On congested motorways, average operating speeds are frequently less than 50 km/h in stop-and-go traffic, resulting in delays that are not only frustrating but also extremely costly and potentially unsafe. Motorway congestion occurs when traffic demand approaches or exceeds the available capacity of the motorway system. Traffic demands vary significantly depending on the season of the year, the day of the week, and even the time of the day. Also, the capacity, often mistakenly considered to be constant, can vary due to weather, construction zones, or traffic incidents. Traffic congestion can be recurrent or non-recurrent, depending on whether the demand or the capacity factor is out of balance.

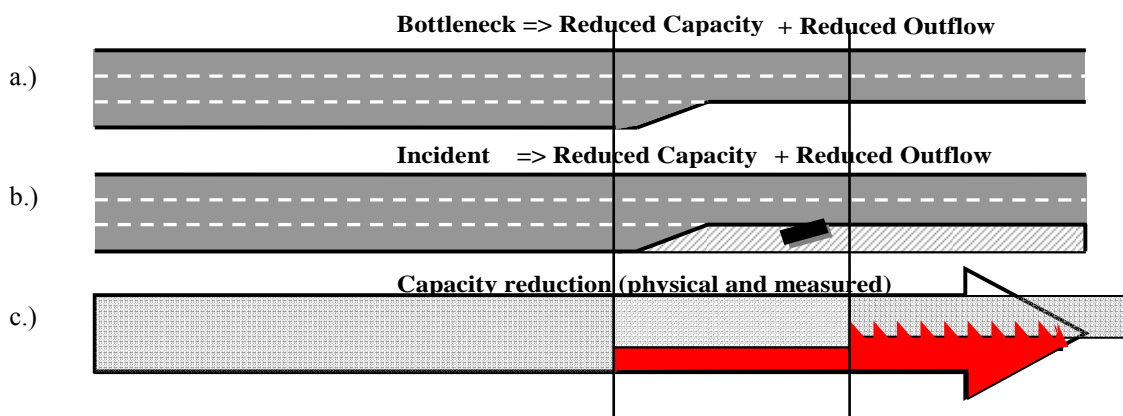
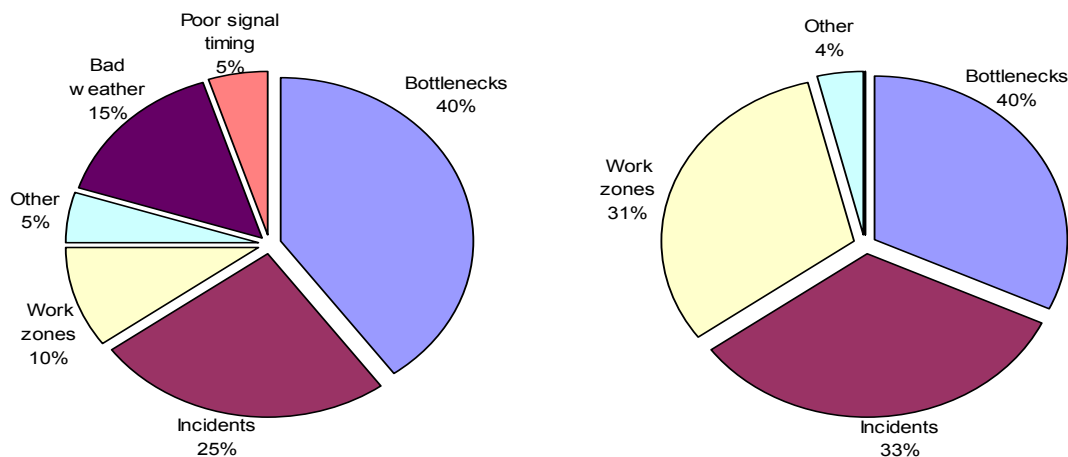


Figure 2-1: Recurrent and non-recurrent congestion

**Recurrent congestion** occurs when demand increases beyond the available capacity. It is usually associated with the morning and afternoon work commutes, when demand exceeds the

physical geometry capacity of the motorway mainlines along the roadway. Physical bottlenecks are locations where the physical capacity is restricted (usually due to reduced number of available lanes, the curvature of the motorway, on- and off-ramp design), with flows from upstream sections (with higher capacities) being funnelled into them (Figure 2-1, a).

**Non-recurrent congestion** results from random or irregular traffic-influencing events that decrease available capacity while the demand usually remains unchanged<sup>7</sup> (see Figure 2-1, b). Roughly 60% of the congestion in the USA (FHWA, 2004A) can be attributed to this type of congestion. ADAC reported (BOGENBERGER, 2002B) that in Germany the three main causes of non-recurring congestion are (see Figure 2-2): various incidents (33% of congestion), work zones (31% of congestion), and other (mainly weather) (4% of congestion). When these events occur, their main impact is to "steal" physical capacity from the roadway, and thus greatly reduce the average throughput and reliability of the entire transportation system. A stopped vehicle, for example, can take a lane out of service, but the same number of vehicles expects to travel through. Speed and throughput drop until the lane is reopened, and then they return to full capacity. Studies (GOOLSBY, 1971) show that by one-lane blockage, even though physical reduction in available capacity is "only" 33%, the measured capacity is reduced to 49% of its normal value (Figure 2-1 c).



**Figure 2-2: The sources of congestion<sup>8</sup>. Right: in USA, Left: in Germany**

Other factors, such as weather, traffic composition, driver population and speed variance, might cause non-recurrent congestions to form. For example, it has been found that light rain showers may reduce motorway capacity by 15-20% (FGSV, 2001).

When the sensitive demand-capacity balance is disturbed, either due to recurrent events or incidents, congestion is formed, outflow decreases, queue grows faster, and consequently

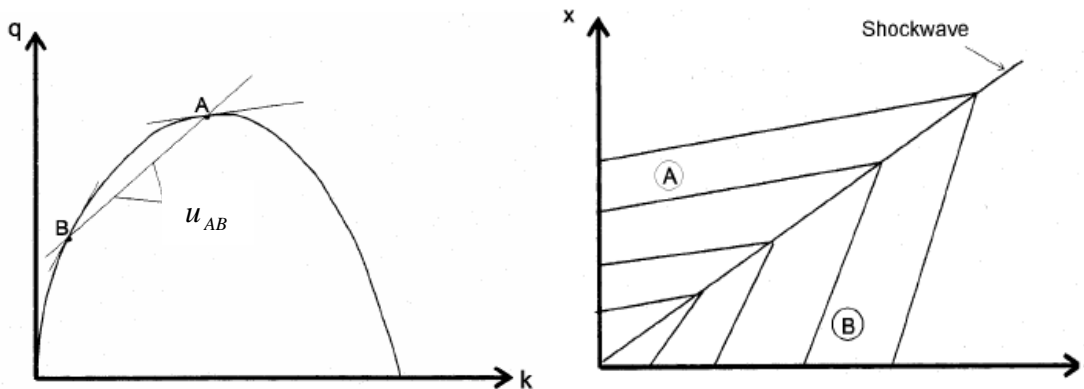
<sup>7</sup> Please note that these events may also cause changes in traffic demand by causing travellers to rethink their trips (e.g. severe weather conditions)

<sup>8</sup> It should be noted that these estimations are composite of many past and ongoing researches and are rough approximations.

congestion lasts longer than the overloaded period, and ends only after the storage queue has been dissipated. In general, it is known that in free-flow condition state information propagates in the same direction as traffic with a (wave) speed that is (at least in first order models (LIGHTHILL AND WHITHAM, 1955) assumed equal or lower than the average vehicle speed. However, if the incoming traffic has smaller densities and thus higher wave speeds than the wave in front, these two waves will meet somewhere along the road and at the point of their “collision” the shockwave will form. A shock represents an abrupt change in density ( $k$ ), flow ( $q$ ), or speed ( $v$ ). If the difference in densities of these two “colliding” regimes is large enough the shockwave will move in the upstream direction. The speed of the shockwave ( $u$ ) is:

$$u_{AB} = \frac{q_A - q_B}{k_A - k_B} \quad \text{Eq. 2-1}$$

Where  $k_A$  and  $q_A$  represent downstream traffic flow, and  $k_B$  and  $q_B$  upstream traffic flow conditions. In the flow density curve (Figure 2-3 left) point A represents a situation where traffic flows near to capacity with the speed well below the free-flow speed. Point B represents an uncongested condition characterised by lower densities and higher speeds. Tangents at points A and B represent the wave speeds for these two situations. If we assume that the faster flow point B occurs later in time than that of point A, then the waves of point B will eventually intersect with those of point A. In the right figure the areas where conditions A and B prevail are represented by the waves. The shockwave speed is represented by the slope of the line connecting the two conditions and represents the path of the shock wave shown in the right figure.



**Figure 2-3: Forming of shockwave. Left figure: a flow-concentration curve. Right figure: trajectories of the traffic waves (KELLER, 2002A; FHWA, 2004B).**

Additionally, once traffic flow breaks down to stop-and-go conditions, capacity is actually reduced — fewer cars can get through the bottleneck because of the extra turbulence (Figure 2-2, c). It is well-known that the congestion after a breakdown usually has an outflow that is 5-10% lower than the available capacity<sup>9</sup> (HALL AND HALL, 1990; BANKS, 1990).

<sup>9</sup> Capacity-drop phenomenon

Whereas recurrent and non-recurrent congestion have different causes, their solutions have many elements in common. Ramp metering, i.e. regulating the amount of incoming traffic at entrance ramps, is a common and widely accepted measure for fighting motorway congestion. Link control systems are another measure that, if properly used, could reduce occurrences and impacts of congestion on traffic flow performances.

### 2.1.2 Safety

Safety can be defined in a number of ways, including the official World Health Organisation (WHO) safety definition ‘freedom from unacceptable risk of harm’. When we speak about traffic safety we usually think about accidents. Accidents can be defined as (KELLER, 2002B): “Any event that due to moving traffic at opened roads and places resulted in fatalities, injuries or/and damages”. Safe road traffic is characterised, in an ideal case, by the absence of crashes, injuries and fatalities.

Accidents have a great effect on the safety of responders and on the mobility of the travelling public. Over 40,000 people are killed and 1, 7 million people injured on roads in the EU every year (ETSC, 2001). In the USA studies reported 41,000 people killed and more than 5 million injured. It is estimated that 18% of fatalities on motorways is due to secondary accidents only. Representing human life in numbers is not a favourable practice, however, the following accident cost estimates could be found in literature. In the USA, in addition to the delay costs, there is close to \$200 billion per year of direct economic loss due to accidents and fatalities (FHWA, 2004A). In the EU the cost to society has been estimated at 160 billion Euros annually.

Safety performances of different networks (length of the road, average daily traffic) are usually expressed and compared through calculation of their corresponding accident rates ( $A_r$ ):

$$A_r = \frac{N_{acc} * 10^6}{ADT * L * T} \left[ \frac{Accidents}{1Mio.Veh * km} \right], \quad Eq. 2-2$$

$N_{acc}$  - total number of specific accident types

$ADT$  - average daily traffic (veh/24h)

$L$  - length of the road

$T$  - time period in which data have been collected

The fatal accident rate is well accepted as one of the most accurate measures of motorway safety. Like accident rate in general, it is determined by dividing the number of fatal accidents by the number of kilometres travelled. This formula eliminates the possibility that short-term anomalies (e.g. multi-passenger accidents on a certain location) cause fluctuations in the rate that are not related to the true cause of the accident. Comparison of accidents with fatality rates for different European countries is given in Table 2-1.

**Table 2-1: Number of fatalities on motorways per billion km's travelled on motorways (Source: OECD-IRTAD database)<sup>10</sup>**

European countries	Max speed limit motorways (km/h)	Killed persons on motorway per billion travelled km's on motorways			
		1997	1998	1999	2002
Austria	130	9,0	9,0	7,2	
Belgium	120	7,8	7,1	6,2	
Denmark	110	0,7	1,0	4,9	
Finland	120	3,0	2,4	4,1	
France	130 (rain: 110)	5,1	4,8	5,0	
Germany	no general limit	-	-	4,1	
Great Britain	112	2,1	2,2	2,0	
Ireland	97	-	-	7,4	
Italy	130	-	-	9,9	
The Netherlands	100/120	2,3	2,7	1,7	
Portugal	120	14,0	15,1	15,1	
Sweden	110	2,6	2,5	2,5	

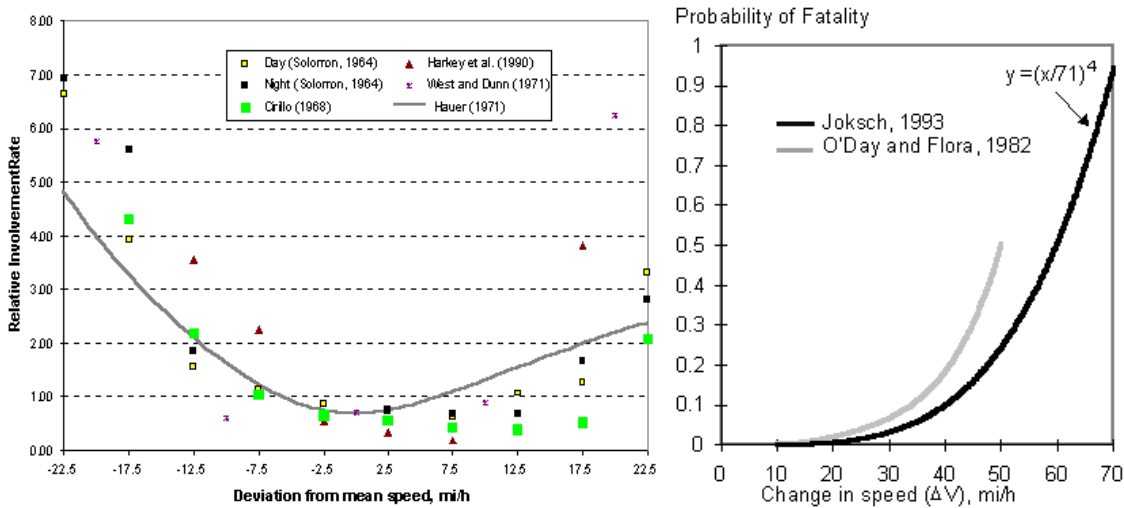
Safe driver behaviour is the result of successful interaction between the following components: the driver, the vehicle, traffic conditions, and the road environment (e.g. RÄMÄ AND KULMALA, 2000; LABS, 1987; KELLER, 2002A). The negative effects of all above-mentioned components, with exception of the vehicle, could be reduced through an appropriate collective traffic control strategy.

According to many studies, 95% of all accidents are due to driver error, such as speeding, driving under influence of alcohol/drugs and driver fatigue. When traffic operating characteristics deteriorate (due to congestions or poor weather) the driver is responsible for behaving adequately and reacting to such situations. However, this is difficult and not always possible because of a number of reasons. (ETSC, 2001; RÄMÄ AND KULMALA, 2000) Firstly, drivers are not always safety-oriented and it could be argued that safety is a secondary consideration for every driver, at least once in a while. Drivers have several parallel goals while driving which may compete with, or even contradict, the safety goal which in itself is abstract and distant. Secondly, the information available to the driver is seldom sufficient. For example, the driver could not anticipate that congestion has occurred somewhere on the road and he/she will face sharp speed drops. Additionally, weather conditions, such as slippery road, are often underestimated by drivers. Thirdly, inappropriate behaviour usually does not provide immediate or sufficient feedback on the threat of low friction, which the driver does not detect timely. Additionally, there are usually several road users or obstacles to be taken into account at any given time; human resources such as attention and energy are limited and there is probably a substantial variation in driver abilities and reaction in risky situations.

<sup>10</sup> Although the presented results originate from the same database, there are significant differences in the way different countries collect accident data (e.g. estimating the number of fatalities due to the accident), and therefore differences between countries should be taken with caution.

There is strong empirical evidence of the direct influence of traffic flow density and speed variation on accident rates, where the drop in density and speed variation reduced the number of accidents (HALL AND PENDLETON, 1989; JANSSON, 1994; SULLIVAN, 1990; VICKERY, 1969; CEDER AND LIVNEH, 1982; GOLOB AND RECKER., 2002; KOCKELMAN AND MA, 2004). The relation between speed variation and the number of accidents (probability of accident involvement) is usually expressed by a U-shape curve (CIRILLO, 1968; WEST AND DUNE, 1971; FILDES AND LEE, 1993). It has been reported that accident rates are lowest for a travel speed close to the mean speed of traffic, and that they increase with greater deviations above and below the mean speed, whereas later studies show that this increase is more significant for the drivers driving well above rather than well below the mean speed. These curves are in correspondence with HAUER'S, 1971 theoretical analysis of overtaking demonstrating that the number of vehicle interactions in terms of passing or being passed is also a U-shaped curve with the minimum at the median speed. The number of vehicles that a driver catches up with and overtakes increases with speed and the number of times a driver is passed by others decreases with speed. Thus, the increased risk of crash involvement is a result of potential conflicts from faster traffic catching up with and passing slower vehicles. This is illustrated in Figure 2-4, which compares relative overtaking rates for a 100-km/h road with a standard deviation of 10 percent with the crash risk from various studies.

The relation between driven speeds and accident rates is more complex and is probably the most-commonly debated component. While it is well understood that, as a result of laws of physics, higher driven speeds produce more severe crashes (e.g., KOCKELMAN AND MA, 2004), generally no clear relationship could be found between the number of accidents and speed policy (ETSC, 2001, FHWA, 1998). For example, even though it is generally believed that a 1% reduction in speed (if appropriately applied) might result in a 3-5% reduction in injuries and fatalities, it would not necessarily result in a reduced accident rates in general. While static speed limits might be sufficient in situations when there are no disturbances in traffic flow and its surroundings, in situations when unexpected events (e.g. shock waves, incidents, bad weather) occur and result in drastic speed drops, even the driver that obeys to static speed limit would be surprised. For example, imagine the driver driving at a speed of 100 km/h (as limited) who suddenly reaches the point of traffic incident where speeds are e.g. 50 km/h. The surprise that the driver would experience in such a situation could lead to swift and thus unsafe reaction even though the driver had obeyed the speed limit. Furthermore, the severity of the situation increases as the difference in speeds of the approaching and downstream traffic flow increases. JOKSCH, 1993 found that the risk of a driver being killed in a crash increased with the change in speed to the fourth power, as shown in Figure 2-4. The risk of a fatality begins to rise when the change in speed at the moment of impact exceeds 48 km/h, and is more than 50 percent likely to be fatal when the change exceeds 96 km/h. The probability of death from an impact speed of 50 mi/h (80 km/h) is 15 times higher than the probability of death resulting from an impact speed of 25 mi/h (40 km/h).



**Figure 2-4: Left: Accident involvement and overtaking rates relative to the average rate and speed (U-curves) from various studies<sup>11</sup>. Right: Effects of change in speed at impact on fatality risk<sup>12</sup> (FHWA, 1998).**

It should be noted that even though the above listed analysis provide invaluable insight into the interplay between flow characteristics and traffic flow safety, systematic studies that close the control-flow-safety loop are still scarce (this can be seen in chapter 2.5, where the reported effects of LCS are presented). However this is not an easy task for at least two reasons, Firstly, usually the only relevant and available information source about accidents and their characteristics is police reports. However, these reports often contain imprecise information regarding accident location and the precise time of its occurrence. This information is critical for systematic and detailed accident analysis. Secondly, the available traffic data is usually aggregated in one-minute intervals, thus also “aggregating” accident impacts.

## 2.2 Components of link control system

Link control systems are designed to improve safety and efficiency of existing motorway network through effective information and control mechanisms based on real-time *incident* and *congestion* detection/prediction/estimation. The focus of this thesis is the LCS ability to improve motorway performances by means of variable speed limits. At a high level of abstraction link control systems could be described with four components connected in a closed loop (Figure 2-6):

- Traffic flow (plant) is the system that is supposed to be controlled. It has its own behaviour, which depends on various disturbances (demand, capacity, as well as incident, weather, etc.), human behaviour and control inputs.

<sup>11</sup> Some of the differences between earlier and new studies could be contributed to change of traffic and vehicle characteristics as well as to the characteristics of data

<sup>12</sup> The shift in the curve to the right can be explained in part by improvements in vehicle crashworthiness, seat-belt use, and emergency medical care over time



- Data collection systems measure the characteristics of the traffic flow such as speed, flow, but also of the ambient factors such as fog or rain. They represent the input to the control station.
- Communication components “communicate” the control action to the driver. The possible control input can be selected from a set of available control actions that are subject to technical and operational constraints of the corresponding control device (e.g. VMS and control that could be displayed) (PAPAGEORGIOU, 2001).
- The control centre has several tasks: data management, surveillance and deciding the next control action based on the pre-defined control state.

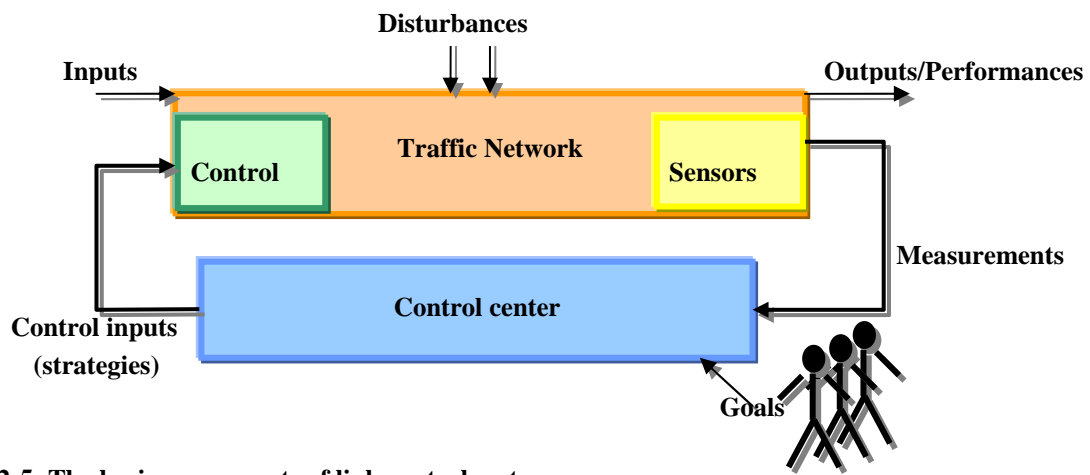


Figure 2-5: The basic components of link control system

The disturbances that influence traffic flow behaviour, in terms of safety and efficiency, have been described in 2.1. Although they cannot be manipulated they may be measurable or detectable (e.g. incident), or predictable (PAPAGEORGIOU, 2001). Control inputs are another factor that indirectly affects traffic flow behaviour. Drivers adjust their behaviour (not always in predictable manner) in response to the control input, which in turn influences traffic flow and is *reflected in the new sensor measurements*.

The core of the control loop is the control strategy. The task of the control strategy is to specify control inputs in real-time based on available measurement/estimation/prediction, so as to achieve the predefined goals despite the effects of various disturbances. In **manual control system** the operator carries out this task, while in **automated control** systems this task is undertaken by an algorithm.

Automatic control can be realized through the application of a static predefined control action (fixed control), it may change according to some fixed rules (e.g. on a time-of-day basis) or it may be calculated in real-time according to the measured/estimated/forecasted plant characteristics, which is the imperative for modern control systems. Hence, the task of a well designed control model is to **choose control action** that will fulfil the predefined *objectives* in the best way, with the *available data* and a known/estimated/forecasted plant dynamics, which are subject to various *disturbances*. In an automated control system the capability to

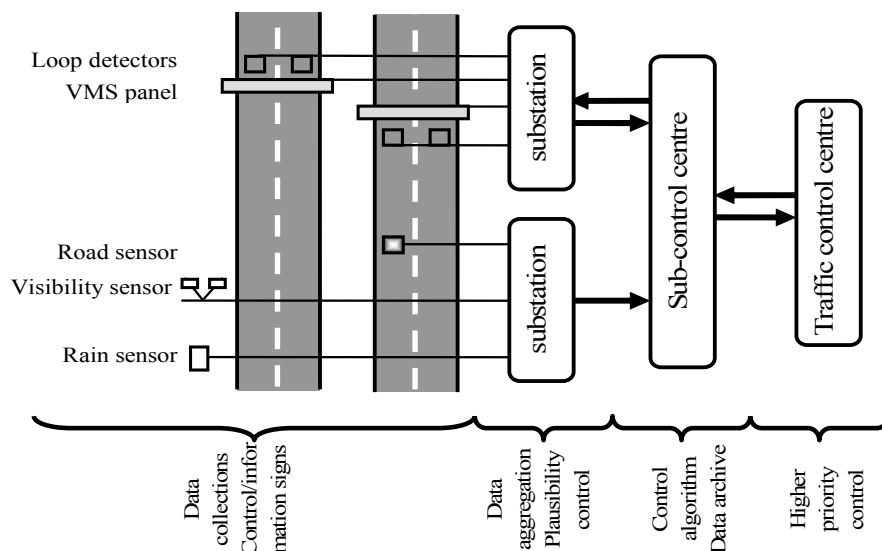
automatically *evaluate performances* and re-calibrate control strategy parameters plays an important role in the system's ability to adapt to changed traffic flow conditions.

### 2.2.1 Data collection systems

The backbone of any traffic control system is the data sources which provide for real-time measurements, based on which the link control algorithm can decide the most "profitable" control action. In Figure 2-5 the data sources are denoted as "sensors" and may be classified into:

1. Traffic data collection systems that measures actual traffic conditions and
2. Data collection systems that measure "surrounding factors" affecting these traffic conditions, such as weather, work zones, etc.

In this thesis the link control framework is built on a typical example of such a traffic data collection system in Germany, whose simplified representation is given in Figure 2-6.



**Figure 2-6: Data collection, communication and processing components in German context (according to SCHICK, 2002)**

The LCS installations in Germany are supported by loop, passive-infrared, radar, laser and microwave detectors (FGSV, 2000). Among these, loop detectors are most frequently used in existing systems. A detector collects data from a single lane. A set of detectors covering different lanes at one location is called a detector station. Detectors are usually placed at distance of 1.5 to 3 km, not far downstream from the control station (VMS panel). According to the TLS (BAST, 2002) each detector should be able to detect a passing vehicle and its driven speed. The detected vehicle should be classified in one of eight predefined categories but in practice only classification as passenger cars or trucks is usually made. In order to reduce the amount of information that should be transferred, at the substations aggregation of these data in one minute values and basic plausibility check are made, as specified by TLS. The following data is then transferred to the spatially distributed sub-control centres: number

of vehicle passages per minute (in two major classes), mean speeds (also per minute), occupancy, and detection status (if the detector was operating correctly or not). The data is checked at sub-control stations, according to MARZ specifications and their plausibility status is determined. In addition, calculations of derived values, such as truck percentage, densities, etc. are made. This data is then stored together with original (from detector) and actual (time of reception) time stamps.

To collect detailed information about weather conditions, visibility and rainfall detectors, as well as detectors measuring pavement conditions, are installed.

On data quality depends the quality of all other derived information and ultimately of the control models itself. Missing data and data corruption is a particular relevant issue in light of robustness and consequently reliability. Even more advanced approaches (e.g. fuzzy logic, neural networks, etc.), would most likely not be applicable in practice without such components. Unfortunately, delivered data is often corrupted due to hardware (detector) failure, communication breakdowns, communication network overload, incorrectly calibrated detectors, etc. In the case of German systems (VUKANOVIC ET AL., 2004), as well as reported in the Netherlands (VAN LINT, 2004), an average of 10% of the data produced by traffic detectors is missing, with frequent extremes reaching up to 20-25% and some detectors being completely out of service. Taking into account that link control systems control long motorway corridors, it is to be expected that it will happen frequently that at least one detector produces incorrect data. It happened on a number of occasions that systems were put out of operation due to bad data quality and consequently bad control decisions (e.g. in Denmark, USA). Furthermore, due to corrupt data, the link control system could produce false warnings or react too late. Thus, it is important to develop a studious concept for control behaviour in such circumstances (e.g. operation in safe-mode). It is not realistic to expect for any model, regardless of how advanced, to be able to produce valid control with corrupted data. Even though we cannot influence detector operation, the better plausibility and correction mechanisms could significantly improve the performances of today's systems. Therefore, the data pre-processing component, which will take the task of data checking and corrections, is a requirement for any system that aims to be applicable in practice.

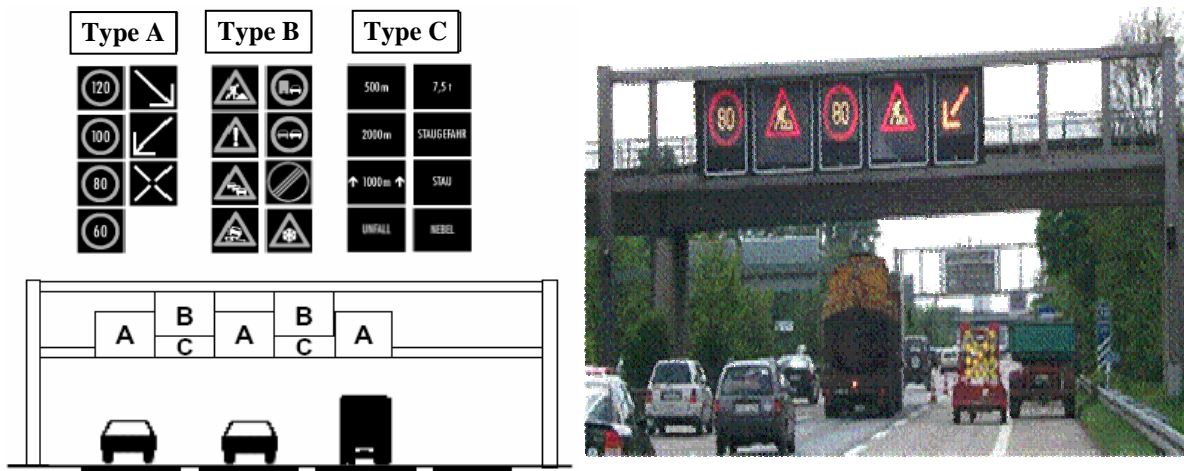
### **2.2.2 Communication component (variable message signs)**

In Germany, link control systems influence traffic flow by way of VMS panels. On each VMS different signs can be displayed, which could be grouped as (BMV, 1997A; BMV, 1997B):

- A – signs include prohibitive or restrictive controls (signs with arrows, red crosses, speed limit, etc.). Speed limits from 60 to 120 km/h, in steps of 20 km/h, and in some cases 130 km/h can be displayed. A – signs are lane specific.

- B – signs include different warning signs (specific immediate dangers related to weather conditions or traffic status) and restrictive controls (e.g. prohibiting truck passing). These signs are not lane specific but placed between A – signs.
- C – signs include different information that should support road users in their driving task. These signs are placed beneath B – signs.

A detail overview of possible signs and their positions is given in Figure 2-7.



**Figure 2-7: Left: One typical LCS panel for a three-lane motorway in Germany (SCHICK, 2002 according to BMV, 1997A and BMV, 1997B). Right: Example of VMS panel as used in Germany**

Many studies (e.g. STEINHOFF ET AL., 2001A; KATES ET AL., 1999; BOLTE, 1984) have been undertaken to investigate unambiguity and driver attitudes toward different signs and their combinations. Here we will only conclude that in general these studies show that these signs are correctly perceived and interpreted by a vast majority of drivers and thus are considered unambiguous. In addition to other findings it is interesting to note that it has been also shown that beside speed limit additional information such as “danger” is preferred by drivers and would increase driver compliance with the displayed control.

### 2.2.3 Importance of human factor

Understanding traffic flow dynamics and problems is one of the imperatives for a successful traffic control system (chapter 2.1). However, in addition to its known dynamics, traffic flow is made up of people that have beliefs, attitudes and moods and who do not always behave as expected. Therefore, the general problem with collective control systems is that their impact strongly depends on the driver attitudes and general public acceptance. Recent studies have shown that VMS signs are correctly perceived and interpreted by the vast majority of drivers. However, interviews with the authorities in (STEINHOFF ET AL., 2001B) indicate that compliance is not constant and seems to deteriorate in time. Although this is subjective evidence, and could be partly attributed to the complexity of the information and relatively infrequent “punishment” of inappropriate behaviour, it supports the inference that poor compliance could be attributed to the cumulative effect of high false-alarm rates (too

restrictive or “wrong” control), which were indeed measured in the systems under consideration. Remember the story about “boy who cried wolf”, where in the end, due to previous numerous cases of false information, people did not run to the boy’s aid when the wolf really did show up. Hence, the success of the link control strongly depends on:

- quality of control (how well it corresponds to the real traffic situation and LCS goals – objective validity), and
- control effects (subjective validity).

These two aspects are interrelated: the effects of control rise with driver acceptance, and acceptance in turn depends, in the long run, on the quality of the control.

Even though it could be argued if all critical situations may be really obvious to the driver, undoubtedly the success of the LCS controls depends strongly on long-term control reliability. Controls that are too restrictive, as well as too flexible, could have serious negative safety effects: they could lead to “the boy that cried wolf” phenomenon or to lowering driver attention in situations when the opposite is desirable. The changes in user acceptance could for example be an indicator of the control validity (STEINHOFF, 2002). Finally, in current analysis of link control effects, the problem to estimate what are the control effects and what are the effects of some external factors is identified. For example, if drivers changed their behaviour due to the control input or the traffic was already too slow.

Thus, the core of this work will be the development of a reliable (objectively valid and robust) control model. To accomplish this it is important to be able to *measure objective validity*. As it will be shown in chapters 2.4 and 2.4.3, the current implementations are not supported by a measure that can capture, if not all than at least the major control effects. Instead, validity of the system is usually proven off-line, through the comparison of similar motorway corridors, or on-line by an expert (again subjective). Therefore, it is also difficult to prove and convince road users that LCS control actions are for their benefit. Part of this thesis is the development of such a function (inspired by the earlier works of BUSCH, 1986; HOOPS ET AL., 2000; KATES ET AL., 1999) which could capture or at least approximate LCS impacts in the way useful to the control model and authorities and understandable to the user.

#### **2.2.4 Warning and harmonisation strategies**

Static speed limits alone (usually determined as 85 percentage of designed speed), cannot cope with highly dynamical traffic flow characteristics and complex driver behaviour. Their impact on safety should be considered in relation to other, abovementioned (chapter 2.1) ambient- and traffic-related factors. From a traffic management perspective, the potential for improving safety lies in timely and dynamically controlling driver (and informing them) according to the prevailing traffic conditions that are not visually available or could be judged wrongly, as well as reducing the disproportion in driver behaviour (speed deviation, lane changing, etc.). Appropriate dynamic control should also have a positive influence on traffic flow efficiency.

The strategies that could be applied via dynamic speed limit, for the purpose of improving traffic flow characteristics<sup>13</sup>, can be seen as being either reactive or proactive (see Figure 2-8), depending on whether they are triggered by an event that *already* happened or by an event that *might* happen.

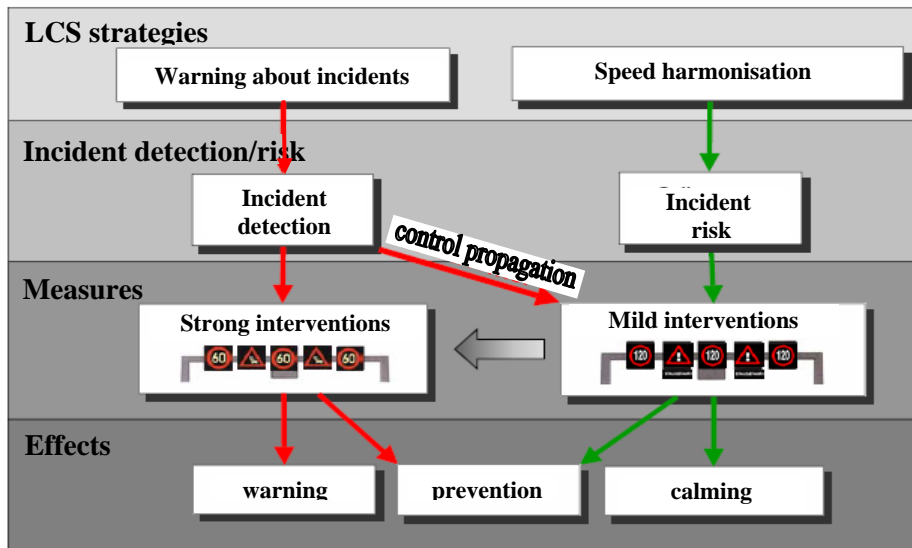


Figure 2-8: Harmonisation and warning strategies (STEINHOFF, 2002)

**Warning or reactive strategies:** are based on the idea that LCS can *warn* and timely adjust driven speeds with dynamic speeds limiting, in order to prepare drivers for the events they are not able to otherwise anticipate (e.g. incident in front) or could judge incorrectly (e.g. underestimation of road slipperiness). These events usually result in lower speeds at the location of the event than on upstream road sections, thus surprising approaching drivers and forcing them to swiftly change their behaviour (sudden breaking, lane changing, etc.). Any such event significantly increases the threat of injury (even psychological). Although warning strategies may utilise the complete range of available speed limits (40 km/h – 120 km/h), they usually use the one lower than the speed that corresponds to the critical capacity of the fundamental diagram. In order to be able to react to critical events, such events must first be detected. Since critical traffic flow conditions may have various effects on traffic flow, their detection is not a simple task. Consequently, there is a vast amount of literature on the automatic incident detection and estimation algorithms that attempt to detect as many incidents as possible, with a small number of false alarms (Appendix A.1). Most of the researches concluded that, due to the high complexity and dynamic of the problem, there is probably no single algorithm that could successfully take over this task. Hence, the distinct characteristic of warning strategies is that they are based on the *detection* of different abrupt

<sup>13</sup> Here, only strategies with respect to the changes in traffic flow are analysed. Weather related controls are not explicitly analysed (but will be if they influence flow characteristics) since they are not the focus of this thesis.

traffic conditions along the roadway, and *warning* the drivers about them so that they can *timely* adjust their driving behaviour.

The time between an accident occurrence and its detection (and thus timely warning) plays an important role in saving lives. According to EVANCO, 1996 effects of a change in accident detection time ( $\Delta ADT$ ) in number of fatalities ( $\Delta NF$ ) have following form:

$$\frac{\Delta NF}{NF} = 0.27 * \frac{\Delta ADT}{ADT}, \quad \text{Eq. 2-3}$$

Using the above equation, EVANCO, 1996 concluded that if the 5.2 minutes detection time is reduced to 3 minutes, there would be an 11% reduction in the number of fatalities nationally. Clearly, human operator requires some time, that should be added to the detection time, to process information and make the complex decision about the appropriate control action. This time reduces the potential of LCS to increase safety and thus, whenever possible, LCS should be controlled automatically by the control model and monitored by the operator (instead of controlled only manually).

**Harmonisation or proactive strategies** are based on the anticipation or prediction of these events before they actually happen (or might happen) and are trying to prevent them. Harmonization strategies usually use speed limits higher than the speed corresponding to the critical capacity of the fundamental diagram (80 km/h to 120 km/h). They are usually applied to traffic situation characterized by high density but by still high speeds, where any inappropriate speeding or slowing, or lane-changing manoeuvre could lead to incidents or congestion. Since it is extremely difficult to prove whether harmonization of traffic flow has lead to prevention of incidents or whether it had no effect whatsoever (similar to “egg and chicken” problem), different views on and approaches to the fulfilment of the harmonization goal may be found in literature. The *harmonization approach, as used here*, results in a more uniform speed profile (measured longitudinally) (STEINHOFF, 2002; KATES ET AL., 1999) and thus to a safer traffic situation. In addition to increasing safety, harmonisation strategies may be used to distribute traffic along motorway, thus postponing traffic break-down occurrence and increasing network throughput.

In some situations, it seems reasonable to use the so-called “***control propagation***” strategy. This strategy is based on idea that drivers further upstream should also be timely informed about an incident ahead, thus also reducing the probability of secondary accidents. Even though this strategy is motivated by safety considerations, it triggers higher speed controls and has effects closer to the harmonization one (STEINHOFF, 2002).

Clearly *not all strategies make sense in all situations*. For example in free-flow, the only control that is needed is warning about the incident in front, but there is no need for harmonization. The same applies to a congested state, where harmonization cannot prevent an incident that has already occurred, but lower speed limits could homogenize traffic flow and

thus reduce further disturbances and possible secondary accidents. Due to the traffic complexity, quality of the detection/estimation of various incidents and their effects might also vary from situation to situation.

## **2.3 Brief history and overview of world practice**

### **2.3.1 History**

Traffic management and control development has always depended on advances in information and communication technology. Thus, ideas of using advance information and communication technology to improve traffic performances have been put forward and/or advanced with each new technological wave. The first wave came in the technologically optimistic late fifties (as a result of advances in technology and the field of operations research). In Germany, the first link control systems were put in operation in 1965, on 30 km long corridor of BAB A8 between Salzburg and Munich. Mechanical variable message signs were placed on the road side, at a distance of 1.5 km to 2 km, and could display speed limits of 60km/h, 80km/h and 100km/h as well as “danger” and “accident” messages. Control was manual, based on images from video cameras. The first installation in USA, on the New Jersey Turnpike New Jersey Turnpike, also dates from the late 1960s.

The second wave came in the seventies, when computer development was very intensive. In 1970 the German Federal Ministry of Transportation introduced the first „Framework for link control systems“, in which guidelines for increasing the throughput and safety at critical and heavily loaded corridors were given. In this year the first German electronic link control systems were installed (e.g. Aichelberg, between Stuttgart and Munich). Unfortunately, at this time, in many cases expectations exceeded the level of technical development of the components (hardware and software): either components could not carry out what was expected of them, or the costs were too high. This has slowed down development worldwide. This is probably the reason why some installations from this period (for example in Michigan, USA) have been put out of operation later, and are still inactive.

Finally, in the third telematics wave, in the eighties, technological development matched the expectations of vehicles, road and traffic control experts. Intensive research and development started immediately even though some technologies were still very expensive. Considering future development (BUSCH, 1971), BMW’s “Program for link control systems” was introduced in 1980 through the installation of 111 variable message signs which have been permanently improved upon since. Weather detection was first introduced in 1983 on a 7 km long corridor (both directions) of A8 near Stuttgart. The corridor was covered with modern installations that could display lane specific speed limits, ranging from 60km/h to 120km/h, in 10km/h steps. For the first time LCS was automatically controlled with the help of complex algorithm. At this time, the first larger LCS installations were introduced in Holland (MIDDELHAM, 2002) and England.



The shape that LCS have in Germany was established in the early 90s. LCS installation included inductive loop detectors (today it also includes radar and infrared detectors) delivering one minute aggregated measurements of speed and flow (per vehicle type), and weather detectors delivering measurement of visibility, wetness, ice, etc. Different informative signs (“congestion”, “danger”, “work zones”, etc.) and texts (“1000 m in front”, “fog”, etc.) could be displayed in addition to speed limits (60 km/h to 120 km/h, in 20 km/h steps, as well as 130 km/h) and prohibition signs (e.g. prohibiting truck passing). During this period LCS installations in Holland and England also expanded extensively and achieved their present shape (1993 and 1995 respectively).

In 2001, 850 km of Germany’s 11,700 km motorway network was covered with LCS (BMVBF, 2002A), and total investments up to the present have amounted to 450 million euro. By the year 2007 an additional 200 million euros are to be invested and 1,200 km of motorway network is to be covered with LCS installations.

In Germany the crucial steps toward harmonisation of LCS deployment practice were also made in the 1990s. During that period, the first standardisations, which are still in effect, were introduced:

- Richtlinie für Wechselverkehrszeichenanlagen (RWVA), 1997
- Richtlinie für Wechselverkehrszeichen (RWVZ), 1997
- Technische Lieferbedingungen für Streckenstationen (TLS), 1993
- Merkblatt für Verkehrsrechnerzentralen und Unterzentralen (MARZ), 1997

The first three standards are concerning the technical implementation of LCS installations: VMS message location, distance and context (RWVA); messages description, types, size and order (RWVZ); details regarding available traffic and weather data, and protocols and telegrams (TLS). In MARZ, the basis for the implementation of control algorithm for an existing LCS is given, and will be described in detail in chapter 2.4.2.

### **2.3.2 Practice**

Germany, Austria, Switzerland, Holland, UK, Finland, Italy, USA, Australia, Israel, and Japan are some of the countries where LCS have been implemented. Of these, probably the most particular is Germany, where there is no general speed limit on motorways. Although all installations have mainly the same goals (increasing safety and efficiency), the wide variety of possible controls and information that LCS can display and the variety of circumstances in which they can be applied, has resulted in the heterogeneity of the approaches between countries, and occasionally even within same country. Actually, each country or region tried to solve its own critical issues and traffic problems (e.g. Finland where LCS is used for the purpose of increasing safety during bad weather conditions) according to its specific traffic management and control policies and philosophies.

The most obvious difference is apparent between European and practices on other continents. While the approaches throughout the world focus on fast incident management, the attention in Europe focuses on both - detection and prevention. Thus, most of the approaches worldwide use link control systems as a part of broader incident management concept, where the operator (alone or with the help of incident management tool) manually controls variable message signs. In Europe however, the systems are mostly automatically controlled with the help of the control model and their main aim is to automatically warn the drivers about the danger ahead.

In Austria, on the A2 south-bound motorway, link control systems are used to reduce speed (80 km/h and 100 km/h for cars, 60 km/h and 80 km/h for trucks) when noise reaches a certain predefined allowed limit. In Switzerland, on the N1 near Bern (SCHICK, 2002), there is a LCS similar to the German model, but with special way of defining the control action. Microscopic data are used instead of macroscopic - each vehicle is followed separately (identified at each detector station), individual travel times are calculated and the control is determined accordingly. There are also approaches that do not utilize speed limits but only lane closure or opening (Brisbane, Australia). Another specific approach is the one in Colorado, USA where, for each individual truck, a vehicle-specific safe speed is recommended, based on truck weight, speed and axel configuration.

In addition to differences in philosophies and approaches in link control systems control, there are also considerable differences in the spacing of VMS panels and utilized detectors. For example, in Germany the distance between VMS panels can vary from 0.5 km to 3 km and usually each VMS panel is supported by traffic data collection equipment. On the other hand in the Netherlands and UK, detectors are placed every 500m and VMS every 700/1000m.

## **2.4 Link control models: State-of-the-art**

Although link control systems have existed since the late 1960s there are only a few automatic control approaches in the world, and even less available literature. One of the reasons is that in countries outside of Europe the link control problem is considered to be a part of broader incident management concepts and the focus is on the development of incident management plans rather than on automatic link control. For this reason many link control systems are actually still manually controlled, where the automatic incident detection algorithm alerts the operator and the operator decides what control action should be taken.

Since we are focusing in automatic link control applying variable speed limits, here we will give an overview of such approaches. Generally, based on the cause for the control (the nature of the event that is identified as “dangerous”), one could distinguish between the following algorithms:

1. Traffic-based link control algorithms activate variable speed limits and other controls when a disturbance in traffic flow has been detected or predicted. These approaches usually focus on disturbances caused by incidents but there are also integrated

approaches that are interested in any kind of disturbances regardless of the cause (for example, it may be an incident as well as bad weather).

2. Weather based link control algorithms activate variable speed limits when weather conditions deteriorate. These are variable speed limits due to rain, snow, slippery road surfaces, and most frequently due to the low visibility (fog). All approaches are based on comparative algorithms. When the measured weather conditions (e.g. visibility) reaches the threshold value the speed control is activated, and gradually decreased with further deterioration of weather conditions. These systems have proven very successful when weather detectors operated properly. However, due to the simple comparative algorithm logic, even small malfunctions of the weather detectors could lead to frequent “wrong” controls. This has been noted in the case of systems in operation, for example in Denmark and Germany. A very detailed investigation of weather based LCS can be found in RÄMÄ AND KULMALA, 2000.

Basically, in order to be able to prevent or reduce negative impacts of an incident at least one part of the link control algorithm has the ability to detect an incident and, thus is (in some way) incident detection algorithm. The difference consists of the fact that an automatic incident detection algorithms (AID) produce an “alarm”, while a link control algorithms produce a control. Furthermore, a link control algorithms should be able to determine the incident severity and, if possible, assess the control effects in order to be able to choose the most appropriate control. Another important difference is that link control algorithms should also pursuit other, essentially different goals (such as harmonisation). Traffic flow harmonisation is driven by other circumstances that are not related to incident detection. However, as the automatic incident detection algorithms could offer invaluable information to the link control model if properly used, the brief overview of AID is given in Appendix A.1.

Traffic based link control systems are supported by a control algorithm that makes the decision about the most appropriate speed control, based on real-time traffic data (and sometimes weather detection data), for the purpose of increasing traffic flow performance. Due to LCS complexity and our limited knowledge about their effects on traffic flow, the proposed link control algorithms are mostly based on “heuristic” approaches. It could be said that link control is one of the least investigated areas of the traffic control field. By analysing the quite sparse available literature these could be dividend into:

1. *“Heuristic” or decision based approaches* that utilise one or more incident detection and prevention algorithms,
2. *Optimal control approaches* that utilise the traffic flow model.

#### **2.4.1 “Heuristic” approaches**

Under “heuristic” approaches we consider algorithms that utilize predefined rules for information integration and control decision making. The algorithms vary in the manner in which the information is combined, but they have one thing in common - they *do not have a*

*measure of performance* integrated into the system. Instead, to each algorithm output (or combination of outputs), a corresponding control action is assigned, rules are defined and algorithms are initialised in advance. Rules are changed based on expert knowledge and extensive off-line analysis. This is why they are called “heuristic” approaches. These approaches are the only ones used in practice. Depending on the way in which information is combined one could distinguish between:

1. Look-up tables
2. Smoothing algorithms
3. Artificial intelligence approaches utilising two or more algorithms
4. Complex IF-THEN algorithms (utilising two or more algorithms)

*Look-up tables* are probably the most simple of all approaches. To each control station a specific look-up table is assigned. In each look-up table, for each combination of input values, a corresponding control action is specified. Usual input values are speed, flow, density and/or occupancy, where the frequency of data acquisition may vary. At each measuring interval, the new data is compared to the look-up table and the control action is determined. Most of the approaches also use additional rules for coordination with successive control stations upstream by posting the speed limits in step-wised manner. The background of this control approach is the flow-density diagram, where to each segment in the diagram a corresponding control is assigned. Threshold values should be determined separately for each detector station. Using solely local data makes the system very sensitive to the data quality and high fluctuation of traffic flow variables. Thus, decision borders are usually created to be lower than the corresponding control. For example, the control action of 60 km/h will be activated only when the speed drops below 60 km/h. Control is stabilised in this way. However, although fluctuations are reduced, when using only local data disturbances could be detected only when they already influenced traffic flow so greatly that the queue already propagated back to the detector station. Bearing in mind the importance of timely warning, most of LCS benefits are lost this way. Some safety benefit can be expected at upstream control station, if the “control propagation” rule is used. In order to have any positive safety effects this approach requires very dense detector infrastructure. An incident will not be detected if incident happens between detectors, and does not propagate back to the upstream detector. Furthermore, these approaches do not use LCS harmonisation potentials.

The Dutch “*smoothing*” MCSS algorithm is used for the control at Holland motorways (MIDDELHAM, 2002). It analyzes data for an individual section of motorway (one instance of the algorithm is assigned to each VMS panel), and is independent of conditions upstream and downstream. It calculates the filtered average speed for the carriageway by means of an exponential smoothing filter. If the smoothed speed drops below 35 km/h, at the corresponding control station the control of 50 km/h will be activated. At the consecutive

upstream control station the control of 70 km/h will be displayed<sup>14</sup>. The control is deactivated when smoothed speed exceeds 50 km/h in all lanes. In this way the fluctuation in traffic flow variables are “smoothed”. It has a concept of graceful degradation, where in case of communication failure each control station is able to continue to work alone. However, everything mentioned for previous look-up tables also applies to this approach, although slightly less severely: sensitivity to data quality, it does not use all the safety benefits (some safety benefit could be expected only if it is supported by a dense detector infrastructure), harmonisation benefits cannot be expected, and it does not have a performance measure integrated into the system.

The HIOCC incident detection algorithm, together with the look-up table of smoothed speeds and flows, is used in England to locate the presence of queues or slow-moving traffic. The look-up table is responsible for control activation due to congestion, while the HIOCC is responsible for control activations due to incidents. The premise of the HIOCC algorithm is that traffic will stop or decelerate considerably if there is an incident. The algorithm (MARTIN ET AL., 2001) takes occupancy data every tenth of a second, and gives a value of 0 to 10 for every second. Zero means that no vehicles have occupied the sensor that second, and 10 means the sensor was occupied for the entire second. An alarm is raised if two values of 10 are given consecutively. It will then automatically set a 40 mph (~65 km/h) speed limit at the control station before traffic joins the end of the queue, and a 60 mph (~96 km/h) limit at the preceding control station. The algorithm is also designed to terminate an alarm. It does this by taking the smoothed values of occupancy over the last five minutes of data preceding the alarm and comparing it to the instantaneous smoothed occupancy values. The algorithm is terminated when the instantaneous values drop below pre-alarm levels. Field results from M1 and M4 in London, along with the Boulevard Peripherique in Paris, show that the algorithm works well under congested conditions, but its effectiveness in light to moderate flows is yet to be seen (MARTIN ET AL., 2001). Given the nature of the algorithm, performance is expected to drop in light flows. A false-alarm rate is difficult to obtain solidly since slow-moving vehicles initiated one third of the false alarms. Finally, this approach has the same problem regarding data quality, no integrated performance measure, no harmonization benefit, and requires a dense detector infrastructure in order to produce any safety benefit.

In BARCELO ET AL., 2001 the outputs from incident detection module (Catastrophy theory algorithm) and estimating incident probability module (hierarchical logit model) are combined. The outputs from these algorithms are sent to the decision support system (table with pre-defined rules – management strategies) and the appropriate speed control and re-routing decision are determined. The algorithm has been tested with AIMSUN (TSS, 1998) and the improvements of travel time spent in the system have been reported, despite the fact

---

<sup>14</sup> Earlier control propagation of up to 90 km/h was used, but according to the latest studies from the Netherlands this has not been proven to be very effective and is usually not used anymore (MIDDELHAM, 2002).

that estimating incident probability module produced more false alarms. The simulation model assumed that drivers obey the speed limit and that there are alternative routes that drivers would take if re-routing is activated. With this approach the earlier detection of a wider spectrum of incidents is possible and a larger distance between detectors can be handled. However, combining the incident detection algorithms with fixed rules without encountering their variable reliability under different situations and data quality, suffers the same problems as the previous approaches. Finally, although the system calculates total travel time spent in the system, it does not have a mechanism that could include information about its performances into the process of making decisions about further control actions.

An algorithm that attempts to pursue both warning and harmonisation benefits is implemented in Germany (MARZ) (BAST, 1999). It utilises various incident detection (Stau1, Stau2, Stau3, Stau4), harmonisation (Belastung, prohibiting truck passing) and weather detection algorithms in order to automatically determine the best speed control. The decision is made in three steps:

1. Functional level: Each algorithm is calculated separately. To the output of each algorithm a corresponding control (speed limit) is assigned.
2. Priority level: the most restrictive of all the calculated programs is taken (with exception when manual control is activated since it has absolute priority)
3. Coordination level: coordination of the control signs at one VMS and between consecutive VMS. Here the main rule is the so-called “control propagation” rule. According to this rule if the speed of 60 km/h or sign “congestion” is to be posted at the local control station, the control will be gradually increased at the upstream control stations. The most restrictive proposed control action will be taken as the system output.

The algorithm attempts to react in all critical situations and to use warning as well as harmonisation potentials of LCS. The logical rule base does not encounter the different reliability of the information and thus could lead to more false alarms than necessary. Although the combination logic is based on simple *IF-THEN rules*, due to the large quantity of information involved and no integrated measure of performance, it is very difficult to calibrate and interpret why the control action has been activated. Finally, from the same reason, if one algorithm is not well calibrated it could have serious negative effect on the performances of the entire system. Sensitivity to corrupt data further increases this danger.

In BEALE, 2002 two algorithms have been combined with the Bayesian Belief Networks (BBN), in order to account for variable algorithm reliability, and thus allowing for the detection of a wider spectrum of incidents with fewer false alarms. The algorithm that should detect incidents under low traffic condition was developed using the Genetic algorithm. The HIOCC algorithm is used for detection under other traffic conditions. The outputs from these two algorithms are then combined with BBN. BBN have the ability to take into account the fact that there will be uncertainty or different confidence levels associated with information.

The reasoning carried out in BBN can be compared to the logical inference used in traditional rule-based systems. In traditional rule-based systems inference has the following form “if A and B are true, then C is true”. The equivalent terminology for a BBN would be “given evidence about A and/or B, what is belief in C”. The advantage is that rules in BBN can be easily formulated and new information can be fairly easily included in the system. However, with the inclusion of the rules and new information, the system might lose its initial transparency and it might be difficult to track. Interpreting the decision process would require a high level of expert knowledge particularly with all the probabilistic inference rules. The system is developed with the aim of detecting an incident and it includes harmonisation strategies. Reports have shown promising results and very high detection of incidents under low traffic conditions. However, to the author’s best knowledge there are no reports on how the system was evaluated and how control actions are determined. Finally, the current version of the system does not include a performance measure. This does not only make the calibration process difficult but does not use all advantages and mechanisms of artificial intelligence methods (e.g. self-learning). Therefore, the simple application of these mechanisms, without deeper insight and analyses of the control problem would produce a “black-box” solution, probably good in particular implementation (e.g. location), but it might not be transferable, tractable, and ultimately may not learn the system appropriate behaviour (e.g. if system performance measure does not reflect genuine system requirements).

**In conclusion** it could be said that these models are used “as they are”: they do not have integrated performance measures that would allow system calibration and, if possible, automatic adaptation to changing traffic conditions. Inherently, monitoring model performances relies on expert knowledge. However, to detect which algorithm is responsible for the final control is not always an easy task and requires experience and detailed knowledge about model functioning. Malfunctioning of the model is usually solved by exclusion of “suspicious” algorithms and/or by (intuitive) changes of algorithm parameters. Alternatively, but also more costly, an off-line study of individual algorithm performances could be carried out. Subsequently, the algorithms that are proven to be objectively inaccurate would be excluded (e.g. as in Bavaria (DENAES, 2002)). Furthermore, without a systematic framework for monitoring and measuring model performances, appropriate inclusion/exclusion of new information into a complex model is very difficult and demanding. Crisp IF-THEN rules are designed for the algorithm where it is assumed that available information is always valid<sup>15</sup> (with the exception of a BBN approach). However, most of the available information is neither completely true nor completely false, but somewhere in-between, and (as this is not complicated enough) their validity changes with the changes in traffic conditions and raw data quality. Finally, algorithms (with exception of MARZ) are developed to detect an incident when it has already happened and to warn drivers about it.

---

<sup>15</sup> Similar could be said for optimal control approaches.

They require a dense detector infrastructure in order to use the entire warning potential and do not consider LCS harmonisation potentials.

#### 2.4.2 Optimal control approaches

Optimal control approaches *use the traffic flow model* to estimate influences of different speed limits and to make the final decision. Several such approaches could be found in literature, such as the multi-layer control PAPAGEORGIOU 1981, sliding-mode control, optimal control, approximation of optimal control with neural network in rolling horizon framework, or optimal control with model predictive control by HEGYI ET AL., 2003. At each decision interval, the effects of the different control actions can be examined and the optimal control can be determined in a systematic manner. Most optimal control approaches model speed limit effects by scaling the fundamental diagram by some predefined scale factor. This however might give too optimistic results (HEGYI ET AL., 2003). Therefore, in HEGYI ET AL., 2003 the METANET model is used and extended with the term that should provide a better model of the speed limit effects. An equation is also introduced to express the variations in driver anticipation of the increase or decrease of downstream densities and the constraints that should be taken into account, such as the need for step-wised sequence of control actions. HEGYI ET AL., 2003 reported a decrease in the severity of the incident influence on traffic flow and shock-wave propagation, and 21% reduction in total time spent in the system (TTS).

The optimal control approaches provide a detailed representation of the problem and invaluable insight into traffic flow dynamics. However, they are subject to several constraints that, from the aspect of LCS, could be critical for their application. While the effects of link control systems are observable only later in time, the optimal control model requires prediction of future demands. Although the approaches of HEGYI ET AL., 2003 and DI FEBBRARO ET AL., 2001 include the prediction component its performances are highly dependent on these (predicted) inputs. However, in practice data is often missing or corrupted. There is strong possibility that in the case of the control of long motorway corridors there will be always certain amount of corrupted data. Furthermore, traffic processes are generally highly complex and dynamic due to the individual participants that all behave stochastically and inconsistently over time. Driver responses to the link control system is difficult to model and the knowledge about it is still very limited (ZACKOR, 1972; RĂMĂ AND KULMALA, 2000; STEINHOFF, 2002; SCHICK, 2002). It is known that user response to the link control input can vary significantly but that does not provide us with sufficient information for appropriate modelling. Up to now optimal control approaches have incorporated the influence of variable speed limits by reducing/increasing critical capacity of the fundamental diagram and extending the model with several additional constraints. However it remains highly questionable whether such an approach replicates flow behaviour under speed limit control. Hence, traffic flow models might deviate from practice not only due to random errors but also due to the fact that people sometimes behave unpredictably. In general, the majority of the systems that use traffic flow models optimise travel time spent in the system, which is not the



only (and the most important) criteria for link control systems. Moreover, for the link control system model it is necessary to react in *all* traffic situations, regardless of whether they are the result of demand increase, or incident or environmental conditions.

Undoubtedly, the optimal control approach with a traffic flow model provides invaluable information regarding possible control actions, but it suffers from the same limitations as the traffic flow model itself. Therefore, the outputs of such a model, if available, should be integrated (probably with greater importance) with other available information (for example, incident detection algorithms) before the final control decision is made.

### 2.4.3 Existing operational measure of performances

The goal of the link control systems is to increase traffic safety and efficiency and therefore, in the ideal case, measure of link control performances (objective function) should include both of these elements. However, to the author's best knowledge there is no such measure in literature. The proposed link control models do not include a performance measure that would capture the major link control effects (safety and efficiency). Instead, the proposed systems usually deal either with traffic flow efficiency or with safety performances. Furthermore, only the optimal control approaches have a performance measure integrated into the control model. Other approaches are, if at all, optimised off-line, mostly by trial-and-error approach, which requires expert supervision and ultimately rely on the expert knowledge. The lack of a performance measure that would integrate both LCS goals is not surprising because, on the one hand, the knowledge on LCS impacts is still quite limited, and combining travel time information with safety gains is not a straightforward process. It must be noted that there are approaches attempting to consider both aspects, usually by including some additional, more or less, heuristic constraints to the system (one of optimal control approaches) (HEGYI ET AL., 2003).

#### 2.4.3.1 Efficiency: total time spent in the system

In literature control ability to increase traffic flow efficiency is usually assessed in terms of travel time spent in the system (TTS). This approach (more or less modified) is used in the optimal control methods presented in 2.4.2. There are offline studies that analyse throughputs, mean speeds, etc. but, they are not applicable to system optimisation but rather to assessing its effects (see chapter 2.5). TTS gives the measure of control "quality", whereas better control should have less delay and consequently lower TTS. The goal of the link control model is defined as minimisation of TTS on the mainline motorway:

$$\text{minimise } TTS_{f_{wy}} = T \sum_{h=1}^H L^t * K_{f_{wy}}(h) \quad , \quad \text{Eq. 2-4}$$

$L = (L_1; L_2; \dots; L_m)$  - the vector of the section lengths,

$K_{f_{wy}}(h) = (k_1(h); \dots; k_m(h))$  - the densities on the motorway during time interval  $h$ ,

T - the length of each time interval,

H - the total number of time intervals that define the control interval.

Delays at on-ramps or some other extensions could be added to this definition, depending on the particular focus of the control model. Managing efficiency means preventing and postponing congestion. Pursuing efficiency is appropriate as long as there are no safety critical events present in the network. However, if increase in demand goes on for a longer time, no matter how good the control is, congestion can probably not be avoided. Moreover, incidents or accidents can occur even if the demand is not increased, but rather due to some other circumstances. In such situation increasing safety is the first priority, even at the cost of TTS.

### 2.4.3.2 Safety: Detection versus false alarm rate

In literature, the common way of assessing the link control safety performances is accident rate analysis (see chapter 2.1.2) prior and subsequent to the installation of the system. However, such analysis requires years of data and could be subject to confounding factors, and therefore it is not considered here to be a possible operational measure of traffic control performances. Another (safety oriented) approach for assessing link control performances, based on the belief that the major strategy for increasing safety is appropriate warning of drivers, is to use detection versus false alarm curves (DR-FAR) (KATES ET AL., 2002; HOOPS ET AL., 2000; BUSCH, 1986). The importance of two-dimensional (DR-FAR) representation has been recognized some time ago (BUSCH, 1986) and it still continues to be a useful tool for users and traffic operators. It confronts the system ability to detect safety critical events (and thus to give warnings) with false alarms that have a negative influence on traffic efficiency<sup>16</sup>:

$$DR = \frac{n_d}{N_i} * 100\%, \quad FAR = \frac{n_{far}}{N_a} * 100\%, \quad Eq. 2-5$$

$n_d$  - number of detected incident cases,

$N_i$  - total number of incident cases,

$n_{far}$  - number of false alarms,

$N_a$  - total number of alarms (decisions).

A link control algorithms generally involves dynamic comparison of some quantity derived from the incoming traffic data with one or more threshold values. A shift of threshold results in a

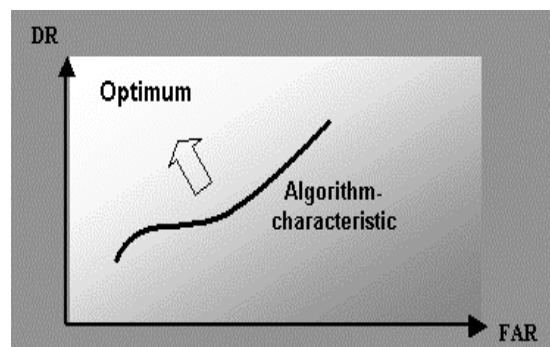


Figure 2-9: DR-FAR curve

<sup>16</sup> Please note that these definitions may vary in literature (the above definition is according to HOOPS ET AL., 2000).

trade-off between DR and FAR. This trade-off is related to the well-known statistical concept of receiver operating characteristics as illustrated in the Figure 2-10, in which the trade-off between FAR and DR is expressed as a curve in a two-dimensional representation connecting FAR-DR points for different values of specific control parameters. The nearer the curve to the top left corner of the graph the better the intrinsic quality of the link control warning capability. Due to the importance of timely warning, in addition to DR and FAR, it is necessary to add the measure that would reflect control efficiency. As a measure of control efficiency a total or mean time to detect is usually used (TTD or MTTD respectively). Although useful, the DR-FAR measure has several potential pitfalls (VUKANOVIC ET AL., 2003A) which make it difficult to apply (especially in on-line operations): treating all incidents with equal importance; there is no precise agreement what constitutes an incident and what is a successful detection; it is a Pareto optimal problem making optimisation more difficult; it does not measure link control harmonization effects, etc. Since the DR-FAR concept is important for assessing link control effects, a more detail description and different approaches for its application are given in Appendix A.2.

### 2.4.3.3 Integrated performance measures

Although the need for expressing link control (safety and efficiency) performances through a one-dimensional cost function has been identified some time ago, there are only a few papers on the subject. Probably the most thoughtful one is presented in EVERTS AND ZACKOR, 1976, where measuring of link control warning performances ( $Z(t)$ ) by transforming successful detection into time costs ( $Z_u(t)$ ) and their integration with time losses ( $Z_t(t)$ ), has been proposed:

$$Z(t) = Z_t(t) + Z_u(t);$$

$$Z_t(t) = x_t * \int \frac{q(x,t)}{v^*(x,t)} dx; \quad Z_u(t) = x_u * \sum_{(X_u)} u(X_u(t)), \quad \text{Eq. 2-6}$$

$x_t$  - time costs pro vehicle and time

$q$  - flow

$v^*$  - speed modified by the control action factor (in case of no control the control action factor = 1)

$x_u$  - costs pro incident detection

$X_u$  - location of one shockwave (only the end of the disturbances should be taken into consideration)

In the work it has been recognised that the especially critical situations are the one at the end of queue. Therefore safety performances are calculated with respect to the queue ends only ( $X_u$ ). To each successful detection a specific monetary gain is assigned. Travel time is calculated as the integral sum over time and estimated speeds and densities (that are to be experienced after the control action). The calculation is done with respect to the forecasted

traffic densities, to allow the best decision for the future traffic conditions. Although it represents a rare approach for capturing different LCS goals, the approach has certain pitfalls, such as estimating future densities, speeds and control effects by scaling the fundamental diagram, treating all incidents with equal importance, not capturing harmonisation effects, etc. Another interesting approach, although in light of incident management strategies, can be found in PETTY ET AL., 2001, where the proposed function for optimising AID performances balances algorithm (good or bad) activations and the cost of sending the tow truck or police, etc.

## 2.5 Effects of link control systems

According to available literature (ZACKOR AND KELLER, 1999; SCHICK, 2002; RÄMÄ AND KULMALA, 2000; STEINHOFF, 2002; etc.) the reported effects of link control systems could be grouped into:

- Effects on traffic safety
  - Direct safety effects (comparison of accident rates)
  - Indirect safety effects (analysis of traffic flow characteristics that are believed to influence safety - e.g. speed deviation, headways, etc.)
- Effects on traffic efficiency (comparison of throughput, capacity and operating speeds).

### 2.5.1 Traffic safety

In the last 20 years a number of studies on the subject of effects of link control systems on reduction in accident rates have been carried out. LCS has proven to be successful, and accident reductions have been observed. It must be noted that the reported results are the result of different accident databases and methodologies, and therefore should be taken as significant evidence, but still critically approached.

In Bavaria (ABDS) it has been estimated that accidents with fatalities have been reduced by 31%, accidents with injured by 30%, and accidents in general by 35%. Even though these high reductions may include bias caused by the regression-to-the-mean effect they are still significant. An accident rate reduction of 15% has been reported in the UK, and in the Netherlands the number is down by 28% (MIDDELHAM, 2002). Furthermore, a 54% reduction in pileup (multi-car) accidents has been reported in the UK (STEINHOFF, 2002). In general, it has been observed that the effects are more beneficial with respect to secondary accidents (RÄMÄ AND KULMALA, 2000). The empirical research that reported an increase in traffic accidents was STEINAUER ET AL., 1999 (STEINHOFF, 2002), where LCS installations on the A8/East direction Salzburg were investigated and a 7% increase in general accident rates, 43% increase in accidents with injured, and 51% increase in the rate of accidents with fatalities was reported. It is assumed that that bad quality of data at this road segment resulted in frequent changes of controls and thus contributed to an increased accident rates.

An extremely huge reduction by 80% in accidents due to fog has been reported in Germany (SIEGENER ET AL. 2000). BALZ AND ZHU, 1994 reported a reduction in the number of injury accidents by 20% due to fog warnings, while in Finland it is estimated that link control systems connected with warning of slippery road conditions reduced number of injury accidents by 10% (RÄMÄ AND KULMALA, 2002).

As already discussed (Chapter 2.1.2), the potential of link control systems also lies in their positive effect on the reduction in accident probabilities. In literature (PERSAUD AND DZBIK, 1993; KATES ET AL., 1999; ZACKOR AND SCHWENZER, 1988; STEINHOFF, 2002) several factors that influence accident probability have been identified, such as the percentage of small headways, speed deviation or even mean speeds. German and Finnish studies reported a significant reductions in mean speeds (3 to 9 km/h) in adverse weather conditions, and the latter also reported a significant decrease in speed variation. A Dutch fog warning system including a text warning (“fog”) and dynamic speed limit VMS signs on a motorway, reduced speeds in fog by 8 to 10 km/h, although in extremely dense fog, the system had an adverse effect on speed. This was due to the excessively high “lowest possible speed limit” display in the VMS (60 km/h). It has been also reported that more uniform speed behaviour was achieved with the introduction of the system. A reduction of mean speed of approximately 3 km/h has been reported in England, but there have been problems with system reliability.

ZACKOR, 1972 carried out one of the first systematic studies of the effects of LCS on traffic flow characteristics by investigating its effects on speed distribution, fundamental diagram and headway distribution. The reduction in critical small headways (<1 sec), by density interval of 18-21 veh/km, from 18% to 13-16% has been reported (ZACKOR AND SCHWENZER, 1988). In the same density interval percentage of extremely short headways (<0.5sec) was reduced from 2% to 1%. Furthermore, reduction in speed deviation, resulting from fast drivers driving slower and slower drivers trying to catch up with speed limit, and mean speed has been observed.

**Table 2-2: STEINHOFF, 2002: Percentage of headways smaller then 1 sec on A8 Ost, Q33, direction Salzburg, middle lane by control of 120km/h against no control (reference) case. Significant are the fields with  $\chi^2 > 2.7$ .**

120km/h V class	Q class from...to	I 0 to 600	II 601 to 1200	III 1201 to 1800	IV 1801 to 2400
100 km/h	reference	7,2	24,3	36,6	45,2
	T 120	9,8	16,3	26,3	23,1
	$\chi^2$	0,379	4,890	10,525	2,544
120 km/h	reference	6,2	17,7	27,6	37,7
	T 120	3,4	11,6	20,8	36,8
	$\chi^2$	4,372	16,187	19,131	0,017
140 km/h	reference	5,8	14,2	21,7	30,3
	T 120	4,8	7,5	15,4	29,0
	$\chi^2$	0,516	17,833	6,776	0,024
160 km/h	reference	4,7	11,4	17,6	29,4
	T 120	1,3	5,1	24,2	0,0
	$\chi^2$	1,959	3,993	1,018	1,018

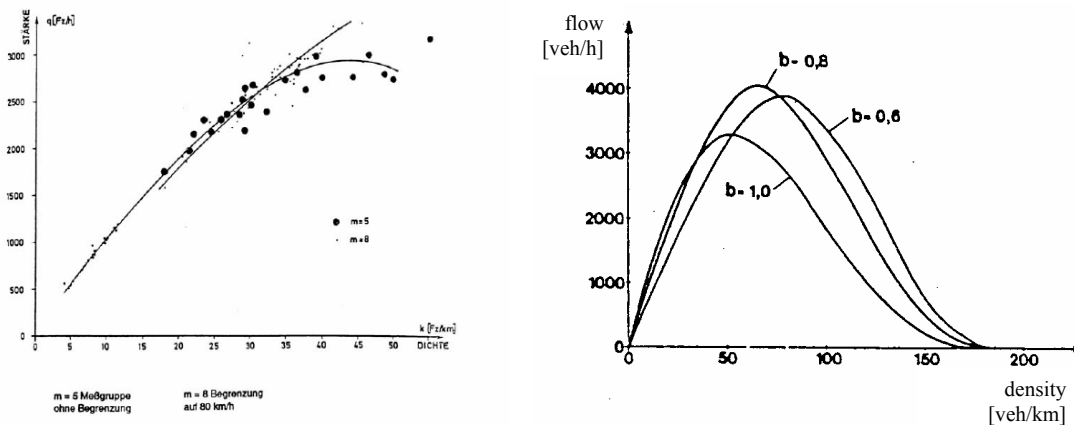
Recently, in STEINHOFF, 2002, very detailed study was carried out on the influence of various control actions, under various circumstances (lane, flow, weather conditions, etc.) on

headways, speed deviations, total time to collision, speed distribution, etc. Different situations in which control actions can have positive or negative effects were analysed. In general, a reduction of for some 4 km/h in speed deviations in controlled and uncontrolled case was observed, which represents approximately a 30% reduction in speed deviation. The general positive effect of controls on the reduction of critically small headways (<1sec) has also been proven (see Table 2-2).

## 2.5.2 Efficiency

In the work of ZACKOR, 1972, the positive influence of speed limits on capacity has been identified (Figure 2-10), in addition to the abovementioned positive safety effects. According to the research the best network utilisation would be achieved by speed limit of 80km/h. The characteristics of vehicles and motorways have changed in the meantime (and probably these quantitative measures) but it is assumed that this result is still qualitatively relevant. The observed phenomenon could be included in the fundamental diagram with specific coefficients for each speed limit to estimate their effects (ZACKOR, 1972; PAPAGEORGIU, 1981).

BALZ, 1995 reported a 10% increase in mean speeds under very dense traffic conditions. ZACKOR AND KELLER, 1999 estimated a 15% increase in mean speeds during peak hours. In their model for the analysis of LCS effects, in situations where high traffic demand was the cause of congestion, they recorded an increase in mean speeds from 60 km/h to 70 km/h.



**Figure 2-10 Left: Fundamental diagram without and with speed control of 80 km/h (source: ZACKOR, 1972); Right: Effects of variable speed limits on fundamental diagram (source: PAPAGEORGIU, 1981). In right figure effects of the speed control are expressed through factor  $b$  which shrinks the fundamental diagram (ZACKOR, 1972):  $b=1$  means there is no speed control; smaller values are assigned in case of speed control –where  $b$  describes corresponding increase of capacity and can be derived from empirical data.**

According to early estimations according to ZACKOR ET AL., 1988; FERRARI 1991, LCS could increase throughput from network by 5% to 10%, compared to non-control case. KÜHNE ET AL., 1995; SCHICK, 2002; DINKEL, 2004 reported the positive effect of LCS on moving traffic flow from the unstable to the meta-stable area of the fundamental diagram and interpreted that

as an increase in performance. Other research by SCHICK, 2002 reported the positive impact of LCS on reduction in traffic flow break-down probability, especially by medium and high traffic flow. There was no clear increase in capacity observed in this paper, but instead, a significant increase in mean speeds (10 km/h to 15 km/h) in the area of medium and high traffic flows has been reported. In STEINHOFF, 2002 was reported that the control 100 km/h in area of high flows resulted in an increase of mean speeds by 7.1 km/h. However, in MIDDELHAM, 2002, it is reported that no positive effects of harmonisation strategies on traffic efficiency have been measured in the Netherlands. In KELLER AND KÄMPF, 2001 a study on perspectives of telematic applications estimates that by 2010 a 10% (or higher) capacity increase should be expected due to appropriate LCS application. It is also estimated that a 5% increase in throughput through improvements of the existing systems should be expected. Research by BRILON AND DREWS, 1996, emphasized the potential of increasing capacity (and safety) by reducing lane-changing manoeuvres, especially in situations when traffic flow is close to capacity. KÜHNE ET AL., 1995 has reported a positive effect of prohibiting truck passing on stabilisation of traffic flow, and throughput increase of between 5% to 10%.

## 2.6 Conclusion

Link control systems and variable speed limits have a distinct advantage over static speed limits in increasing traffic safety and efficiency. Dynamic responses to dynamic changes in traffic flow offer an additional possibility to smooth driver behaviour and prepare them for event in front which is believed to reduce the probability of disturbances and/or their negative effects on traffic flow. Indeed, as demonstrated by implemented link control systems, a well-designed and maintained LCS is an efficient motorway traffic management unit (OBB, 1993; DG-TREN, 2002) in increasing traffic safety and reducing motorway congestion. However, link control is a highly sensitive and complex subject and probably one of the least studied areas within traffic control (PAPAGEORGIU, 2001).

Modern link control systems usually control long motorway corridors, utilising several tens of VMS panels, and pursue multiply, sometimes even conflicting goals. They should identify and react to various events (e.g. weather, congestions, accidents, etc) that, depending on the prevailing traffic situation, might affect traffic flow in various ways. Thus even only identifying these events is a highly complex task that, as numerous studies suggest, cannot be successful and equally reliably undertaken by a single algorithm. Additionally, once the critical event has been identified, determining the appropriate control response is even more difficult. Although many studies have been undertaken to investigate the LCS effect on traffic flow it is still not always clear what will be effects of different speed limits under different circumstances. Inherently, corresponding validated mathematical models currently lack and existing objective functions usually emphasise either efficiency or safety aspects, but fail to integrate the two.

All this, in addition to the fact that models should be able to work with limited and often corrupted macroscopic data, make obtaining of reliable (robust, adaptive and accurate) control model very difficult. This is probably the reason why most of the existing systems utilize several incident detection algorithms whose parameter setting depends on expert knowledge and judgement of what the right control is (partly due to the lack of integration of appropriate objective function into the system, and partly due to their too often false alarms). Therefore, these parameters are usually set in such a way to detect incidents when they have already happened and strongly affected traffic flow, thus taking advantage of only a smaller portion of LCS potentials. At other side, due to the lack of corresponding models, approaches based on the traffic flow model are forced to work on assumptions that may or may not be true and since the effects can be observed only later in time, they strongly rely on the quality of the predictive inputs. Furthermore, link control along long motorway corridors combined with often corrupted or missed data could seriously affect control system robustness. Hence, the evidence strongly suggests that there is considerable additional LCS potential and they may prove much more useful than at present by more careful studying their impacts and incorporating this into the systematic control framework, able to cope with practical requirements as well.

Before we discuss the possible solution and chosen approach let us summarize here the requirements upon successful control system (see also chapter 1.3). Given a clear set of control objectives and technologies, an ideal control methodology should possess the following characteristics (JIN AND ZHANG, 2000; BOGENBERGER, 2001A):

- **Theoretical Foundations** - i.e. reasonable assumptions and objectives, rigorous problem formulation, and efficient and accurate solution methods.
- **Accuracy and Robustness** - The control actions should be effective to achieve the control objective, and degrade gracefully when a part of the system, such as input links, is down.
- **Flexibility and Expandability** - The control method should be easy to implement, modify and expand to account for more complex and perhaps more realistic situations encountered in the motorway systems.
- **Adaptiveness** - Ability to adapt to changing traffic patterns and to handle special situations.
- **Simplicity** - The simplest logic structure possible to reconcile the demands on realism and theoretical elegance.
- **Proactivity and Balance** – Warning drivers before they have reached the “dangerous” area, rather than when they are already in the incident area, preventing accidents and congestion rather than reacting to them, avoiding over-congestion concentrated in one particular component of the system.
- **Reliable System Model** - The model should be able to accurately describe both the operations and the control measures in the motorway system. It should capture major



traffic flow phenomena that are critical to control design, such as shock waves, and driver response to control.

- **Computational Efficiency** - The algorithm should run fast, and require a moderate amount of memory.
- And we will add to this, practical requirements such as **transferability and understandability** – The model should be transferable to a new location and understandable to the user.

Based on our previous discussion and the above listed requirements for a successful link control model the development of a new control model should go in one of the following directions:

- Utilization of a *traffic flow model* expanded with better formulation of speed effects and combined with accurate and computationally efficient forecasting of boundary conditions, careful handling of input failures, and better formulation of the objective function (Chapter 2.4.2)
- Utilization of a *data-driven model* that is able to learn the complex dynamics of traffic flow from data according to a well-formulated objective function (so-called data driven approach).

The high dynamics of traffic flow, its non-linear nature and our limited knowledge about interactions between speed control and traffic flow together with frequently corrupted data makes it, at the moment, probably impossible to capture all these dynamics in a traffic flow model (even a highly-sophisticated one). Moreover, since the link control should react to a variety of situations, where some of them are hardly observable, all information is potentially significant. However, analogous to the real life, this significance depends on the context under which it is made and the amount of information that is already available (in the context of LCS information the significance depends on the traffic state under which information is provided and data quality – Chapter 2.4). Hence, particularly today, when we have more information than ever, mechanisms for their automatic inclusion, exclusion and combination, with the aim of increasing a predefined goal could bring much more benefit to the complex systems like LCS.

Therefore, the author believes that in the context of LCS it seems more advantageous to focus on separate *modelling of control effects* (objective function) on one hand, and developing a sufficiently flexible and model for *information<sup>17</sup> fusion* on the other. Objective function should capture the aforementioned major impacts of LCS control (warning and harmonization). From the user's point of view, the benefits of control such as warning and harmonization, etc., need to be weighed against the travel time cost of speed restrictions. This need is not simply a matter of convenience since compliance suffers from excessive speed

---

<sup>17</sup> When we speak about data we have in mind raw data, while information considers various indicators and outputs from detection/prediction/estimation algorithms

limits, and false alarms have a negative long-term effect on traffic safety as well. A key problem is to reduce these multiple, and sometimes conflicting, criteria to a one-dimensional value appropriate for optimisation.

With these two components it would be possible to realize the so-called data-driven control model able to automatically integrate and distinguish between varying information reliability and importance. Contrary to the approaches based on a physical model of the system, data-driven models attempt to approximate complex system behaviour from data according to the given objective function. Another compelling argument in favour of data-driven approaches is that feedback processes are also automatically dealt with, since these are also “present” in the data due to user responses to the system. Unless the driver responds with completely different driver behaviour (speeds, headways), it is most likely that the response to the link control system will yield traffic condition which is also observed in the data used to calibrate the model. The traffic flow simulation model might produce different results for different forecasted boundary conditions, while data-driven approach “has no concern” for the underlying forecasted values, given that it is familiar with the resulting traffic patterns observable in the data.

However, it is often mistakenly believed that it is simply enough to apply some advanced technique and that results will be good. This is of course not true. Just like the physical-based approaches depend on the model of the physical system, so the data-driven models depend on the definition of “the good decision” (objective function) and the chosen “approximator” structure. The system can be only as good as the objective function and therefore it is necessary to carefully design it. Furthermore, a sufficient amount of data and its appropriate treatment is prerequisite for properly learning the desired behaviour. Since data-driven models can have huge amount of parameters, and consequently approximate highly non-linear processes, it is important to carefully handle system complexity in order to prevent an over-fitting effect. Furthermore, different techniques (from statistics to artificial intelligence field) of different shapes and complexities could be used for deriving the model. Although the data driven approaches do not aim to describe the physical system, but rather to “learn” the relations between its parameters so to approximate its behaviour, the derived model must be tractable and understandable to the user (operator). Therefore, a careful trade-off between accuracy and simplicity should be made when choosing the model.

In this thesis the comprehensive, practically applicable, adaptive link control framework called INCA, based on the data-driven approach, has been developed and will be described in the following chapter. The crucial components of the system are a *cost-benefit-based objective function*, a *flexible information fusion model*, and an *optimisation approach that supports the derivation of the general control model*. The objective function quantifies the basic warning and harmonisation benefits against the travel-time losses, by referencing the incident risk probabilities (potential or real) in different situations/controls, and considering

the information propagation in upstream direction. Information is combined with the (Logit-based) regression model. The model offers the possibility for information “weighting”; additional mechanism for determining the appropriate trade-off between model bias and variance can be easily incorporated; optimisation can be done computationally efficient, and it is flexible and understandable even with large number of control parameters. It is suitable for modelling processes in which discrete decisions should be made, as is the case with link control systems. In order to handle model complexity and attribute correlation mechanisms, such as Ridge regression and Bootstrap, have been incorporated into the system. Control decision is made in two steps, first at each control station and then coordinated along the corridor. The framework is supported by additional data correction and imputation mechanisms to increase system robustness. Finally, the system modular structure allows easy extension, re-configuration and further independent development of each component.

### **3 Data driven model for link control on motorways using an empirical objective function: INCA**

In the previous chapter, we illustrated the complex link control problem and its potential for increasing traffic safety and efficiency. We concluded that, although existing approaches have shown to increase safety and efficiency, greater benefit could be expected from the development of a more precise objective function and systematic control framework. Such systematic control framework could be developed either by using a sophisticated traffic flow model combined with a better model of speed control effects and prediction of boundary conditions, or by applying a carefully designed, “intelligent” data-driven approach. Both approaches are supported by a large body of theoretical literature on analogous problems. This thesis concentrates on the latter of the two methodologies: the “intelligent” paradigm of a system that approximates the model structure by “learning” from data. “Learning” in this context refers to the estimation of model parameters (and inherently structure) in such a way that would maximize an appropriate objective function that captures safety and efficiency. The goal was to develop a comprehensive link control framework that is based on theoretical knowledge (about traffic dynamic and data driven approaches) and applicable in the German context. To this end, the model must satisfy a range of strict theoretical to practical requirements such as (chapter 1.2) reliability, transferability, interpretability, etc.

Since there are no general speed limits on motorways in Germany, the model should be able not only to detect and warn the driver about already existing danger on the road, but also to “smooth” highly dense but still flowing traffic so to postpone or prevent disturbances. This need must be (and is) reflected in the developed control model and in the utilised objective function. Since there is no such function in literature (chapter 2.4.3), in this work the new one has been developed based on the work of KATES ET AL., 1999; STEINHOFF, 2002; BUSCH, 1983; EVERTS AND ZACKOR, 1976. The major challenge is to determine which of several,

sometimes conflicting, goals should be pursued under dynamic traffic conditions. Due to our still limited knowledge about the direct impacts of link control measures on traffic flow, the new function is focused on modelling the speed limit effects. Inherently, the developed control model focuses on warning drivers and “smoothing” traffic flow by means of variable speed limits. Other controls, such as weather related or manual controls for example, are also integrated but as an external input and thus are not optimised within the system. However, if our knowledge about the effects of other controls expands, the flexible system structure would allow further extension of the objective function and consequently extension of the control model.

Numerous incident detection/estimation/prediction algorithms have been proposed in literature. However, the usually lacking connection with the objective function and problem complexity often lead to their underutilisation in practice. The thesis is developed based on the belief that these algorithms offer valuable information if properly utilised (combined). However, this information is different in nature, frequency and reliability. Furthermore, its reliability and importance vary under different traffic conditions and this knowledge should be also incorporated in the model. Clearly, combining different information from long motorway corridors might result in a system that is too complex to be tractable and, inherently, to be used. Therefore the derived model must be tractable, interpretable and understandable to the user (scientist, engineer, operator, etc.).

A too complex system may also negatively influence its reliability. A reliable model is at the same time robust, accurate, valid and adaptive. However, there are usually trade-offs between these properties. Robustness does not necessarily encompass accuracy, and vice versa: the most accurate model will probably not be the most robust one due to its sophistication and sensitivity to fault, incomplete and noisy information, which are often present in practice. In practical applications, in the trade-off between accuracy and robustness, robustness is usually far more critical. The choice and performance of control solutions must therefore be measured not only by its accuracy but also by its robustness, interpretability, simplicity, user-friendliness, overall cost, etc. Therefore, in the context of the link control problem and data driven approaches, it seems that the most parsimonious model possible (simple but accurate enough) would best satisfy above consideration.

Thus of the various, more or less sophisticated data-driven approaches, the one that should be used is the one that is able to approximate the behaviour but is still understandable. Preferably, the knowledge about system dynamics and characteristics should be included into the determination of model structure. After all, no matter which data driven method is used the systematic data driven framework will always<sup>18</sup> require the same components, while individual used techniques may vary.

---

<sup>18</sup> There is a slight difference for neural networks, which have the methods for parameter optimisation (e.g. back propagation) integrated in them, while other approaches require a combination with some optimisation

In addition to theoretical considerations, practical constraints in the field also pose important challenges. The data collection systems mostly deployed in practice, consist of local detection equipment resulting in local aggregated traffic flow characteristics (flows, local mean speeds). Hence, both the new model and objective function should be able to work with this aggregated, macroscopic data. Furthermore, control should be able to face problems such as detection at widely spaced points, the problem of traffic data that is frequently incomplete or faulty, the necessity to establish confidence during a transition period, etc.

*Therefore, the comprehensive link control framework developed in this thesis (INCA) is focused on answering the following questions:*

- What is the appropriate traffic management response (the optimal speed on a variable message sign) to a complex traffic situation?
- How to assess the quality of this response, taking into account the overall influence on motorway traffic?
- How to optimise the quality of this response and integrate this optimisation into the control process?
- Which model for information integration is sufficiently flexible to allow optimisation and is capable of representing variable reliability and the priority of the available information, but is still understandable and interpretable?
- Which mechanisms should be incorporated into the model to enable automatic handling of complexity (trade-off between bias and variance, thus accuracy and robustness)?
- Which mechanisms should be incorporated to handle often corrupted and missing data?

As a result INCA, a new, practicably applicable, LCS model with an adaptive data and information fusion approach based on optimisation of an appropriately defined one-dimensional objective function, is developed.

The following chapters are organised as follows. First, in order to focus the control modelling problem, the available strategies, their characteristics, and definition of critical events, as they are seen in this work, will be given. This will be followed by the basics of data-driven approaches and considerations that should be taken into account when determining the model (which led to the deriving of the INCA structure). Afterwards, the general hypothesis about the control aims (concepts regarding control strategies and definition of incident) and constraints upon the developed system (some rules that control decision should fulfil) will be addressed. The general control framework and basic components will be briefly described next. In subsequent chapters, each component of control framework is described in details.

---

technique. However, this is a difference in the specific technique but not in the concept itself – optimisation is needed.

### 3.1 Data driven approaches

The INCA is based on the knowledge from different overlapping fields that are in literature usually referred to as data-driven approaches. While the physical based control approaches try to model and explain the underlying process (e.g. traffic-flow model), data-driven approaches are based on limited knowledge about the process under investigation and rely on data describing input and output characteristics. Data-driven approaches use results from such overlapping fields as statistics, data mining, artificial neural networks (ANN)<sup>19</sup>, rule-based type approaches such as expert systems, fuzzy logic concepts, rule-induction and machine learning systems (SOLOMATINE, 2002). They are able not only to approximate practically any given function, but also to generalise, providing correct output for previously "unseen" inputs, provided that sufficient data is available and training is carried out properly. Thus data-driven approaches are a very powerful tool for handling non-linear systems with unknown models, or models characterised by uncertainty, vagueness and imprecision.

Another compelling argument in favour of data-driven approaches is that feedback processes are automatically dealt with as well, since these are also "present" in the data due to user response to the system. It simply does not matter for a model whether a specific traffic condition is the result of a feedback process following its installation, or not, as long as it is familiar with the resulting data. Unless the driver responds with completely different behaviour (speeds, headways), it is most likely that a response to the link control system will yield a traffic condition also observed in the data used to calibrate the model.

As long as a measure of performance is available and supported by appropriate mechanisms, a properly designed data-driven framework will be able to automatically decide whether the new available information is relevant and what its importance is. This concept is also important in the context of future development where even more data, different in nature and frequency (e.g. from FCD, AVI, etc), will be available.

However, in order to use the advantages of the data driven approaches, it is necessary to have a sufficient amount of data available and that the "learning" is done properly. Data-driven models, if complex enough, could approximate any given behaviour. However, such approximation is done with the help of many parameters and is based on data that usually does not represent all the possible situations. This can lead to a model that is very sensitive to changes in the characteristics of input data set. Thus it may frequently happen that the developed model performs well on known (training) data but is very unstable under new (test) data. This is also known as the problem of model bias and variance and leads us back to the discussion about the trade-off between accuracy and robustness. Therefore it is necessary to carefully determine model complexity, the number of parameters and use of available training

---

<sup>19</sup> Please note that many classical statistical models could be also classified as ANN models (SARLE, 1998; BISHOP, 1995)

data. Furthermore, as already stated, for the control systems it is of great importance to be understandable and interpretable, and preferably based on prior knowledge about dynamics of the controlled system. Hence, just as the physical-based approaches depend on the model of the physical system (e.g. traffic flow model), the data-driven approaches depend on the chosen model structure and methods for the optimisation of model parameters and complexity.

### 3.1.1 Model bias, variance and Parsimony principal

Although data-driven models could theoretically approximate almost any non-linear function, their robustness strongly depends on the complexity of the derived model and proper treatment of available data. This is due to the well-known trade-off between model bias and variance (BOX AND JENKINS, 1976). While the bias error is purely due to the structural inflexibility of the model, the variance error is part of the model error which is due to uncertainties in the estimated parameters. Clearly the number of parameters crucially influences this decomposition. This is especially, but not only, important for data-driven approaches as they can have many parameters fitted to approximate the behaviour observed in data used for optimisation<sup>20</sup>. It is clear that the bias error<sup>21</sup> is large for inflexible models and decreases with the increase in model complexity. One modelling goal must be to make bias error small, which implies to make the model very flexible. However as the number of parameters increases, the benefit of incorporating additional parameters decreases. This is due to variance error. Since in practice one always deals with a finite and noisy data set from which the model parameter are estimated, these parameters usually deviate from their optimal values as they are optimised so to fit to this specific data set (and the noise contained in them). Thus, the variance error increases with the number of parameters in the model and if ignored could lead to over-fitting effect, where slight modification in the input can lead to drastic changes in system behaviour. The *parsimony principal* stems directly from this fact (BOX AND JENKINS, 1976), and states that from all models that can accurately describe the process, the simplest one is the best.

In conclusion: the model should not be too simple because then it would not be capable of capturing the process behaviour with a reasonable degree of accuracy. On the other hand, the model should not be too complex because then it would contain too many parameters to be properly estimated with the available finite data set. Thus, the “optimal model complexity” should be chosen – one that makes *the best trade-off between model bias and variance*.

---

<sup>20</sup> In literature on artificial intelligence also called training data.

<sup>21</sup> Non-linear processes usually cannot be modelled without a bias error. The only exception occurs when the true nonlinear structure of the process is known, e.g. through physical modelling

### 3.1.2 Optimising model parameters and model complexity

With the use of the parameter optimisation procedure data-driven model will be able to “learn” the best behaviour for **given** (seen) data set. In literature this data set is usually called the training data set. Although this is a nice feature, we are interested in the model behaviour on **unknown** or real (unseen) data set. This data set is usually called the test data set. If the learned model behaves approximately equally well on unseen as on the test data set then the model could be considered to be general. However, the **generalisation** property is not given to the data driven models<sup>22</sup> by default, but it must be learned as well. For determining the best bias vs. variance trade-off (most general model), the model quality on previously unseen data (not the one used for the “learning” of the model) must be assessed. Otherwise the over-fitting cannot be detected – a too complex model that adapts well to the known data and noise contained within it, but performs much poorer on unseen data set. For this reason a test set or/and some complexity penalty term must be applied.

The good bias-variance trade-off could be realised through explicit and/or implicit structure optimisation (NELLES, 1999). Explicit means that optimisation is carried out by examining models with different number of parameters. In other words, model performance is examined with explicit exclusion/inclusion of the new parameters, neurons, rules, etc. They are computationally expensive but result in smaller models where all parameters are significant. Implicit structure optimisation, also called regularisation, does not change the nominal number of parameters but nevertheless model complexity varies. This is done by choosing a relatively complex model where not all degrees of freedom are actually used. Although the model can have many parameters, the effective number of parameters can be significantly smaller than the nominal ones. Thus, regularisation increases bias error, and decreases variance error. This is done by regularisation techniques such as non-smoothness penalties or constraints (curvature penalty, Ridge regression, weight decay), training with early stopping, and local or staggered optimisation.

The simplest way to determine model quality/generality is to “learn” it using the training data set and evaluate its performances using a different test data set (sometimes a validation data set can be also formed). If a sufficient amount of data is available then the data could be split into these separate data sets. It is important that both data sets are representative – equally well covering all considered process operating regimes. However this is usually not the case and data cannot be separated in the representative groups. In such cases more sophisticated approaches are needed, such as (NELLES, 1999): Cross validation; Jackknife, Bootstrapping; Information criteria; Multi-objective optimisation; Statistical tests; Correlation-based methods, etc.

---

<sup>22</sup> Generalisation is often claimed to be characteristic of data-driven approaches (e.g. “neural networks could generalize any behaviour” (SARLE, 2001)) but without really analysing model stability.



The model complexity could be also reduced by reduction of the “course of dimensionality”. It is based on the fact that, in general, models become more difficult to solve as the dimensionality of the input space increases. It is the intrinsic property of the problem and it is independent on the specific model employed. In high-dimensional space it is almost impossible to approximate general nonlinear function. However, in practice the course of dimensionality is much less severe because in reality some of the following is fulfilled: there are non-reachable regions of the input-space (for example high pressure and low temperature may be contradictory in physical systems) that does not have to be covered by the model; inputs are correlated or redundant; the behaviour is very smooth in some regions in the input-space and thus the nonlinear function can be very simple (even linear), in some regions of the input-space; in some regions of the input-space very accurate model behaviour is not necessary. Therefore, the amount of data required for the solution of real-world problems usually does not increase exponentially with input-space dimensionality. However, it is a question of how model complexity increases with input-space dimensionality. Several models (e.g. classical look-up tables or fuzzy models) might suffer from the course of dimensionality, i.e. their complexity might increase exponentially with the number of inputs. The most common mechanisms for overcoming or reducing the course of dimensionality are additive, hierarchical, tree structures, or projection-based structures.

### 3.1.3 Data driven methods in the context of link control

Since presenting all the data-driven methods goes beyond the scope of this work, in following text several methods that were considered as possible solutions to the link control problem are described. In general, data-driven link control problems can be described as follows. Given  $K$  (continuous and discrete) values  $X_k$ ,  $k = 1 \dots K$  from different detection/estimation/prediction algorithms that are available as a time series (one value per minute), make the best decision for each control station (VMS), in tractable and robust manner.

Different methods could be used for solving the abovementioned problem. The following three methods, with different complexity and philosophies, will be analysed here:

- Regression analysis
- Neural networks
- Fuzzy logic

#### Regression analysis

In its simplest form regression analysis involves finding the best straight line relationship to explain how the variation in an outcome (or dependent) variable  $Y$  depends on the variation in a predictor (or independent or explanatory) variable  $X$ . However, in most cases, the outcome will depend on more than one explanatory variable. This leads to the multiple regression, in which the outcome is derived through linear combination of the possible explanatory variable:

$$Z = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k, \quad \text{Eq. 3-1}$$

Multiple regression can be easily extended to deal with situations where the response consists of  $p > 1$  different variables  $Z$ . Such an extended model is known as multivariate regression analysis (or general linear models). More details on distinct advantages of multivariate regression can be found in (e.g. BOX AND JENKINS, 1976).

Parameters  $\beta_k$  could be *determined* through optimisation using empirical data, usually with the maximisation of the so-called statistical likelihood function. If necessary, it is also possible to determine value other than  $Z$  and/or to optimise some objective function other than the likelihood function. For example, in the case of link control probably the best estimate is not the accuracy of probability estimation, but instead the ability to produce the greatest benefit.

In order to include model nonlinearity in the system, it could be additionally extended with a further combination of available data (e.g. their products, cubes, quadrates, etc) or use of generalised additive models. In generalised models the dependent variable values are predicted from a linear combination of explanatory variables, which are "connected" to the dependent variable via a *link function*. Depending on the assumed distribution of the  $y$  variable values, various link functions can be chosen (MCCULLAGH AND NELDER, 1983): identity link ( $f(z) = z$ ); Log link ( $f(z) = \log(z)$ ); logit link ( $f(z) = \log(z/(1-z))$ ), etc.

However, each further non-linearisation would make interpretation of the beta-coefficients more difficult. Moreover, correlation between input values, although without influence on the performances when assessed on the same set, might negatively influence performances of the regression models. Firstly, the variance of the regression coefficients may be so high that individual coefficients are not statistically significant. Secondly, the relative magnitudes and even the signs of the coefficients may defy interpretation. Thirdly, the values of the individual regression coefficients may change radically with the elimination or addition of a predictor variable in the equation. In fact, the sign of the coefficient might even change.

### **Artificial neural networks**

Neural networks (NN) are a wide class of flexible nonlinear regression and discriminant analysis, data reduction models, and nonlinear dynamical systems. Their development was originally motivated by the biological brain structure and its ability to carry out complex tasks (such as information processing, learning, adapting) (SARLE, 2001). Analogous to the brain structure, the major (processing) unit of the neural network is a "neuron". The "neuron" carries out linear or nonlinear transformation of input values into one or more output signals

(see Figure 3-1, left). However, while the intelligence of the human brain is preserved by about one hundred billion neurons, the typical NN rarely have more than a few hundred<sup>23</sup>.

Input values  $X$  forwarded to the “neuron” can be discrete and continuous. They are modified with the positive or negative weights of the link  $\beta_{jk}$  (whose function is analogue to the synaptic junction of biological neurons). The neuron  $j$  is made up of two parts. The first simply aggregates weighted  $K$  inputs resulting in the quantity  $Z$ ; the second part is a linear or nonlinear filter, usually called activation function  $f$ , through which combined output signal  $y_j$  flows:

$$Z_j = \sum_{k=0}^n \beta_{jk} * x_k; \quad y_j = f(Z_j), \quad \text{Eq. 3-2}$$

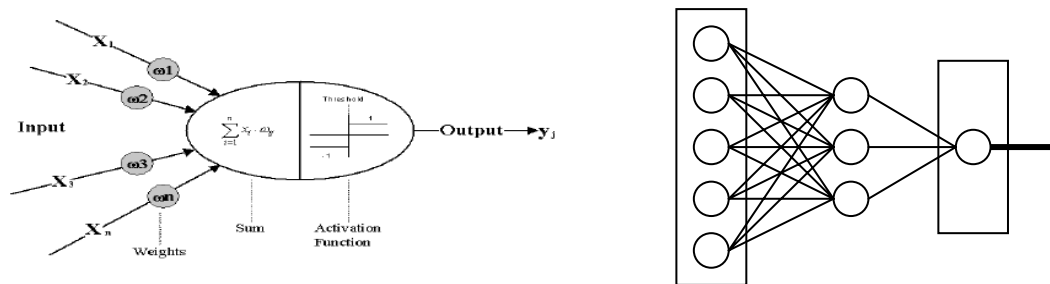
An activation function  $f$  (*analogous to the link function* in generalised regression models) maps any real input into a usually bounded range, often 0 to 1, or -1 to 1, usually called sigmoid functions. The most common activation function is probably the logistic function. Beside, other functions such as threshold, linear or identity, hyperbolic tangent, Gaussian are also often used.

Neural networks often consist of a large number of such neurons, usually interconnected in complex ways and often organised into layers (see Figure 3-1, right). The values coming from outside world usually form the so-called input layer. The final output from the NN is calculated by the above described “neurons” that are usually grouped into the output layer. A NN can have one or more hidden layers between the two, which consist of one or more “neurons” and where further transformations are made. ***The NN without hidden layers and one neuron in the output layer, which is activated by a logistic function, is the well known logistic regression from the previous chapter.*** Indeed, many neural networks (e.g. multilayer perceptrons MLP, see BISHOP 1995) are actually similar or identical to popular statistical techniques such as generalised linear models, nonlinear or linear regression, etc. A very interesting overview of the relation between NN and statistical fields is given in RIPLEY, 1993; SARLE 1995.

The ability of NN to approximate any non-linear function, their flexible structure, and integrated “learning” mechanisms makes them very popular for solving complex non-linear problems where a system model is not available. However they are very hard to interpret due to the many parameters that do not necessarily have “physical” meaning. This might not be a problem in the application where we are interested only in the prediction of some value (e.g. travel time prediction) but this is one of the critical points for control model interpretability. Additionally, too many parameters, due to probably their higher variance, could lead to an unstable system, if it has not been “learned” properly.

---

<sup>23</sup> The largest NN typically have several thousand neurons and less than a million connections (synaptic junctions)



**Figure 3-1: Neural networks. Left: a typical neuron (BOGENBERGER, 2002). Right: architecture of a multilayer neural network**

## Fuzzy logic

The fuzzy set theory was first introduced by ZADEH, 1965, inspired by the process of uncertainty, imprecision and vagueness in human perception and communication. Fuzzy set theory can be seen as a generalisation of the classical set theory. In the classical set theory very precise (crisp) bounds separate the elements that belong to certain set from the elements that do not. In human communication and description of the processes usually terms such as “large”, “medium”, “small” are used. However, each person has a different understanding of what is “large” and what is “small”. Moreover, some values can be more “small” than others. Fuzzy sets are inspired by these considerations. They are classes with unsharp boundaries in which the membership is a matter of a degree and can take any value from the closed interval  $(0,1)$ .

Fuzzy logic refers to all theories and technologies that employ fuzzy sets. Thus, in fuzzy logic, unlike in classical logic, the truth of any statement becomes a matter of degree. Fuzzy logic is based on a linguistic variable, characterised by a linguistic term (i.e. name of the fuzzy set) and corresponding membership function (express the meaning of the fuzzy set). *Fuzzy control is a control systems based on fuzzy logic.* As fuzzy logic can be defined as reasoning with words rather than numbers, so fuzzy control can be identified as control with sentences rather than equations. Fuzzy control is a knowledge-based control strategy. It is particularly useful when a sufficiently accurate and yet reasonably complex model of the physical system is not available, or a performance measure is not meaningful or practical. The control designed problem is usually empirically acquired knowledge regarding the operation of the process, instead of being constrained within a strictly analytic framework. This knowledge, cast into linguistic or rule-based form, constitutes the core of a fuzzy control system. Its many successful implementation in other fields have inspired its (now wide-ranging) use in the field of traffic engineering (e.g. BUSCH, 1994A; TEODOROVIC, 1999; BOGENBERGER ET AL., 2001B).

The basic elements of a fuzzy controller are: fuzzification, inference (rule-base) and defuzzification. In the fuzzification part membership functions are calculated for each crisp input value. Inference represents the core of the fuzzy controller, where fuzzy IF THEN rules are specified and represent the decision process. Fuzzy rules usually represent human

expertise and knowledge. Defuzzification is a part where such fuzzy outputs are transferred into crisp output, usable for control. A detail description of fuzzy set theory, logic and control and its application in traffic engineering can be found in TEODOROVIC, 1999.

Fuzzy logic and control has the distinct feature of being easily interpretable and tractable by the user, since it represents human reasoning. On the down side, when the number of variables and rules used exceeds a certain point, fuzzy control models are known for the problem of exponentially growing course of dimensionality. There are various methods proposed in TEODOROVIC, 2005A; WANG AND MANDEL, 1992, for reducing model complexity by elimination of less important rules. However, in the context of the link control problem where at least twenty pieces of information can be assigned to each control station, and due to our desire to develop a tractable model even with a theoretically unlimited quantity of data, is it assumed that fuzzy logic would lose this key property.

### 3.1.4 Common components of data-driven approaches

Data-driven methods vary in the level of complexity, approximation capability, generalisation characteristics, etc. What is the best model depends on the specific problem under consideration. However, regardless of which approach is used, *each comprehensive data-driven approach should include the following components:*

- a mechanism for the optimisation of model parameters,
- a mechanism for the detection and optimisation of the model complexity,
- a measure of performance able to represent the model effects with available data.

Additionally, for the control problem it is important to have a model that is:

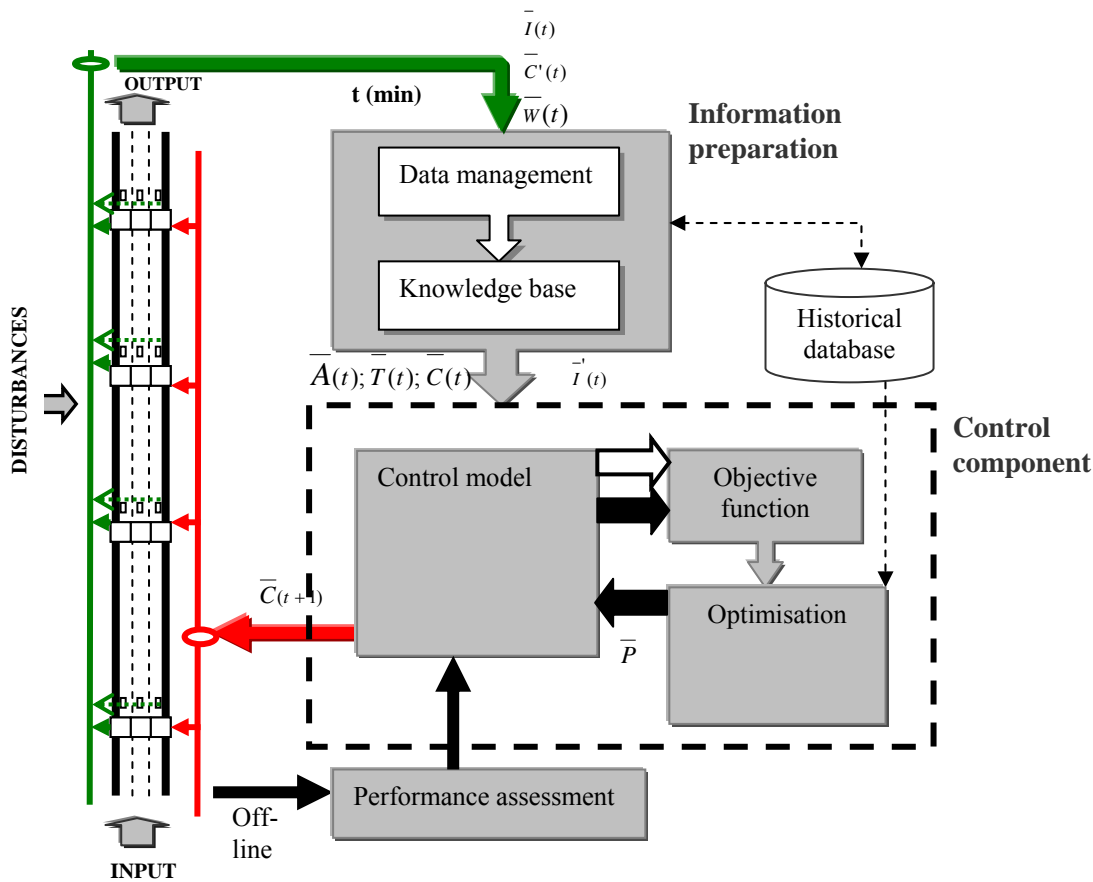
- general, tractable, interpretable and configurable.

## 3.2 INCA general link control framework

In view of the history of seemingly promising theoretical approaches that failed to meet performance expectations in practice, the data-driven method in this thesis was implemented in the INCA framework and was tested extensively and successfully. An overview of the control framework is shown in Figure 3-2.

The framework is divided into information preparation and control components. Information preparation part could be portrayed as an external component, where any data, indicator or algorithm can be integrated. The control component is responsible for the appropriate integration of the information and producing the control decision. Thus, it is dependable and cannot function without information preparation part. A short description of each component and their basic functionalities is given in the following text. Detailed specification of each component is given in subsequent chapters.

The link control systems are supplied with data collection equipment for collecting traffic and sometimes weather data in the area. Data is delivered usually in one minute intervals. Each minute  $t$  data about traffic flow characteristics  $\bar{I}(t)$ , weather  $\bar{W}(t)$ , and active controls  $\bar{C}(t)$  are forwarded to the INCA *data management* component. The data management module analyses quality of raw data, makes corrections if possible and necessary, and archives the data ( $\bar{I}(t)$ ). For the sake of simplicity we show the database as a local storage but the data can also be saved in files (ASCII, XML) as well. Since the data is often corrupted or missing (chapter 2.2.1), the data checking and correction process plays an important role in the overall system.



**Figure 3-2: Comprehensive link control framework: INCA**

Once pre-processed, data can be archived and forwarded to the *knowledge base* component. In the knowledge based component information is derived from reports, and appropriate indicators and detection/estimation/prediction algorithms ( $\bar{A}(t)$ ) are computed and recorded. This is the part where any new algorithm can be implemented. Furthermore, it analyses the aspects from which potential differences in control applicability or algorithm reliability could arise – e.g. data quality, traffic context, weather, truck percentage, etc. – and thus provides a basis for improving control quality and robustness. In INCA, determining the current traffic

conditions ( $\bar{T}(t)$ ) is believed to provide an additional input for more intelligent data integration and it is thus implemented in the knowledge base. The information produced in the knowledge base component is forwarded to the control component.

Information from data management and knowledge base ( $\bar{I}(t)$ ,  $\bar{C}(t)$ ,  $\bar{A}(t)$ ,  $\bar{T}(t)$ ) represent the input for the control component. The control component consists of three main parts: the objective function, the control model, and optimisation. An *objective function* is used for optimisation as well as for monitoring system performance. It quantifies the warning and harmonisation benefits of the imposed control and induced travel-time losses. One could observe two data flows: white line that represents the on-line process which is activated each minute and black line that represent the off-line optimisation loop. During each measuring interval the *control decision* for the next time interval is determined based on the data from the knowledge base and the state of influencing factors (traffic context). Due to the problem complexity, the speed posted by the LCS is determined in a hierarchical manner, by a two-stage process beginning with a “local” “segment-based” control decision in which one “intelligent” local controller is assigned to each control station. The local decision model is called segment-based since it utilises normalised data from the local but also from neighbouring detectors, and as such should be able to make the best decision for the motorway segment in front (up to the next control station). Warning and harmonisation decisions are determined using different models: warning decisions are based on (logit based) regression model, and harmonisation decisions are based on the Belastung algorithm. In the second, global stage, the local decisions are forwarded to the global layer which adjusts them if necessary. Adjustments are done either due to some external higher priority control (e.g. manual, weather, etc.) or due to the control “coordination/smoothing” rules. Mechanisms for preventing too frequent changes of the control actions are also implemented. The control decisions that are derived in this manner  $\bar{C}(t+1)$  are sent back to the VMS panels.

The parameters of the warning and harmonisation component are the object of an optimisation process. For a typical LCS, the data supply permits testing and re-calibration several times per year. Additionally, by continuously monitoring performance, significant structural changes in traffic patterns or severe technical problems can be detected. Misdetection and successful detection may be seen after several minutes. However, in order to reach a final conclusion about control performances it is necessary to observe control behaviour during several subsequent events that are not happening frequently. The optimisation component is made up of two parts – parameter optimisation and model complexity optimisation. In INCA, parameter ( $\bar{P}$ ) optimisation is achieved using the downhill simplex method, where model complexity optimisation and the handling of correlation between parameters are supported by the Bootstrap and the Ridge regression approaches, respectively.

A modular structure of the developed framework allows for easy modifications/changes of the applied techniques, and it is highly configurable and easily interpretable. The model has been

installed at Munich traffic control centre (chapter 4), and extensively tested (chapter 5) in the field (in on-line open-loop tests).

### 3.3 Objective function

Existence of the *different goals (efficiency and safety)* and importance of user *acceptance* differentiate the existing problem from the known applications of objective functions in traffic networks, because by LCS installations it is not only the travel time that is in focus but also traffic safety. Link control systems have the great potential to reduce accident risk through *timely* detection and warning drivers about the danger ahead. Furthermore, harmonising traffic flow reduces the probability of an accident and positively affects traffic performance (KATES ET AL, 2002). From both the system and user points of view, the benefits of control such as warning and harmonisation, etc., need to be weighed against the travel time cost of speed restrictions. This need is not simply a matter of convenience since compliance suffers from excessive speed limits, and false alarms also have a negative long-term effect on traffic safety (KATES ET AL., 1999, STEINHOFF, 2001A; ZACKOR AND SCHWENZER, 1988). Hence, including travel time in the objective function is appropriate, even if traffic safety is taken as the absolute priority. A key problem is to define operational measure of traffic safety and to reduce these multiple, and sometimes conflicting, criteria to a one-dimensional value appropriate for optimisation.

*The objective function presented here rewards control (algorithms) with low false-alarm rates, high detection rates, as well as controls inducing an appropriate level of harmonisation, and expresses the aggregated model performance in terms of time savings.* The objective function can be used for optimisation, monitoring, and for off-line and on-line evaluation of the entire system. Additionally, it could also be used for the optimisation and evaluation of individual algorithms. Furthermore, a “more meaningful” function should positively influence public awareness and thus acceptance. It has various configurable parameters whose specification should correspond to our knowledge about traffic dynamic, warning and harmonisation benefits, and site-specific needs.

The further text first gives the precise definition of what constitutes an incident, and what can be considered as appropriate or inappropriate warning. Subsequently different control situations and possible control costs and benefits will be analysed. The definition of a basic accident exposure and costs and benefits derived from it will be given next. The chapter will be concluded with an overview of objective function parameters and the algorithm for its calculation.

#### 3.3.1 Definition of an incident

A precise definition of an incident in this work is necessary in order to have unique understanding of warning strategies and the inherently unified approach to the evaluation of link control effects. Different definitions of incidents can be found in literature. They can be



microscopic such as the one in (BUSCH, 1986): “Deviation of one or more road user from currently intended driver behaviour”, or macroscopic such as the one in (ZHANG ET AL., 1995): “*An unexpected event that decreases the road capacity and/or reduces the traffic safety level and that may also disturb traffic flow condition*”. The incident pattern is governed by several factors, the most important of which are traffic condition prior to the incident, number of lanes closed, density of incoming traffic (demand), incident type, incident location relative to entrance/exist ramps, etc.

In this thesis a definition that can be used in real-time, allowing automatic identification of critical situations and quantifying of system performances is needed. Motorway control systems that operate in real-time (on-line) usually only have aggregated traffic measurements available, and sometimes weather data. Hence, in real-time operation (or optimisation process) incident can be automatically *identified* only if it noticeably influences the available macroscopic traffic data.

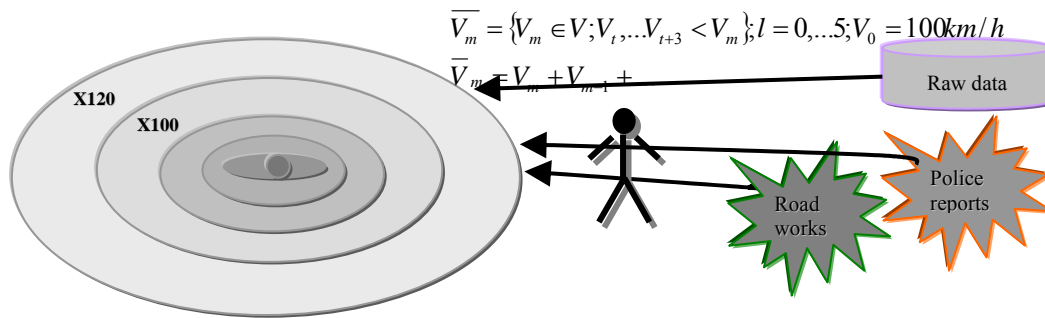
Variable speed limits as a mean of collective control can be used to adjust driven speeds so that they would be in better accordance with the downstream situation. Therefore, for link control systems it is important to identify changes in the traffic flow, rather than their causes. As shown in chapter 2.1.2, one of the most critical events that the driver collective could face, and it is also observable in traffic data, is an unexpected difference in driven speeds along the motorway. According to above considerations, the definition of an incident will be narrowed and focus will be shifted from cause to consequence. Thus, the definition in BUSCH, 1986 is modified using the one in KATES ET AL., 2002, and following one will be used: “***an incident is defined as a sharp speed drop along a section***<sup>24</sup>”. This implies that in this context recurrent congestions could be (and are) also seen as incidents as they may have the same effects - they result in lower speeds at the place of an event and could induce shockwaves.

Not all incidents have the same importance or are equally severe. Since detailed incident information is not available in the detector data, a “surrogate” represented as a sequence of *m speed levels*  $V(m)$  (HOOPS ET AL., 2000) from 20km/h to 100km/h at step 20km/h, ***is introduced in the above definition of an incident***. Introduction of speed levels makes it possible to distinguish (even approximately) between less and more important incidents.

An event will be classified as a speed drop of level *m* if speeds are lower than this specific speed level  $m(V(m))$  for more than 3 minutes. Once the speed drop under some specific level has been identified, the incident is considered to last until the measured speed reaches 80km/h (100km/h) again. If the speed drop below 60km/h is identified, it will also belong to speed drop group of 80km/h. With these definitions it is possible to extract information about the number of events ( $n(m)$ ) of some specific speed level directly from data (see Figure 3-3).

---

<sup>24</sup> It is important to observe if the speed drop lasts for several minutes in order to filter out situations that are simply due to the frequent changes of traffic measures.



**Figure 3-3: Incidents and their severities – generated from data and reports**

Clearly, the given definition of incident does not capture all dangerous events. There are other safety critical events (e.g. incident in free flow) that would probably not cause speed drops to occur, and therefore will not be observable in loop detector data. However, since it is important to have a rule which would enable automatic quantification of system performances for monitoring and optimisation purposes, it is necessary to have the above definitions. For the offline optimisation (e.g. if police reports would be available) it would be possible to add this information and artificially introduce corresponding severity information (see Figure 3-3).

It is important to emphasize that an incident usually happens somewhere between two detector stations and some time is needed for its effects to propagate to the upstream detector. On the basis of empirical data from A80st motorway the congested-region boundary propagates upstream at speed of 4m/s, where the exact value depends on incident characteristics, road geometry, and traffic level (see chapter 2.1.2). In a setting with detectors placed at a distance of up to 3 km, propagation of information in upstream direction can take several minutes. Hence, reacting when lower speeds are already measured in data (and drivers are driving slowly anyway) is too late. The driver should be *informed in advance*, before the incident area is reached in order to induce indirect safety benefits. This is an important fact that was taken into account in the development of LCS objective function.

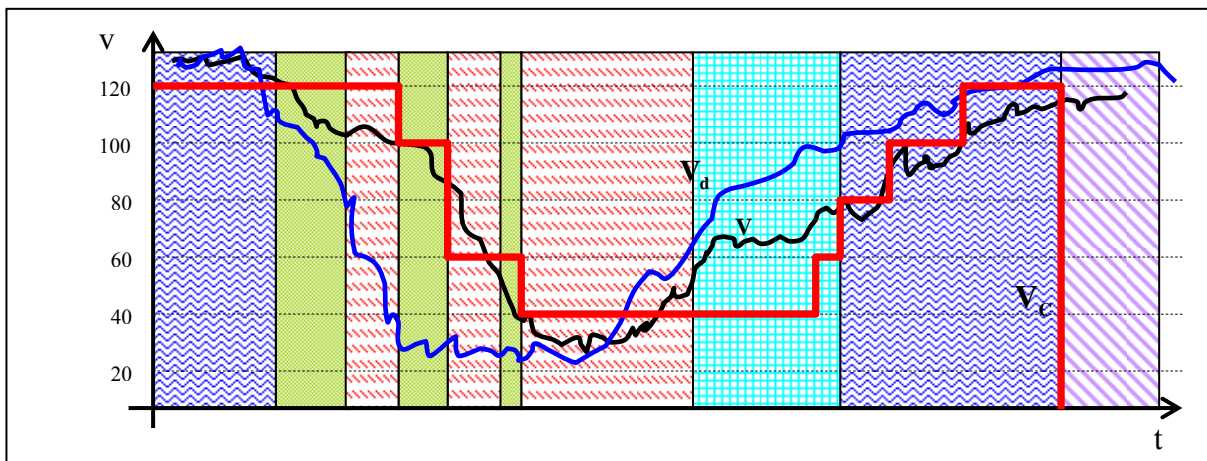
### 3.3.2 Analysis of typical link control situations

In order to unambiguously define the situations in which control can be beneficial and in which it cannot, the typical control situations are analysed and potential benefits and costs have been identified. This analysis also represents the basis of the modelling approach used for the development of the objective function. Since the objective function should be able to assess performance based on the available (macroscopic) traffic flow data, the analysis is carried out using only data that is usually available: speed and flow. Even though according to the TLS specifications (BAST, 2002), measures of speed deviation should also be provided by the traffic data collection systems, these are often corrupted and therefore will not be considered here. The control situation  $S(j, t)$  at one VMS  $j$  in time  $t$  could be expressed as a state vector of measured speed ( $V(j, t)$ ), lowest speed on the segment in front ( $V_d(j, t)$ ),

control speed ( $V_c(j,t)$ ), difference between the control and measured speed ( $\Delta V(j,t)$ ) and measured flow ( $Q(j,t)$ ):



$$S(j,t) = f((V, V_c, V_d, \Delta V, Q)_{(j,t)}) \quad \text{Eq. 3-3}$$

The Figure 3-4 illustrates qualitatively typical situations with respect to relation between measured speeds and controls in time. Each typical situation is marked with a different colour. The speed profile at the location is shown by the black ( $V$ ), while downstream speed ( $V_d$ ) is shown by the blue line. The control profile at the location ( $V_c$ ) is shown by the red line.






**Figure 3-4: Typical link control situations**

In total five typical situations have been identified.

1.   $Q \ll$  and  $V, V_d \gg$ : Since traffic demand is low and there is no incident (danger) at the section, there is *no need for a control*. Any control activated during this time will produce only unnecessary *costs* (and degradation of acceptance). It should be emphasised that an accident could happen in this situation as well, but since this is usually due to driver or vehicle error, such accidents cannot be prevented by LCS.
2.   $V \gg V_d$ : Situations in which an *incident*, according to the definition given in chapter 3.3.1, is present at some downstream location (e.g. somewhere between two detector stations) but it is still not observable at the local control station. In this situation the LCS has a potential to *warn* the drivers and thus reduce the surprise. If drivers are warned within a certain period of time before incident propagate upstream to the VMS then the warning potential has been exploited and *benefits* should be calculated. Since this control is lower than driven speeds, it will also produce *costs*, although much lower than induced benefits. This situation lasts until the incident become measurable at the control location, which is usually equal to time needed for congestion to propagate upstream ( $t_p$ ). Controls resulting in a speed restriction only *after* the congestion has already reached the VMS are rather useless and are thus best

counted as an “indifferent” control – it does not bring any additional costs but no great benefits either.

3.   $V_C \sim V$  : There is no great difference between the measured and control speed. If the flow is close to the dense one and speeds are still high, *harmonisation* benefits should be expected. In the case of lower speeds, a slight *warning* benefit should be expected. In both cases costs induced by the control are very low.
4.   $V_C < V$  : The control speed is much lower than the driven speed. Since the control is not followed by the speed drop (in time or space) it will produce high *costs* without any benefit. These situations have a most seriously damaging effect on acceptance.
5.   $V < V_C$  : The control failed to react to changes in traffic flow. Such a control will not cause any benefits. If the difference is too drastic, also with respect to the  $V_d$ , it can have negative impacts on safety, thus producing *costs*. However, this situation must not be directly included in the objective function. It could be seen as “*zero*” or reference point, since any control that performs better under such a situation will also have better performance.

The time needed for congestion to propagate ( $t_p$ ) could be calculated for each individual situation using the LW traffic flow model (LIGHTHILL AND WHITHAM, 1955) and shockwave theory, given that the exact speeds, flows, and distance between locations are known (chapter 2.1.1). However, the exact location of an incident is rarely known. Furthermore, for the reconstruction of the driver experience with local data in unified manner, having an aggregated approximated propagation time would be advantageous. Therefore, we are interested in the average propagation time ( $T_{p_j}$ ) for each control station  $j$ , which can be derived in the following manner. Instead of determining the exact location of the incident, the distance between subsequent VMS panels ( $VMS_j$  and  $VMS_{j+1}$ ) can be taken as a reference distance ( $\Delta X_j$ ). Instead of calculating each shockwave speed separately, empirical evidences about average shockwave propagation speed  $V_s$  (according to empirical data from A8Ost the shockwave speed is in average 4m/s) can be used. Therefore,  $T_{p_j}$  can be estimated as the average time needed for the propagation of the congestion in the upstream direction between two subsequent VMS panels at the distance of  $\Delta X_j$ :

$$T_{p_j} = \Delta X_j / V_s \quad \text{Eq. 3-4}$$

The value  $T_{p_j}$  can be then used to reconstruct the experience of the driver with local data, and to estimate contributions to benefits and costs.

### 3.3.3 Approach for estimation of link control effects

If it is possible to approximate the accident probabilities in each of the above defined case (chapter 3.3.2), together with the influence of the control on the reduction of these probabilities, then as the starting model for describing a typical situation on a motorway with VMS in Germany a cost (or loss) model of the general form can be used:

$$G[\text{€}/\text{km}] = C_T / V + p * C_A \quad \text{Eq. 3-5}$$

$C_T$  - value of time in Euro/h,

$C_A$  - represents the cost of an accident in Euros and per kilometre,

$p$  - is the risk (probability of an accident under different conditions),

$V$  - is the speed in km/h.

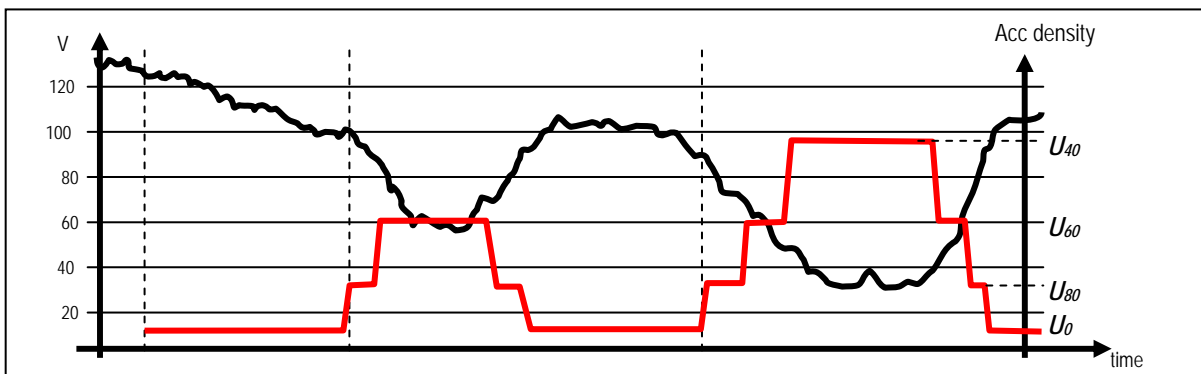
The model (Eq. 3-5) is clearly only an approximation to the true costs either from a public point of view or as perceived by an individual. For example, the costs of fuel and environmental costs are neglected, as are “irrational” motivations resulting from frustration in traffic jams, etc. Nonetheless, it gives a good approximation of the behaviour of a large ensemble of passenger cars on German motorways.

The challenge in applying such a function is to make the best possible estimate using the available data and still limited knowledge about traffic dynamic and corresponding safety risks. In most applications, one has loop detector data and possibly police accident statistics. Hence, the idea is to estimate the contributions to (Eq. 3-5) that would have been obtained by posting a certain control (e.g., 80 km/h) from this limited information and to compare these to a *reference situation* (e.g., posting no speed limit). In other words, the model tries to compare the risk with and without control (warning and/or harmonisation) and to calculate the benefits of reduced risk due to the control effect. For this purpose, first the risks in a no-control situation should be determined and then the contributions of different controls to its reduction should be estimated. For estimating risks, information about accident at the roadway can be used.

The idea behind this is that on each motorway, even under free flow traffic conditions, there is always a small probability (“danger”) that the driver will be involved in accident. Terms used in literature such as accident exposure or rates actually represent a measure of such average “danger”. This average risk is also perceived by the drivers and can be observed in their behaviour. For example, at the free traffic flow not all drivers drive 200 km/h but at an average around 150 km/h. This is due to the fact that drivers do not drive at the maximal possible speed even when they could, but at a speed that represents a balance between their perception of risk and possible speed. Since we are interested in combining safety and efficiency measures, expressing of this risk as *travel time per vehicle and kilometre* would yield an appropriate “monetary unit”. Therefore, this small but always present risk will be referred to as the *average accident time density* and denoted hereinafter as  $U_0$  (see Figure 3-

5). Transforming of the accident rates or exposure into above unit, yields a measure of the average, always present, cost.

The average “danger” grows as traffic conditions worsen. In the context of LCS and according to the incident definition in this work (chapter 3.3.1), these are situations where sharp speed drops are present at some downstream location but cannot be perceived by the driver, which increases the threat of injury (even psychological). The model can estimate this increased exposure by *reconstructing the speed propagation in time*, identifying such situations, and assigning higher accident time densities to them (and thus potential costs). The objective function thus includes a statistical model for the relationship between accidents and sharp speed drops, which constitute an *acute hazard*. The risk corresponding to the acute hazard will be referred to as the *acute accident time density*, and is denoted as  $U_m$  (see Figure 3-5). Since different locations might have different acuteness (for example due some specific geometrical properties) value of  $U_m$  varies along the motorway and is calculated for each control station  $j$  separately. Using the measure of accident (average and acute) densities, the risks present in a situation without control (reference case) can be derived. This risk can be perceived as the existing potential cost, which can be reduced by an appropriate control action.



**Figure 3-5: Average and acute accident densities**

The *effects of each control* are then estimated by calculation of relative “benefits” (i.e., reduction of losses) associated with the reduction in the probability of an accident due to a timely warning or to harmonisation of traffic. However, in order to achieve this it is necessary (for each time period  $t$ ) to reconstruct expected situation on the section in front, using available local traffic data. Only with such a reconstructed situation it is possible to estimate the accident time densities and effects in the above mentioned control situations.

### 3.3.4 Estimation of the accident time densities

The accident time density can be seen as driver “exposure” to a risk. With higher accident time density, the risk and “potential costs” are also higher. Determination of the *average accident time densities* from accident rates was introduced in KATES ET AL., 1999 and will be

applied here. According to this average accident time density and with the help of *empirical model* the acute accident time densities in cases of sharp speed drops are estimated.

In KATES ET AL., 1999, the *average accident time density* is determined by analysing injury and fatality accident rates, accident costs and value of time. Accident rates ( $A_r$ ) are calculated as the number of accidents per 10 million travel kilometres ( $10^7$  veh/km). According to MAYCOCK, 1997 the average cost of injury and fatality accidents ( $A_C$ ) is estimated to be 50,000€, which is considered to be an underestimation rather than an overestimation. The value of time ( $C_T$ ) is estimated using the average hourly neto income that is ~10 €/h. Basically, the fatality accident rate is multiplied by accident costs, divided through kilometres travelled and value of time:

$$U_0[\text{sec/veh/km}] = (((A_r * A_C) / C_T) * 3600) / 10^7 \quad \text{Eq. 3-6}$$

In this way the measure represented in sec/veh/km is derived. The average accident time density  $U_0$ , for the motorway with the injury accident rate of ~2 (per 10 million travel kilometres), is:

$$U_0 = 3.6 \quad [\text{sec/veh/km}] \quad \text{Eq. 3-7}$$

i.e. for each vehicle and kilometre the average (potential) economic accident cost is ~3.6 seconds. The value of  $U_0$  can be separately calculated for each motorway. Since the average accident rates on motorways around Munich is around 2.1 (KATES ET AL., 2002) the above value will be used here.

However, there is an *acute hazard* for vehicles approaching a tailback without warning, which is associated mostly with the steepness of the negative speed gradient and which is *much larger* than  $U_0$ . Since this information is not generally available in detector data, we consider as a surrogate the “severity” of the speed drop, which is recorded in detector data (according to the definition of incidents given in chapter 3.3.1). The severity is then quantified by introducing a sequence of speed levels  $V(m)$  for  $m = 1, 2, 3, \dots$  where we  $V(m)$  define in the declining sequence, as given in the Table 3-1<sup>25</sup>:

**Table 3-1: Speed drop levels and rates**

m	$V(m)$	Example for $f(m)$
1	80 km/h	0.5
2	60 km/h	0.4
3	40 km/h	0.2

<sup>25</sup> m can be considered as continuous variable and  $V(m)$  as monotone declining function of m.

We associate an *acute accident time density*  $U_m$  with speed drops below each such level  $V(m)$ . We introduce the  $f(m)$  measure, which represents the part of all accidents that can be contributed to the speed drops of the level  $V(m)$ , where  $V < V(m)$ . In other words,  $f(m)$  is a part of the all accidents that can be contributed to the speed drops *at least* of the level  $V(m)$ . According to this definition,  $f(m)$  is a monotone decreasing function where  $0 \leq f(m) \leq 1$ . Since not all accidents are due to a speed drop, the maximum of  $f(m)$  is considerably lower than 1 (see Table 3-1). The average daily number of such events is denoted as  $n(m)$ .

The acute danger is present only in situations where downstream speed is much lower than the speed at the location  $j$ . This situation lasts for the period of time  $T_{P_j}$  which is needed for information to propagate in the upstream direction until it reaches the local detector station.  $T_{P_j}$  is estimated as described in chapter 3.3.5 (Eq. 3-4). Determination of the duration time of the acute danger is important in order to distribute acute densities only across this time period. Since in one day, at one VMS ( $VMS_j$ ), there is  $n_j(m)$  speed drops of level  $V(m)$ , the major part of the overall accident danger (called acute danger) will be concentrated in smaller part of the day ( $T_{P_j}/1440$ ). Since only the part  $f_j(m)$  of accidents (not all) could be assigned to such speed drops, acute accident time densities  $U_{j,m}$  at the location  $j$  can be calculated as:

$$U_{j,m} = FA_j(m) * U_0, \text{ with} \quad \text{Eq. 3-8}$$

$$FA_j(m) = (1440 * f_j(m)) / (n_j(m) * T_{P_j}) \quad \text{Eq. 3-9}$$

$n_j(m)$  - number of events of severity level  $m$

$f_j(m)$  - part of all accidents that are caused by speed drops of level  $m$

$FA_j(m)$  - acuteness of speed drop of severity level  $m$

Hence,  $U_{j,m}$  represents the “danger” (cost) that an accident could happen at some speed level  $V(m)$  at the location  $j$ . It is different for different speed drop levels and represents a unique weighting factor for accident severity at specific location.

### 3.3.5 Reconstruction of the traffic situation at the section

In Figure 3-6 a space-time diagram illustrating a hypothetical traffic situation between the upstream  $VMS_j$  and the downstream  $VMS_{j+1}$  control stations is given. Based on the above discussion and incident definition, we are especially interested in the lowest value of speed  $V_{down}(j, t_0)$  that driver, who enters section ( $VMS_j$ ) at time  $t_0$ , will “experience” during travel to  $VMS_{j+1}$ . The speed  $V_{down}(j, t_0)$  cannot be directly observed. Instead it is estimated using shockwave theory, under the assumption that speed drop that is observed at detector  $d_j$  at



time  $t_n$  was generated in some previous time further downstream and has propagated in the upstream direction. The time needed for the incident to propagate upstream to the local control station has been denoted as  $T_{P_j}$  and is estimated by Eq. 3-4. In this case, one could calculate this propagation “backward” in time, and reconstruct position of the congestion area in earlier time.

The incident that has been formed at time  $t_0$  at the downstream location should reach the upstream location at time  $t_n$ , where:

$$t_n = t_0 + T_{P_j} \quad \text{Eq. 3-10}$$

The minimum speed ( $V_{down}(j, t_0)$ ) that the driver would experience during travel along the section can be estimated as:

$$V_{down}(j, t_0) = \min((V(j, t_0), \dots, V(j, t_n))) \quad \text{Eq. 3-11}$$

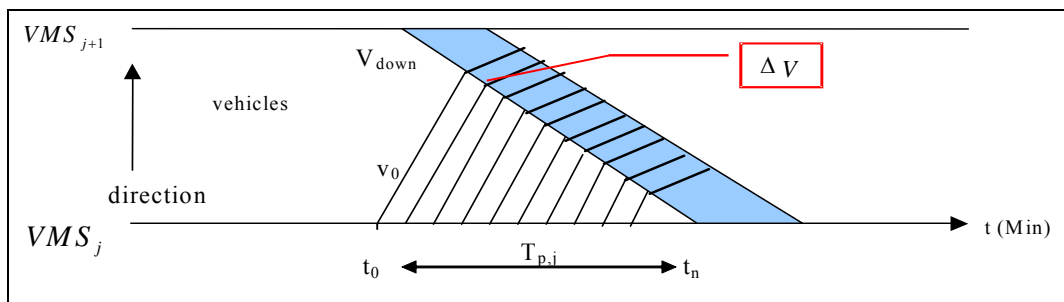


Figure 3-6: Time-space diagram (reconstruction of the speed profile)

### 3.3.6 Modelling of link control effects

For each minute  $t$  in the control period  $T$ , the objective function estimates the contribution of the particular control on the reduction of the above derived accident time densities. Basically, the control action that reduces danger should be “rewarded” with the value that corresponds to the value of this reduced “danger”. Since the goal of link control action is to increase the safety and efficiency within the section, estimation of control contributions is done with respect to the situation at the section and not only the local one. This information is usually not available, and therefore is estimated (reconstructed) from the available local data as given in chapter 3.3.5. Once this situation is reconstructed, the effects of different control actions can be estimated for each minute  $t$ . The total effects of the control ( $G$ ) can be expressed as the sum of benefits ( $B_T$ ) induced by harmonisation ( $B_H$ ) and warning ( $B_W$ ), against the travel time loss caused by the control ( $C_{TT}$ ):

$$G = B_T + C_{TT}, \text{ with} \quad \text{Eq. 3-12}$$

$$B_T = \alpha * B_W + (1 - \alpha) * B_H,$$

where  $\alpha$  represents the importance of the specific control strategy. It can range from 0 to 1, where 0 would mean that only harmonisation is important, while value 1 would mean that

only warning is important. By default  $\alpha$  is set to 0.5 which corresponds to their equal importance. The function can be further extended by assigning additional weights to the benefit and cost components, so to emphasize or ignore particular control effects.

In the case of a warning, the benefit ( $B_w$ ) is attributable to removing the “surprise” that would occur in the case of a sharp speed drop without warning, thus reducing the acute hazard. In the case of harmonisation ( $B_H$ ), we incorporated a model based on the results of KATES ET AL., 1999 and STEINHOFF ET AL., 2001B to estimate a variable accident prevention rate. Note that in contrast to warning drivers upstream of the sharp speed drops, harmonisation strategies are generally not used when there are acute hazards but rather include mild intervention with a mild effect, and thus their benefits are of the order of magnitude of the average accident density. The control negative effects on travel time (caused by lower allowed speeds) ( $C_{TT}$ ) are confronted to these benefits. With this approach, different controls can be compared and the optimal control that achieves the best balance between benefits and costs can be determined.

Both harmonisation and warning effects are derived from accident densities and therefore can be also expressed in *sec/veh/km*. Representing all the effects with a unique measurement unit allows the straightforward derivation of the total control effects at time  $t$ .

### 3.3.6.1 Calculation of time costs

As far as travel time is concerned, for each minute  $t$  we may estimate the “cost” of a speed limit ( $V_C$ ) compared to no speed limit by calculating the *delay* that would result from drivers obeying the control:

$$C_{TT}(t) = TT_{control}(t) - TT_{nocontrol}(t) \quad [\text{sec/veh/km}]$$

$$TT_{control}(t) = \frac{\Delta X}{V_C(t)} \quad \text{Eq. 3-13}$$

$$TT_{nocontrol}(t) = \sum_{i=0}^{T_p} dt_i$$

Although we know that compliance is *not* 100 %, we model the losses to effectiveness resulting from lack of compliance as equal to the time that *would have been* lost if all drivers had complied. This is due to the fact that *acceptance*, which plays an important role in LCS success, strongly depends on the “displayed delay” (even when people do not obey the control) and not on the “actually experienced delay”. This aspect is also important in *avoiding false alarms*.

Calculating of the travel time without control is slightly different from the one with control. This has been done with the goal of capturing possible changes in the traffic flow along the section. To this end, the travel time without control is based on integral calculation (e.g. EVERTS AND ZACKOR, 1976), where the section of length  $L$  is divided into  $I$  number of sub-

segments, where  $I = T_p$ . Next, for each vehicle entering the section at time  $t$  the travel time is calculated as the sum of travel times  $dt$  at each sub-segment  $i$  ( $i = 0..T_p$ ) estimated with speed measured in time  $t - i$ .

The cost will be integrated into the final control performance calculation  $G$ , only if  $TT_{control} > TT_{nocontrol}$ , i.e. when  $C_{TT} > 0$ .

### 3.3.6.2 Warning

The potential of the improvement of the traffic safety using a warning is very great for all drivers that are warned on time – in the period between incident formation at  $VMS_{j+1}(t_0)$  and its propagation to the upstream location  $VMS_j(t_n)$ . Starting with time  $t_n$ , the drivers are already in the incident area and there is no situation they should be warned against.

Identification whether the incident has reached the location is made by comparing the measured speed with predefined value of incident speed  $V_I$  (at the moment  $V_I = 40 \text{ km/h}$ ). If the measured speed is lower than this value, the incident reached the location. Since drivers are already driving slowly, they are no longer in acute danger of being confronted with a sharp speed drop. In this situation control is actually not needed, since the drivers are travelling already slowly, at the speed of the incident wave. Intuitively, it could be estimated that warning benefit in situations where  $V(t) < V_I$  should be equal to zero. However, in this situation, controls that are in better accordance with driver experience would increase user trust and thus acceptance, and should therefore be rewarded with some benefit, though considerably smaller than in acute situations. Therefore, in this situation the average accident time density measure  $U_0$  will be used instead of acute accident density.

In situations where downstream speed  $V_{down}$  drops below a certain acute speed drop level  $V(m)$ , such as 80 km/h or lower for example (chapter 3.3.1), the warning benefits can be calculated as follows. First, the speed drop level  $V(m)$  of  $V_{down}$  is determined. If the measured speed is still higher than  $V_I$  ( $V(t) > V_I$ ), the acute accident time density  $U_m$  is assigned to it. Otherwise, if  $V(t) < V_I$ , the  $U_0$  value will be used.

Since the severity of the situation and warning effects further depend on the difference between the downstream and controlled speed (JOKSCH, 1993), with each warning level (displayed speed) a reduction in the “surprise” is associated. For example, if a driver sees a displayed speed of 80 km/h before entering a speed drop of 60 km/h, then there is only a moderate residual “surprise”. Thus, the surprise ( $sp$ ) can be expressed as a function of the lowest speed on the section ahead ( $V_{down}$ ) and (measured or controlled) speeds at the location ( $V$  and  $V_C$ ). As the contribution to the objective function we then consider a “surprise” which exists without the control  $sp_0$ , and compare it with the surprise by a given control action  $sp_C$ .

The warning benefits for  $Q(t)$  vehicles [veh/h] entering the section at time  $t$  can be calculated as follows:

$$B_w(t) \begin{cases} = \frac{q(t)}{60} * U_m * (sp_0(t) - sp_c(t)), & \text{if } sp_0(t) > 0 \text{ and } sp_c(t) < sp_0(t) \\ = \frac{q(t)}{60} * U_0 & \text{if } V(t) \leq V_{down}(t) \text{ and } V_C \leq V_{down}(t) \\ = 0 & \text{otherwise (if } sp_c(t) > sp_0(t) \text{ or } sp_0(t) > 0) \end{cases} \quad \text{Eq. 3-14}$$

with:

$$sp_0(t) = x^\gamma, \text{ where } x = (V(t) - V_{down}(t)) / V_N \quad \text{Eq. 3-15}$$

$$sp_c(t) = x^\gamma, \text{ where } x = (V_C(t) - V_{down}(t)) / V_N \quad \text{Eq. 3-16}$$

where  $\gamma=4$  according to JOKSCH, 1993 analysis of impact speeds on accident severity (chapter 2.1.2), and  $V_N = 60 \text{ km/h}$

Since the benefit is calculated only in cases where  $sp_0 > 0$  ( $V \geq V_{down}$  and  $V_C \geq V_{down}$ ), too restrictive controls ( $V_C < V_{down}$  and  $V \geq V_{down}$ ) will produce only costs. Besides, contributions to the benefits are also zero when the surprise is not decreased (displayed speed control is higher than the travelled speed), i.e. when  $V_C > V \Rightarrow sp_c > sp_0$ .

### 3.3.6.3 Harmonisation

Under a harmonisation benefit is considered an *accident prevention rate* associated with smoothing traffic flow, if the traffic context is such that the harmonisation is considered to be effective. In practice harmonisation controls result in a lower change of the average speed, but also in a significant reduction of speed deviations. Therefore, the harmonisation safety potential is basically seen in the reduction of the speed deviation (longitudinally). At the moment the implemented version of the evaluation have no access to the speed deviations, and therefore express the benefits of the harmonisation based on the relation between the displayed and driven speeds, as reported in STEINHOFF ET AL., 2001B; KATES ET AL, 1999.

Harmonisation benefits can be expected only at the situation where traffic volume ( $q(t)$ ) exceeds the volume of free flow traffic ( $Q_{min}$ ) but the speeds are still higher than the minimum speed for which the harmonisation benefit should be expected ( $V_{Hmin}$ ). In this situation, if the speeds observed in subsequent time periods do not deviate significantly from the speed limit due to harmonisation, it can be considered that the harmonisation control smoothed the traffic flow and thus reduced deviations along the section. Benefits of traffic safety improvement through harmonisation exist when the following condition is met:

$$\mathbf{IF} \ A_1 \ \mathbf{AND} \ A_2 \ \mathbf{AND} \ (A_3 \ \mathbf{OR} \ A_4) \quad \text{Eq. 3-17}$$

where A-conditions are:

$A_1$ :  $q(t) \geq Q_{\min}$ : Traffic flow  $q(t)$  exceeds minimum flow  $Q_{\min}$  for which harmonisation benefit can be expected (according to KATES ET AL., 1999  $Q_{\min} = 600$  veh/h/lane)

$A_2$ :  $V(t) \geq V_{H \min}$ : The real driven speed  $V(t)$  exceeds the minimum speed  $V_{H \min}$  (according to STEINHOFF ET AL., 2001B  $V_{H \min} = 60$  km/h),

$A_3$ : The difference between the real driven measured speed  $V(t)$  and displayed speed  $V_C(t)$  lies within the permitted area

$A_4$ : The difference between the downstream experienced speed  $V_{down}(t)$  and displayed speed  $V_C(t)$  lies within the permitted area

A wider area for deviation of the driven speeds from the controlled one should be allowed, since we know that harmonisation control leads to the reduction of the speed deviations even though it leads to only slight reduction of driven speeds. This permitted area is defined by the measure of range  $V_{range}$  (20 km/h) and “bias” or “offset”  $V_{bias}$  (+5 km/h). From this, for  $A_3$  and  $A_4$  follows either:

$$\begin{aligned} V_C(t) - V_{bias} - V_{range} < V(t) < V_C(t) + V_{bias} + V_{range} \\ \text{OR} \\ V_C(t) - V_{bias} - V_{range} < V_{down}(t) < V_C(t) + V_{bias} + V_{range} \end{aligned} \quad \text{Eq. 3-18}$$

According to KATES ET AL., 1999 for harmonisation programs without warning effects the accident potential is not in the acute area, but is estimated to be of the order of magnitude of  $U_0$ . Therefore, the safety effect of harmonisation is essentially to reduce the cost density below its “background” value ( $U_0$ ). Thus, if the conditions are fulfilled in one minute (success case), then the simplified model (without data for standard deviation) of harmonisation benefit per vehicle and km can be calculated as:

$$B_H(t) = h * q(t) / 60, \text{ where } h = U_0 \quad \text{Eq. 3-19}$$

As defined earlier, the parameter  $U_0$  is the average accident time density in sec/veh/km. In the successful case where  $q(t) = 1800$  veh/h, with VMS panels at a distance of 1 km and  $U_0 = 3.6$  sec/veh/km, a total benefit of 108 seconds per vehicle will be generated.

### 3.3.7 Calculation of the objective function – summary

Since all control effects are represented as the unique measure [sec/veh/km], the final control performances  $G$  over period  $T$  can be calculated as a sum of the control performances at each

time  $t$ . Hence, the calculation of the objective function is done according to the following steps:

1. For each control station (VMS), calculate the time needed for the incident to propagate from the downstream control station to the VMS (chapter 3.3.2)
2. Calculate the average and acute accident time density for the specific motorway (chapter 3.3.4)
3. Reconstruct the traffic situation at the section for the entire period  $T$  (chapter 3.3.5)
4. For each time  $t$  calculate the following (chapter 3.3.6)
  - a. control costs:  $C_{TT}(t)$
  - b. warning benefits:  $B_w(t)$
  - c. harmonisation benefits:  $B_H(t)$
  - d. end performances at time  $t$ , by taking into account the specific importance of each control effect:  $G(t)$
5. Calculate the overall control performance  $G = \sum_{t=0}^T G(t)$  (chapter 3.3.6)

### 3.3.8 Objective function parameters

In the INCA objective function various parameters have been used for modelling LCS effects. Part of these parameters are derived from historical data (e.g. number of speed drops  $n(m)$  of specific speed level  $V(m)$ ) and should not be changed manually. However, the major part of parameters are configurable and can be updated with new knowledge about system effects or could be used for fine-tuning of the system behaviour and better representation of site-specific needs (for example, to better correspond with the preferences of the authorities).

Configurable parameters could be classified into the following three groups and are summarised, together with their default values, based on the works of KATES ET AL., 1999; STEINHOFF ET AL., 2001B; JOKSCH, 1993; in Table 3-2:

1. General parameters
2. Parameters for modelling of the harmonisation effects
3. Parameters for modelling of the warning effects

**Table 3-2: Performance function parameters**

Variable	Description	Default value
<b>General Parameters</b>		
$U_0$	Average accident time density	3.6 (sec/veh/km)
$\alpha$	Importance of the warning strategy	0.5 (0,1)
$Q_{\min}$	SBA control benefits could be expected only if the measured flow per lane is above this value.	600 (veh/h/lane)
$V_S$	Speed of shockwave propagation in upstream direction	4 (m/s)
$T_{P\_MAX}$	Maximum shockwave propagation time between two detectors	20 (min)
<b>Warning effects</b>		
$\gamma$	Warning benefit shrinking factor	4
$V_N$	Speed used for the normalisation of the warning effects	60 (km/h)
$V_I$	If the measured speed falls below $V_I$ then warning effects will no longer be significant since drivers are already driving very slowly	40 (km/h)
<b>Harmonisation effects</b>		
$V_{range}$	If the difference between measured and controlled speeds is greater then this value no harmonisation benefit can be expected	20 (km/h)
$V_{bias}$	Allowed bias for $V_{range}$	5 (km/h)
$V_{H\ min}$	If the measured speed goes below this value harmonisation benefits should not be expected	60 (km/h)

### 3.4 Data management

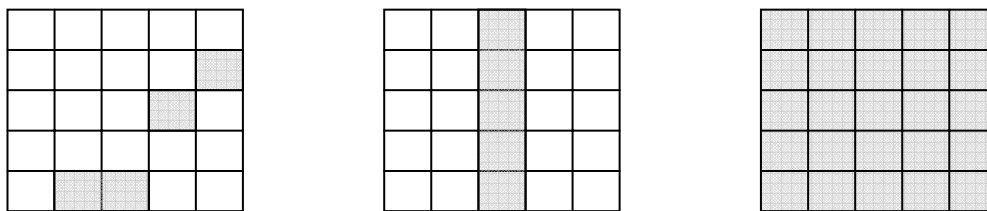
In general, the data management module performs following tasks:

1. Data checking: checking whether data is missing or is implausible (check validity and/or consistency).
2. Data completion: filling possible gaps in the data with reasonable replacements.
3. Data archiving: Archiving data in unique way (referenced to the same “coordinated” system). In INCA data is archived in simple ASCII files. For each detector and control station one file per day is created, where input data and their quality status is saved. Reference tables with the road geometry, location and name of each detector and control station serves INCA to find the appropriate data in files.

#### 3.4.1 Data checking

Usually, there are different causes of the data corruption that could reflect in the similar behaviour (frequency of corruption) in the available data set. Since in this work we are more interested in the frequency of the missing data then in the cause of the corruption, the following classification will be used (VUKANOVIC ET AL., 2004) (Figure 3-7):

1. Random failures are seen as one to several minute long gaps in the input data vector. Gaps in the input data vector could form due to the communication failure or implausible data.
2. Structural failure usually occurs due to physical damage of the detection system or due to maintenance works. Although physical damage can cause a detector not to deliver data for only few minutes (see previous point) in this case we have in mind extreme situations when data is missing for entire day or many days.
3. Systematic failure occurs usually when detection device produce measurement with bias and noise due to, for example, device calibration errors, false counts of vehicles, round off errors, etc.<sup>26</sup>



**Figure 3-7: Three types of input failures: random, structural and systematic (VAN LINT, 2004)**

In practice a mix of input failure types will occur. In general, it is much easier to detect corrupted data in off-line operation, where subsequent data sets are known. In on-line operation however, not only that it is more difficult to detect whether data is corrupted, but the detection (and correction) mechanisms must be computationally efficient to keep system operating in real time.

Random failures as well as structural failures can be easily detected by comparing time stamps. Detection of systematic failures may be very difficult and requires more advance plausibility check techniques that are not the subject of this work. It is generally believed (BISHOP, 1995) that data-driven approaches are not very sensitive to small disturbances (noise) and some bias in its inputs. Luckily, more severe systematic failures usually result in several detectors producing noticeably biased values when compared with neighbouring detectors or simple logic.

### 3.4.2 Plausibility checks

Different algorithms for data plausibility checks have been proposed and tested (BAST, 1999; HAJ-SALEM AND LEBACQUE, 2002; VAN LINT, 2004; etc. They range from simple threshold based, via “balanced” and statistically based to model based. The most common approach is to compare deliver values (occupancy, speed, flow) with pre-defined threshold values. In this

---

<sup>26</sup> Actually, also usually available measure of speed, which is the average mean speed, could be seen as a kind of systematic failure since it has been shown (e.g. VAN LINT, 2004) that it represents the overestimation of the actual driven speeds.



way significant errors can be detected. More advance approaches compare detectors averages on-time to that average across all detectors at the detector station. Even though they can detect substantial failures, they required extensive calibration and have decreased reliability in situations where traffic flow behaviour has changed. Finally, neural networks, genetic algorithms, the Fourier transform and different advanced, in nature prognostic, algorithms have been also proposed and tested. In general it could be said that with complexity their performance rises, but the potential for applying them in on-line operations drops.

In INCA we are interested in detecting random and possibly structural, plausibility failures in on-line operation. The random plausibility failures are usually easy observable, reflected in data that is significantly different from what is to be expected. Most of these failures can be captured by threshold based algorithms and logical tests. Therefore, in INCA several simple threshold and logical tests are performed to check for data plausibility. First the values of all input variables are checked whether they are within their expected boundaries. The lowest boundary for all measures is set to be zero. The highest boundaries are different for different measures. The measured value is considered implausible if it is lower ( $<$ ) than the lowest boundary, or higher than or equal ( $\geq$ ) to the highest boundary:

$$V_{up} = 250[km/h]; \quad Q_{up} = 60[veh/min]; \quad Occ_{up} = 100[\%]$$

If all values satisfy the threshold conditions, the logical tests will be performed. Basically, logical tests analyse the consistency of different measures by comparing them. For example: if the measured flow is equal to zero, the measured speed must be also equal to zero, and vice versa; if the measured flow is greater than zero, occupancy cannot be zero, etc.

For both, missing and implausible data, the status of data is set to a value other than 0 so to indicate to the system that the data are corrupted. Such data will be recognised in the data replacement part and will be appropriately corrected.

### 3.4.3 Data replacement (completion)

Replacement of missing data is, as plausibility tests, a field for itself with many of discussions and proposed solutions. In general following approaches can be identified (VAN LINT, 2004):

1. zero replacement: actually this is a “leave as is” approach, where data is replaced by simply setting input values to zero or some predefined value that would be a sign that the data is corrupted. This approach is sometimes used by data-driven approaches to “learn” the model how to behave when data are of lower quality and as such to ensure the so-called “graceful-degradation”.
2. simple imputation: replace missing values with some ad-hock (statistical) procedure: the last known value, spatial interpolate, exponential smoothing, forecasted value by means of time series, etc. If the model is robust regarding changes in the statistical properties of data (that could be changed by these approaches), simple imputation may produce good results.

3. model based imputation: replace missing data by procedures related to knowledge of the (physical) process generating the data, such as traffic flow models (HAJ-SALEM AND LEBACQUE, 2002), Kalman filter, combination of the previous two (VAN LINT, 2004), cross-correlation algorithm, etc. These are very powerful as they address the spatiotemporal characteristics of the traffic process. On the down side they require a much greater modelling effort - specifying nodes, sections, calibration of parameters of fundamental diagram, etc. and are computationally expensive (filling in missing data on some route, requires the model to run at least one measurement period in real-time).
4. There are also multiple imputation approaches used in the field of neuro-computing, pattern-recognition, etc. These are based on the idea that corrupted data is replaced a number of times, say  $N > 1$  times. In each replacement the missing data is corrected through some simple or model based imputation method. Then, with  $N$  “complete” data sets,  $N$  predictions or inferences can be made, which can be statistically summarised. With such an approach the statistical properties of the source data (inherit to missing data as well) are preserved. However, in the filed of traffic control, model-based imputation approaches are more appropriate.

Generally, as for plausibility tests, more complex algorithms would have better performances but are more difficult to apply in on-line operation (see point 3). We hypothesise that INCA may still yields valid results in operation, given a data-cleaning procedure based on simple imputation. The spatial interpolation would be a good solution when the distance between detectors is not too great. However, since detectors are usually at a distance greater than 1500m, the data is estimated from neighbouring lane detector, as proposed in MARZ. Missing data  $M$  at detector  $d$ , in time interval  $t$ , is estimated from neighbouring detector  $d_n$  using the following equation:

$$M(d, t) = M(d, t-1) * M(d_n, t) / M(d_n, t-1). \quad \text{Eq. 3-20}$$

The estimation of number of heavy vehicles that are estimated from data from upstream lane detector is an exception:

$$Q_{truck}(d, t) = Q_{truck}(d, t-1) * Q_{truck}(d-1, t) / Q_{truck}(d-1, t-1). \quad \text{Eq. 3-21}$$

If the previous measures from lane detector are not available (implausible) the values will be set to be equal to the one from the neighbouring/downstream detector. At other hand, if data from neighbouring detector is not available but the one from local detector is, then the forecasted values from the previous time interval can be used. In that case, corrupted measurements  $M(d, t+1)$  from detector  $d$  at time instant  $t+1$  can be replaced by smoothed (forecasted) value  $M_f(d, t+1)$  from the previous time interval:

$$M(d, t+1) = M_f(d, t+1) = M_s(d, t) + \Delta M(d, t), \text{ where} \quad \text{Eq. 3-22}$$

$$\begin{aligned}\Delta M(d,t) &= \beta * M(d,t) - M(d,t-1) + (1-\beta) * \Delta M(d,t-1) = trend \\ M_s(d,t) &= \alpha * M(d,t) + (1-\alpha) * M(d,t-1) = smoothed value \quad \text{Eq. 3-23} \\ \alpha, \beta &\in (0,1)\end{aligned}$$

The usual value for smoothing factor  $\alpha$  is 0,2 and 0,3, while the usual value for trend factor  $\beta$  is 0,1 and 0,2. These values can be separately configured for each detector data.

### 3.5 Knowledge base

The knowledge base represents, as name suggests, the place where all information that could be used in the process of control decision making is calculated/prepared. The basic idea behind is that existing *detection/prediction/estimation algorithms* offer valuable information that, if properly used, could improve link control performance. In other words, the INCA is developed based on the idea that integration of this information would make it possible to create a synergistic process in which the consolidation of individual information creates a combined resource with a productive value greater than the sum of its parts. However, since we cannot create productive value from bad or no information, INCA final performances still depends on this information.

In many, but not all data-driven applications it is recommended to transform input data into the *normalised/scaled* values. Normalising input variables tends to improve the training process (faster convergence and reducing chances of getting stuck at a local minimum) by improving the numerical condition of the optimisation problem and ensuring that various default values involved in initialisation and termination are appropriate. They could lead to prevention of saturation and positively influence weights initialisation. Normalisation is also important in the light of methods such as Ridge regression, which is used in the INCA optimisation block. Therefore several methods for the data normalisation are provided in INCA, which can be applied if necessary. These are the simple min-max normalisation and more advance, statistically based zero-mean normalisation. Zero-mean normalisation utilizes the statistical measure of central tendency and variance to help remove outliers, and spread out the data distribution. The mean and standard deviation of the initial set of data values are required for this.

Another function of the knowledge base is to determine the actual *traffic state (context)*. Determination of the traffic state is important for several reasons, some of which have already been mentioned in previous chapters:

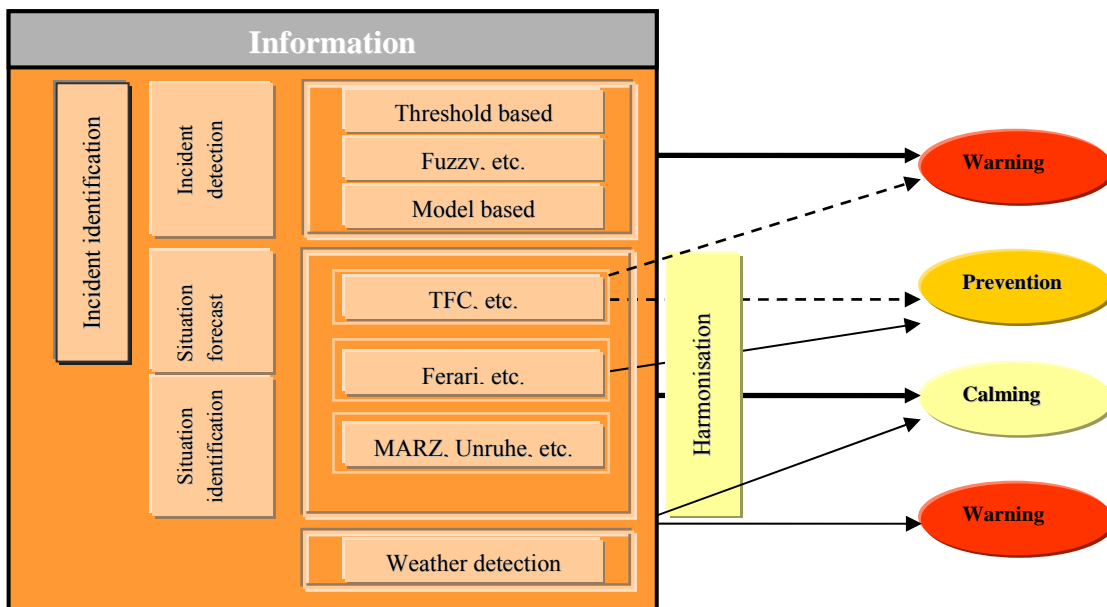
- Not all control strategies make sense under different traffic conditions. For example, there is no need for the harmonisation strategy in the situations when traffic is already congested.
- Algorithms performances might vary under different traffic conditions (Appendix A.1). This is not surprising as most algorithms are basically designed to detect a specific type of incident - one that happens under low traffic flow, or one that arises in the stop and go

traffic, etc. This variability is different for each algorithm as it depends on their specific design. Based on these studies it is to expect that algorithms could be much better utilised if these considerations are included into the decision-making process.

- Theoretically the parameters of some algorithms could also be adjusted according to changes of traffic conditions.

### 3.5.1 Utilised information

The information used in INCA comes from various algorithms and indicators. When we speak about algorithms we have in mind the formulas and equations combined with certain rules, which produce a statement (discrete interpretable value) as the output. Usually the statement is “there is” or “there is no” incident, or the probability of incident is 20%. On other hand indicators are various (less or more meaningful) continuous values usually, but not necessarily, produced using simple formulas. Hence, this can be raw data as well as smoothed values, prognostic values, or their combination. Indicators cannot be deterministically interpreted as algorithms can (they do not deliver categorical output), but they still represent a source that could provide the control module with valuable information. Generally, it is assumed that continuous values could bring more benefit to INCA as it is not the same if some indicator is a little or much over the threshold value. Ultimately, the control module should be able to distinguish between more valuable and less valuable information.



**Figure 3-8: Algorithm classification and strategies that could profit from their outputs**

In the context of LCS, bearing in mind its goals and strategies, algorithms could be classified as follows (Figure 3-8):

- **Detection:** detection of already present incidents – they determine the need for warning strategies

- Estimation: estimation of instability of traffic situation – they generally determine the need for harmonisation strategies
- Prognostic: forecast of possible disturbances – they could be used for warning as well as for harmonisation

**Incident detection algorithms**

In INCA, MARZ algorithms (BAST, 1999) such as Stau1, Stau2, Stau3 and Stau4 are implemented. Initially, California8 was also implemented but as it performed poorly in preliminary tests and it is excluded from further description. It is believed that, due to limited literature, its implementation or/and calibration of its threshold values was not carried out appropriately. Here a short overview of the implemented algorithms will be given. For more details on MARZ algorithms please refer to BAST, 1999. A detailed general overview of incident detection algorithms is given in Appendix A.1.

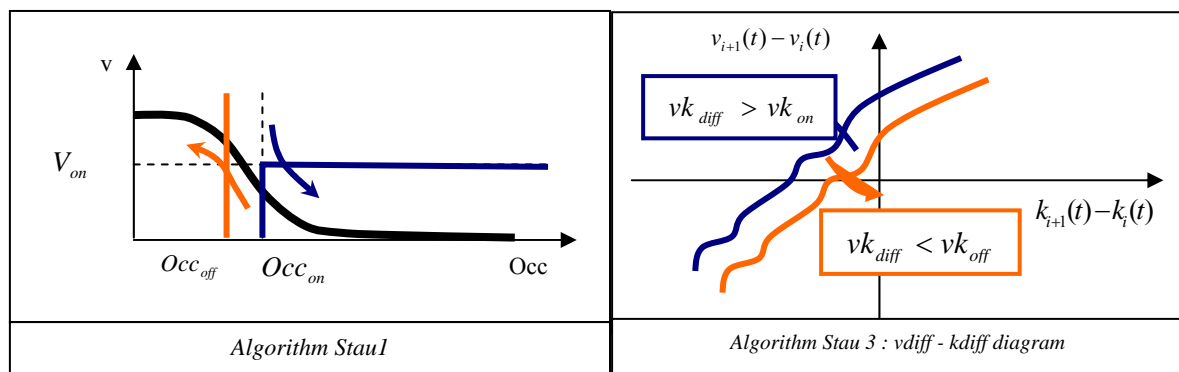
Stau1 and Stau2 are threshold-based incident detection algorithms (see Appendix A.1). *Stau1* is based on comparison of occupancy of each lane ( $Occ(d)$ ) and forecasted mean speed value at detector station  $i$  ( $V_f(i)$ ). The logic for activation and deactivation has the following form:

**Activation:** IF at one lane ( $(Occ(d) > Occ_{on})$  AND  $(V_f(i) > V_{on})$ ) THEN (Stau1) ↗

**Deactivation:** IF at all lanes ( $Occ(d) < Occ_{off}$ ) THEN (Stau1) ↘ Eq. 3-24

The algorithm logic is based on the relations in the v-b fundamental diagram, and is shown in Figure 3-9, left. Basically the problem is that it depends strongly on the data quality (it is enough that data are not plausible at one detector station), it reports congestion when it is already present, it can last too long because it waits for occupancy to drop, etc.

The *Stau2* algorithm compares the forecasted mean speed values and traffic flow at detector station. Basically, only speed plays a role, and the flow value is there to ensure that the algorithm is not activated by night, in the case when only one slow vehicle has passed. Due to its focus on measured speed at the local detector it can detect an incident only when it is already observed in the data and therefore has a lower warning benefit.



**Figure 3-9: Illustration of Stau1 and Stau3 algorithm logic**

Basically, *Stau3* and *Stau4* have the same logic behind with the difference that *Stau3* uses speed and density relations ( $vk_{diff}$  indicator), while *Stau4* uses speed and occupancy relation ( $vb_{diff}$  indicator) of two subsequent detector stations and compares them with predefined threshold boundaries. The  $vk_{diff}$  and  $vb_{diff}$  indicators are based on the shockwave theory, where after the incident two traffic flow regimes are formed: denser and slower wave upstream ( $i$ ) and lower density wave downstream ( $i+1$ ) (BUSCH, 1994B). They are calculated by analysing the forecasted speed ( $V_f$ ), forecasted density ( $k_f$ ) and occupancy values ( $Occ$ ), where the forecasted values are calculated as described in chapter 3.4.3:

$$vk_{diff} = \left( \sqrt{\left( \frac{V_{free}(i) - V_f(i)}{V_{free}(i)} \right)^2 + \left( \frac{k_f(i)}{k_{max}(i)} \right)^2} \right) - \left( \sqrt{\left( \frac{V_{free}(i+1) - V_f(i+1)}{V_{free}(i+1)} \right)^2 + \left( \frac{k_f(i+1)}{k_{max}(i+1)} \right)^2} \right)$$

Eq. 3-25

$v_{free}$  is the free flow speed,  $k_{max}$  is the critical density.

Due to the fact that they consider v-k or v-occupancy relations, *Stau3* and *Stau4* are considered to belong to the model-based group of algorithms. The *Stau3* algorithm (see Figure 3-9, right) will be activated if at detector station  $i$  the corresponding calculated  $vk_{diff}(i)$  value and measured flow  $Q(i)$  reach their given threshold values. The algorithm will be deactivated if one of these values drops below the threshold value:

**Activation:** IF ( $vk_{diff}(i) > vk_{on}(i)$ ) AND ( $Q(i) > Q_{on}(i)$ ) THEN (*Stau3*) ↗

**Deactivation:** IF ( $vk_{diff}(i) < vk_{off}(i)$ ) OR ( $Q(i) < Q_{off}(i)$ ) THEN (*Stau3*) ↘ Eq. 3-26

Although the *Stau3* and *Stau4* algorithms are based on more knowledge about traffic flow characteristics, in practice they induce too many false alarms. It is assumed that unsynchronised data (timely) from different detectors, as well as the choice of threshold values, are some of the causes for such performances.

### Estimation algorithms

In this work, estimation algorithms identify the situations that could be potentially unsafe (e.g. great speed deviations) or could lead to traffic flow breakdown (e.g. high densities). Thus, estimation algorithms could provide valuable information for the harmonisation strategy and even might have predictive capability beneficial for the warning strategy. The following estimation algorithms, taken from MARZ, have been implemented in INCA: *Unhrue*, *TruckPercentage*, and *Belastung*.

The *Unhrue* is based on the belief that higher speed deviation could lead to instabilities and thus should influence harmonisation decision. It compares the speed deviation in the left lane

$l$  ( $s_v(l)$ ) with the threshold values. If speed deviation is higher than preferred and the flow in the left lane ( $q(l)$ ) and forecasted flow at detector station ( $q_f(i)$ ) are also higher, the algorithm will be activated. It will be deactivated when all these values drop below their threshold borders again:

**Activation:**  $IF (s_v(l) > s_{v_{on}}) AND (q(l) > Q_{on}) AND (q_f(i) > Q_{f_{on}}) THEN (Unhrue) \nearrow$

**Deactivation:**  $IF (s_v(l) < s_{v_{off}}) AND (q(l) < Q_{off}) AND (q_f(i) < Q_{f_{on}}) THEN (Unhrue) \searrow$

Eq. 3-27

Where usually:  $Q_{on} = 20$  veh/min,  $Q_{off} = 15$  veh/min,  $s_{v_{on}} = 20$  km/h,  $s_{v_{off}} = 15$  km/h,  $Q_{f_{on}} = 3000$  veh/h and  $Q_{f_{off}} = 2300$  veh/h for a 3-lane motorway.

The *TruckPercentage* algorithm is based on the idea that higher percentage of trucks in the traffic flow could create disturbances or reduce capacity in traffic flow due to the expected higher number of lane changing manoeuvres. It is activated if the percentage of truck traffic together with forecasted weighted traffic volume  $Q_b$  (usually one truck counts as 2 passenger cars) reaches predefined threshold values (for truck percentage this is usually 20%, for flow this is usually 4000km/h). The algorithm is currently used to activate only the truck overtaking prohibition sign but not to set speed limits.

Unlike in previous algorithms, the *Belastung* algorithm does not detect potential disturbance factors. Instead it identifies dense flow situations where it is believed that harmonisation strategies could smooth traffic flow and thus increase throughput. It relies on the q-v fundamental diagram and compares the values of the forecasted weighted traffic volume ( $Q_b$ ), mean speed of passenger cars and density, with threshold values. In other words it maps these values on the fundamental diagram and determines the level of criticality. Although it is generally based on the comparison of the measured and threshold values, since it relies on the fundamental diagram it could be also seen as model-based algorithm. It delivers four levels of alarms where each should be treated with a specific speed limit (from 120km/h to 60km/h in 20km/h steps). Since it does not detect disturbances or disturbance factors, but is based on the fundamental diagram, it is considered the most straightforward algorithm for the harmonisation decision.

### Additional indicators

Generally, in addition to the above described algorithms, the control model could combine any other available information such as raw data or smoothed (predicted) values or indicators. Since in this work we focus on the combination of information and not raw data (though this could be also possibility) we derived several additional indicators that could be used by the control module.

$$X(1) = \begin{cases} = V_f(i) / V_{on}(i) & \text{if } |V_{car_f}(i) - V_{truck_f}(i)| \leq V_{diff} \\ = 0 & \text{otherwise} \end{cases},$$

where  $V_{on}(i), V_{diff}$  are threshold values from Stau2 algorithm

$$X(2) = Q(i) / Q_{on}(i), \text{ where } Q_{on}(i) \text{ is threshold value from Stau3 algorithm (Eq. 3-26)}$$

$$X(3) = Q_b(i) / Q_{bon}(i),$$

$$X(4) = \frac{k(i)}{k_{on}(i)} * \frac{V_{car_f}(i)}{V_{fon}(i)},$$

where  $Q_{bon}(i), v_{fon}(i)$  and  $k_{on}(i)$  are threshold values from Belastung algorithm for control 80 km/h

$$X(5) = V_{norm} / V(i), \text{ where } V_{norm} = 80 \text{ km/h}$$

Eq. 3-28

The  $vk_{diff}$  (X(6)) and  $vb_{diff}$  (X(7)) values used for the calculation of Stau3 and Stau4 are considered here as available indicators as well.

With the above indicators it will be possible to:

- derive probably more valuable (continues) information then the (discrete) information from the above described algorithms
- test INCA for its generality, as well as flexibility and extendibility.

### 3.5.2 Determining traffic state

Under determination of the traffic state or context we consider the classification of the prevailing traffic conditions into one of the predefined groups. The definition of these groups (states) can be done in various, more or less detailed, ways. Traffic state can be classified by the considering traffic measures or/and weather conditions, special events, work zones, etc. Furthermore, traffic measurements from each lane separately or aggregated across lanes could be considered. The traffic state can be determined for a location from where data is collected or for a given motorway segment (link), etc. Clearly, determining the traffic context is not the trivial task and approaches with different level of details and classes can be found in literature (KIM, 2002; STEINHOFF, 2002; BAST, 1999; FGSV, 2001, etc). Generally, the chosen approach and its level of detail usually depend on the specific purpose of the classification.

Since the special focus of this work is determining the speed control based on traffic flow characteristics (not weather-related) a classification based on the aggregated traffic flow measurements is needed. This can be achieved by partitioning the fundamental diagram into several classes. The number of classes should correspond with our prior knowledge about specific purposes and characteristics of different control strategies. Additionally, classification should, precisely enough, correspond to our knowledge about different algorithm reliability



under varying traffic conditions. Although not the initial motivation, the utilisation of traffic state identification actually leads to the partitioning of the model's input space, offering the possibility for reduction of the control problem complexity and allowing for simpler models to be used for each space (chapter 3.1.2).

Jamed As the basis for decision upon a number of traffic state classes and their boundaries, the results in SIEBER, 2003 and VUKANOVIC ET AL., 2004 are used. In SIEBER, 2003, VUKANOVIC ET AL., 2004, two months of data have been used for the analysis of the performances of several incident detection algorithms, based on the classification of traffic states by MARZ (BAST, 1999) and KIM, 2002 (Figure 3-10).

MARZ (Figure 3-10, right) classifies the traffic into four states: free, dense, congested and jammed states depending on the forecasted (smoothed) values of speed and density. KIM, 2002 used the traffic state estimation (based on speed, flow and density values) as an input in his traffic flow model (KIM, 2002; KIM AND KELLER, 2001). In work by KIM, 2002 traffic has been classified into 5 states, with the help of fuzzy rules: free, impeded, synchronised, congested and jammed traffic. By rough representation of the fuzzy borders as crisp values, KIM, 2002 classification could be presented as given in the Figure 3-10, left. The algorithm alarms and speed drops are classified based on MARZ (BAST, 1999) and KIM, 2002 classifications. For each of the classes performances of MARZ and California8 algorithms are calculated using DR-FAR from HOOPS ET AL., 2000 (Appendix A.2).

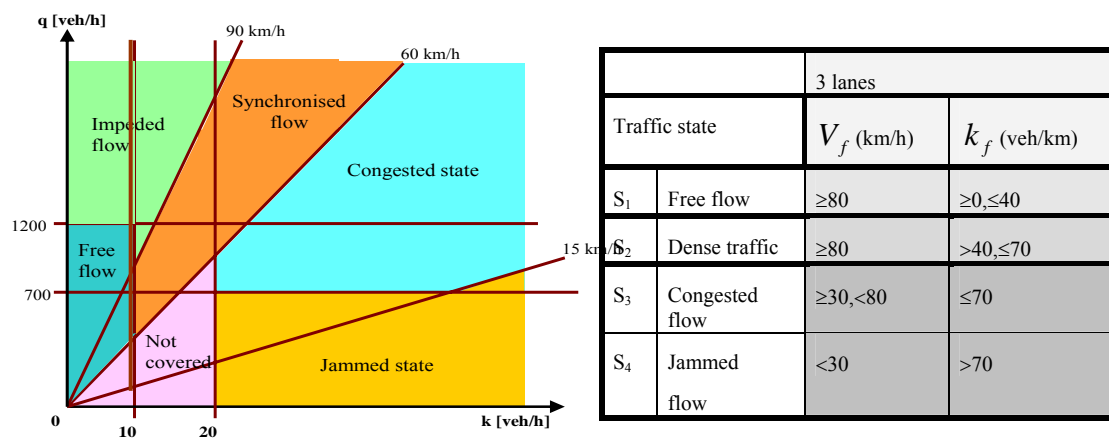


Figure 3-10: Classification of traffic states by KIM, 2002 (left) and BAST, 1999 (right)

In SIEBER, 2003 it has been shown that algorithms perform well under already congested conditions, in the free state are producing only false alarms, and have the most unstable (as expected) performances in the synchronised traffic state. It could be intuitively concluded that if, for example, algorithm Stau1 or Stau3 made the alarm under free traffic condition this information should not be considered at all. According to the results, in the light of link control problem, classification into four or five traffic states produces boundaries that are redundant (e.g. 4<sup>th</sup> and 5<sup>th</sup> state by KIM, 2002). Based on the MARZ and KIM, 2002 classifications, above presented results from SIEBER, 2003 and VUKANOVIC ET AL., 2004, and

consideration of link control strategies, *in INCA classification into three states has been utilised:*

- Free flow: low dense traffic characterised with low traffic flow and high speeds.
- Synchronised: state with higher density but still high speeds
- Congested: state with highly dense traffic and low speeds

The default values for classification of the fundamental diagram are provided in Table 3-3. They are characterised by a quite wide synchronised area. This is due to the aim to make these thresholds as universal as possible. The given values are representative for one typical 3-lane motorway section, without on- and off-ramps and special geometric characteristics. Since the borders between states are quite distinct it is considered that such classification could be used for each location without inducing too much instability. However, since the classification is based on the fundamental diagram, separate determinations of the thresholds for each fundamental diagram are preferable.

**Table 3-3: Borders of traffic states employed in INCA**

Traffic state		V (km/h)	k (veh/km)
S <sub>1</sub>	Free flow	≥100	≥0, ≤18
S <sub>2</sub>	Synchronised flow	≥60	>18, ≤70
S <sub>3</sub>	Congested flow	<60	-

In INCA control decision is made with respect to these traffic states, but it could be also done without it. In addition to the above classification, classifications by MARZ (BAST, 1999) and KIM, 2002 are provided in INCA as well. Which classification, if any, should be used can be configured in the control module.

### 3.6 Control model

The control model has the highly complex task of integrating information, in transparent manner, from the knowledge base in order to produce the control decisions for each VMS along the motorway. Since the objective function models effects of speed controls, only systematic optimisation and monitoring of speed controls is possible. ***Therefore, the focus of the INCA control module is on the development of the model for variable speed limiting under consideration of the German case.*** In Germany speed limits from 120 km/h to 60 km/h, in steps of 20 km/h<sup>27</sup>, can be displayed. The speed control of 40km/h does not exist as such but instead, usually sign “congestion” is shown. Since it has been shown that the sign “congestion” has the most alerting effect on drivers (STEINHOFF ET AL., 2001A), it will be considered here as being a speed limit of 40km/h. Although other controls are not the part of

<sup>27</sup> Sometimes also control of 130km/h can be displayed. However, since control of 130km/h is usually not used in existing applications it will be not considered here.

the optimisation process and monitoring, since they are necessary for system operation in practice, they are considered as an external “adjustment” component and as such included in the model (chapter 3.6.2).

Since one link control system consists of a sequence of VMS the link control model must take into the consideration their interactions. Furthermore, control at one VMS should control and inform the driver with respect to the situation at the segment in front (!), rather than to the situation at location (local detector)<sup>28</sup>. For example, although a speed limit of 100 km/h could be regarded as an adequate local strategy if there is no incident within a location, the presence of a severe downstream incident could imply that from a *global* point of view it is better to post a speed limit of 80 km/h, thus smoothing speed gradients. Hence, controls cannot be observed as independent actions, but should be considered in a broader (coordinated) context. Additionally, transitions of subsequent speed controls must be smooth in:

- time (the sequence of the controls at one VMS), and
- space (controls at the same time  $t$ , but subsequent VMS)

Interactions in time arise due to the changes of the control in time at one VMS panel. It is basically related to the problem of how often and smoothly the control at one VMS is switched on and off. If controls are not *smooth in time* they could lead to dangerous situations by creating subsequent traffic waves  $u(t)$  and  $u(t+1)$ , which are characterised by different speeds. In such way, not only unnecessary heterogenisation of the traffic flow would be produced, but it might even lead to creation of their “collisions”, which would not be otherwise experienced. This phenomenon is known in literature also as “hysteresis”. The common practice is to regulate *termination of the control* by introducing additional “switch-off” parameters. However such an approach is often prone to failures either due to too restrictive parameters or due to corrupted data. Therefore, in INCA *a novel approach for the control termination has been introduced: each control lasts for, at least, some predefined fixed time period*. If during this period no other activations have been observed, it is smoothly switched off in steps of 20 km/h. As such, this information is automatically part of the optimisation process and there is no need for additional parameters.

Interactions in space arise from different controls at different VMS panels at the same (similar) time. If the controls are not smooth in space (from higher to lower speeds) they might also lead to the undesirable “collision” effect: formation of the upstream wave  $u(j)$  with higher speeds that could meet downstream low speed wave  $u(j+1)$ . With smooth speed controls more sophisticated potentials of LCS (such as faster discharge of congestion area)

---

<sup>28</sup> Although this is intuitively clear, as it has been shown in chapter 2.4, most of existing approaches are determining their control action based on the local situation (local detector) only and sometimes are using control propagation rule, thus not using all prevention potentials of LCS.

could be also achieved. Therefore, in Germany it is often required that the driver does not encounter a decrease in the speed limit larger than a predefined amount, which is 20 km/h.

Handling of VMS interactions could be done by applying either the global or the hierarchical control approach. Global approaches control all control stations using one control model. This implies that the model should be derived from all the data and for all control stations, which results in a highly complex data-driven model, with many parameters. In addition to the high burden on optimisation process (with limited available training data), the high complexity of the model would make it most probably not tractable or interpretable<sup>29</sup>. Therefore INCA uses the *hierarchical control approach*, where one “intelligent” controller is assigned to each control unit, and local controls are further “coordinated” in the global control layer. It should be noted that in transportation literature applications of global data driven approaches have been successfully used in areas where tractability does not have a high priority (e.g. state space neural networks for the travel time prediction (VAN LINT, 2004)).

### 3.6.1 Hierarchical control approach

The hierarchical control approach used in INCA has the following two layers:

- ***a local intelligent “segment-based” layer:*** In the local layer, one “intelligent” local controller is assigned to each control station. However, the traffic state utilised for “local” decisions is constructed using neighbouring detectors and is thus segment-based. Furthermore, the local decision is also optimised using the objective function which integrates the knowledge about information propagation in the upstream direction.
- ***a supervisor or “global” layer:*** local decisions are forwarded to the global layer which adjusts them if necessary.

The local “segment-based” module, also called the “intelligent” component, is the core of the INCA link control model, where different information is combined in order to produce the best speed control decision for a single VMS. The architecture of the local “segment-based” module is the same for each VMS. Control is performed in a two-stage process. Traffic state estimation and information derivation is performed in a knowledge-base block. Information is derived from reports and algorithms, and the appropriate indicators are computed and recorded. This information is forwarded to the “intelligent” component. The model combines all this available information, taking into account the present traffic state and corresponding control parameters, and derives the local control decision. It is believed that, by including all relevant the information from section in front, each control station will be able to make the best control decision from a global point of view as well. The control will be “smoothly”

---

<sup>29</sup> Assuming that there are 20 control stations and that we are using (only) 5 variables for each of them, it would imply that the model would consist of (at least) 100 parameters.

switched off, in steps of 20 km/h unless re-activated by a new control decision. In this way, the switch-off parameters can be excluded (first stage).

These local control decisions are subject to modification if required at the global control level (second stage). The global control level has several responsibilities:

- to monitor and integrate operator control (e.g., due to accidents, road works, etc.) and further high priority or external information;
- to “smooth” local decisions taking into account their global influence.

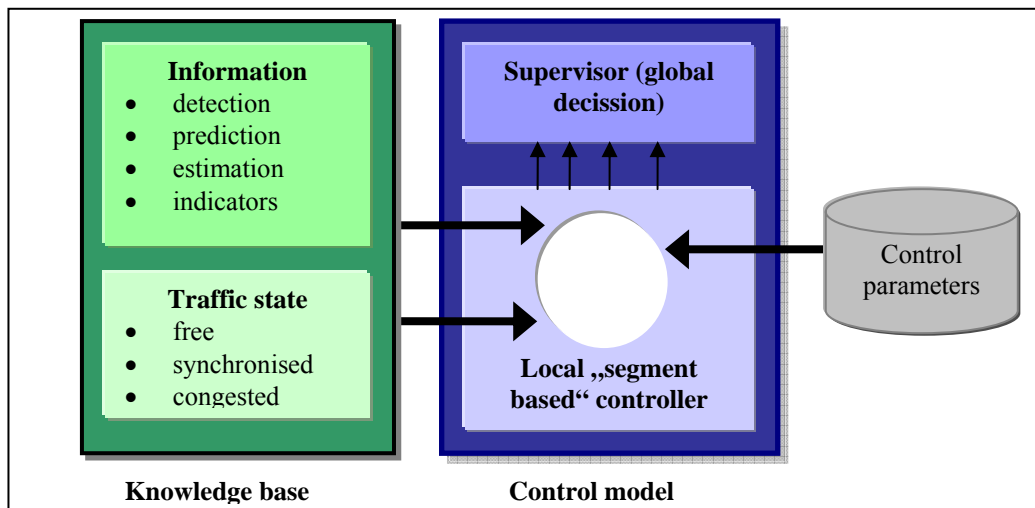


Figure 3-11: Architecture of the control model

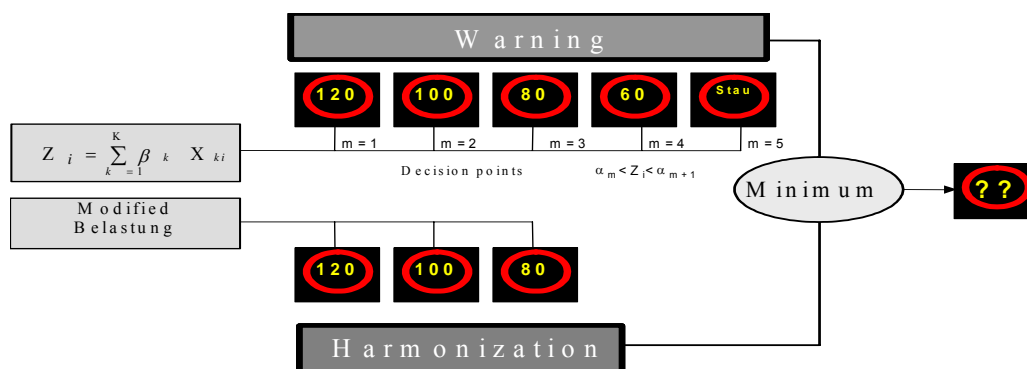
### 3.6.2 Local “segment-based” control

The local “segment-based” control is the core of the control decision. It is assumed that, if the “segment-based” model properly captures the traffic flow dynamics and if the information about the situation at the segment is good enough, such module would also make the best decision from the global point of view. The major LCS strategies considered here are (chapter 2.2.4):

- harmonisation (flow smoothing), and
- warning (incident and congestion detection).

The motivations of these two strategies are different and algorithms for their recognition are based on quite distinct models. Therefore, *the harmonisation and warning components are separated into two independent logical blocks, both contributing to the final local control.* The warning component can activate any available speed controls (from 120 km/h down to the “congestion” sign). Since in HOOPS ET AL., 2000 and STEINHOFF ET AL., 2001A it has been shown that controls lower than 80km/h have little to do with harmonisation, the harmonisation component in INCA can activate controls of 120 km/h, 100 km/h and 80 km/h (Figure 3-12). An additional advantage of this decomposition is the ability to reduce the burden upon optimisation process by separately optimising each component (see chapter 3.7).

The data-driven models which could be used for the information combination in the context of link control problem have been introduced in chapter 3.1.3. These models are characterised by different levels of complexity and interpretability. However, the choice of the model here should not be seen as critical. Firstly, the control problem is divided into fine logical blocks (from hierarchical approach to distinguishing between different states and strategies) in which it is expected that simpler model will perform equally well. Secondly, the complete framework is supported by the complex optimisation process and advanced objective function which should lead to an objectively valid model. Finally, since for optimisation process usually only limited data is available and interpretability is seen as a high priority goal, the approach with relatively lower complexity (parsimony principal) has the best chance to demonstrate all the benefits of an intelligent approach.



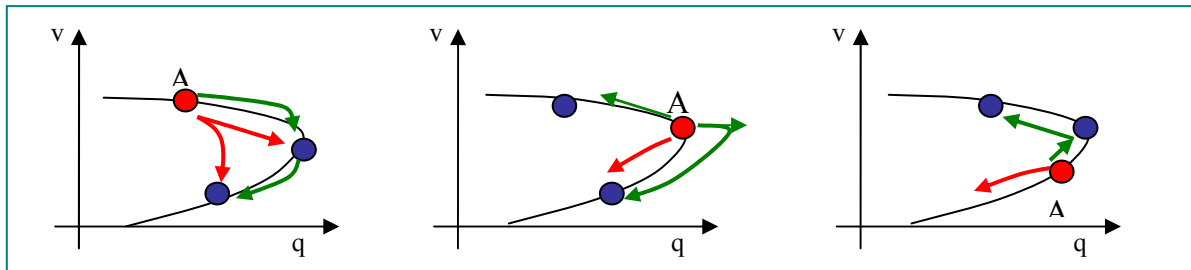
**Figure 3-12: Harmonisation and warning controls and their integration**

Therefore, the (logit-based) regression model is used for combining the available information and determining the warning control decision (Figure 3-12). Basically, the harmonisation decision could be made by integrating information in the same manner as in the warning block. However, from available information only Belastung algorithm (MARZ) is considered to be a real harmonisation algorithm based on the traffic flow model (flow-density relationship). Other algorithms are more suitable for warning decision and are included in the same. Another motivation of choosing only one algorithm for the harmonisation decision was to test and illustrate INCA capability to optimise the parameters of an individual algorithm and not only of the integrated control model.

### 3.6.2.1 Usage of traffic state identification

In INCA three traffic states, which have different traffic flow dynamics and require different control strategies, have been identified: free, synchronised and congested traffic state. According to the knowledge about traffic flow dynamics several assumptions have been made and are illustrated in Figure 3-13. In Figure 3-13, examples of different traffic states and the possible role of the link control are illustrated. The red circle illustrates traffic state A currently observed at the control station. The blue circles illustrate traffic states in which traffic could move. Each figure shows transitions that should be avoided (red arrows) and the

one that should be persuaded (green arrows) with the link control. The summary of the needed control strategies under different traffic states is given in Table 3-4.



**Figure 3-13: Traffic flow states and their transitions. Red arrows show transitions and “collisions” that should be avoided, green arrows show transitions that should be forced with link control. Left: transition from local free flow; middle: transitions from synchronised flow; right: transitions from congested flow.**

**Table 3-4: Needs for control actions under consideration of traffic states**

Traffic state		Harmonisation	Warning
S <sub>1</sub>	Free	X	X
S <sub>2</sub>	Synchronised	X	X
S <sub>3</sub>	Congested		X

Since harmonisation and warning strategies are modelled in separate logical blocks, the realisation of the rules given in Table 3-4 is fairly easy. If the control decision is made under congested state this will activate only one (warning) component, which will then alone contribute to the final local control decision. However, in synchronised and sometimes free state, where both strategies might be needed, the *more restrictive control* (of the two) is chosen as the final local control decision. In INCA the user has the possibility to decide whether the warning control model will be assigned to each state, or control will be based only on one control model.

### 3.6.2.2 Model for warning

In INCA the warning decision is made by a combination of information produced at the local and neighbouring detectors. The warning decision is made with the help of the *multivariate ordered regression model* linked with the logistic probability model<sup>30</sup>. It includes a variable weighting scheme for multiple indicators and algorithms as well as a probability model for detection and false-alarm rates. Since we are interested in the costs and benefits of the control and not the accuracy of the probability estimation, the decision is made based on the variable weighting scheme for multiply indicators, which actually represents the multivariate regression model. Since the output from the warning module should be the control measure (discrete) the outputs from multivariate regression model are compared with threshold values thus resulting in the multivariate ordered regression model. With the introduction of the random (error) term and logit link function (instead of the identity function in the regression

<sup>30</sup> This regression model can also be viewed as the simple MLP neural network (BISHOP, 1995)

model) an ordered logit model can be derived (chapter 3.1.3). As such it is possible to calculate detection or false alarm probabilities, which however does not influence the control decision. For each traffic state one such warning model is assigned.

First, from the various  $K$  information from the knowledge base (e.g.  $Stau1$ ,  $vk_{diff}$ , etc.) the vector  $X$  is formed. Under assumption that information has some specific importance (under a specific traffic situation), for each minute  $t$  continuous “fused” time depended variable  $Z_t$  can be computed as sum of algorithm outputs multiplied by parameters  $\beta_k$ :

$$Z_t = \sum_{k=1}^K \beta_k X_{kt} \quad \text{Eq. 3-29}$$

Where, the  $K$  parameters  $\beta_k$  can be interpreted as the “weight” of the  $k$ -th algorithm,  $k = 1, \dots, K$ . The large value of  $\beta_k$  means that information  $X_k$  has great influence on the control decision.

In the following,  $Z$  will be referred to as the *warning indicator*.  $Z_t$  represents the basic value for the control decision and plays a central role in optimisation process. From warning indicator  $Z_t$ , with the help of so-called ordered logit model, the “latent” variable  $D_t$  can be estimated. This is achieved by introducing random term  $\varepsilon_t$ , which describes estimation error. The resulting model for  $D_t$  has the form:

$$D_t = Z_t + \varepsilon_t \quad , \quad \text{Eq. 3-30}$$

where,  $\varepsilon_t$  is assumed to follow logistic distribution:

$$F(\varepsilon) = \frac{1}{1 + \exp(-\varepsilon / \sigma)} \quad \text{Eq. 3-31}$$

Distribution  $F(\varepsilon_t)$  has a mean value 0 and variance  $\pi^2 \sigma^2 / 6$

**Table 3-5: Warning model decision points**

m	Condition	Control decision
0	$Z < \alpha_1$	no control
1	$\alpha_1 < Z < \alpha_2$	120 km/h
2	$\alpha_2 < Z < \alpha_3$	100 km/h
3	$\alpha_3 < Z < \alpha_4$	80 km/h
4	$\alpha_4 < Z < \alpha_5$	60 km/h
5	$\alpha_5 < Z < \alpha_6$	sign „congestion“

In order to determine the control logic, the continuous warning indicator  $Z_t$  has to be linked to “manifest” variable (warning control decision)  $V_w$ . Therefore a sequence of the decision points is introduced in the model. Decision points are defined as threshold values  $\alpha_m$ , where



$m=1,2,\dots,M+1$  and  $\alpha_m < \alpha_{m+1}$ . The local control decision  $V_w(m)$  is then made by comparing the incident indicator with decision points  $\alpha$  :

$$V_w = m, \leftrightarrow \alpha_m \leq Z_t < \alpha_{m+1} \quad (m = 0,1,2,\dots, M) \quad \text{Eq. 3-32}$$

The values of  $m$  correspond to the speed controls and are ordered in Table 3-5 ( $M=5$ ). Since parameters  $\alpha_m$  must be monotonically increasing they could be expressed as:

$$\alpha_m = \alpha_{m-1} + \exp(\lambda_m), \text{ where} \quad \text{Eq. 3-33}$$

$$\alpha_1 = \text{constant}$$

In this way,  $\alpha$  could be calculated by estimating  $\lambda$ . This will lead to the elimination of unnecessary constraints from the optimisation since condition  $\alpha_m < \alpha_{m+1}$  will be automatically satisfied.

In this manner, the unique warning control decision is produced, which is in accordance with existing German specifications (BAST, 1999). The resulting model is a multivariate ordered regression model and is schematically presented in Figure 3-14. For each traffic state one such unique model can be utilised. The difference in models originates from different values of algorithm “weights” and decision boundaries. Theoretically, any information from the knowledge base could be included in the model so to produce the warning decision. The optimisation process automatically determines the contribution (if any) of this information to the final decision by estimation of:

- K parameters (“weights”)  $\beta_k$ , where  $k=1,\dots,K$
- M parameters  $\lambda_m$ , where  $m=1,\dots,M$

for each traffic state.

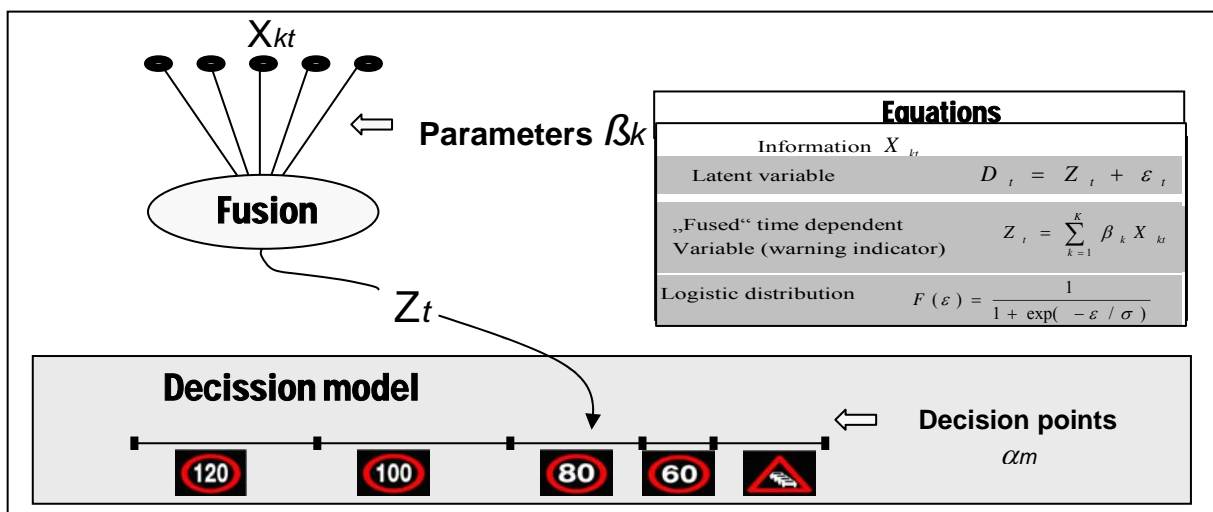


Figure 3-14: Warning decision model

Using Eq. 3-31 it is possible to quantify the uncertainty of the classification. From the model, the probability of control action  $V_w(m)$  being a “hit” (successful detection) can be estimated by:

$$\begin{aligned}
 P(1) &= \frac{1}{1 + \exp(Z_t - \alpha_1)} \\
 P(2) &= \frac{1}{1 + \exp(Z_t - \alpha_2)} - \frac{1}{1 + \exp(Z_t - \alpha_1)} \\
 &\dots
 \end{aligned}
 \tag{Eq. 3-34}$$

Basically, the model nonlinearity is due to calculations of decision borders  $\alpha$  and the estimations of these probabilities. Probabilities are not of direct interest to the control decision since we are interested in calculating the benefits and costs and determining the control, and not the precision of the hit probability. These estimates might however be useful for further statistical analysis.

### 3.6.2.3 Model for harmonisation decision

Based on the above considerations (chapter 3.6.2), for harmonisation decision a modified Belastung algorithm is used. The control decision is made by comparing the smoothed/forecasted values of speed of cars ( $Vcar_f$ ), traffic density ( $k$ ) and weighted flow ( $Q_b$ ) with the corresponding threshold values of the fundamental diagram. Schematic representation of the input variables in the model and their interaction (as given in MARZ (BAST, 1999)) is given in Figure 3-15.

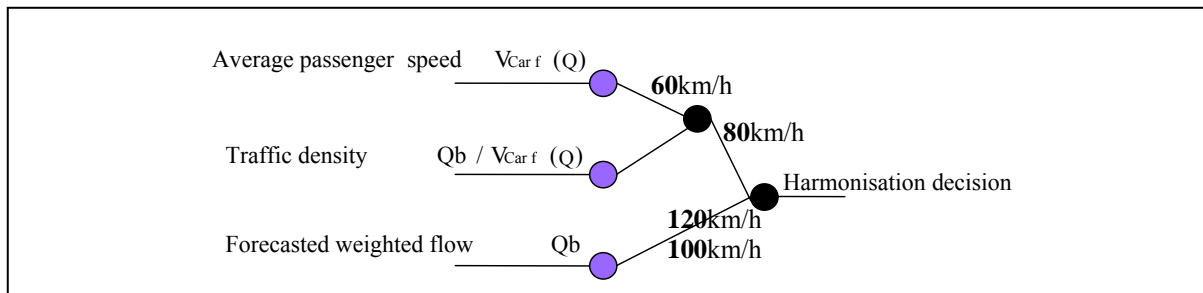
In its original form the Belastung algorithm could be responsible for the activation of controls between 120km/h and 60km/h. A different switch-on and switch-off threshold values for  $Q_b$ ,  $k$ , and  $Vcar_f$  are used for each control level  $m$ . Generally, control decision  $V_H(m)$  will be suggested if the measured values satisfy following condition:

$$\begin{aligned}
 &\text{IF at one detector station: } (Q_b > Q_{b_{on}}(m)) \text{ OR } ((Vcar_f < Vcar_{f_{on}}(m)) \text{ AND } (k > k_{on}(m))) \\
 &\quad \text{THEN } V_H(m) \text{ **ACIVATE**}
 \end{aligned}
 \tag{Eq. 3-35}$$

The controls will be terminated when the measured traffic flow values satisfy the following condition based on the switch-off threshold values:

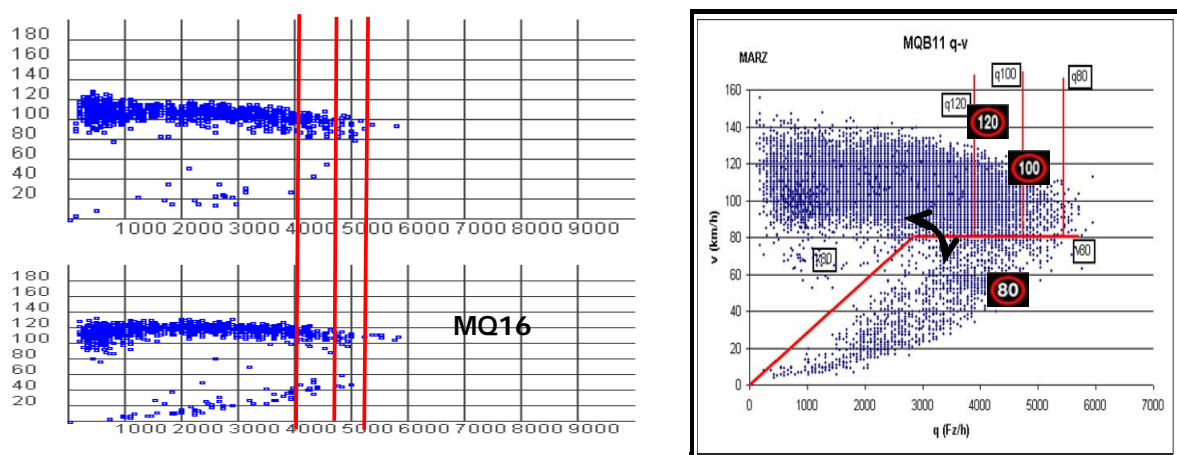
$$\begin{aligned}
 &\text{IF at one detector station: } (Q_b < Q_{b_{off}}(m)) \text{ AND } (Vcar_f > Vcar_{f_{off}}(m)) \text{ AND } (k < k_{off}(m)) \\
 &\quad \text{THEN } V_H(m) \text{ **DEACTIVATE**}
 \end{aligned}
 \tag{Eq. 3-36}$$

The decision for control 60km/h is made by checking the right term only ( $k$  and  $v$  values); for control 80 km/h by checking of the complete term ( $Q_b$ ,  $k$ , and  $Vcar_f$  values); for controls 100 km/h and 120 km/h it is made by proving only the left term ( $Q_b$  values).



**Figure 3-15: The Belastung algorithm- model structure**

The default values of these thresholds are given in MARZ (BAST, 1999). Since the Belastung algorithm is based on the fundamental diagram, which varies from location to location, in MARZ it is recommended that, for each control station the given default threshold values should be changed according to the traffic flow dynamics at the location. However, this is not easy to do and it usually is not done at all, since there is no objective framework which would give the operator the possibility to compare the effects of different threshold values. The complexity of the control decision makes it difficult to determine these values solely through expert knowledge. Therefore, usually default recommended values are used in practice. The exceptions are locations where the Belastung behaves strangely, and forces the authorities (operator) to change the parameters. However, in such cases the parameters are usually set very high, resulting in Belastung activations either when it is too late, or not at all (DENAES ET AL., 2001). For example, control of 120 km/h will be activated only when traffic is already travelling at a speed lower than this, e.g. 110 km/h. However each above-mentioned situation (using of default values or too high values) would most probably lead to underutilisation of the Belastung algorithm. Examples are given in Figure 3-16: left is an example of different fundamental diagrams (for detector stations MQ14 and MQ16 at BAB A8/Ost) and the default flow threshold value, right is the example of Belastung threshold values for one detector station.



**Figure 3-16: Left: default threshold values for different fundamental diagrams (MQ14 and MQ16); Right: possible unstable transitions from no control or 120km/h to control 80km/h for detector station MQ11 at BAB A8/Ost – where default threshold values seem to be appropriate**

Another potentially problematic fact is that for the 100km/h decision only flow threshold values ( $Q_b$ ) are used. Previous researches (KATES ET AL., 2002), as well as motorway authorities, evidenced that determining the control decisions of 100 km/h by comparison of flow values only could lead to too frequent control changes and/or the unstable transitions between the control outputs (Figure 3-16, right).

According to the previous discussion and based on the INCA control model specification, in INCA several *modifications* of the above described Belastung algorithm have been made:

1. According to the INCA, the local “segment-based” model harmonisation module can be responsible only for the decisions higher or equal to 80km/h. Therefore the control suggestion of 60km/h, and consequently threshold borders, is excluded from the model.
2. Due to the novel approach for control termination (next chapter) the switch-off threshold values are no longer necessary. Instead each control action will “live” for a predefined period of time  $T_H$ . This value could be the part of the optimisation process, but it is here taken to be constant, i.e.  $T_H = 5$  minutes.
3. As shown in KATES ET AL., 2002 determining the control decision of 100 km/h with the help of additional density and speed threshold values would lead to finer partitioning of the fundamental diagram and would result in more stable control decisions. Therefore, these threshold borders are added to the INCA Belastung implementation and are used by default. However, the user has the option to decide if these parameters should be applied or not.

Based on the above discussion, for each control station it will be necessary to determine the following Belastung *parameters*:

- for the control of 80km/h 3 threshold values ( $Q_{bon}(80), k_{on}(80), Vcar_{fon}(80)$ ) are needed
- the control of 100 km/h can be determined based on 3 or 1 threshold values, where  $Q_{bon}(100)$  is necessary and  $Vcar_{fon}(100)$  and  $k_{on}(100)$  are optional
- the control of 120 km/h is determined by comparing flow values only ( $Q_{bon}(120)$ ).

If  $k_{on}(100)$  and  $Vcar_{fon}(100)$  threshold values are not utilised, the harmonisation control model will result in a 5-, or otherwise in a 7-*parameter-approach*. The threshold borders are determined by the optimisation process. In this way, for each control station a separate location specific threshold values can be determined in a systematic and objective way. It should be emphasised that using 7 parameters is recommended since the optimisation procedure should be able to decide alone if all these parameters are actually necessary. In other words, the optimisation process should determine if the harmonisation control dynamics

can be effectively described using fewer parameters. So it can happen that, even if the model started with 7 parameters, some of them are set to be zero or less. These parameters are not considered to be important for the Belastung control decision and thus are automatically excluded from the harmonisation model.

**Table 3-6: Original and INCA threshold values for a three-lane motorway**

$V_H$	$Q_b$ OR		$(Vcar_f$ AND		$k$ )	
	MARZ	INCA	MARZ	INCA	MARZ	INCA
120 km/h	> 4000	$Q_{bon}(120)$	-		-	
100 km/h	> 4800	$Q_{bon}(100)$	-	$(Vcar_{f_{on}}(100))$	-	$(k_{on}(100))$
80 km/h	> 5400	$Q_{bon}(80)$	< 70	$Vcar_{f_{on}}(80)$	> 50	$k_{on}(80)$
60 km/h	-		< 50		> 50	

For better overview and comparison, in Table 3-6 the default Belastung threshold values recommended by MARZ, and parameters for optimisation used in INCA are shown (marked in blue). INCA harmonisation variables, whose inclusion is optional ( $k$  and  $v$  threshold values for control 100 km/h) are shown in brackets. Switch-off threshold values are not shown since these are not used in INCA.

#### 3.6.2.4 Control termination rule

As we already said one of the important characteristics of the control model is its ability to produce stable control actions. Stable control actions are ones that do not change too often and rarely allow unstable transitions. Unfortunately, the common practice, to keep a speed control activated until a (parameterised) switch-off condition is satisfied, is prone to failure either due to inappropriate thresholds or sensitivity to data errors. Usually these threshold values are set high as to prevent the too frequent changes. In this case the so-called “hysteresis” is too high and the control might be activated for an unnecessarily long time. On the other side, if the threshold values are too liberal the “hysteresis” will be too low (on and off values are too close) and the stability problem could arise. Moreover, the corrupted data could lead to the situations where it cannot be recognised that conditions for termination of the control have been fulfilled. In each case the control will have a negative effect on traffic flow either by introducing unnecessary delays, or by producing unnecessary and unsafe fluctuations, and will have a negative effect on acceptance.

In INCA however, the control action will be “smoothly” switched off, in steps of 20km/h unless re-activated by a new control decision. This “smooth” switching off is realised in the following manner. To each warning  $V_w(m)$  and harmonisation  $V_H(m)$  control a predefined duration time is assigned. In other words, the control will be “in charge” during a fixed period of time,  $T_w(m)$ ,  $m=1,..,5$  and  $T_H(m)$ ,  $m=1,..,3$ . If the control has not been re-activated (there were no new control suggestions) during this fixed time period, the control will be “smoothly” switched off by one level (increase of 20 km/h). Each of these control levels will also last for this fixed period of time. However, if during this time the control is activated

again its duration will be prolonged. Then the control decision will be the minimum of the new calculated control and the one that has been previously activated but it is “still in charge”:

$$V_w(t) = \min(V_w(t), V_w(t-1)), \quad \text{Eq. 3-37}$$

where,

$V_w(t)$  - the new control suggestion of the warning component (produced at time  $t$ )

$V_w(t-1)$  - the control that was active at time  $t-1$  (and maybe  $t-2$ , etc.) and, according to the above described procedure, should be still kept “alive”.

For harmonisation an analogous procedure is used.

Obviously, this rule leads to a more stable controls, since the fluctuations can be only as large as the abovementioned predefined control duration time. Additionally, the changes in controls larger than 20 km/h (ascending) cannot be experienced. Under the assumption that acceptance is perfect (100%), the time sequence of the controls will correspond to the space development of speeds. In this way the control will not produce unnecessary disturbances (“collision” of traffic waves) and the condition of having controls smooth in time is fulfilled. In the approaches used up to now (the one with threshold values) this criterion was not automatically fulfilled, since it was possible to have unsmooth termination of controls if the threshold values suggested so.

In this way, smoothing in time is independently guaranteed and the switch-off parameters can be excluded from the optimisation process. Basically, the control duration parameters  $T_w(m)$  and  $T_H(m)$  can be either taken as predefined constant fixed values, or could be part of the optimisation process. In case of their optimisation, the initial values could be set either to the above defined values or to the time needed for the shockwave to propagate from the downstream location to the control station.

### 3.6.3 Global “supervisor” layer

Although it has been assumed that the local “intelligent” controller will be able to produce the best control for the complete segment and, inherently, for the complete stretch (global view), there are several situations in which a global supervisor layer may be needed. Local control decisions are subject to modification if required at the global control level. The global control level has several responsibilities (given in the order in which adjustments are made):

- it monitors and integrates external information of high priority;
- it monitors and integrates operator control (e.g., due to accidents, road works, etc.) and

- it “propagates” and “funnels” local decisions under consideration of their global effects (e.g. differences in upstream detection of speed limits at two consecutive VMS should not exceed 20km/h);

The description of the adjustment process is explained in the order in which they are made.

First the inclusion of controls that are not determined by the local controller is considered. These can be various speed limits controls due to the adverse weather conditions, prohibiting truck passing sign, “attention” signs, different information, etc.<sup>31</sup>. These external controls  $V_{EXT}(j,t)$  and the reasons for their activation are produced by the sub-control centre and are available at each minute  $t$  for each control station  $j$ . If the control  $V_{EXT}(t)$  has been specified as one that should be integrated with INCA outputs<sup>32</sup>, it will be compared with the output from the local control model  $V_C(j,t)$  and the smaller of these two will be taken as the final control for station  $j$  at time  $t$   $V_{FC}(j,t)$ :

$$V_{FC}(j,t) = \min(V_C(j,t), V_{EXT}(j,t)) \quad \text{Eq. 3-38}$$

Manual control following work zones, accidents or some other special events should be also taken into consideration. In LCS implementations and thus in INCA as well, manual controls are seen as controls of the highest priority and will “overwrite” any (even more restrictive) control produced by the control model. Therefore, if at time  $t$  the operator (manually) activates control  $V_M(j,t)$ , the local control suggestion  $V_C(j,t)$  will be ignored and the final control for location  $j$  at time  $t$ ,  $V_{FC}(j,t) = V_M(j,t)$

Global adjustments can be made when all the local controls have been determined. The global adjustments are made due to safety reasons and are in accordance with German recommendations given in BAST, 1999. Basically, there are two possible adjustments:

- upstream control propagation rule - mainly due to warning controls, and
- funnelling of traffic in the downstream direction - due to the harmonisation controls.

A control propagation rule in the upstream direction is the result of the need for controls that are smooth in space. The propagation rule is activated only for  $V_C(j,t)$  lower than 100 km/h. If at some control location  $j$ , at time  $t$ , such speed control  $V_{FC}(j,t)$  is activated, then at the upstream control stations  $j-1$  control  $V_{FC}(j-1,t)$  should be posted in increments of 20 km/h. However, controls at upstream control stations will be modified only if they are higher than the “propagated” control.

---

<sup>31</sup> Detailed specification of available controls is given in TLS (BAST, 2002).

<sup>32</sup> In INCA by default only prohibiting truck passing sign and speed controls due to the low visibility are included.

$$\text{IF } V_{FC}(j,t) \leq 80 \text{ km/h THEN } V_{FC}(j-1,t) = \min(V_{FC}(j,t) + 20 \text{ km/h}, V_{FC}(j-1,t)) \quad \text{Eq. 3-39}$$

The propagation is performed until the speed control reaches 100 km/h. In such way, the driver will be “gently” led to the congested area. It should be noted that there is one additional motivation to this rule (especially in current applications). Namely, if the local control fails to detect incident further downstream, not only that the drivers will be smoothly slowed down but they will be also warned on time. Furthermore, with this rule eventually faster discharge of the congested area could be achieved.

Funnelling of traffic flow is due to controls of 100 km/h and 120 km/h. It represents the propagation of the control decision in the downstream direction. In other words, control  $V_{FC}(j,t)$  produced at station  $j$  at time  $t$  will be also applied at the downstream control station  $j+1$   $V_{FC}(j+1,t)$ , given it is not greater than the downstream local control.

$$\text{IF } V_{FC}(j,t) \geq 100 \text{ km/h THEN } V_{FC}(j+1,t) = \min(V_{FC}(j,t), V_{FC}(j+1,t)) \quad \text{Eq. 3-40}$$

This is due to the wish to homogenise traffic flow longitudinally, and for this reason consistent controls in harmonisation area are considered to be crucial. Even though in MARZ it is recommended that such control should propagate up to two downstream control locations, since in Germany distance between subsequent control locations is quite large, in INCA the control is funnelled only to the first downstream location. However this is again reconfigurable.

### 3.7 Optimisation

Parameter optimisation is an essential part of any data-driven approach. However, we argue that parameter optimisation alone is not enough for deriving a *general* data-driven model. One could use the best optimisation technique available and learn the model, but this would not guarantee that the model will be equally good on new unseen data. Therefore it is necessary (in addition to the parameter optimisation process) to integrate the process for determining and using the most stable model structure (for known and unknown data sets), i.e. the most general one. The optimisation process implemented in INCA integrates both:

- parameter optimisation, and
- model complexity optimisation (model selection)

Parameter optimisation implies the process of searching for a parameter vector that satisfies given constraints and leads to the “optimal” value of the objective function. Parameter optimisation is a well researched area, and depending on the particular problem various more or less advanced techniques can be used. Although one could use the advance parameter optimisation technique which could lead to finding the global optima, this is the global optima for the training data set (!). In other words, for most data-driven approaches there is no guarantee that the global optimum for the training data set is the global optimum for an



unseen data set also (chapter 3.1.2). Indeed, much more often it could be the case that the global optimum for the particular data set is far from the global optimum for a new unseen data set. This is mostly due to the many parameters used in the model and the limited amount of available data used for training. ***Therefore, for the data-driven approaches it is of higher interest to find model that is general, i.e. the optimum that would lead to an equally good performance for any data set, than finding the global optimum for the training data set. Hence, we are interested in finding a good enough but stable optimum, rather than the global optimum.***

The problem of model generalisation property is linked to model bias and variance (discussed in chapter 3.1.1). Let us assume that a certain regression model has been optimised several times, each time producing the estimated values  $f(x)$ . The square error ( $SE$ ) between estimated values  $f(x)$  and real values  $y(x)$  could be separated into bias and variance error as follows (NELLES, 1999):

$$SE = [y(x) - \langle f(x) \rangle]^2 + \langle [f(x) - \langle f(x) \rangle]^2 \rangle \quad \text{Eq. 3-41}$$

The first part corresponds to the model „*bias*“, the second one to “*variance*”. While the bias error is purely due to the structural inflexibility of the model, the variance error is part of the model error which is due to uncertainties in the estimated parameters. If the data set used for the optimisation covers all possible situations, the variance problem will not emerge (NELLES, 1999). Simply the parameter vector would be the best for all situations. However, in real life we hardly ever have training data set that covers all, or even enough representative situations. In this case, applying parameter optimisation only, would make bias error small, but could neglect model variance. Therefore, the optimised parameters may be the best for this single optimisation data set, but might perform much worse on an unknown data set. In INCA, in addition to parameter optimisation, the focus of the optimisation process is on incorporating and using mechanisms that will make detection of model variance and finding the appropriate trade-off between bias and variance possible, and lead to a general control model.

Since the available amount of training data is usually limited, it is not feasible to divide the training data set into several training sets (subsets) that would be separately optimised, and in such way to estimate model variance. To overcome this problem, so-called re-sampling approaches like cross-validation, jackknife, Bootstrap (EFRON AND TIBSHIRANI 1993; SARLE, 2001), have been proposed in literature. Re-sampling approaches make it possible to extract information about model stability (variance) from limited available data. Of these, Bootstrap usually provides better estimates of the generalisation error (SARLE, 2001) and is used in INCA optimisation framework.

Finding the appropriate trade-off between model variance and bias, implies the problem of finding the appropriate number of parameters (included information) of the warning control model. The basic problem with regression-based models is the correlation of the used

information. Although the INCA warning model could work well even with some correlation in parameters<sup>33</sup>, such a condition could lead to a model that is difficult to interpret. For example, due to the correlation between information, the “weights” of some information might be negative although it is known that this information is important for the control. Since the benefit of such parameter is redundant and small in comparison to the variance they induce, for the sake of model stability it would be beneficial to trade their variance for a slightly higher bias, by excluding them from the system. Mechanisms that allow this trade-off are called regularisation techniques. In INCA the Ridge regression regularisation technique is used. It is also known as the “shrinking” technique and, in the context of neural networks, it is called weight decay. By implementing the Ridge regression the parameters that are correlated or not significant can be eliminated from the model. The optimisation approach utilising Ridge regression and re-sampling (Bootstrap) should lead to a general model.

Ridge regression is related to the warning component. Bootstrap is useful for both modules, although clearly higher benefit should be expected for the warning component which has a greater number of parameters for optimisation and thus possibly a higher variance.

### 3.7.1 Parameters and constraints

The parameter vector contains parameters from the warning and harmonisation components. For optimisation the numbers of parameters and additional constraints are of interest. According to the INCA the control model parameter space is defined as follows:

- The parameters of the harmonisation component are the threshold values of the modified situation estimation algorithm Belastung. These are  $Q_{bon}(m)$ ,  $Vcar_{fon}(m)$ ,  $k_{on}(m)$  thresholds, where  $m=1,2,3$  correspond to the control recommendations of 120 km/h, 100 km/h, and 80 km/h, respectively. Depending on the chosen approach Belastung parameter set can be of dimension 5 or 7. Parameters may take any value and there are no additional constraints. A negative parameter value means that the parameter should not be considered in the decision making process, and can be excluded from the model. If the optimised boundaries for the control of 100 km/h are higher than 120 km/h this means that 120 km/h control will never be activated but rather 100 km/h instead, and the thresholds for the control of 120 km/h could be excluded from the model.
- In the warning component, parameters  $\beta_k$  corresponding to input values  $X(k)$ ,  $k=1,..K$  should be optimised. For example, parameter  $\beta_{Stau1}$  is assigned to input value  $X(Stau1)$ . According to the model definition, these parameters can be interpreted as “weights” of the specific input sources (in this case the Stau1 algorithm) in the warning model, where a

---

<sup>33</sup> For models whose decision is made upon calculated probabilities, such as discrete choice models, correlation could result in implausible calculated probabilities (BEN-AKIVA, 1998)

higher “weights” implies greater algorithm importance. The parameters  $\beta_k$  could also take any value (positive or negative). Negative values do not necessarily mean that the algorithm output negatively influences the control decision but might also mean that this algorithm is in correlation with some other algorithm.

- In addition to the  $\beta_k$  parameters in the warning component, decision points  $\alpha_m$  should also be determined, where  $m=2,\dots,M$  and  $M=5$ .  $\alpha_1$  is set to be constant and therefore should not be optimised. Other  $\alpha$  parameters must satisfy following constraint  $\alpha_m < \alpha_{m+1}$ . However, since  $\alpha_m$  can be determined as  $\alpha_m = \alpha_{m-1} + \exp(\lambda_m)$ , instead of optimising  $\alpha_m$  one should optimise  $\lambda_m$  and this constraint will automatically be satisfied (and thus can be eliminated from the optimisation process).

It follows that optimising the harmonisation and warning parameters can be formulated without explicit constraints.

### 3.7.2 Parameter optimisation using Downhill simplex method

The INCA control model consists from two logical control blocks (harmonisation and warning) each with its own parameters that should be optimised. To keep this logical separation, and considering their different control motivations, it has been decided to separately optimise the parameters of these two components. This could lead to certain overlapping in the area of the 100 km/h and 80 km/h controls. However, such overlapping and, consequently, possible deviation of the best optimum, will be rare since these controls are triggered by quite different events – one are warnings about already present incidents, others attempt to prevent or postpone them. The slight overlapping that might be expected is not considered critical for the application, especially as the goal of the optimisation is not to find the global but good enough and stable optimum. Separate optimisation is interesting for the analysis of system behaviour and its individual components. Furthermore, the influence of the changes in model parameters on the calibrated values could be investigated. To conclude, parameters could also be optimised together, if necessary.

Since in INCA, the gradient of the objective function is non-linear in parameters, a non-linear parameter optimisation technique is required. Therefore, for optimisation of INCA parameters we need the *non-linear unconstrained multidimensional optimisation technique*. Non-linear optimisation problems generally have the following properties (NELLES, 1999): there are many local optima; the surface in the neighbourhood of a local optimum can be approximated; no analytic solution exists; an iterative algorithm is recommended; they can hardly be applied online.

Generally, a distinction can be made between local and global non-linear optimisation techniques. The local optimisation techniques start at one initial point and examine the neighbourhood of this point. Global searches start with several different points and thus have

the ability to explore different regions of the parameter space. Generally, global searches (e.g. genetic algorithms or simulated annealing) usually have some stochastic elements, and might sometimes find the global optimum of non-linear optimisation problem but are computationally much more expensive. Since *the goal in INCA is not to find the global optimum, but optimum that is good enough and stable*, this can also be achieved with less demanding local optimisation techniques.

Local optimisation approaches can be grouped into the following two groups – gradient-based methods, and others that are based solely on the evaluation of the objective function, i.e. the so-called direct search methods. Generally, gradient search methods are believed to have better performances against direct search methods. However, gradient searches require analytic partial derivatives with respect to all parameters. Since these are not available in the present case a direct search method is needed. Several such methods can be found in literature (PRESS ET AL., 2002; NELLES, 1999), where we used the downhill simplex method due to Nelder and Mead (PRESS ET AL., 2002). The downhill simplex method should not be confused with simplex method for linear programming (!). Since this is well-known model, only the short description will be given here while more details may be found in Appendix A.3.

The goal of the downhill simplex method is to find an n-dimensional (multidimensional) parameter vector that minimizes the objective function. The simplex method does not use derivatives (analytic or numeric) but only function evaluation. The downhill simplex method is an efficient iterative algorithm for numerical solving of unconstrained minimisation problems. The downhill simplex method just crawls downhill in a straightforward fashion that makes almost no assumptions about the objective function. It is completely self-constrained. This can be extremely slow but in some cases also extremely robust (PRESS ET AL., 2002). The methods could be terminated when the decrease in the function value in the terminating step is fractionally smaller than some tolerance  $f_{tol}$  (usually 0.00001), or when a predefined number of iterations is reached. However, the  $f_{tol}$  (or some other) criteria might be fooled by a single anomalous step that, for some reason, failed to get anywhere. Therefore, it is usually recommended to restart the multidimensional minimisation procedure at the point where it claims to have found a minimum.

In the optimisation of INCA parameters the downhill simplex method performs well (chapter 5.4.7). However, it should be noted that with the increase in the number of parameters, the number of iterations and evaluated functions will increase also, which will slow down the optimisation process. Therefore, for more complex approaches with many more parameters (e.g. neural networks or fuzzy systems approaches) optimisation using other search techniques is recommended. The same is recommended in the cases where finding the global minimum is the main priority.

### 3.7.3 Estimation of the generalisation error using Bootstrap

Given only one training data set Bootstrap (e.g. EFRON AND TIBSHIRANI 1993) makes it possible to acquire a more or less detailed estimation of parameter variance by creating internal subsamples. The Bootstrapping implementation that is used in INCA repeatedly analyses data subsamples instead of repeatedly analysing data subsets. Each subsample is a random sample with replacement from the full sample. It allows and it is used in INCA for:

- Estimation of the parameter (model) variance (generalisation error) with a limited amount of available training data.
- Model structure optimisation. The Bootstrap method offers the possibility of determining parameter variance on different training subsamples. Parameters shown to be unstable on several training subsamples, can be then excluded from the model.
- Consequently, Bootstrap could further increase robustness.

For the Bootstrap application in this work, the available optimisation data (training data set) is considered to be the collection of days. This means that each complete day will be considered as one (“the smallest”) unit. The days are then numerated as  $d = 1, \dots, D$ . The next step is to draw  $B$  new subsamples from available  $D$  days. Each  $B$  subsample is a random sample of numbers between 1 and  $D$ , generated with replacement from the full sample. Sampling with replacement essentially means that each random choice of the number between 1 and  $D$  is done on the full data set (already chosen numbers are not eliminated from the further choice process). This implies that in almost each new  $B$  subsample some values (in this case days) from initial real set will be repeated (e.g. two times day 3 or four times day 21). The  $B$  subsamples are called Bootstrap subsamples.

The use of the Bootstrap method will be explained using the example of the warning component. We start from the data sets  $(X_t, Y_t)$ , where  $X_t$  is the parameter vector and  $Y_t$  is related to the measured traffic flow values  $(v, q, k)$  needed for the evaluation of the objective function. With the Bootstrap, in addition to the values of  $(X_t, Y_t)$  of initial data set (original training data set) which will be denoted as  $0^{th}$  bootstrap ( $b=0$ ), many new samples of  $(X_t(b), Y_t(b))$ , where  $b = 0, \dots, B$ , will be obtained. Statistics interesting to us (optimal values of control parameters:  $M$  parameters of  $\lambda_m$ ,  $K$  parameters of  $\beta_k$ ) are estimated for each Bootstrap sample with the optimisation of the objective function. This results in  $\beta_k(b)$  and  $\lambda_m(b)$  number of estimations, where  $b = 0, \dots, B$ . Hence, in total there will be  $B+1$  estimated values of variables of interest. Using the Bootstrap for the harmonisation parameters is analogous to the case of the warning procedure.

According to the Bootstrap, the (known) empirical distribution can be used as an approximation of the true distribution of the independent variable. This is done by calculating the statistical measures such as mean, variance, etc., of parameter values produced using

different samples. This estimation provides information about parameter variance (generalisation). In order to make this estimation a sufficient number of observations (in our case days) is still needed. Therefore, the Bootstrap sample size should be as large as the original sample size (e.g.  $D$ ). The number of optimal number of replications (subsamples) depends on the problem itself. The fact is that on the formal level Bootstrap would require an infinite number of replications (in our case number of replication is marked as  $B$ ) (SARLE, 1995). However, the key to the usefulness of the Bootstrap is that it converges very quickly in terms of the number of replications, and so running a finite number of replications is sufficient. In other words, with the increase in the number of Bootstrap samples, the precision of the estimated variance grows. The question is how accurate these estimates should be for the particular problem. In our case, a few digits of precision is sufficient, because we are not interested in a perfect calculation but good enough approximations (we do not get more valuable information by estimating the variance to be exactly e.g. 0.6754398 then by of e.g. 0.68). Since Bootstrap converges quickly, the estimation we need can be produced with several Bootstraps (10-20).

From these  $B$  Bootstrap samples it is possible to approximate distributions of values, which would otherwise be obtainable only if we had many optimisation subsets. Moreover, it is possible to estimate the variance of the estimated parameters even though we have only a limited amount of data. Finally, from the produced  $B$  parameter values, one **aggregated control model** can be obtained, which is believed to be more robust on one hand and on the other makes parameter imprecision evident. The process of obtaining the aggregated control module and measures of imprecision (variance) are given in the following section.

### 3.7.3.1 Aggregation of the results from Bootstrap samples

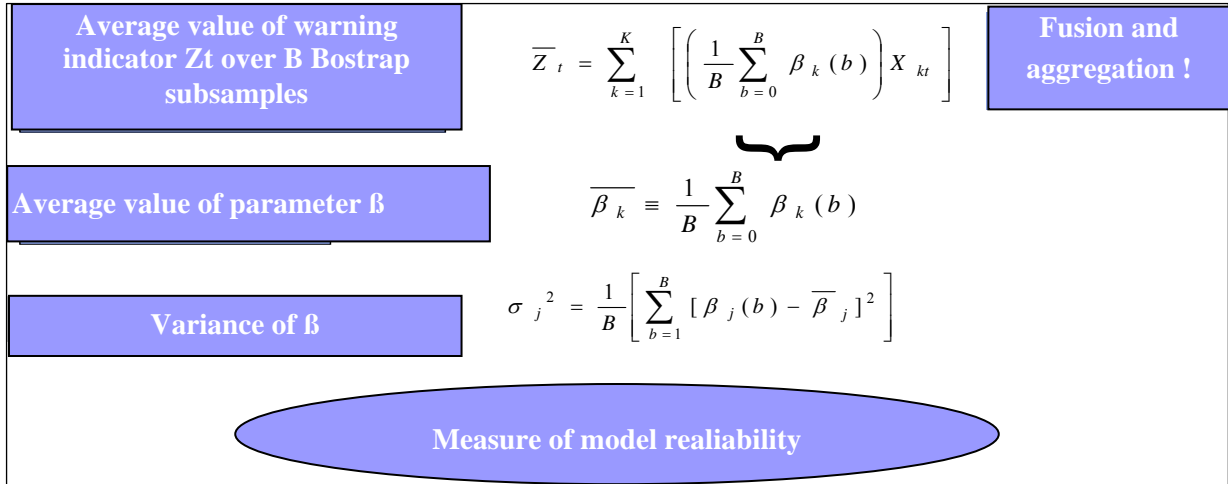
Through estimation of parameters from different  $B$  Bootstrap subsamples  $(X_i(b), V_i(b))$ ,  $b = 0, \dots, B$ , many values of the warning parameters  $\{\lambda_m, \beta_k\}$  will be produced, which will be referred to as  $\{\lambda_m(b), \beta_k(b)\}$ . The goal is to obtain, from these  $B$  estimations, a measure of the parameter variance and aggregated control model, which is believed to be more general (thus more robust).

For the acquisition of the parameter variance and aggregated control model INCA uses so-called Bootstrap aggregation (BREIMAN, 1996). Bootstrap aggregation (bagging) fits the aggregated model by using the above  $B$  Bootstrap subsamples and combining them by averaging the model outputs and inputs. The illustration of how the simple averaging could be used will be given using the example of the warning module (see Figure 3-17). Harmonisation averaging can be performed analogously.

From Bootstrapping we acquired the “collection” of models – one for each Bootstrap sample. This means (BREIMAN, 1996) that for each minute  $t$  an estimation of the corresponding

aggregated value  $\bar{Z}_t$  (*bagged average*) can be obtained by calculating its average value over Bootstrap samples  $Z_t(b)$  using

$$\bar{Z}_t = \frac{1}{B} \sum_{b=0}^B Z_t(b) \quad \text{Eq. 3-42}$$



**Figure 3-17: Bootstrap aggregation of the warning control model**

Based on the warning model and equation for the warning indicator  $Z_t$ , it yields:

$$\bar{Z}_t = \frac{1}{B} \sum_{b=0}^B \sum_{k=1}^K \beta_k(b) X_{kt} , \quad \text{Eq. 3-43}$$

This can be also expressed as:

$$\bar{Z}_t = \sum_{k=1}^K \left[ \left( \frac{1}{B} \sum_{b=0}^B \beta_k(b) \right) X_{kt} \right], \text{ which can be written as:}$$

$$\bar{Z}_t = \sum_{k=1}^K \bar{\beta}_k X_{kt} , \text{ where } \bar{\beta}_k \equiv \frac{1}{B} \sum_{b=0}^B \beta_k(b) \quad \text{Eq. 3-44 a,b,c}$$

The above transformation implies that the average of warning indicator  $Z_t$  can be obtained using the average of model parameters  $\beta_k$ . Since the combined parameters can be compared and estimated directly from the model, this leads to significant simplification of the problem.

According to BREIMAN, 1996, under constraint  $\alpha_1 = \text{constant}$ , the same averaging procedure can be undertaken for decision points

$$\bar{\alpha}_m \equiv \frac{1}{B} \sum_{b=0}^B \alpha_m(b) \quad \text{Eq. 3-45}$$

Variance of the warning indicator in minute  $t$  is

$$\sigma_Z^2 = \frac{1}{B} \left[ \sum_{b=0}^B [Z_t(b) - \bar{Z}_t]^2 \right] \quad \text{Eq. 3-46}$$

Variance of the parameter  $\beta_k$  can be estimated with the help of

$$\sigma_k^2 = \frac{1}{B} \left[ \sum_{b=1}^B [\beta_k(b) - \bar{\beta}_k]^2 \right] \quad \text{Eq. 3-47}$$

### 3.7.4 Reduction of the degrees of freedom using Ridge regression

The problem of correlation or “unnecessary” parameters increases with the dimension of the parameter vector. Therefore, the variance of the worst estimated parameter increases with the increase in model flexibility. Hence, according to the previous discussion, there is fundamental trade-off between the benefits of additional independent variable (due to higher model flexibility) and drawbacks due to increasing of estimation variance. Hence, for the robustness and transferability, it would be a great advantage if this “redundant” and/or “unnecessary” information is not considered, i.e. if parameters  $\beta_k$  could be automatically set to zero. Parameters  $\alpha_m$  derived from  $\lambda_m$  are necessary and cannot be eliminated.

Ridge regression is the regularisation technique which makes possible automatic controlling of this trade-off, and inherently the reduction of parameters, with the help of regularisation term. In Ridge regression, the extension

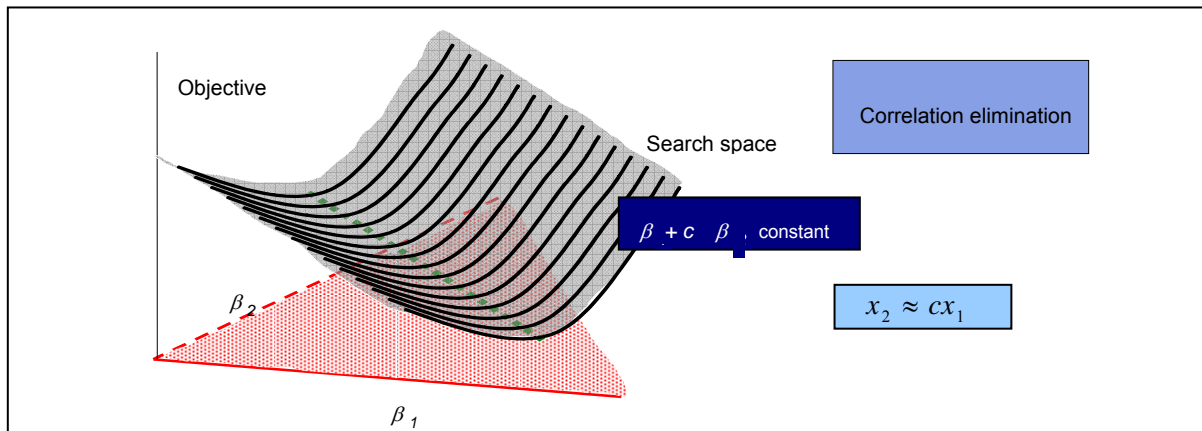
$$MSE' = MSE + \mu * \sum_{k=1}^K \beta_k^2 \quad \text{Eq. 3-48}$$

is added to the classical objective function. The idea behind the additional term is remarkably simple. Basically,  $\mu$  rewards smaller values of coefficients, in this case  $\beta_k$ . Those parameters that are not important for solving the optimisation problem tend toward zero in order to decrease the penalty term. Therefore only the significant parameters will be used since their error reduction effect is greater than their penalty term (NELLES, 1999). The parameter  $\mu$  regulates the desired level of coefficients and „smoothness“. For  $\mu \rightarrow 0$  the optimisation problem is recovered, while  $\mu \rightarrow \infty$  forces all parameters to zero. Such an objective function will (for  $\mu > 0$ ) always lead to a solution with increased bias but smaller variance than without it.

If, for example, the input variables  $x_1$  and  $x_2$  are strongly correlated ( $x_2 \approx c x_1$ ), in application without the Ridge regression only small differences in the value of objective function will be observable, as long as the linear combination of corresponding parameters (for example  $\beta_1 + c \beta_2$ ) remains constant. This is due to the correlation where the sum  $\beta_1 x_1 + \beta_2 x_2$  also remains constant in this case. Solely due to the randomness of available data, in many situations this could lead to found values of  $\beta_1$  and  $\beta_2$  that are “opposite to each other”. In



other words, we might obtain, for example, negative values of  $\beta_2$  only due to its correlation with the positive value of  $\beta_1$ . This of course could significantly increase model variance.



**Figure 3-18: Multicolinearity**

If parameters are strongly correlated the Ridge regression will force one of these two parameters,  $\beta_1$  and  $\beta_2$ , to 0, and in such way will eliminate the corresponding input variables from the optimisation. Furthermore, in this way parameters  $\beta$ , whose contribution to the good decision is smaller than the penalty term, will be „punished“ (eliminated), and the degree of freedom (number of parameters) will be reduced.

According to the chapter 3.6.2, there are  $K$   $\beta_k$  parameters. Theoretically, after application of the Ridge regression some values of  $\beta$  parameters (e.g.  $\beta_1$ ) might disappear. Due to the finite error tolerance of the optimisation termination criteria, the result of the Ridge regression will not be  $\beta_1 = 0$ , but  $\beta_1 \approx \Delta \ll 1$ . Where  $\Delta$  is a small value, approximately of the same size as the error tolerance itself (around  $10^{-2}$ ). Therefore, all input values  $X_k$  where  $\beta_1 \leq \Delta$  should be eliminated from the model. After the elimination of these values, the new optimisation procedure (for this reduced number of independent variables) could be started again.

The Ridge regression procedure used in INCA leads to the desired reduction of variance and the dimension of parameter space. However, the price to be paid is an increasing estimation bias with increasing  $\mu$ , and therefore an iterative approach in order to find “the best” value for  $\mu$  is sometimes required.

### 3.7.5 Adjustment of the objective function

In chapter 3.3.6 the new one dimensional objective function appropriate for optimising and monitoring of the link control model has been described. The objective function  $G$  has following general form:

$$G = B_T - C_{TT}, \text{ with } B_T = \alpha * B_W + (1 - \alpha) * B_H \quad \text{Eq. 3-49}$$

Where,  $B_W$  is the warning benefit,  $B_H$  is harmonisation benefit and  $C_{TT}$  is travel time cost. The higher the positive difference between the benefits and costs the better the control model is. Therefore, the model that leads to the maximisation of the above function  $G$  would be considered as “optimal”.

However, since we extended our model using the Ridge regression, the objective function should be extended with the regularisation “penalty” term from q. 3-48 which should be calculated for each time period  $t$   $\mu \sum_{t=1}^T \sum_{k=1}^K \beta_k^2$ . The penalty term represent the costs of including each input variable and therefore represents the costs of the control. Since  $G$  initially represents the benefit, where higher values are better, the penalty term should be subtracted (analogous to the costs) from the initial function  $G$  which results in modified objective function  $G'$ :

$$G' = G - \mu \sum_{t=1}^T \sum_{k=1}^K \beta_k^2 \quad \text{Eq. 3-50}$$

The downhill simplex method was used, which searches its way downhill (minimum) for the best solution. We convert this maximisation into the minimisation problem by trivial transformation of the form:

$$\min(-G') = -G + \mu \sum_{t=1}^T \sum_{k=1}^K \beta_k^2 = ((B_W + B_H) - C_{TT}) + \mu \sum_{t=1}^T \sum_{k=1}^K \beta_k^2 \quad \text{Eq. 3-51}$$

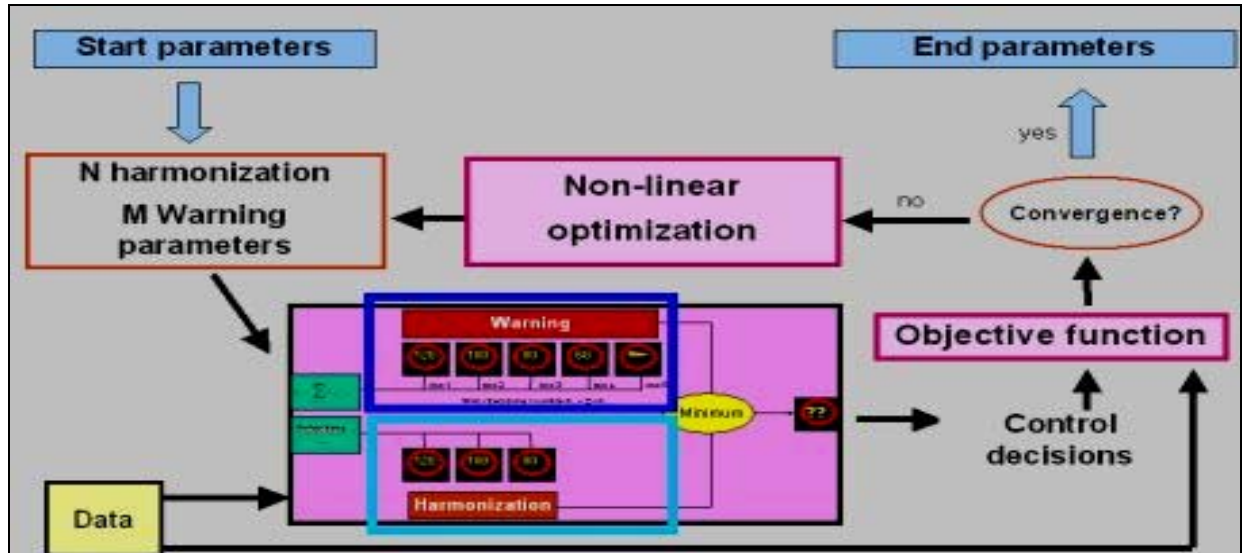
In literature, different approaches for determining and choosing the appropriate  $\mu$  value are proposed (NELLES, 1999). Generally, parameter  $\mu$  can be determined according to the order of magnitude of average  $G$ . Since the measure of  $G$  is expressed as travel time per kilometre per vehicle (i.e. sec/veh/km) and the input values  $X_{kt}$  are scaled in such a way that the values have the order of magnitude of 1 (due to the normalisation process described in chapter 3.5), the value of  $\mu$  is estimated to be  $\mu \approx 0.1$ . By using of Ridge regression it is usually useful (and sometimes also necessary) to test several values of  $\mu$  ( $\Rightarrow$  several optimisation iterations) and analyse model behaviour.

### 3.7.6 Optimisation process

For a typical link control system, the data supply permits testing and re-calibration several times per year. In addition, by continuously monitoring performance, significant structural changes in traffic patterns or severe technical problems can be detected within a relatively short time and optimisation could be started again.

*The optimisation process integrates (in the previous chapter described): parameters, parameter optimisation procedure, Ridge regression, Bootstrapping, and the objective function.* The optimisation is done in an integrated manner – optimising parameters and model

structure in “a single” process. The parameter optimisation could be perceived as the “smallest” general unit of the entire optimisation process, on top of which extensions such as Ridge regression and Bootstrapping are added. Therefore, as well as for sake of simplicity, first the typical parameter optimisation run will be explained.



**Figure 3-19: Parameter optimisation process**

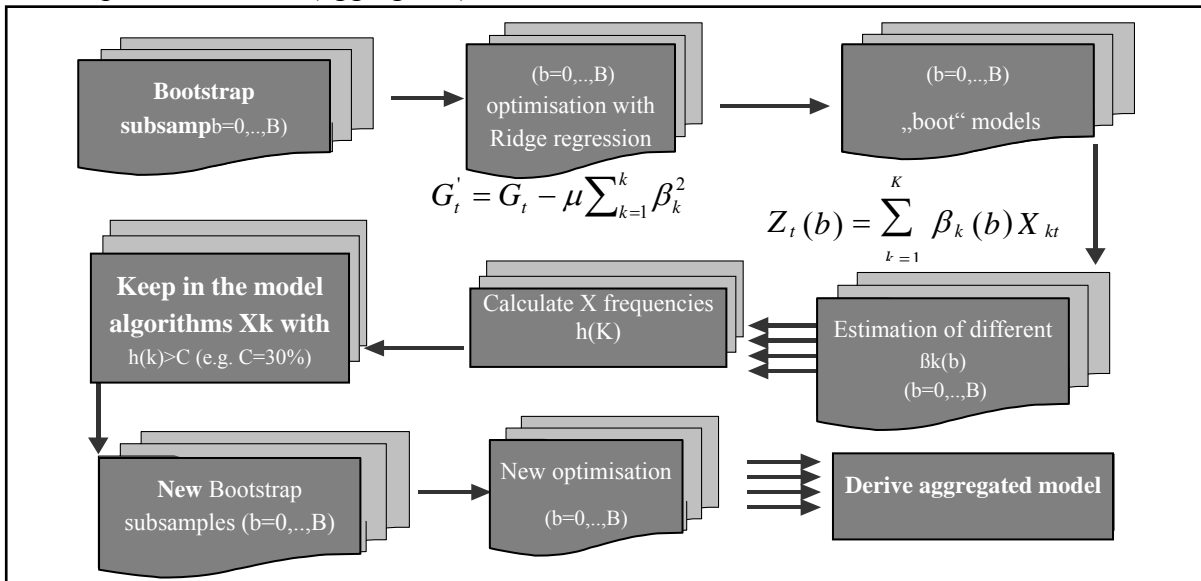
One typical parameter optimisation procedure includes the parameters that should be optimised (chapter 3.7.1), objective function (chapter 3.7.5) and chosen parameter optimisation technique (3.7.2) and is shown in Figure 3-19.

The existing detection/estimation/prediction algorithms and indicators provide outputs (chapter 3.5.1), some of which are continuously varying quantities, while others are discrete. In order to use these in INCA model, we first transform the indicators to an appropriate normalised scale (chapter 3.5). The parameters of the warning model are associated with the weight of these transformed indicators and are coded accordingly. The parameters of the harmonisation model are associated with the threshold borders of the Belastung algorithm. Due to the reasons given in chapter 3.7.2, in INCA optimisation of warning and harmonisation control parameters is done separately, but it can also be done together, if needed.

For each initial parameter vector the sequence of the control actions that would be produced with them is calculated, and their performances are measured using modified objective function. If optimisation without Ridge regression is required, the parameter  $\mu$  is set to be equal to zero. Otherwise,  $\mu$  is set to be  $> 0$  (in our case = 0.1). The parameter optimisation procedure further proceeds as described in chapter 3.7.2. According to these parameter vectors and their performances the downhill simplex creates the new parameter vectors which should be evaluated. The controls for this vector are estimated, the objective function is calculated and new transformations are made. The procedure iterates until the process converges, or until the predefined number of the iterations is reached (chapter 3.7.2). In order

to reduce the possibility of staking in a local minimum, the optimisation procedure is restarted several times. The restart is performed by setting the “best” parameter vector from the previous iteration as the initial vector of the new optimisation run.

The extension of the optimisation process with Bootstrap is shown in Figure 3-20. By using Bootstrap instead of having a single-parameter optimisation run described above, we will have  $B + 1$  optimisation runs for each data subsample as well as for initial one. The next thing needed is a procedure for the decision on variables which should be selected from these  $B + 1$  subsamples in the final (aggregated) model.



**Figure 3-20: Model selection using Bootstrap**

The following steps should be performed to eliminate superficial control model parameters and derive aggregated control model:

1. For each Bootstrap sample start parameter optimisation using Ridge regression.
2. Calculate  $X_k$  frequency  $h(X_k)$
3. Identify input values  $X_k$  with  $h(X_k) > c$  ( $c = 30\%$ ) and create the new input vector  $K'$  from them
4. Create new  $B$  Bootstrap subsamples and run the parameter optimisation procedure for remaining  $K'$  input variables  $X_k$  without Ridge regression.
5. Derive the aggregated control model from the produced  $B + 1$  values of  $K'$  parameters as described in chapter 3.7.3.

When deriving the aggregated harmonisation control model, only steps 1 and 5 are required.

### 3.7.7 Generating the ideal control

Generating the ideal control is based on TEODOROVIC AND EDARA, 2005. Basically, one could see this process as playing God: if we know, at each moment, what will happen next and what are our exact goals and the consequences of our actions (objective function) then it would be possible to make an ideal decisions (sequence of control actions). Deriving such ideal control sequence would make possible to test the following (at least):

- Whether the objective function appropriately optimises link control goals? This can be proven by measuring performances of ideal control with some other appropriate measure of performance (chapter 5.3)
- What amount of link control potential has not been used by the INCA control model? This can be done by comparing the performances of the INCA controls and ideal controls, where the difference represents the potential that has not been used by INCA

Finding the ideal control decisions can be considered an optimisation or search process. Therefore, ideal control is generated with the help of a real valued genetic algorithm and with the aim of maximising the objective function. Genetic algorithms (introduced by John Holland in the 1960s) belong to the group of evolutionary algorithms and are a widely used global search technique with the ability to explore and exploit a given operating space using available performance measures. The idea behind genetic algorithms was to evolve a population of candidate solutions to a given problem, using operators inspired by natural genetic variation and natural selection. The terminology of genetic algorithms with the corresponding classical optimisation expressions is compared in Table 3-7 (NELLES, 1999).

Since genetic algorithms are not the focus of this work, we will not explain in details the backgrounds and principle of their functioning. Instead we will give only a short overview of its use in generating the ideal control sequence in INCA. For more details on genetic algorithms see NELLES, 1999, etc.

**Table 3-7: Comparison of Terminology of Genetic Algorithm and Classical Optimisation**

Evolutionary Computation	Classical Optimisation
Individual	Parameter Vector
Population	Set of Parameter Vector
Fitness	Inverse of Loss Function Value
Fitness Function	Inverse of Loss Function
Generation	Iteration
Application of Genetic Operators	Parameter Vector Update

For generating the optimal control sequence, control decisions have to be converted into a real valued string called a chromosome, with a gene representing control in each minute. Each individual is characterised by one such chromosome. Since the controls in each minute are real values, this transformation is straightforward. However, the sequence of the control actions must fulfil several constraints given in chapter 3.6. For example: the control must last at least 3 minutes, the control must be smoothly terminated, in steps of 20 km/h, etc. These

constraints must be included when building the chromosomes in order to realistically represent the ideal allowed control. Several individuals generated in such a way make up the population. The optimisation is performed in an iterative process, where transformations called selection, crossover and mutation are performed in each iteration. In genetic algorithm literature an iteration is called a generation.

In addition to initialisation at different location of parameter space, the major power of the genetic algorithms lies in processes like mutation, recombination and selection, etc.. These are done analogously to the natural evolution. Please note that there are many variations in the implementation of genetic algorithm and the explanation given here is in accordance with the settings used for generating of INCA ideal control sequence. The fitness of each individual (sequence of control actions) is determined with the help of the objective function. When the fitness of each individual is known the selection can be performed. Selection can be performed in several ways. The roulette-wheel selection is the most frequent one and it is used in INCA. In roulette-wheel selection the strongest individual has the greatest chance to be chosen randomly, and thus to survive. Selection ensures that the population tends to evolve towards better-performing individuals, thereby solving the optimisation problem. From these randomly chosen individuals (also called parents), with the help of recombination and mutation, new individuals will be generated. In the discrete recombination (used in INCA) the offspring (new individual) inherits some parameters from one parent and the other parameters from the other parent. In mutation, genes of the individuals generated in this manner are eventually changed (mutated). The number of genes that will be mutated depends on the mutation probability. If the gene can be mutated, it will be replaced with a randomly chosen value from the set of allowed values (speed control levels in the case of INCA). It should be noted that after these transformations the general constraints upon the sequence of control actions (e.g. 3 minute control duration, etc.) should be re-checked and, if these constraints are not satisfied, the control sequence should be adjusted. The process is completed when the number of iterations reaches the maximal generation size or when the process converges.

**Table 3-8: Configurable parameters for generating the optimal control using the genetic algorithm**

Variable	Description	Default value
GAPop	Population size	20
GAGeneration	Maximum number of generations = number of iterations. Active only if the convergence rule is set to 0	50000
GAProbMut	Mutation probability	0.0001 (0,1)
GAProbCross	Crossover probability	0.8 (0,1)
GAProbRepl	Replication probability	0.7 (0,1)
GAConvergenceType	Convergence type: 0 – number of generations 1 – function convergence	1

The genetic algorithm optimisation parameters that should be specified are the population size, number of generations, probability of crossover and mutation, and finally the

convergence type. Values used for the generation of the ideal control in INCA are given in Table 3-8.

With the settings given in Table 3-8, the determination of the ideal link control sequence for one complete day (1440 minutes => 1440 genes) converges after 1000 - 1200 iterations.

### 3.8 Assessment of control performances

Assessment of the control model performances and/or impacts is an important part of model development (KELLER, 2002C). The chosen assessment approach depends on the *assessment objective* and available *assessment (test) environment*. In other words, it depends on the answer to the question: what is the final goal of our assessment and how we can achieve it? When this question is answered, it will be possible to determine the appropriate *performance indicators* with which control performances should be measured. The indicators must be able to properly capture assessment objectives and to work within the given assessment environment.

#### 3.8.1 Assessment objectives and indicators

Depending on the objective of the assessment the following classification can be made (KELLER, 2002C):

- *Technical assessment,*
- *Operational assessment,*
- *Socio-economical assessment (e.g. SCHICK, 2002; FGSV, 1997)*
- *Technology assessment (e.g. ZACKOR AND KELLER, 1999)*

Obviously, the above approaches are not independent, e.g. an economic assessment is directly linked to the operational assessment while technical and operational assessments are correlated.

In the context of this work, we are interested in the *operational assessment* of INCA performances. These can be further extended into economical assessment if necessary. Similar to other link control models it is expected that by appropriate warning and harmonising of traffic flow INCA will lead to increased safety and efficiency on the motorway. Therefore, the INCA operational assessment can be done either by measuring its direct impacts on safety and efficiency (e.g. accident rates), or by measuring its indirect impacts – that is ability to warn and harmonise (e.g. DR-FAR curves).

Another important point is that we are interested in the assessment of INCA performances under real-life conditions, i.e. INCA robustness. For example, in practice corrupted data and/or communication failures are frequently present and we are interested in testing INCA performances under such conditions too. The model that is able to better handle real-life situations will most probably perform better in real-life applications as well.

### 3.8.2 Assessment analysis

The model performances can be determined by using one (or more) of the following analyses:

- before-after analysis,
- with-without analysis,
- simulation analysis,
- empirical analysis.

The *before-after* analysis is probably the most straightforward one. It compares the situations before and after the control has been implemented via appropriate indicators (e.g. accident rates). Since the same motorway, characterised with the same dynamics, is considered, the obtained results are consistent and thus comparable. However, this analysis has to be done over longer time periods in order to produce stable (not depended on the temporal fluctuations) and statistically significant results. Moreover, it requires that the control model operate in practice for some time. Since LCS are very complex systems with strong influence on traffic participants, experimenting with the LCS controls and testing of the new system is not seen as a favourable practice by authorities and thus usually not possible.

The *with-without* analysis compares similar traffic situations, where in one case there was control and in another case there was no control. It can be done by analysing data from the same location, grouping them into with and without control classes, and then comparing the similar situations (e.g. ZACKOR, 1972; STEINHOFF ET AL., 2002). Another possibility is to compare motorways with similar characteristics, where again one has and other does not have LCS installed (e.g. SCHICK, 2002). The problem with the first alternative is that usually we have very little uninfluenced data available. More precisely, uninfluenced data is usually the one under free flow, while we are interested in dense or congested situations. Hence, the analysis of unstable situations will be possible only when the incident has not been detected or has been detected too late, which still represents only a small portion of data. Moreover, the control case might influence the following non-controlled situation and opposite. Comparing two motorways is questionable, since each motorway has its own characteristics, dynamics, and weather characteristics. Furthermore, traffic composition and commuter traffic is the point that is very difficult to determine and take into consideration, but they might influence traffic dynamics significantly. Therefore, even by very detailed filtering of the data, it is still questionable whether the obtained cases are representative and consistent. If there is no consistency in the data, a consistency in the results cannot be expected.

In the case when before-after analysis is not possible, *simulation or empirical data* analysis are the only solution. Both include some approximations and assumptions and have their pros and cons. Which one will be chosen depends on the particular control measure and available data. Simulation is probably the most frequently used way of assessing the control effects. Both approaches offer the possibility of testing controls under different traffic conditions in a (approximated) consistent manner. Furthermore, they offer wide area for experimenting with



different control settings, etc. However, the quality of the results obtained using simulation strongly depends on the simulation validity and ability to model control effects. This might be critical in the case of link control systems. Simulation of the control effects is based<sup>34</sup> on scaling the fundamental diagram by some predefined scale factor. Acceptance is usually assumed to be 100% with the exception of some tools where the acceptance factor can be calibrated. However, such calibration usually only further scales the fundamental diagram. Therefore, by using the simulation it is very difficult to determine what really the control effect is, and what is the error of the simulation model. Moreover, in order to assess LCS effects many various situations have to be covered and investigated, which implies that all these must be created and carefully calibrated.

In the context of link control, the analysis with empirical data offers probably the more consistent environment for control assessment – especially of the indirect control impacts. Since data in empirical analysis is “finished” and cannot be changed, this analysis is suitable for testing control ability to fulfil predefined goals – assessing indirect control performances. In the case of LCS it is especially suitable in testing the control ability to (detect and) warn the driver, which is also seen as the primary goal of LCS. Assuming that an incident will happen anyway, the importance of the LCS is to warn drivers about it and this can be easily tested using empirical data.

*According to the above discussion, the INCA has been “learned” and tested using empirical data. More precisely, “learning” has been done off-line using empirical data, while the performance testing has been done in on-line open-loop manner. This means that the INCA has been installed in the traffic control centre and, at each minute, control decisions are produced according to the data acquired in the real-time. In this way all situations present in practice, such as missing data or communication failure, are included in model testing as well.*

### 3.8.3 Assessment indicators

Although important, measuring direct impacts of link control is difficult (it requires huge amount of data) and sometimes it is even not possible (it is difficult to obtain the before and after, or with and without conditions - chapter 3.8.2). Fortunately, various studies evidence to the link control positive impacts if they warn and harmonise properly. Assessing the INCA capability to warn (implying also detection) and harmonise would represent the indirect measure of its operational performances. *Therefore, the primary focus of the INCA operational assessment is on measuring its **indirect performances** (success in warning and harmonisation).* Based on the above discussion, for assessing the INCA performances using empirical data the objective function presented in chapter 3.3, and DR-FAR curves are used. The objective function estimates harmonisation and warning effects and so induced costs of the control, while DR-FAR curves estimate INCA major warning capabilities.

---

<sup>34</sup> To the author’s knowledge this is the case for all existing simulation tools.

### 3.8.4 Reference cases

In addition to assessing absolute model performances, it is also of interest to compare model quality to other alternatives. In literature, these other alternatives are usually called reference cases. Comparing INCA with several reference cases will lead to better insight into situations where INCA is especially beneficial, where there is space for improvement, and ultimately if it would lead to the better utilisation of link control systems than current model (MARZ). Therefore, in context of INCA, following reference cases will be used:

- no control case,
- current system performances (MARZ), and
- ideal system performances.

The no control case can be seen as null reference (objective function for no control case is 0 sec/veh/km, while DR and FAR are both also equal to zero). Therefore, the obtained measure of INCA performance represents an absolute improvement or deterioration in comparison with no control. Comparing of INCA with MARZ is of interest in order to see if INCA have better ability to timely warn and harmonise traffic flow. Better warning and harmonisation capability imply that it will also lead to an increase in safety and efficiency and, inherently, in public awareness and acceptance. As described in chapter 3.7.7, if we assume that the objective function appropriately models link control effects, and if we know what will happen in the future, it is possible to determine the ideal control sequence (TEODOROVIC AND EDARA, 2005A) by choosing the one that maximises the objective function. The ideal control can then be seen as maximal possible benefit that could be achieved using LCS. Comparison of INCA performances with such an ideal case makes it possible to identify how much “space” is left for improvement. Furthermore, it is possible to identify other special situations – e.g. where INCA was particularly good or bad, etc.

## 3.9 Conclusion

In the previous sections, each component of the general control framework (chapter 3.2) has been analysed in detail and extended (filled) with needed components. The schematic representation of the resulting system – INCA - is given in Figure 3-21.

The crucial components of the system are the cost-benefit based objective function, flexible information fusion model, and (parameter and model) optimisation process. In each measuring interval, the raw data is checked for consistency and plausibility, and corrected if necessary. In the knowledge base, at each control interval, various algorithms and indicators are calculated using available data. The calculated data is further normalised which allows for better optimisation performances and usage of ridge regression. The information derived in this manner, together with the estimated traffic states are forwarded to the control component.

The objective function, quantifies the basic warning and harmonisation benefits against the travel time losses by referencing the incident risk probabilities (potential or real) in different

situations/controls and by considering the information propagation in upstream direction. The costs of the control are estimated as the time that would have been lost if drivers obeyed to a control. The benefits calculation is based on the calculated average accident density and acute accident densities. The control that reduces the costs is “rewarded” with the sum of the cost reduction. The objective function has several configurable parameters which can be adjusted to better suit needs in the field. Expressing effects in terms of time savings is not only more appealing to the user but make possible to utilise the measure in other, for example, cost-benefit calculations.

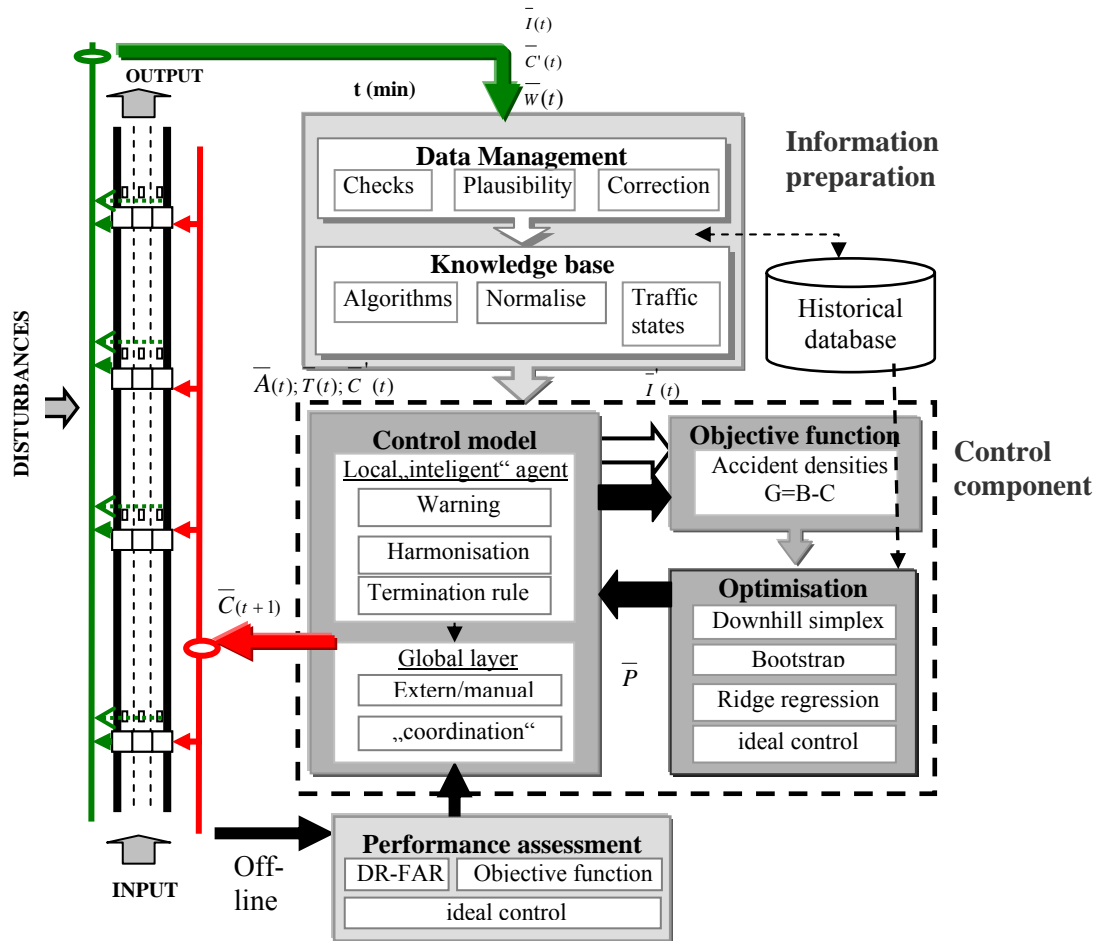


Figure 3-21: Comprehensive link control framework: INCA

The control model combines the available information in a flexible manner, which accounts for their varying reliability under different situations. The major challenge was to find the model which is understandable and tractable but still sufficiently accurate. The control decision is made in two steps: first at each control station and then coordinated along the corridor. In the first step, for each control station (VMS) a local decision is made based on data from the local and neighbouring detectors, with the help of “intelligent” controller. Due to the different motivation of the warning and harmonisation strategies they are modelled using different models. The warning decision is made using a (logit-based) regression model, while the harmonisation decision is made based on modified Belastung algorithms. A warning

model is assigned to each identified specific traffic state. The derived local controls are further modified on the global level, either due to external information, such as manual/weather control, or due to the global “coordination” rules, such as control funnelling and propagation. A novel approach to control termination is also included in the control component, which allows for the smooth switching-off of the control, and the reduction of the number of control parameters by allowing exclusion of switch-off threshold values.

The goal of the optimisation process is to acquire best possible parameter estimates using a limited amount of data and in such way to ensure model robustness and transferability. The uniqueness of the optimisation approach is that the optimisation process is supported by mechanisms for excluding ineffective and redundant information from the model, with limited training data available. The important elements of the optimisation process (in addition to parameter optimisation) are the Ridge regression and the Bootstrap technique. Ridge regression allows for the reduction of unnecessary degrees of freedom, while the Bootstrap technique provides the framework for the

- estimation of the variance,
- model selection, and
- aggregation,

with a limited amount of the available training data. The search for the optimal parameter vector is performed using the downhill simplex optimisation method. As a part of the optimisation component also the ideal control is generated, in order to allow for the testing of the objective function, as well as comparison of INCA performances with the ideal one and identifying the places for improvement. The described steps make up an integrated optimisation framework for the calibration of the new control model, applicable in practice.

The developed model is practical and generic in the sense that it can be applied on any motorway route equipped with a traffic data collection system, given a historical database of measurements for calibration and validation is available.

## 4 Technical implementation

To optimise, test, and evaluate the above described INCA control framework, the control software INCAS has been developed. By using INCAS, any typical link control installation in Germany could (according to INCA model) be **controlled** and **monitored on-line** in the traffic control centre, or *off-line* in the “simulated” mode. By using graphical user interface (GUI), the operator has the possibility to see prevailing traffic conditions, resulting control decisions and control performances in any moment, as well as algorithms that are responsible for control activation. Supported by previously described components, INCAS could be also used for off-line **optimisation** and **evaluation** of the control model. An additional communication component TRANSMIT has been developed to achieve on-line operation at traffic control

centre. TRANSMIT is able to connect to a typical sub-control centre in Bavaria and obtain raw traffic data in real-time. TRANSMIT forwards the data to INCAS via the TCP/IP communication protocol.

At present the INCA control framework, as well as INCAS software, are unique in Germany, since other control approaches (and consequently software) do not offer the possibility of systematic optimisation and evaluation of link control model. INCAS is highly configurable, where initialisation and configuration are performed via ASCII files. The used object-oriented programming approach makes further extension or improvement of each component fairly simple. With supported mechanisms and software design, INCAS could be easily transferred to a new motorway, where adjustment of the model parameters to a new motorway is performed in a systematic manner, with the help of the optimisation procedure.

#### 4.1 INCA control software (INCAS)

INCAS is developed in Java and requires at least 520MB RAM and a Pentium IV processor. A brief overview of the main software features and components is given here. A detail description of the software can be found in VUKANOVIC ET AL., 2003B. Components of the INCAS software are shown in different colours in Figure 4-1, and can be roughly grouped into:

- “intelligence” where previously described algorithms, models, and processes are implemented (white block)
- interfaces that allow communication with “external” world (in yellow)
- data-structure where raw and calculated data is stored in the internal software memory (light blue block)
- event manager which, depending on the inputs from outside, determines the proper sequence of actions that should be taken (dark blue block)

Mechanisms implemented in the “intelligence” part of INCAS could be used for:

- determination and *visualisation* of appropriate link control actions,
- *optimisation* of the control model and algorithms, and
- *evaluation* of the control performances.

Determination of the control actions could be done either on-line using TRANSMIT, or off-line by reading data from ASCII files and “simulating” control actions. The control decision could be made by utilising or not utilising traffic states, and it is made according to the specification given in the initialisation files. At every moment the prevailing traffic conditions, calculated by INCA and MARZ control actions, algorithm activations, and performance of the INCA control model are displayed via the GUI. In addition, the user has the option of seeing the historical propagation of the each value shown on the GUI.

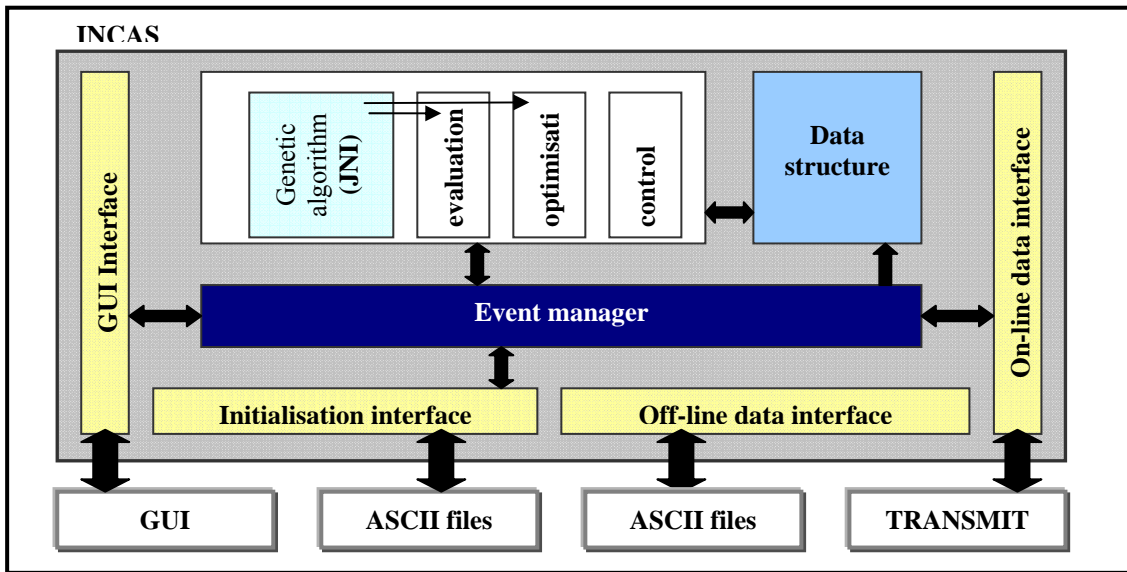


Figure 4-1: INCA control software (INCAS)

The optimisation process integrates all of the above described components: downhill simplex, Bootstrap, Ridge regression and generating of ideal control. The user can choose whether the Bootstrap and Ridge regression are to be used in the optimisation process. The MIT GALib<sup>35</sup> library (WALL, 1996) is used to determine an ideal control. GALib is a C++ library allowing for the programming of problem-specific optimisation using any representation and genetic operators. With the help of GALib the specific coding and above-described genetic operators (chapter 3.7.7) were implemented; the objective function for the fitness calculation was integrated, and the resulting program was compiled as a dynamic link library (DLL). The DLL is integrated into the Java-based software INCAS with the help of Java Native Interface (JNI).

The evaluation provides a measure of performance in terms of objective function and DR-FAR curves (according to HOOPS ET AL., 2000). It can be used for the evaluation of INCA, MARZ and the ideal control. Both optimisation and evaluation are performed off-line and their results are saved in ASCII files.

**Interfaces** have crucial role in *communication* with the external world and are responsible for the appropriate INCA *initialisation* and *configuration*. INCAS has four distinct interfaces:

- Initialisation and configuration interface, is the interface with the ASCII files where controlled motorway and corresponding link control installation, INCA control parameters, as well as optimisation and evaluation parameters are defined.
- On-line data interface, implements the TCP/IP protocol and serves for on-line acquisition of raw data. To this end, different telegrams have been defined, each corresponding to a specific type of raw data: detector data, detector station data, weather data, or control data. It allows for the separation of the control model (INCA)

<sup>35</sup> GALib: Genetic algorithm library - developed by the Massachusetts Institute of Technology in the year 1995

from the field specific communication implementation (e.g. TLS (BAST, 2002)). Communication with TRANSMIT is realised via this interface.

- GUI interface, displays/visualises information to the user.
- Off-line data interface, reads/stores historical data. It is used in off-line operation, either for optimisation, evaluation, or off-line control replication.

## 4.2 INCAS GUI

The major challenge of GUI design is finding the right balance between the information that will be displayed. This is especially critical for link control software, since the control depends on many factors ranging from traffic situation, weather related or/and manual control, to global effects of the control. In order to show sufficient information and still remain tractable, INCAS GUI has been divided into three blocks (Figure 4-2):

- motorway and the belonging link control installation,
- algorithms and warning component values, and
- the objective function.

In addition, a status line is located at the bottom, where messages about the current calculations and eventual errors are displayed.

On GUI, for each moment  $t$ , the following information is displayed:

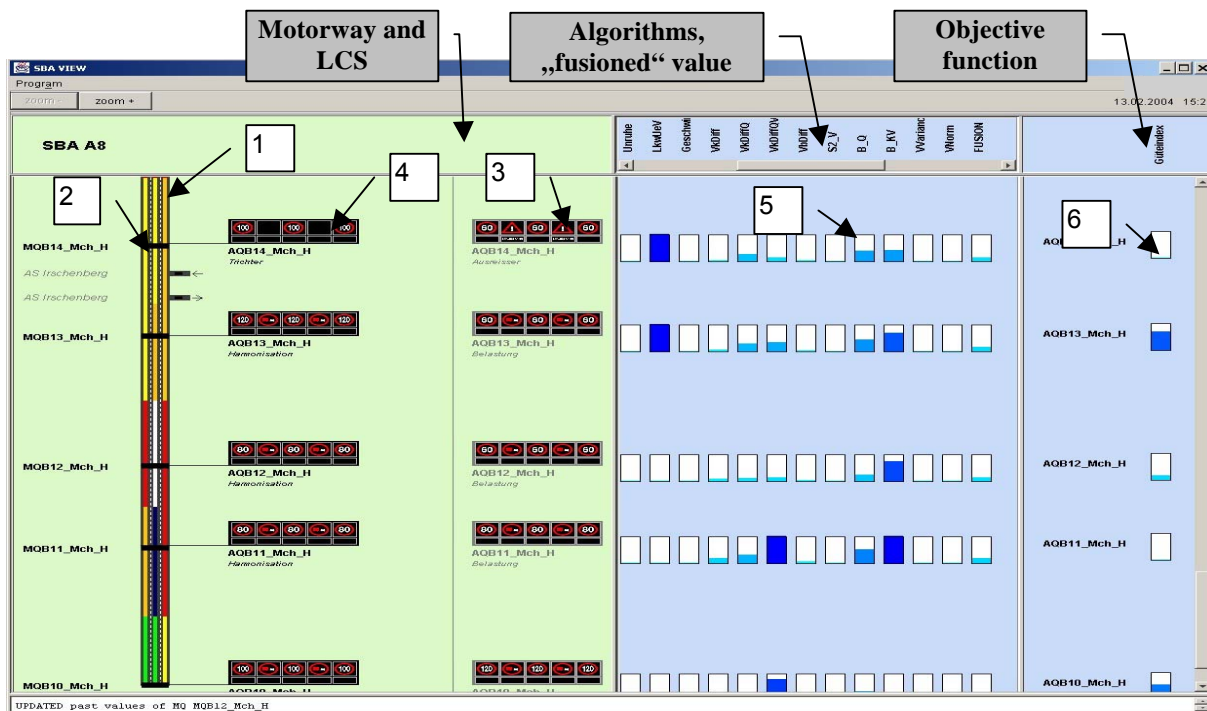
1. Traffic state for each lane within a section,
2. Data from traffic detectors for each lane and aggregated over the lanes,
3. MARZ controls and activation causes,
4. INCA controls and activation causes,
5. Calculated algorithms and warning module values,
6. Value of objective function for INCA controls

On the left side of the screen, motorway and corresponding link control installation are shown. Number of lanes, on- and off-ramps, and positions of VMS and detectors are shown at the distance proportional to the situation in the field. They are *automatically generated* from the configuration file. This implies that for any new link control installation only the configuration file needs to be changed, and the INCAS GUI will be generated in accordance with the new motorway.

Each lane on the motorway is displayed separately and in the colour that represents the corresponding LOS: green colour for free flow, orange and yellow for synchronised, and red for congested. Additionally, blue colour represents cases of implausible data, while white represents cases of missing data.

The black horizontal rectangle on motorway represents the location of VMS and the corresponding detector. Parallel to each black block, two VMS (with A, B, and C panels) are

shown. At first to the left VMS panel the INCA control decision is shown; at second panel the currently active control, calculated by MARZ, is shown. The reason why the control has been triggered is displayed beneath each VMS panel. Under MARZ control the displayed reason corresponds to the MARZ specification (BAST, 1999) and is delivered by the sub-control centre. The reason displayed beneath INCA control consists of information about the local decision (harmonisation and/or warning), and eventual external or global modifications. For example, if at one VMS, harmonisation module produces the control decision, but it is overwritten by the manual work zone control, the message “harmonisation/road works” will be displayed beneath the sign.



**Figure 4-2: INCAS GUI**

In the middle of the screen, outputs from all algorithms from the knowledge-base and output from the warning module are shown. Parallel to each VMS, a series of rectangles is shown, each corresponding to a specific algorithm. If the rectangle is white it implies that the algorithm has not been triggered. With the activation of the algorithm the rectangle will be filled, proportionally to the maximum possible output from the algorithm. All algorithms are shown in the following manner: the higher the algorithm output activation the more severe situation according to the algorithm. To be easily observable, with higher algorithm activations the colour with which the rectangle is filled changes from light to dark blue. At the end of the sequence, the output from the warning module is shown in the same manner.

On the right-hand side of the screen, the calculated INCA control performances (objective function) at each VMS are shown. It is also represented as rectangles and follows the same logic as the algorithm block: the higher the value of the objective function, the bluer the rectangular and better the performance of the control at the location. In addition to monitoring



the current calculations, the user has the possibility to view the historical profiles of all above-mentioned values (Figure 4-3).



Figure 4-3: Historical profiles of different measured and calculated values

### 4.3 On-line operation in Munich motorway control centre

In order to test INCA on-line performances INCAS and TRANSMIT were installed at the Munich traffic control centre (Figure 4-4). In Germany the data collected on the motorway is sent to the sub-control centre at each measuring interval. The standard way to transfer data to the traffic control centre is by way of TLS telegrams. However, in Bavaria, the sub-control centres are usually additionally supported by the SIMENS ObjectManager component which makes acquisition of raw data easier. Each computer in the traffic control centre supported by the “copy” of the ObjectManager can access data at a specific sub-control centre. The communication component, called TRANSMIT was developed for the purpose of acquiring traffic and control data on-line (in real time) via ObjectManager. It contains the ObjectManager ocx component, and is connected to the sub-control centre via instance of ObjectManager. INCAS connects to TRANSMIT as a client and further exchange of data follows the TCP/IP communication protocol. Each time new data is received, the ObjectManager, and consequently TRANSMIT, are notified. TRANSMIT transforms the data into appropriate telegrams and sends it to all registered clients (in this case INCAS). TRANSMIT archives the received data into appropriate ASCII files, which are used by INCAS during its off-line operation. Hence, the following components have to be installed at the control centre: *INCAS, TRANSMIT, and ObjectManager instance*.

Since one sub-control centre can store data from several motorways, TRANSMIT must register for the data of interest. This is done using the TRANSMIT configuration file, where controlled motorway and types of data that should be collected (e.g. detector, weather, etc.) are specified.

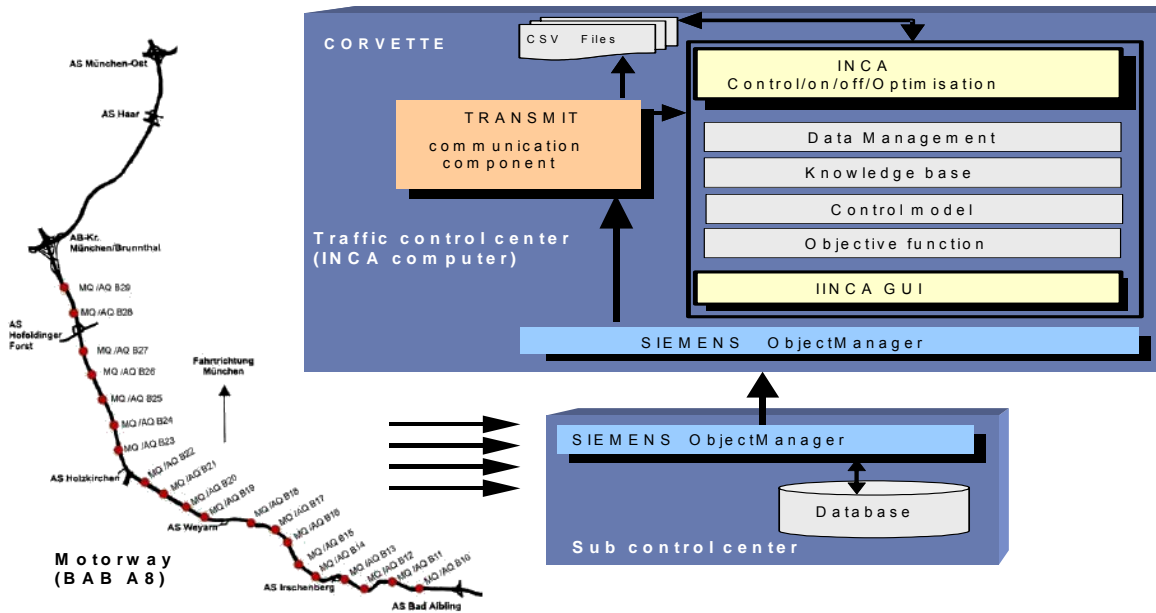


Figure 4-4: INCAS and its communication with the motorway link control system

## 5 Optimisation and evaluation results

INCAS software based on INCA was developed and installed in Munich traffic management centre to test and evaluate INCA performances. Data needed for system calibration was collected during the first two-month period (from 10 July to 10 September, 2004). This data is hereinafter referred to as *training data*.

Testing any new control and optimisations procedure is a complex task, particularly since the new approach is designed to capture several different performance aspects. Therefore, first the ability of the *objective function* to appropriately estimate link control effects is investigated in detail. After it was proven that the function appropriately models the most important link control effects, and taking into account *test site and data characteristics*, the system was optimised. The *optimisation process* is performed in two steps: first, the harmonisation parameters are optimised taking into consideration only *harmonisation* benefits (without the warning part); the resulting parameters are “frozen”, and the *warning* parameters are then optimised taking into account the total benefit.

INCA calibration was done and afterwards it was run in an on-line open-loop mode for 2 months (from 15 September to 25 November 2004). On-line open-loop means that INCA was running in the traffic management centre and produced controls in real time (with data that was available on-line), but did not send controls back to the VMS panel on the road. This is important since the effects of missing or implausible data are automatically reflected in the control performance. This data is used for evaluation and is called *test data*. Controls produced by INCA are *evaluated* using the objective function and receiver operating characteristics obtained according to DR-FAR. INCA performance is compared to the one

from the current system (MARZ) and ideal control. Finally, to illustrate INCA *flexibility* and ability to automatically encounter for the new available information, six new indicators produced by (traffic flow model based) AZTEK algorithm are introduced into INCA system. INCA is re-calibrated with this new available information, the new model is generated and its performances were evaluated and compared with performances from previous INCA model.

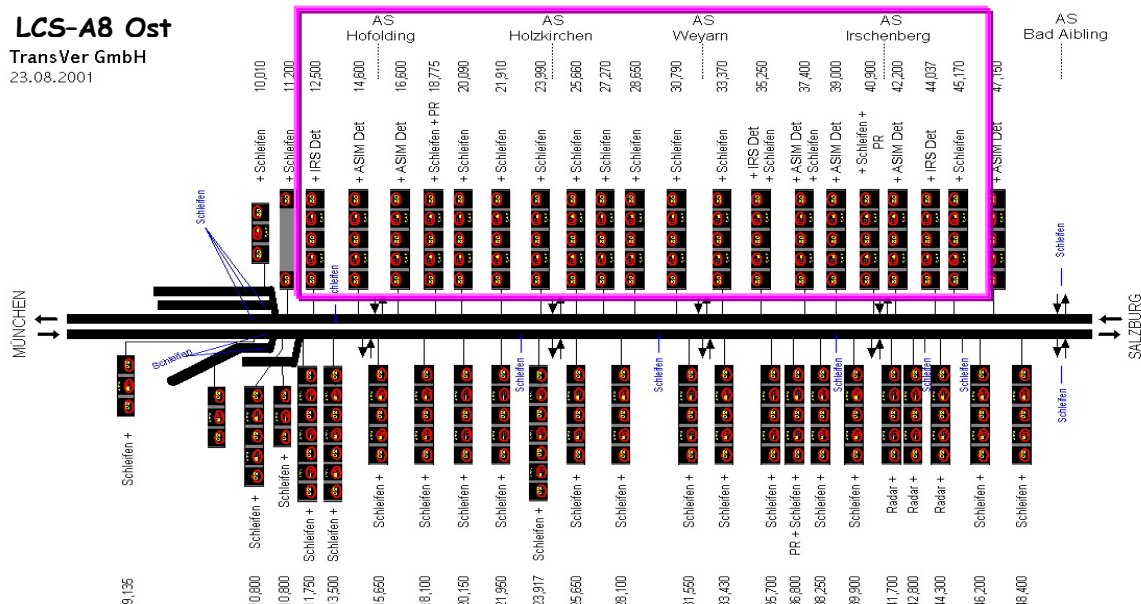


Figure 5-1: Test site: link control installation at BAB A8 Ost, direction Munich (DENAES AND KATES, 2002)

### 5.1 Test site

As a test area a 38 km-long section of A8 motorway Salzburg direction Munich, between AS Bad Aibling and AK München-Süd, (Figure 5-1) was analysed. The section contains 20 VMS panels and 20 detectors, and has 4 on- and off- ramps. The section contains a segment (“Irschenberg”) that is critical due to its profile (see Figure 5-2): This segment has a steep grade, a large excess of incidents, and represents the location where the current control performs poorly. The average daily traffic on the motorway is around 40,000 vehicles. On Fridays and during weekends demand reaches up to 49,000, while on the working days it varies between 33,000 and 39,000 vehicles<sup>36</sup>.

<sup>36</sup> For a more detailed analysis of traffic flow characteristics with respect to traffic demand and truck percentage see DINKEL, 2004 or STEINHOFF, 2002.

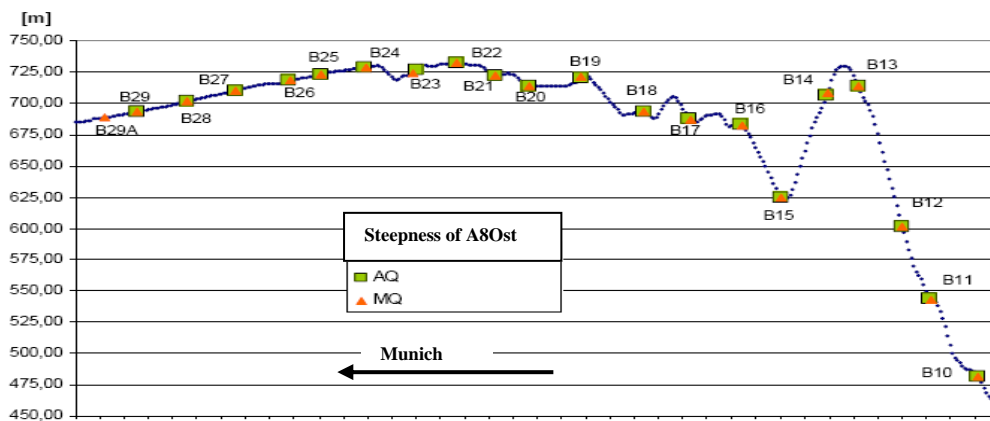


Figure 5-2: Geometric profile of the motorway (STEINHOFF, 2002)

## 5.2 Training and test data characteristics

Data (and consequently control stations) to be used for optimisation was selected based on data-quality analysis. Also, days with significantly corrupted data in the test-data set can be identified and used for studying system robustness. To be able to appropriately estimate optimisation quality and also to interpret the evaluation results, the traffic characteristics at the test site, are considered in terms of number of speed drops, special events, and weather conditions during training and test phase. According to the available weather data there has not been significant difference in weather conditions between the training and test period. The results of the weather analysis are given in Appendix A.4.

### 5.2.1 Data quality

In Figure 5-3 and Figure 5-4 the percentage of corrupted data during the training and test period, for each detector station is shown. These percentages include missing and implausible data as well as data with wrong time stamps, while systematic errors are not analysed (see chapter 3.4.1). Detectors with significantly low data quality are additionally marked.

From the below graphs it can be seen that:

- detector MQB28 was out of operation during the entire (training and test) period,
- detector MQB13 had serious problems with the data quality, especially in the second half of the training period (from 1.8.2004) and first half of the test period (until 16.10.2004),
- detectors MQB25 and MQB20 deliver data sets with a high percentage of corrupted data during the entire training period. Where there was an improvement in data quality during the second half of the test period,
- detector MQB11 sporadically had very poor data.
- during the test period, between 17.9.2004 and 24.9.2004, detectors MQB25, MQB26 and MQB27 sent 0s during the entire day.



Figure 5-3: Data quality during the training period

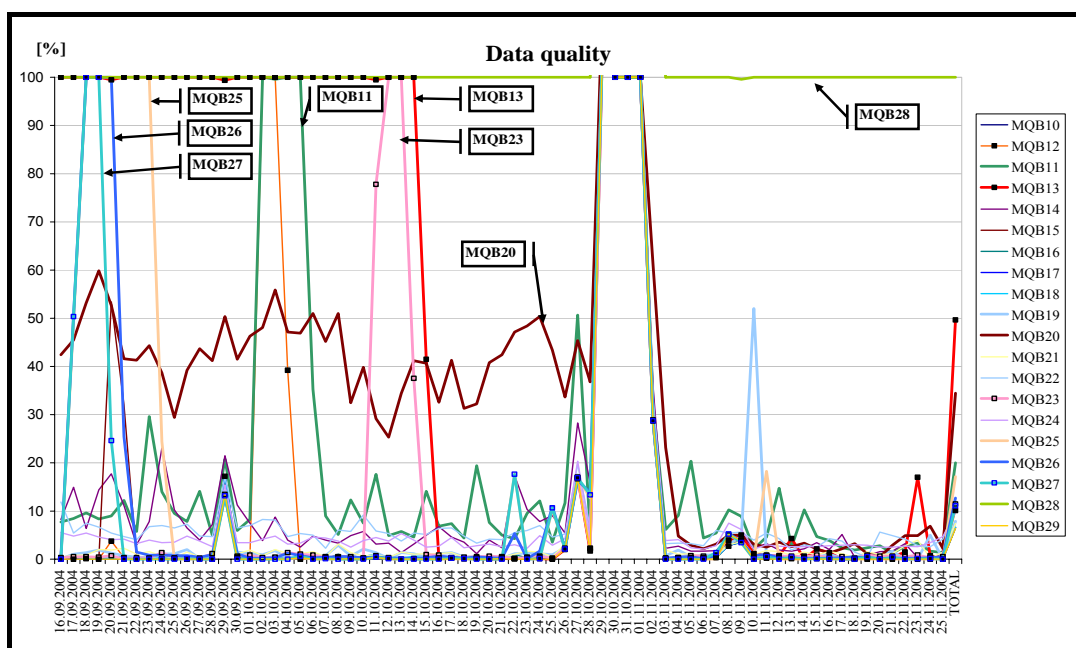


Figure 5-4: Data quality during the test period

Here it should be noted that neither of the above mentioned hardware failures was reported by the data collection system itself. In other words, according to the TLS specification, hardware or basic plausibility failure should be reported by the data collection system itself by setting the data attribute “status” to a value other than 0 (BAST, 2002) (see chapter 2.2.1). However, this was not the case in the investigated system.

### 5.2.2 Number of speed drops

In Figure 5-5 and Figure 5-6 the number of speed drops during the training and test period, for each control station is given.

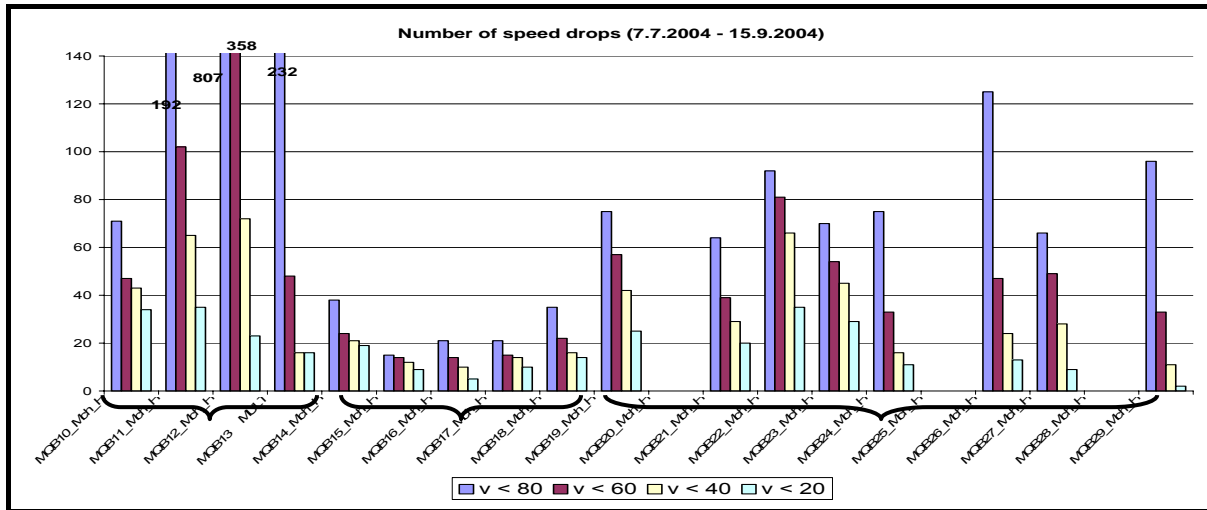


Figure 5-5: Distribution of speed drops along the motorway during the training period

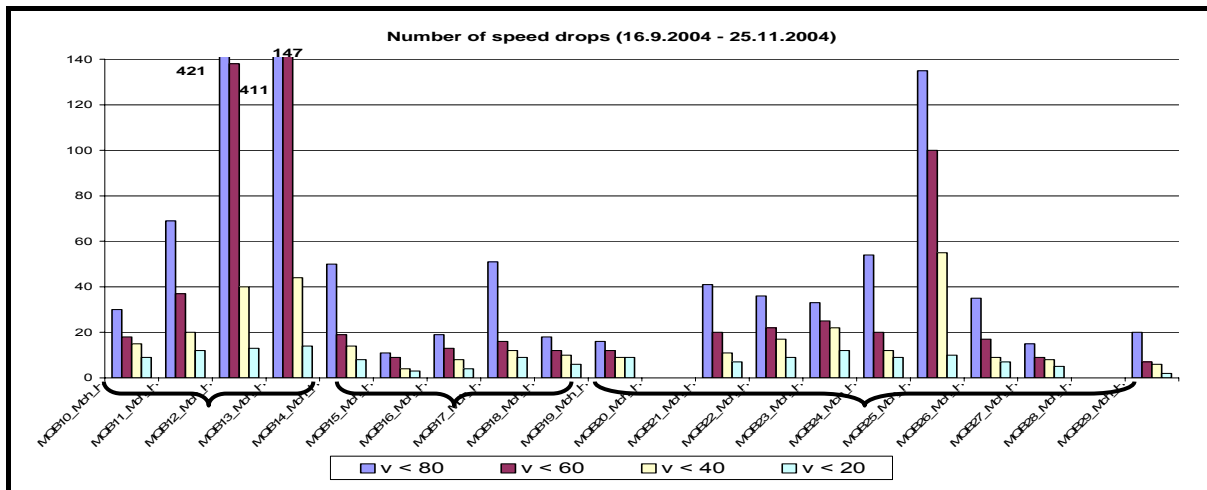


Figure 5-6: Distribution of speed drops along the motorway during the test period

By analysing the figures below **three different characteristic motorway (control) areas** can be identified:

1. *Critical Irschenberg area*: Due to the steep grade it is characterised by a special kind of incidents (many speed drops < 80 and 60 km/h). Detectors MQB10-MQB13 belong to this group.
2. *Calm area*: at km 40, after the Irschenberg on-ramp and before the AS Weyarn on-ramp. Detectors MQB14-MQB18 belong to this group.
3. *Unstable area*: between the AS Weyarn on-ramp and the AK M-Süd motorway junction (relatively many speed drops). All other detectors (MQB19-MQB29) belong to this group.

Instabilities could be observed for MQB14 and MQB19. At MQB19 significantly more speed drops are measured during the training period (when it belongs more to the third detector group) than in the test period (when it belongs more to the second detector group), while at the MQB14 it is opposite. However, due to other similar characteristics and previous research from DENAES AND KATES, 2002 the above classification is considered to be valid.

The number of speed drops during training period is significantly higher (~2 times) than in the test set. This is probably due to summer holidays when traffic on A8 Ost is very dense and subject to frequent traffic breakdowns. A particularly different pattern, reflected in a smaller number of speed drops during the test period, can be observed in the case of detectors in the 3<sup>rd</sup> control group. Another interesting pattern can be observed in the test set by detector pairs MQB16->MQB17 and MQB24->MQB25 where downstream detectors (MQB17, MQB25), even though belonging to the same group, have a much higher number of speed drops of levels 80 km/h and 60 km/h.

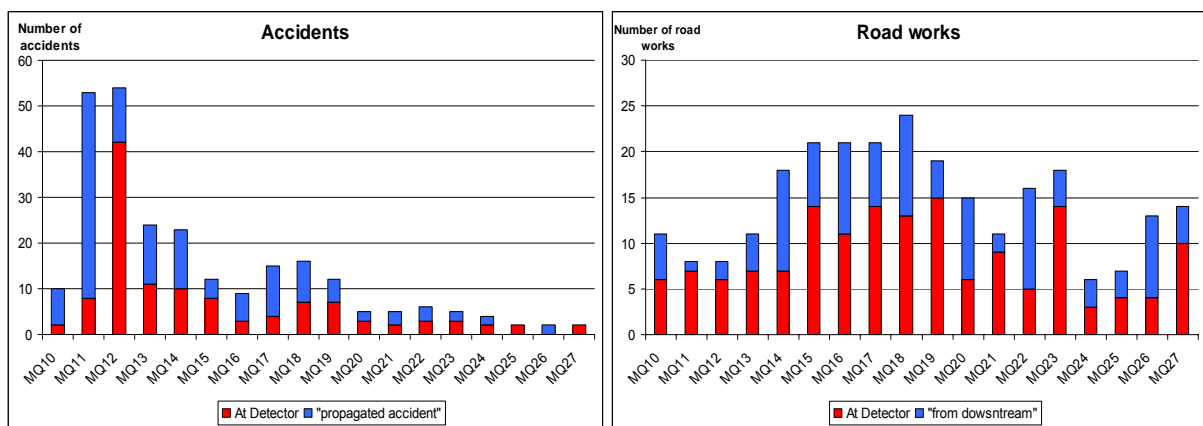


Figure 5-7: Number of reported accidents (left) and work zones (right) per detector

### 5.2.3 Accidents and work zones

Information about accidents and work zones was extracted from the protocols stored at the traffic management centre. Detailed positions and occurring of the accidents and work zones, per day and detector, are provided in Appendix A.4. The total number of manual warnings, for the period between July, 2004 and November 11, 2004 and each detector station, is shown in Figure 5-7. In the figure the number of accidents that have been reported and are located exactly downstream of one specific detector is shown in red. The sum of the red blocks represents the total number of accidents and work zones on the road in analysed period. For each control station (VMS) the values shown in blue represent the accident and work zones “propagation” activations. This means that accident/road work happened further downstream, beyond the next control station, but influenced traffic all the way upstream to the corresponding VMS. The sum of the red and blue represents the total number of manual warnings at one control location.

Generally, the number of accidents corresponds to the number of speed drops of a lower level (Figure 5-5 and Figure 5-6). An extremely high number of accidents was reported in the

“Irschenberg” area, where the segment between MQB12 and MQB13 is especially critical, with 42 reported accidents. Although most of these accidents are of minor character they may significantly influence traffic conditions. It is believed that the “Irschenberg” area represents the segment that could significantly benefit from improved link control. On the other hand, during the same period, the majority of the work zones was performed in the 2<sup>nd</sup> control area.

#### 5.2.4 Optimisation and evaluation – data and approach used

The above investigation affected several important system calibration and evaluation process decisions. The 20% data quality rule influenced the used optimisation approach. The 20% rule means that only the days with less than 20% of corrupted data (more than 80% of correct data) are used for optimisation. Depending on the total number of days with satisfactory data quality, detectors and corresponding control stations (VMS) are classified into the following groups:

1. Data quality is *unsatisfactory*: due to the high percentage of corrupted data; successful system optimisation cannot be achieved. Detectors MQB28, MQB20 and MQB25 belong to this group and they will not be optimised. In the test period, corresponding control stations will be controlled using the model from the first upstream control station.
2. Data quality is *moderate*: control could be optimised but only by re-sampling the available data (Bootstrap). Detectors MQB13 and MQB11 belong to this group.
3. Data quality is *good*: there is enough data available for successful system optimisation. Even though these control stations (VMS) could also be calibrated without additional re-sampling, it is always recommended to use Bootstrap since it will lead to more stable control model. All other detectors belong to this group.

Data that is known to represent low flow situations (night data from 22:00h to 06:00h), which could only mislead optimisation procedure, will not be used in optimisation process. Although the night data is not used for optimisation, INCA controls these periods also.

Since the optimisation process is supported with re-sampling mechanism the *optimisation* will be carried out *separately* for each control station. Generally, it should be expected that areas with more critical events (speed drops) would allow for better optimisation and therefore should have better performances in on-line implementation (e.g. the first group).

*Traffic characteristics in the training period (data used in INCA optimisation) are noticeably different from the one in the test period.* Whether and how these differences influence on-line control performances will be reviewed in the evaluation process. The control stations assigned to the same control group, although with different dynamics, still show similar behaviour and could be evaluated together. Therefore, INCA performances and influence of specific parameter settings will be analysed based on classification of the motorway into control groups (chapter 5.2.2). Furthermore, performance of control stations belonging to the same



group can be compared by taking into account the different settings used for the generation (optimisation) of the control model.

For the purposes of the evaluation procedure, specific days have been identified according to the traffic, weather, or data quality conditions:

- To investigate INCA robustness with respect to the data quality, empirical data for the days with questionable data quality (e.g. September 23 or November 11, etc) will be specially analysed and plotted.
- To investigate and compare INCA and current system detection/reaction performances in case of accidents, detailed speed and control historical plots will be made for some of the following days October 7, October 9, November 3, November 4.
- To investigate INCA behaviour under bad weather conditions, some of the following days September 23, October 1, October 9, November 3, November 8 will be analysed.

## 5.3 Objective function evaluation

### 5.3.1 Introduction

The objective function models the complex, and sometimes conflicting, LCS goals. Therefore, it is important to prove whether the function properly estimates the most important control effects and thus, whether it is appropriate to use it for optimisation and evaluation in the first place.

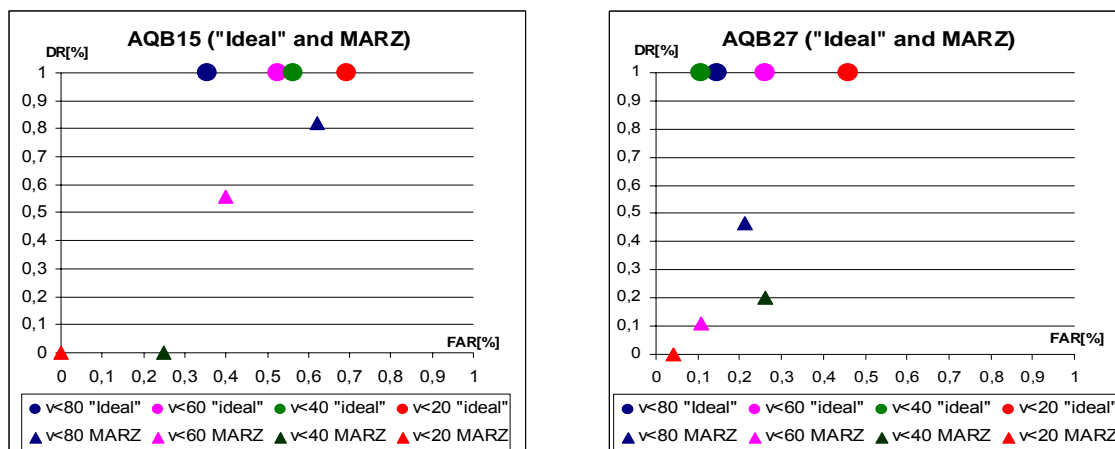
To this end, first the *ideal control* is generated (chapter 3.7.7) using objective function default parameter settings (chapter 3.3.8), and evaluated. If the objective function properly estimates control effects, a high performant ideal control should be produced. Since only a few operational measures of link control performances can be found in literature, the ideal control produced here is *evaluated* using the detection versus false alarm rate curve (DR-FAR) described in Appendix A.2. Even though the DR-FAR measure does not capture all of the LCS effects, it is a helpful indicator for LCS warning performances, which are at the same time the ones bringing the major safety benefits. However, it should be noted that, due to the harmonisation effects of the control, incident propagation in the upstream direction and other factors, the DR-FAR measure might overestimate the number of false alarms and/or underestimate the number of detections. Some cases of the specific ideal control situation and its benefits and losses calculated in each minute are given as *examples* illustrating how objective function quantifies link control performance.

### 5.3.2 Evaluation of the ideal control

The process of investigation of the ideal control performances (and inherently objective function validity) is shown using the examples of VMS AQ15 and AQ27. In Figure 5-8 quality of ideal and current (MARZ) control on these two VMS panels, in DR-FAR form, is shown. Although it is not fair to compare an off-line produced optimised control with the

current not-optimised on-line control, this comparison, together with their benefits and costs estimated by objective function (Figure 5-9), provides better insight into the way the objective function models LCS effects.

Only the points of greatest interest are shown in Figure 5-8. For example, the ability of control 80 km/h to timely detect such speed drops is of major interest for the speed drop of 80 km/h. Therefore, the blue circle corresponds to the quality of control 80 km/h with respect to the speed drops of 80 km/h. The same applies for other speed drops: for speed drop of 60 km/h control of 60 km/h is of the greatest interest, etc. It should be noted that the corresponding control for speed drops below 20 km/h (red circle) is the same as for a drop of 40 km/h, which is the “congestion” sign. Therefore, possible higher values of false alarms in cases of a speed drop of 20 km/h and control “congestion” can be expected.

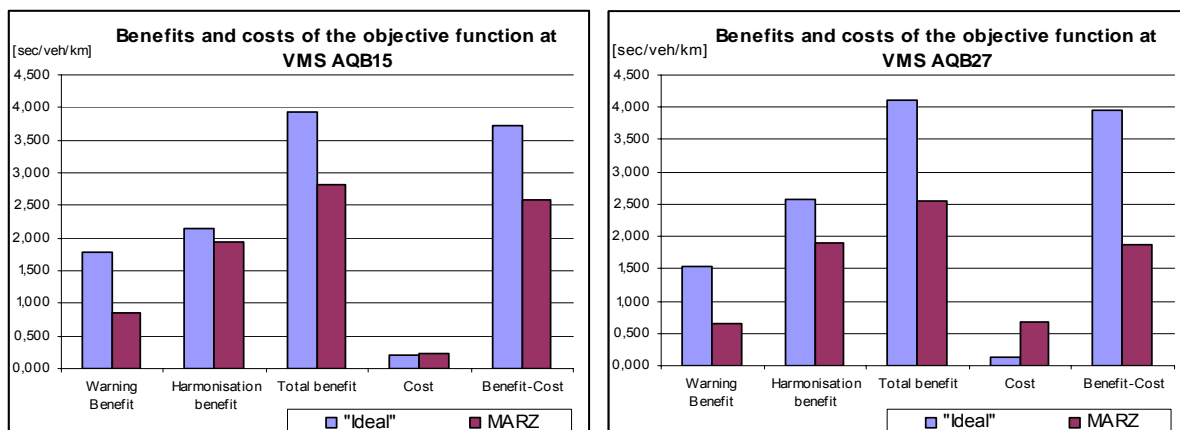


**Figure 5-8: Performances of ideal and current (MARZ) control for VMS AQB15 and AQB27, evaluated using DR-FAR curves**

In the above figures it can be observed that objective function optimises the system (here ideal control) so as to detect as many speed drops as possible, with the lowest possible number of false alarms. It can be considered that the objective function *appropriately* estimates the LCS *warning* effects. Slightly higher FAR at VMS AQB15 does not necessarily mean that the control was inappropriate. This section is characterised by a significant percentage of downstream incidents which do not propagate all the way upstream to the AQB15. In such situations the warning control at AQB15 will be activated, even though no incident has been observed locally. Although such control is considered desirable, DR-FAR curves will regard such situations as false alarms. Furthermore, it is also possible that the measured speed is very close to the some speed limit, but not lower. For example, in a situation when measured speed is around 65km/h, control 60 km/h will usually induce more benefits (reduce disturbances) than travel time losses. However with the DR-FAR method this situation is classified as a false alarm.

The benefits and costs of the ideal and MARZ controls are displayed in Figure 5-9. Several interesting conclusions can be made by observing Figure 5-9. Firstly, the accumulated total harmonisation benefit is higher than the total warning benefit. This is due to the fact that

incidents are rare events, and the corresponding warning controls are seldom activated. Therefore, even though a single warning brings much greater benefit than a single harmonisation, over a longer period of time the cumulative effects of harmonisation controls are greater. Secondly, the existing control might affect the relation between traffic flow and speed (fundamental diagram) in the harmonisation area. This is visible at VMS AQB15, where the speed limit 120 km/h is almost always active. Since the ideal control at VMS AQB15 is generated using this, already influenced data (modified fundamental diagram), the harmonisation benefits of the ideal and the MARZ control are quite similar. The total average benefit is calculated under the assumption that the warning and the harmonisation benefits are of equal importance. The average costs, i.e. travel time loss produced by the control, are small in comparison with the benefits. It is also apparent that the ideal control at AQB15 produces greater costs than the existing control, but also significantly increases the total benefit. Furthermore, by observing the DR-FAR curves and values of the objective function for MARZ control at VMS AQB15, it could be concluded that the current control is activated either very seldom, or too late (when the incident is already present at the location) and, therefore, it produces very low benefit - which is in accordance with field observations.



**Figure 5-9: Quality of the ideal and the current (MARZ) controls measured using the objective function**

The difference between total benefits and costs represents the final measure of the control performance. Using the objective function, the generated ideal control induces the total benefit of 3.7 sec/veh/km and 4 sec/veh/km for AQB15 and AQB27, respectively. For the 1,5 km long section of VMS AQB15, where in investigated period there were 2,461,878 vehicles, total harmonisation and safety benefit (saved lives) corresponds to 3,975 saved hours (~ 5 months). At VMS AQB27, for a 2 km long section and 2,833,999 registered vehicles, 6,298 hours are saved (~8,5 months). If objective function captures all known SBA effects, which is not the case, this value could be seen as maximum possible benefit of the SBA at the specific control station.

### 5.3.3 Empirical examples

To illustrate how the objective function quantifies control quality for each “ideally” controlled minute estimated benefits and costs of ideal control are shown in Figure 5-10.

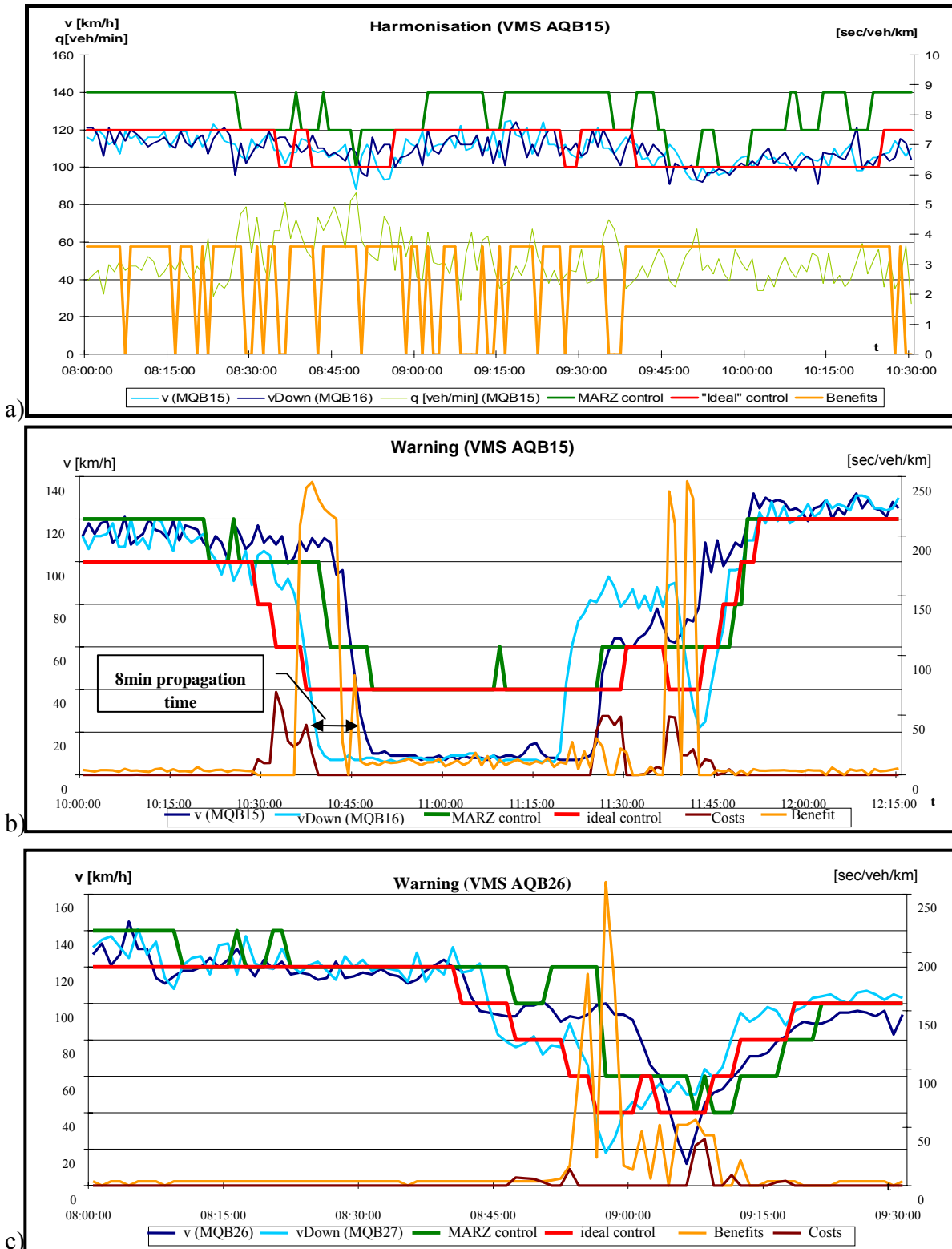


Figure 5-10: Estimated benefits and costs of ideal control. a) harmonisation case for VMS AQB15; b) warning case for VMS AQB15; c) warning case for VMS AQB26

In the below figures it can be observed that in peaks the harmonisation benefits are considerably lesser than the warning ones. It is assumed that warning strategies could prevent severe events from happening, thus saving lives and producing greater benefits. Harmonisation strategies reduce probability of less severe accidents and traffic breakdowns that are more frequent, but not as expensive as fatal accidents. Furthermore, harmonisation strategies usually only slightly slow down incoming traffic, and therefore induce very small costs. Due to the 3-minute-control-termination-rule, the harmonisation controls are applied only when the traffic situation is unstable for a longer period of time. Slightly more frequent changes between controls 100 km/h and 120 km/h can be observed in unstable dense traffic, where incidents are more likely to happen.

Figures illustrating warning related controls (Figure 5-10, b and c) show that the ideal control is optimised to timely warn the drivers of downstream incidents, even though traffic is still “flowing” at the local detector station. In an acute area, the warning strategies induce higher travel time losses but, at the same time, very great safety benefits, which is considered to be appropriate. Figure 5-10 (c), illustrates that the warning benefits increase with the increase of the difference between the driven and lowest speed in the section. Therefore, the potential benefits of the “congestion” control in the situation just prior to 09:00h are much greater than in the situation after, when traffic is already slow. In situations when congestion is already present at the location, the warning still induces benefits, although considerably smaller than in an acute situation (Figure 5-10 (b)). It should be noted that if the situation before the incident, with slightly lower but still high downstream speeds, lasts longer, it can happen that the function underestimates the positive effects of the control. This would designate only slight benefits to the control but high costs, although control has a preventive effect and, therefore, more benefits should be calculated (e.g. in Figure 5-10 (c) control of 80 km/h around 8:50h).

### 5.3.4 Summary

The above discussion supports the assumption that the objective function “learns” the system appropriate behaviour. The resulting quality cannot be completely assessed with DR-FAR curves since the function also quantifies other LCS effects, for example, the one at the section in front. However, DR-FAR curves are significant and they clearly confirm that the optimal control leads to very effective warning controls, where as many situations as possible are detected with the least possible false alarms. According to the empirical examples, where benefits and costs of the ideal control are plotted for each controlled minute, the objective function adequately “rewards” and “punishes” controls. This is especially visible in areas where traffic is still “flowing” at a local station, but an incident has already occurred at a location downstream. Furthermore, it is shown that due to the 3-minute-control-termination-rule, the harmonisation control will be activated only when the traffic situation worsens for a longer period of time. The space for further improvements of the objective function estimation exists in certain areas before an incident, where local and downstream speeds are

lower, but still above the “congested” speed for a longer period of time. If such situations are prolonged, the objective function will underestimate the control benefits. However, the control will be appropriately activated, mainly due to the 3-minute-control-termination-rule.

## 5.4 Control model optimisation

### 5.4.1 Optimisation approach and settings

The optimisation process is performed in two steps: first, the parameters of harmonisation are optimised taking into account only the *harmonisation* benefits (without the warning part); the resulting harmonisation parameters are then “frozen”, and the *warning* parameters are then optimised taking into account the total benefits. Optimisation can be carried out for each VMS separately, or several subsequent VMS panels can be optimised together. Separate optimisation for each VMS would result in parameters better corresponding to the section specific dynamics, but this requires a larger amount of training data. In the context of this work, with two months of traffic data available for optimisation and supported by re-sampling mechanisms, it was decided to perform the optimisations for each control station separately.

The results of the optimisation depend on the chosen initial (warning and harmonisation) control model, as well as on the optimisation settings. For the initial harmonisation control model the 5- or 7- parameter version of the Belastung can be chosen. The initial warning control model depends on the decision whether the classification into traffic states is considered (three-state approach) or not (one-state approach).

The results of this optimisation depend on the configuration (parameters) of the objective function (chapter 3.3.8). For each strategy there are specific configurable objective function parameters. The parameter common in both strategies is the “weight” or “importance” of a particular strategy, which is by default set to be 0.5 (equal harmonisation and warning importance). However, to make system more sensitive to the harmonisation or warning goals, it makes sense to increase this value to 0.7 or, even 1, when optimising the specific control component.

Finally, the optimisation can be performed by using of Bootstrap or Ridge regression. If Bootstrap is used, the smallest number of subsamples should not be less than 5. If the Ridge regression is used, the parameters “lambda” and “tolerance” (lowest tolerable parameter “weight”) should be specified. Information with “weight” lower than value of “tolerance” will be automatically excluded from the warning model.

A summary of the parameters that influence the final optimisation results is given in Table 5-1 (harmonisation) and Table 5-2 (warning). The influences of different optimisation settings on the optimisation results are analysed:

- choice of the initial harmonisation and warning model,
- influence of different parameters of the objective function,
- influence of the Bootstrap approach,

- influence of Ridge regression and its parameters,
- optimisation convergence.

**Table 5-1: Harmonisation optimisation parameters**

Num	Name	Possible values	Default values
1	Harmonisation model (number of Belastung parameters)	5 or 7	7
2	Bootstrap	True or False	True
	Number of subsamples (B)	1-n	5
3-7	Parameters of objective function		
	$1 - \alpha$	(0-1)	1
	$V_{range}$	0-120	20
	$V_{bias}$	0-20	5
	$V_{H\ min}$	10-160	60
	$Q_{min}$	0-2200	600

**Table 5-2: Warning optimisation parameters**

Num	Name	Possible values	Default values
1	Warning model (division into traffic states)	True or False	True
2	Bootstrap	True or False	True
	Number of subsamples (B)	1-n	5
3-4	Ridge Regression	True or False	True
	$\mu$	0-1	0.2
	$\beta_{min}$	0-0.3	0.005
5-8	Parameters of the objective function		
	$\alpha$	(0-1)	0.5
	$\gamma$	0-10	4
	$V_N$	0-120 (km/h)	60
	$V_I$	0-80 (km/h)	40

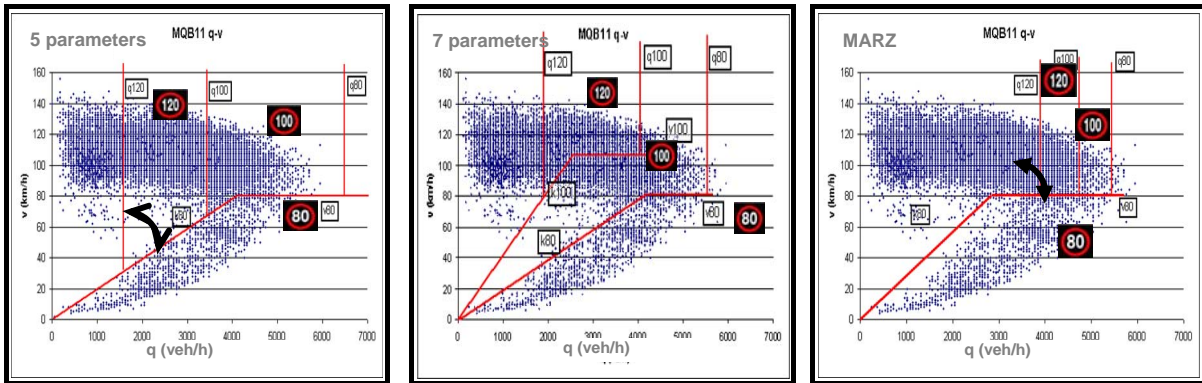
#### 5.4.2 Choice of the harmonisation model

The difference between parameters produced with 5- and 7-parameter approaches, and their comparison with the parameters currently used (MARZ) is illustrated in Figure 5-11, on the example of VMS AQB11<sup>37</sup>. For a better overview, in Table 5-3 a list of parameter values is given.

In the 7-parameter approach the  $Q_{bon}(100)$  value is greater than in 5-parameter approach, since the area with lower flows and speeds is covered by the additionally introduced density and speed threshold values. The difference between the current and INCA 120 km/h control borders is due to the  $Q_{min}$  parameter value of objective function (as defined in chapter

<sup>37</sup> The VMS AQB11 is one of the rare control stations in current system where Belastung threshold values are not set too high and might lead to proactive harmonisation controls.

3.3.6.3). This parameter represents the flow from which harmonisation controls could bring some benefit, and thus it directly influences the optimised  $Q_{bon}$  (120) threshold value.



**Figure 5-11: Optimised harmonisation (Belastung) parameters for VMS AQB11 (corresponding detector is MQB11). Left: 5 parameter model. Middle: 7 parameter model. Right: currently used (MARZ) parameters.**

**Table 5-3: Harmonisation parameters at VMS AQB11 (detector MQB11)**

Model	q120On	k100On	q100On	v100On	k80On	q80On	v80On
MARZ	3900	-1	4700	-1	35	5500	80
INCA (5 Parameter)	1666	-1	3600	-1	47	6600	80
INCA (7 Parameter)	1855	24	4005	108	50	5576	81

It should be noted that optimised harmonisation parameters might depend on the current control, because the harmonisation model is based on the fundamental diagram which might be influenced by the existing control. Several reports (ZACKOR AND SCHWENZER, 1998; STEINHOFF ET AL., 2001) have shown that if the harmonisation control were to differ too much from the measured or downstream speed, an undesired disturbances should be expected, e.g. an increase of standard deviation. In such situations, the harmonisation control would most probably have a negative rather than a positive effect. This knowledge is incorporated into the objective function, and therefore the harmonisation strategies attempt to produce a control that is preventive, but at the same time close to the prevailing (measured) traffic conditions. Therefore, if at one control station a harmonisation control is almost always active (e.g. controls of 100 km/h and 120 km/h in the “Irschenberg” area), this will influence the corresponding fundamental diagram and, consequently, the optimised harmonisation parameters.

### 5.4.3 Choice of warning model

The goal of the warning optimisation process is to find the optimal “weights” of the information used and optimal decision boundaries corresponding to these weights. A list of available algorithms and variable names is given in chapter *Nomenclature*. The warning model for each control station integrates the above information, acquired from local or downstream or both locations (detector station).



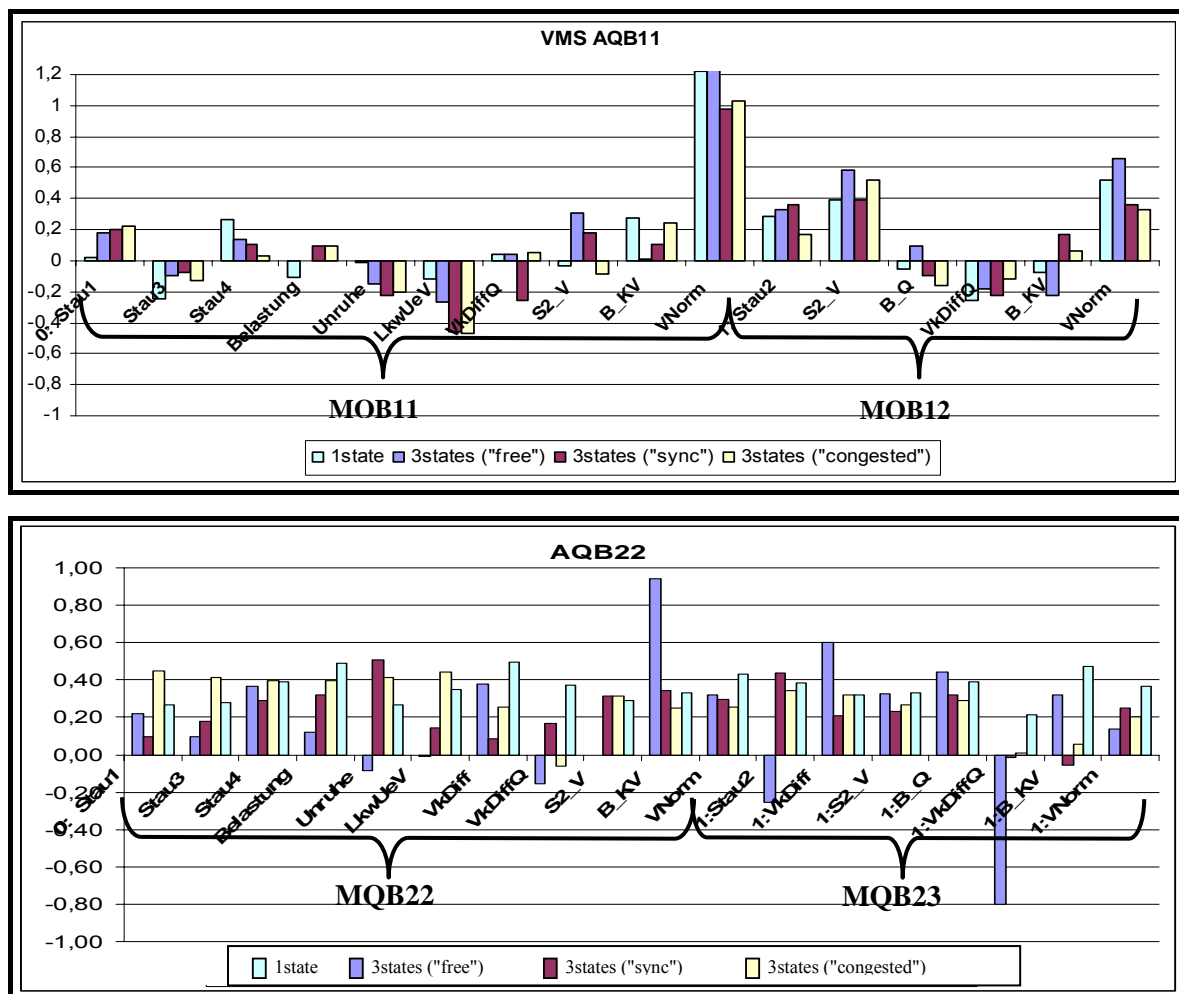


Figure 5-12: Summary of  $\beta$ -coefficients (Y-axis) in model with one and with three states. Top: at VMS AQB11, bottom: at VMS AQB22

Examples of the differences between three- and one-state model approach (with separate models for each traffic state and only one warning model) are given in the above Figure 5-12, on the examples of VMS AQB11 and AQB22, respectively. Interesting observations about traffic dynamics, algorithm performances and warning control can be made by comparing the values produced with one and three state approaches. Here the characteristics of the warning control at VMS AQB11 are analysed (an analogous procedure can be used for other control stations as well):

- The most stable and important algorithms for final warning decision at VMS AQB11, regardless of the chosen approach, turned out to be the vNorm algorithm. This is probably due to the “stable” congestion pattern in the “Irschenberg” area where speed decrease is stable indicator for congestion. At AQB22 however speed is more unstable and therefore vNorm algorithm does not have such importance.

- A lower “weight” of the vNorm algorithm from the downstream detector in synchronised and dense traffic state implies that the traffic dynamic is such that disturbances formed at AQB12 propagate upstream very fast, along a 1 km long section, up to detector AQB11.
- MARZ algorithms Stau3, Stau4 from the local detector, and Stau2 from the downstream detector, have stable and significant influence on the final warning decision. Furthermore, Stau4 and Stau2 have greater significance for the decision under free and synchronised state than under the congested one. This speaks in favour of the Stau4 good detection performances. The significance of the Stau2 algorithm supports the previous statement about fast incident propagation in the upstream direction (due to the vNorm downstream “weights”). On the other hand, the Stau3 and Stau1 algorithms are more important for the decision under the congested state.
- The B\_KV algorithm is very important for the control decision in a congested state, but it has almost no importance for the decision made in free flow state. Its great significance in the one-state approach implies that its exclusion from the decision under a free flow state is not due to its false alarms, but rather due to the fact that it does not deliver any relevant information in this state (the same can be said for the above mentioned Stau2 and Stau4 algorithms). The instability of the S2\_V algorithm from the local detector – different “weights” in the three-state approach and low “weight” in the one-state approach - implies that its less reliable performance in the congested state leads to more false alarms than could be compensated through successful detections under free flow and synchronised state.
- Algorithms that have the “weights” with different signs (positive or negative) in the one- and three-state approaches are correlated with other algorithms (e.g. Belastung and Unruhe).

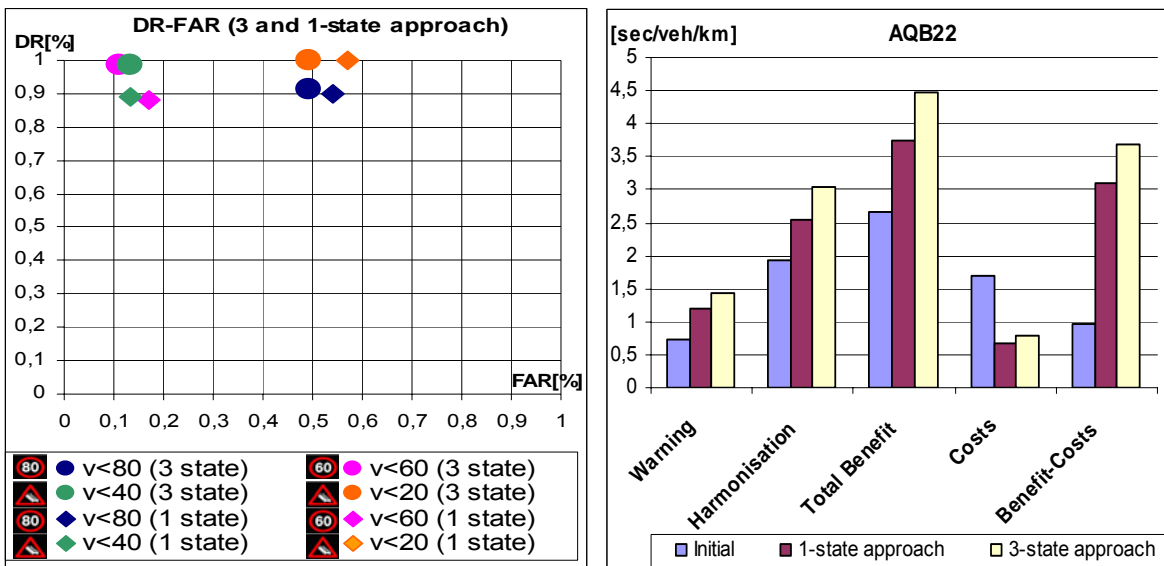


Figure 5-13: Performances of the control produced with one- and three-state approach at AQB22 evaluated with DR-FAR curves (left) and objective function (right).

The performances of controls generated with one- and three-state approaches are evaluated with DR-FAR curves and objective function (Figure 5-13), using AQB22 as the example. Points of interest for the one-state approach are shown as a rectangle in the DR-FAR diagram, while for the three-state approach they are shown as a circle. In the DR-FAR diagrams it can be observed that the three-state approach provides better detection of the speed drops by low or no increase in the number of false alarms. In some cases it learns the system as to have fewer false alarms. It can be seen that the three-state approach leads to controls with higher warning and harmonisation benefits (at least under optimisation set), and at the same time slightly higher costs. Half of the improvement in three-state-approach is due to better warning capabilities of the three-state model.

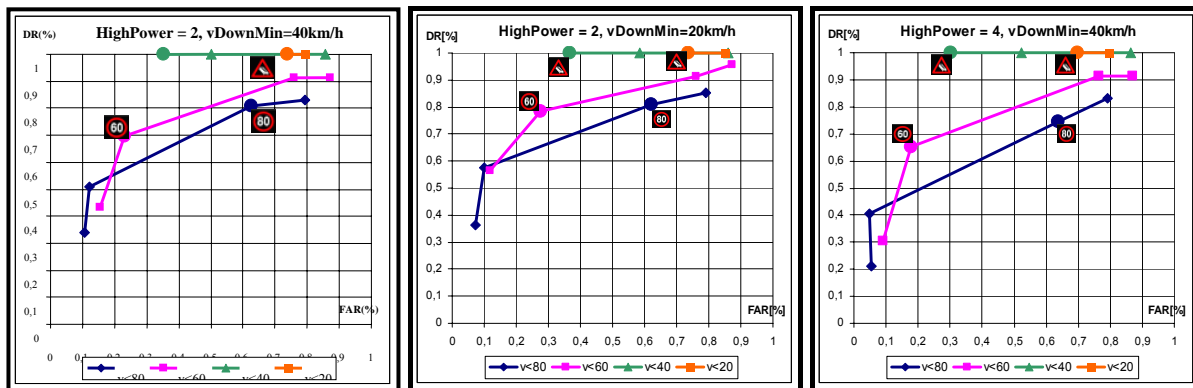
#### 5.4.4 Influence of the parameters of the objective function

The influence of different parameter settings of the objective function on optimisation results is given in Table 5-4, on example of VMS AQB11. The influence of the “weight” ( $1 - \alpha$ ) parameter (which describes the “importance” of harmonisation strategy) can be observed by comparing examples number 4, 6 and 7. As expected, with the increase in the harmonisation “weight” the threshold values decrease, thus leading to more restrictive harmonisation control. The influence of the  $Q_{\min}$  parameter can be seen by comparing examples 1, 2, 3 and 4. Since it determines the traffic volume under which positive effects of harmonisation control can be expected, it directly influences the threshold value  $Q_{\text{bon}}(120)$ . Therefore, the  $Q_{\text{bon}}(120)$  threshold value increases with the increase in the value of  $Q_{\min}$ , while others remain more or less unchanged. The influence of changes of  $V_{\text{range}}$  and  $V_{\text{bias}}$  can be observed by comparing examples number 4 and 5. Generally, with the reduction of their values the control becomes less tolerant to its deviation from the driven speeds and, therefore, less restrictive (the control is activated only when driven speeds are lower for a longer period of time). However, the influence of these parameters is not so straightforward and easily observable as in case of the previous parameters.

**Table 5-4: Different objective function parameter settings and corresponding “optimal” harmonisation parameters (example of VMS AQB11)**

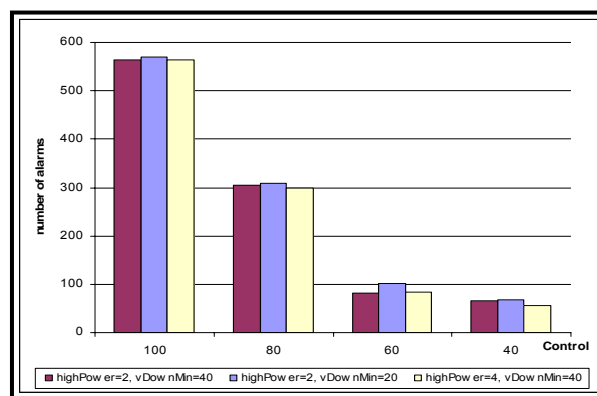
	Harmonisation „weight“ (1 - $\alpha$ )	Function param.			Optimised Belastung parameters							
		$Q_{min}$	$V_{range}$	$V_{bias}$	$Q_{bon}(120)$	$k_{on}(100)$	$Q_{bon}(100)$	$VcaF_{on}(100)$	$k_{on}(80)$	$Q_{bon}(80)$	$VcaF_{on}(80)$	
1	1	300	20	5	902	17	4509	113	44	5530	82	
2	1	800	20	5	1456	26	4500	113	27	5659	81	
3	1	1000	20	5	2324	32	4501	110	34	5508	81	
4	1	600	20	5	1855	24	4005	108	50	5576	81	
5	1	600	10	3	1600	1	4532	117	41	6276	81	
6	0.8	600	20	5	1893	21	4505	112	49	5526	81	
7	0.2	600	20	5	2307	22	4501	106	51	5509	76	

Figure 5-14 shows DR-FAR curves for warning controls produced with different values of the objective function parameters  $\gamma$  and  $V_I$  at AQB18. The number of activations of controls 100 km/h, 80 km/h, 60 km/h and sign “congestion” is given in Figure 5-15.



**Figure 5-14: Influence of objective function parameters on the optimisation of the warning module (DR-FAR). “HighPower” corresponds to parameter  $\mu$  ; “vDownMin” to  $V_I$**

Lowering the  $V_I$  parameter implies that positive effects of warning control will be calculated even for slower traffic. Therefore, the number of control activations will increase with the decrease in  $V_I$ . Consequently, if possible, the number of successful detections of the speed level 60 km/h and lower is increased, too. The increase in the  $\gamma$  parameter slightly reduces the detection rate and partly reduces the false alarm rates, since it somewhat lowers the surprise factor, and inherently the warning benefit.



**Figure 5-15: Number of control activations under different optimisation scenarios**

### 5.4.5 Influence of the Bootstrap

Influence of the Bootstrap on the optimisation of the harmonisation and warning parameters is illustrated on example of VMS AQB13 and AQB22, respectively.

AQB13 had low data-quality during the training period (see chapter 5.2.1). On this section the fixed control of 100 km/h is active. It significantly influences the fundamental diagram where measured speed hardly ever exceeds 110 km/h. Consequently, it should be expected that at this location control of 120 km/h would be needed very rarely or never. The 7-parameter model is used for harmonisation decision and optimised. Optimisation is done by creating 10 Bootstrap replications (subsamples) and finding the optimal parameters for each of them. The results of the harmonisation optimisation are summarised in Table 5-5 and Figure 5-16.

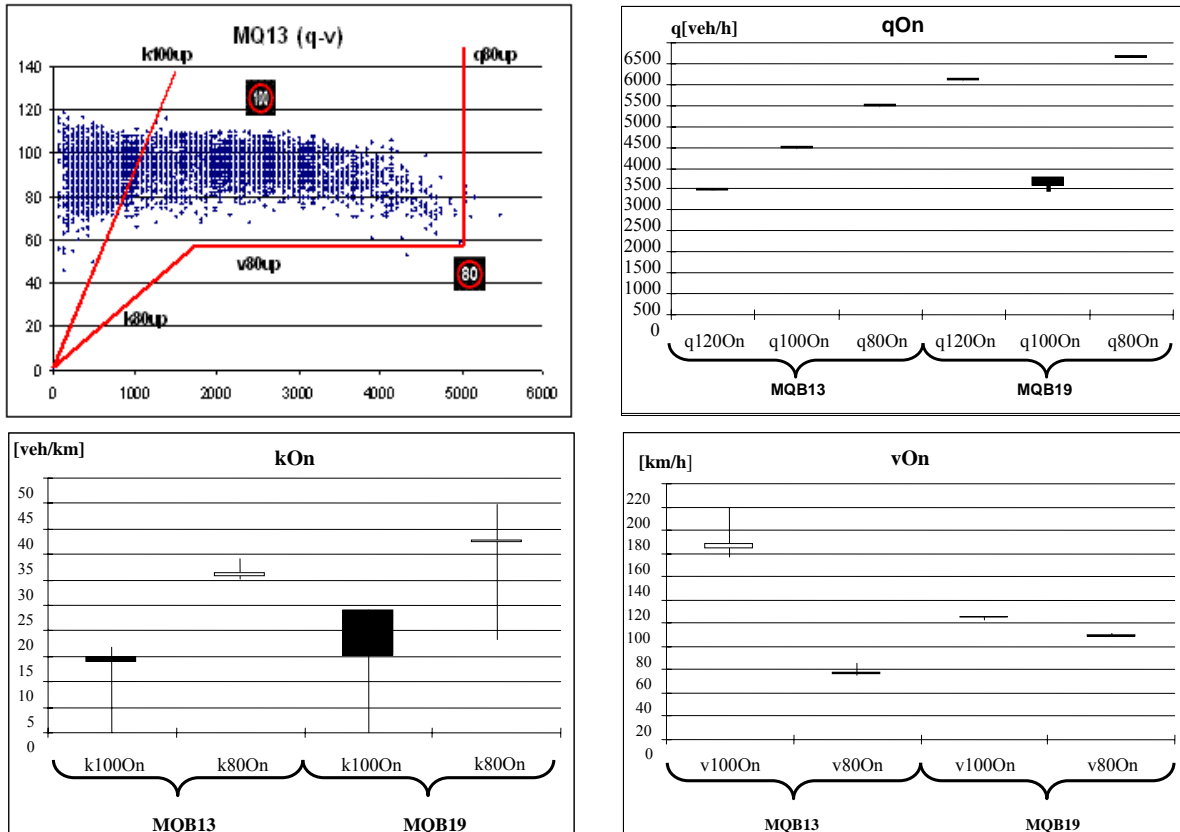
**Table 5-5: Harmonisation parameters for AQB13 produced using the Bootstrap approach (10 Bootstrap-samples): values for each Bootstrap, final aggregated values and standard deviation of parameters. The initial training data set is marked as 0<sup>th</sup> Bootstrap.**

Bootstrap iteration	harmonisation		"weight"=1				
NUM	QUp120	kUp100	qUp100	vUp100	kUp80	qUp80	vUp80
0	2980	15	3999	164	31	5009	56
1	2987	-1	4000	194	34	5004	55
2	2992	13	3999	159	31	5008	57
3	2998	15	3999	157	30	5008	57
4	2976	13	4000	163	31	5009	57
5	2992	-1	4000	199	34	5005	55
6	2979	14	4000	163	31	5009	56
7	2986	14	3999	160	31	5008	56
8	2979	6	4000	161	32	5008	57
9	2978	17	3999	169	30	5001	66
10	2974	17	4000	163	31	5009	57
<b>Average value</b>							
	2984	14	4000	168	31	5007	57
<b>Standard deviation</b>							
	7,76	3,13	0,52	14,29	1,32	2,63	3,03

Small standard deviation means that for different, randomly chosen situations (days), the control model with the same (or similar) threshold values will lead to the best decision. Hence, the smaller the standard deviation of the optimised parameters the more stable the decision point is. In the case of VMS AQB13, all relevant parameters show high stability. At first glance, parameter  $Vcar_{f_{on}}(100)$  seems to be an exception since it has higher deviation. However, this does not negatively influence the control stability, since in each case  $Vcar_{f_{on}}(100)$  is optimised to be much higher than the speed measured: in each situation the measured speed is lower than the  $Vcar_{f_{on}}(100)$  value, and therefore, the decision upon 100 km/h control basically depends on the  $k_{on}(100)$  value (Figure 5-16, top left). Since the  $Q_{b_{on}}(100)$  value is also in this area (redundant), it is not shown in Figure 5-16.

Figure 5-16 shows threshold values produced without Bootstrap, as well as maximum, minimum and average threshold values produced with Bootstrap method, on the examples of AQB13 and AQB19. The vertical line connects the maximum and the minimum value produced during the Bootstrap process. The rectangle marks the area between the value produced without Bootstrap and aggregated one produced with Bootstrap. If the value

produced without the Bootstrap is higher than the final Bootstrap value the rectangle is black, otherwise it is white. The optimised threshold values are stable at both control stations. This is especially the case for the speed and flow threshold values. A small exception is the density threshold value for the 100 km/h control ( $k_{on}(100)$ ) at VMS AQB19. The higher stability of threshold values for control at VMS AQB13 is due to the existing 100 km/h control, which stabilises the fundamental diagram and, thus the harmonisation parameters.



**Figure 5-16: Results of optimisation using Bootstrap. Top left: optimal harmonisation parameters at VMS AQB13 (Detector MQB13) produced using Bootstrap. Other: minimal, maximal and average parameter values produced with Bootstrap and values produced without Bootstrap.**

The influence of the Bootstrap on the optimisation of the warning parameters and their stability is shown in Figure 5-17 and Figure 5-18, on the example of the VMS AQB22. In Figure 5-17 the estimated variance of each warning parameter and average variance value that is 0,016 (red line) are shown. Parameters produced without Bootstrap as well as the minimal, maximal and average value produced with Bootstrap, are shown in Figure 5-18 analogous to the example of the harmonisation model given in Figure 5-16.

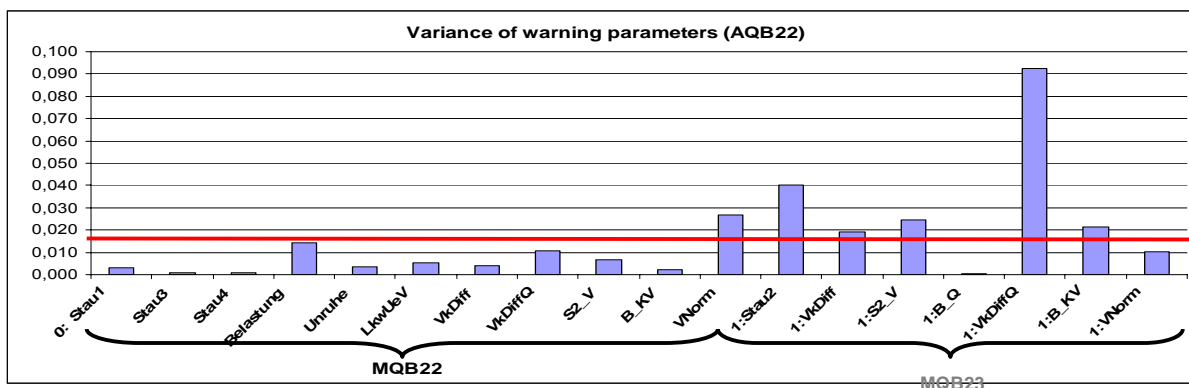


Figure 5-17: Variance of the warning parameters („weights”) at VMS AQB22. estimated using Bootstrap

The long vertical lines in Figure 5-18, e.g. the ones next to the local Belastung and downstream Stau2, S2\_V, vkDiffQ, and vkDiff algorithms, point out the instabilities of algorithm “weights” during the optimisation process, corresponding to the higher variance given in Figure 5-17. However, the small rectangle next to algorithms Stau2, S2\_V and Belastung implies that their variations during the optimisation process should not lead to the instabilities of the produced control model. Contrary, the larger rectangle of the vkDiffQ and vkDiff indicates that the unstable results of the optimisation might be due to the correlation or unreliability of this information and, therefore their exclusion from the model should be considered.

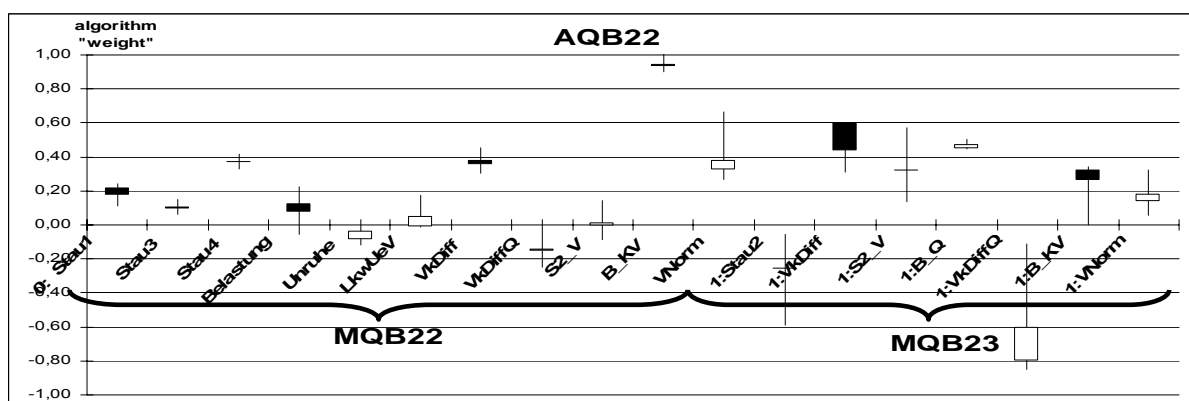


Figure 5-18: Maximal, minimal, and average warning parameter values produced with Bootstrap and values produced without Bootstrap (at example of VMS AQB22).

### 5.4.6 Influence of Ridge regression

By using the Ridge regression strongly correlated information should be automatically eliminated from the model, where criticality of parameter correlation is specified by the parameter  $\mu$  (the higher the value the more critical the correlation). For the sake of simplicity, the influence of Ridge regression and parameter  $\mu$  on reduction in the number of parameters and their “weights” is illustrated on example of the one-state approach for VMS AQB10 (Figure 5-19). Figure 5-19 shows that the value of  $\mu=0.7$  leads to the overall smaller “weight” of the algorithms and ultimately to elimination of correlated information, such as

Stau3 and vkDiffQ from the local detector station, and B\_Q and B\_KV from the downstream location.

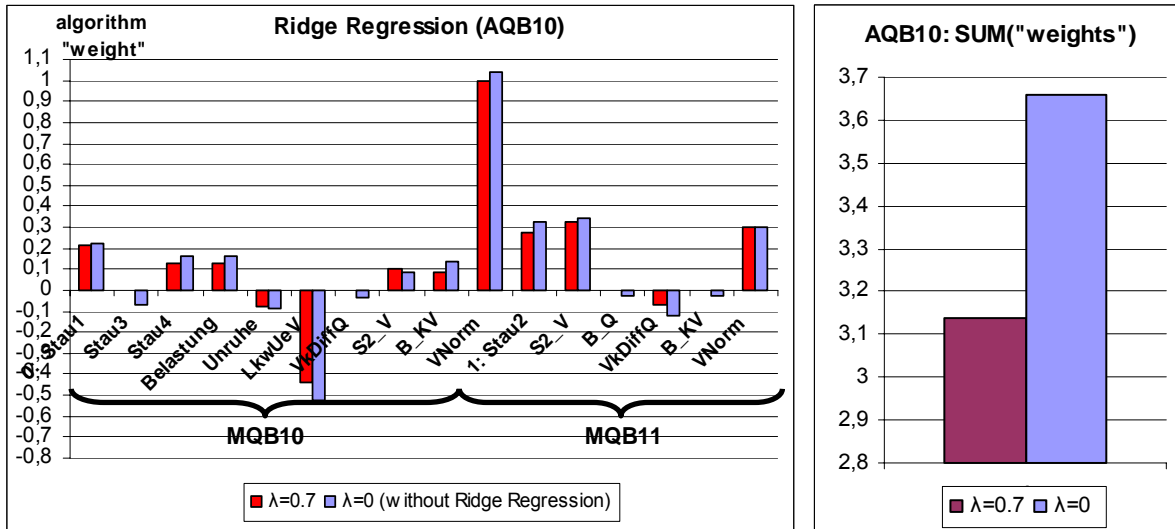


Figure 5-19: Influence of Ridge Regression on the warning model by AQB10. Left: optimised algorithm “weights”. Right: sum of all absolute parameter values with and without Ridge Regression.

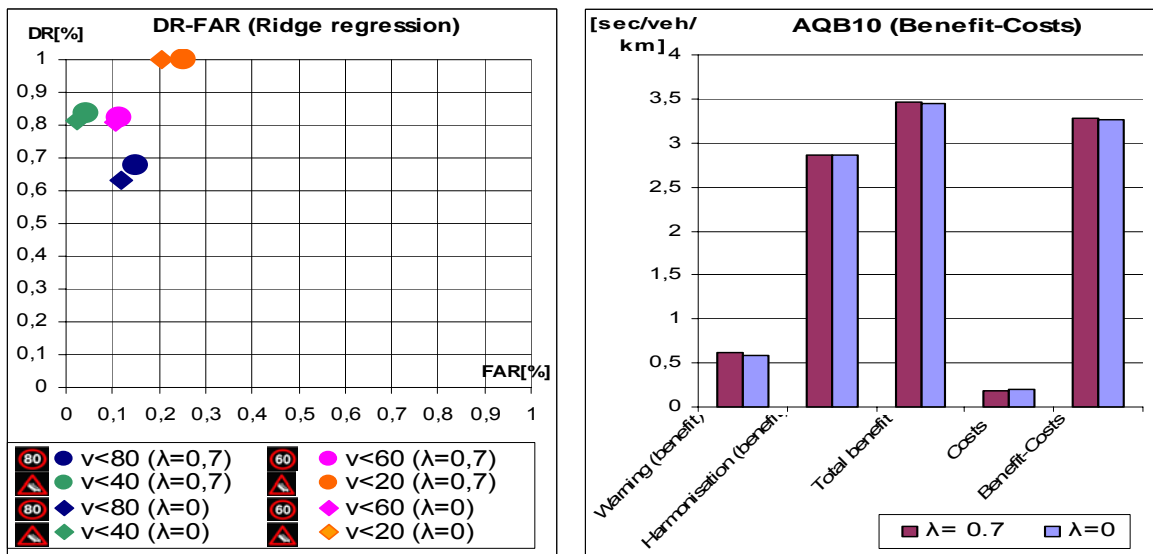


Figure 5-20: DR-FAR curves and values of objective function for controls produced with and without Ridge Regression at AQB10

The controls produced with and without Ridge regression at VMS AQB10, are evaluated using DR-FAR and the objective function (Figure 5-20). The major goal of the Ridge regression is to produce a more stable control model which will be used in on-line operation, and thus it “sacrifices” part of the model precision for the sake of a more stable control. In this case, Ridge regression yields a model which produces a slightly lower warning benefit but, more importantly it noticeably reduces cost as well. From the DR-FAR curves it can be seen that Ridge regression leads to better detection of the warning controls of 40 km/h, 60 km/h and 80 km/h, and induces a slightly higher false alarm rate for the 80 km/h controls. By



observing the DR-FAR curves and objective function together, it could be concluded that the controls optimised using Ridge regression react a slightly later and, therefore, bring lower contributions to the warning benefits, but allow for more events to be detected. The Ridge regression reduces the danger of too many false alarms and thus allows the system optimisation in the direction of detecting more events. On the other hand, in the approach without Ridge regression any further increase in detection would lead to greater system instability, which can be possibly reflected in too often and unpredictable activations of the more restrictive controls (false alarms).

### 5.4.7 Convergence

The convergence of the harmonisation and warning optimisation procedure is shown in Figure 5-21 and Figure 5-22, respectively. On the x-axis the total number of function evaluations is shown. On the y-axis, only optimisation iterations which lead to a new “optimal” point within the search space are shown. Since nonlinear optimisation might stop in the local minimum, four re-iterations (black, orange, pink and blue curves) of the optimisation process are carried out before the optimisation procedure is completed. All these re-optimisations constitute one harmonisation/warning optimisation procedure and their sum is shown by the red curve.

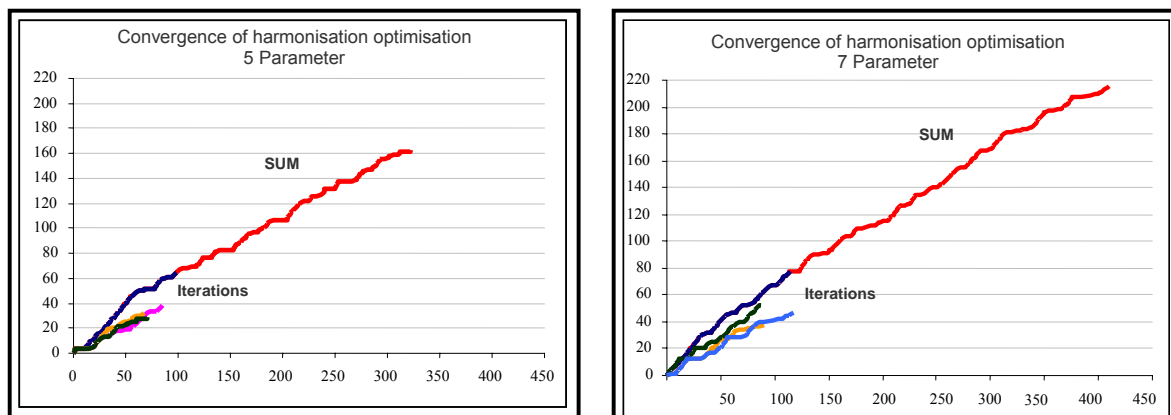


Figure 5-21: Convergence of harmonisation optimisation. Left: 5-parameter model. Right: 7-parameter model.

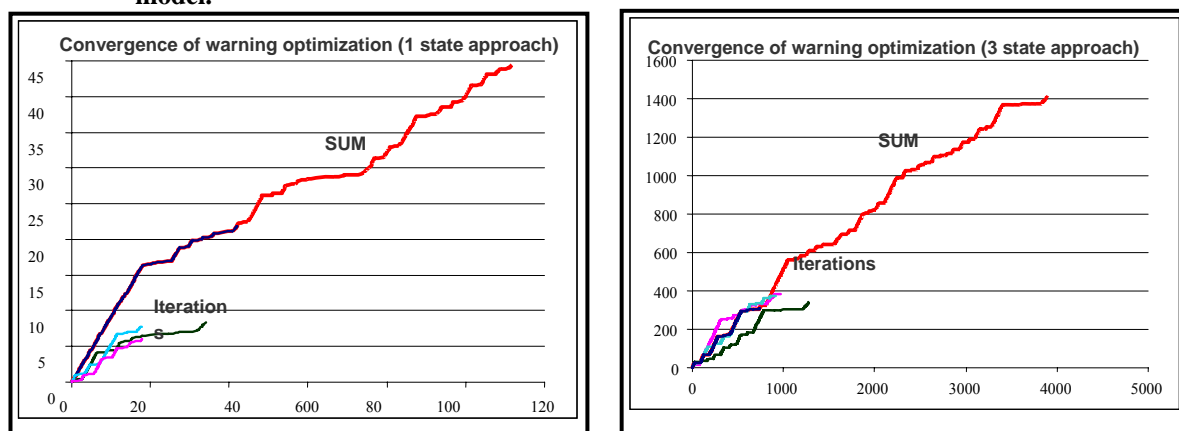


Figure 5-22: Convergence of the warning optimization. Left: convergence of the model that does not consider division into traffic states. Right: convergence of the model with division into traffic state.

### 5.4.8 Overview of the optimised (INCA) and current (MARZ) parameters

An overview of the INCA optimised harmonisation parameters and currently used Belastung parameters (MARZ) for each control station is given in Table 5-6.

**Table 5-6: Optimal Belastung parameters determined using INCA harmonisation optimisation procedure (right) and currently used Belastung parameters (MARZ) (left)**

Optimised INCA values								MARZ							
VMS/DET	QUp120	kUp100	qUp100	vUp100	kUp80	qUp80	vUp80	VMS/DET	QUp120	kUp100	qUp100	vUp100	kUp80	qUp80	vUp80
AQB/MQB10	2624	20	6572	100	10	7157	31	AQB/MQB10	4800	-1	5500	-1	40	6100	80
AQB/MQB11	1855	24	4005	108	50	5576	81	AQB/MQB11	3900	-1	4700	-1	35	5500	80
AQB/MQB12	3977	18	5600	172	36	6113	83	AQB/MQB12	3800	-1	4000	-1	35	5200	80
AQB/MQB13	2984	14	4000	168	31	5007	57	AQB/MQB13	3800	-1	5300	-1	40	6000	80
AQB/MQB14	3987	27	5600	140	32	6113	82	AQB/MQB14	4100	-1	5400	-1	40	6000	80
AQB/MQB15	1830	0	5615	104	89	6139	138	AQB/MQB15	5000	-1	5800	-1	45	6400	80
AQB/MQB16	1872	-1	5930	-1	7	8433	69	AQB/MQB16	5000	-1	5800	-1	45	6400	80
AQB/MQB17	4838	21	5605	110	19	6082	70	AQB/MQB17	5000	-1	5800	-1	45	6400	80
AQB/MQB18	4659	22	5600	107	11	6105	93	AQB/MQB18	5000	-1	5800	-1	45	6400	80
AQB/MQB19	3039	15	5604	104	38	6159	89	AQB/MQB19	5000	-1	5800	-1	45	6400	80
AQB/MQB20	CORRUPTED DATA							AQB/MQB20	5000	-1	5800	-1	45	6400	80
AQB/MQB21	4198	17	5606	106	48	6156	79	AQB/MQB21	5000	-1	5800	-1	45	6400	80
AQB/MQB22	4419	23	5599	109	30	6112	94	AQB/MQB22	5000	-1	5800	-1	45	6400	80
AQB/MQB23	4964	21	5596	109	25	6364	83	AQB/MQB23	5600	-1	6300	-1	45	7000	80
AQB/MQB24	4634	45	6507	115	92	6964	88	AQB/MQB24	5200	-1	5800	-1	45	6400	80
AQB/MQB25	CORRUPTED DATA							AQB/MQB25	5200	-1	5800	-1	45	6400	80
AQB/MQB26	4607	18	5899	105	27	6711	81	AQB/MQB26	5200	-1	6000	-1	45	6600	80
AQB/MQB27	5024	14	6303	106	27	6752	29	AQB/MQB27	4800	-1	6300	-1	45	7000	80
AQB/MQB28	CORRUPTED DATA							AQB/MQB28	5200	-1	6300	-1	45	7000	80
AQB/MQB29	2933		6412		0	8478	0	AQB/MQB29	4800	-1	6600	-1	45	7000	80

By harmonisation optimisation in INCA, for each control station different ideal harmonisation parameters are produced, depending on the fundamental diagram and specific traffic dynamics on the section. Parameter value of  $-1$  implies that this parameter will not be used (it is considered irrelevant) in making of the corresponding control decision. For example, at VMS AQB16 control decision of 100 km/h will be made using flow measurements only, since  $k_{on}(100)$  and  $Vcar_{f_{on}}(100)$  are equal to  $-1$ . Very high or very small threshold values (e.g.  $Vcar_{f_{on}}(80)$ ,  $k_{on}(100)$  by VMS AQB15,  $Vcar_{f_{on}}(100)$  by VMS AQB13, etc.) will have a similar effect - as described in previous section on the example of AQB13 (chapter 5.4.5).

The currently employed parameters (MARZ) are quite similar for all control stations. In addition, they are usually so high that harmonisation controls will be only seldom activated (with the exception of the threshold value for control of 120 km/h), i.e. usually when traffic is already going slower than the displayed speed limit. This is especially visible by the parameters of the control 80 km/h (or 60 km/h which is not shown here). Exceptions are the control station in the “Irschenberg” area (AQB10-AQB13) where, due to the criticality of the location, the operator was forced to adjust them. The current use of the default parameters is not surprising, since at present (contrary to the INCA framework) the user does not have the possibility to systematically assess the effects of different parameters and to choose the best one.

## 5.5 Evaluation of the INCA on-line performances

INCA was tested in the on-line open-loop mode during a two-month period from September 14, 2004 to November 25, 2004. The controls produced during this period were evaluated using the objective function and DR-FAR curves and compared to the performances of the currently used (MARZ) and ideal controls.

### 5.5.1 Applied optimisation approach and derived control models

The warning and harmonisation control models are optimised separately for each control station. The objective function with its default parameter settings (chapter 3.3.8) is used in the optimisation procedure. The same optimisation settings are used for each control station when deriving of harmonisation control models (chapter 5.4.8): the 7-parameter version of the Belastung algorithm is optimised using the Bootstrap method with 10 subsamples.

Table 5-7: Optimisation approach used to generate control parameters for each VMS

VMS	WARNING			HARMONISATION
	Traffic context	Ridge Regression	Bootstrap samples	Number of parameters
AQB10	3	0,7		3* (10 + 3)
AQB11			10	10 + 6
AQB12				10 + 6
AQB13			10	10 + 6
AQB14				10 + 6
AQB15				10 + 6
AQB16	3		10	3*(11 + 6)
AQB17		0.7	10	11 + 5
AQB18			10	11 + 7
AQB19	3		10	3*(11 + 7)
AQB20	LOW DATA QUALITY => AQB21			
AQB21		0.7		11 + 7
AQB22			10	11 + 7
AQB23	3			3*(11 + 6)
AQB24		0.7		11 + 5
AQB25	LOW DATA QUALITY => AQB26			
AQB26			10	11 + 7
AQB27				10 + 7
AQB28	NO DATA			
AQB29				11 + 7

To investigate the influence of the changes in optimisation approach on the derived warning model and its performance in reality, at different control stations a different optimisation approach is used to derive warning control model (Table 5-7). The decision regarding the optimisation approach used is made based on the classification of the motorway into three characteristic segments (control groups) (chapter 5.2.2). Basically, the optimisation approach varied between control stations from the same control group: Ridge regression is used for generation of the control model for VMS AQB10, AQB17, AQB21 and AQB24; Bootstrap for generating the warning control model for AQB11, AQB13, AQB16-AQB19, AQB22, AQB26; consideration of the traffic state is used for deriving the warning control model for VMS AQB10, AQB16, AQB19 and AQB23. The control stations that did not have enough

data for optimisation (AQB20, AQB25, and AQB28) used the control model from the upstream control station. A detailed overview of the optimisation approach used to generate control parameters for each control station, and the resulting number of parameters, is given in Table 5-7.

Although the harmonisation model was generated in the same manner for each control station, in the case of some control stations optimisation resulted in a different (reduced) model (chapter 5.4.8). For example, at VMS AQB16 and AQB29 parameters  $k_{on}(100)$  and  $Vcar_{f_{on}}(100)$  are optimised to be -1 and thus are not used for the harmonisation decision. In this way, initial 7-parameter Belastung model is actually reduced to a 5-parameter model. The total number of parameters that constitutes the warning control model, given in above table, is calculated as the sum of information from the local and downstream control station which is, according to optimisation procedure, considered to be significant for the specific warning model. The number of determined significant information varies from control station to control station. Generally, warning control model for AQB10-AQB15 uses slightly less information than those of other control stations. This is due to the algorithm vkDiff that has proven to be unstable at these control stations and thus is marked as “insignificant”. In addition, at some other control stations, e.g. AQB16, AQB17, and AQB23, some information is identified as “of low significance” and thus it is not included in the warning model.

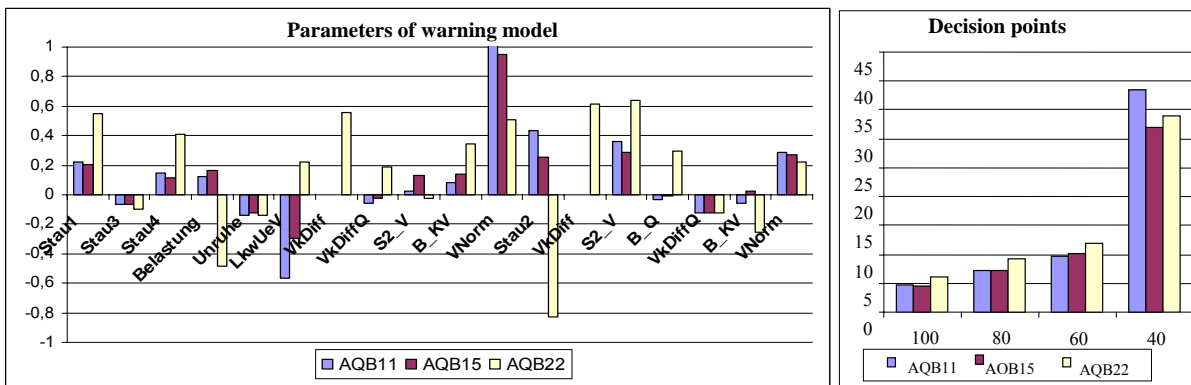


Figure 5-23: Different parameters for different control stations and corresponding decision boundaries

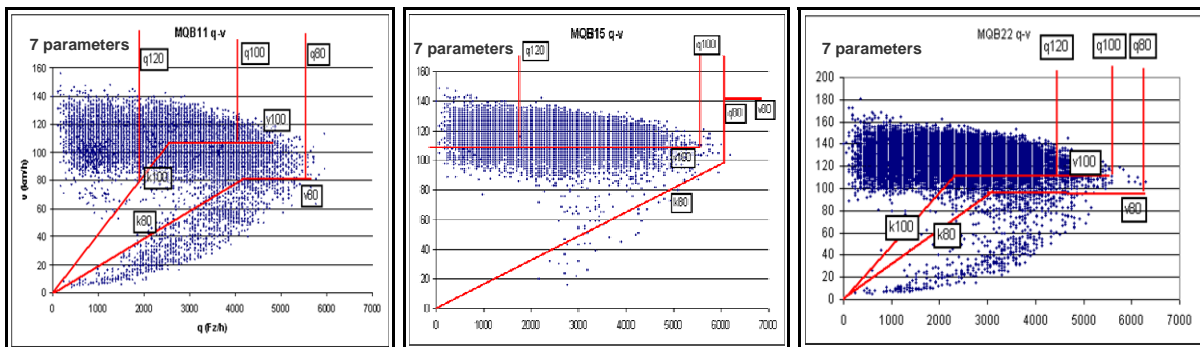


Figure 5-24: Harmonisation decision parameters for AQB11, AQB15 and AQB22.

The variations of the generated parameters are illustrated by showing together one control model from each “control group”. For the sake of the simplicity, the models that do not consider division into traffic states, from VMS AQB11, AQB15 and AQB21, are shown together. Figure 5-23 shows the optimised warning parameters of these control stations, while in Figure 5-24 the harmonisation parameters are plotted on the corresponding fundamental diagram.

### 5.5.2 On-line results

INCA control performance is assessed using the objective function and DR-FAR curves. An overview of INCA performance and its comparison with the current (MARZ) and the ideal control is given for the entire motorway segment (Figure 5-25) and for each control group, as defined in the chapter 5.2.2 (Figure 5-26). Figure 5-27 provides an overview of the INCA and MARZ performances on the entire motorway in terms of DR-FAR curves. Points of interest on the DR-FAR curves are marked with the large circles.

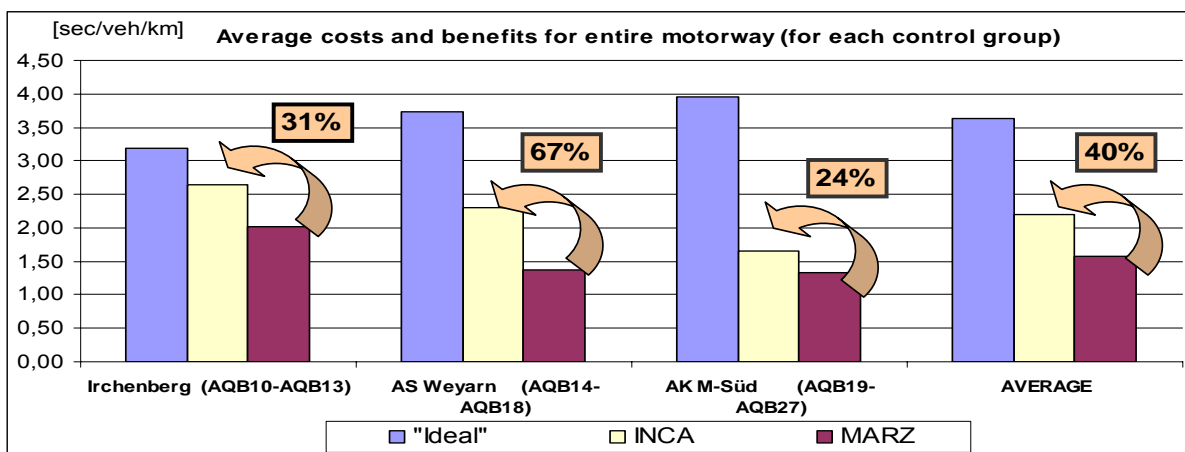


Figure 5-25: Performances of the control for entire motorway estimated using the objective function

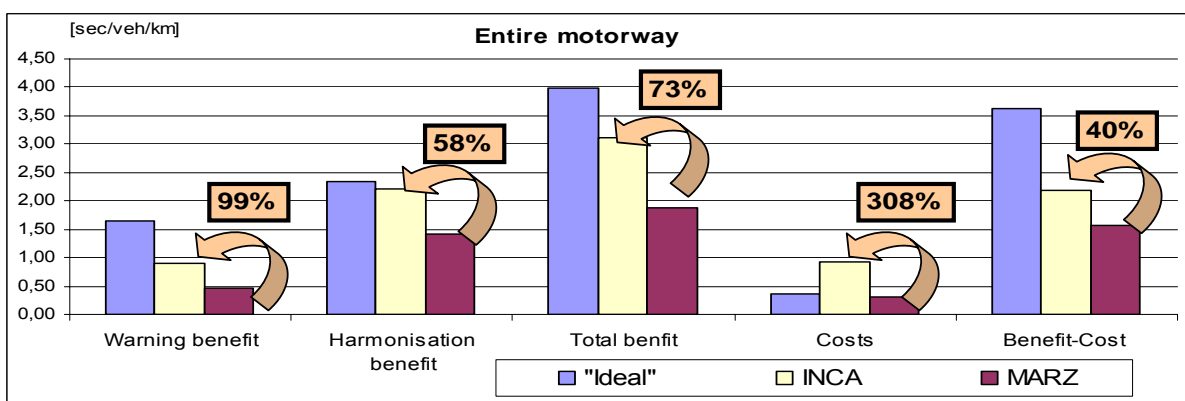
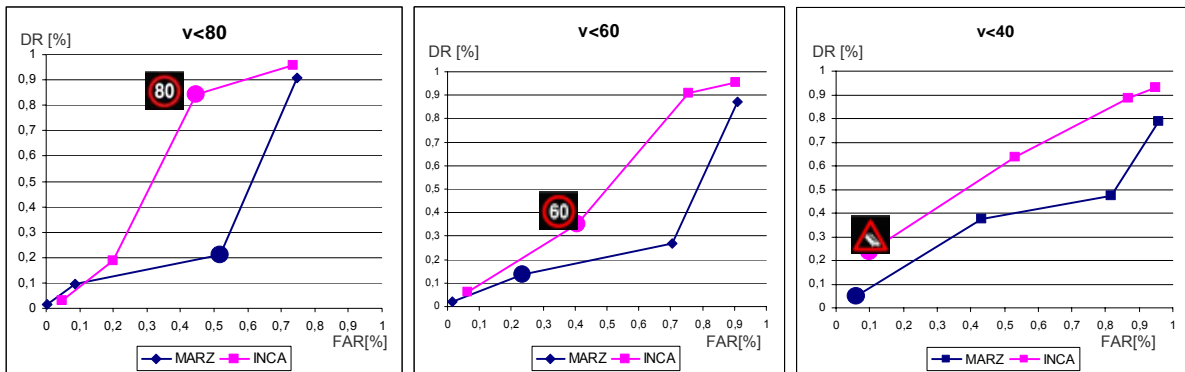


Figure 5-26: Performances of INCA, MARZ and ideal controls for each control group, and overall

INCA outperforms the existing system according to both the objective function and the DR-FAR curves. According to the objective function the average improvement of 43% in control system performances could be induced by introduction of INCA instead of current system.

INCA significantly increases the warning and harmonisation benefit where, due to the proactive nature of INCA; this increase in benefits is accompanied by an increase in costs. However, it should be noted that *percentages in the figures represent improvements produced by INCA in comparison with the current MARZ control*, which usually activates a control after a breakdown is already present at the location, thus, inducing very low costs. The highest increase in overall performances is observable in the second control group. The lower increase by first control group is due to the small possibility to increase harmonisation benefit, due to the always present harmonisation control of 120 km/h or 100 km/h.



**Figure 5-27: Performances of INCA and MARZ at the complete motorway**

The DR-FAR curves show that the increase in the performances of the 80 km/h control has improved greatly. In addition to the 61% increase in the detection rate, the false alarm rate dropped by approximately 10%. It should be noted that the increase in the performances of the 80 km/h control can be observed at all control stations, though not always to the same extent. The control “congestion”, considered here as 40 km/h, also over-performs the existing control at all control stations. According to the DR-FAR curves, the detection rate increases by 16%, while the false alarm rate increases by 4%. For speed drop lower than 40 km/h it is also interesting to observe the control speed of 60 km/h, for which it can be seen that detection rate is increased by 21% followed by an increase in false alarm rate by 10%. Only a slight DR-FAR improvement can be observed for the 60 km/h speed drop, where detection and false alarm rates increased almost equally. Estimating performances for drops of 60 km/h is especially difficult, since the control of 60 km/h is needed in the area of the fundamental diagram with the specific traffic dynamic. The control is often activated due to the control propagation from the downstream control station (warning about incident beyond the downstream control station). However, incidents do not always propagate all way to the upstream control station and with equal speed, which could lead to the calculation of more false alarms or fewer detection rates. Data quality and changed traffic characteristics in the test set (e.g. AQB24) also play a significant role in entire process.

The benefits and costs induced by MARZ and INCA control for each control group, as well as the corresponding DR-FAR curves for the 80 km/h speed drop are shown in Figure 5-28. As expected, the most stable control behaviour (and performance) is achieved in the first control group (Irschenberg). This is mainly due to the higher number of events (incidents) in the

training data set which allowed the INCA to better learn the appropriate behaviour. The proactive character of INCA in this control group resulted in the significant increase in the warning benefit but also induces greater costs than the current control. The harmonisation benefit is almost the same as by the current system. This is due to the abovementioned usually always present controls of 100 km/h or 120 km/h in the “Irschenberg” area, due to which the harmonisation controls have a smaller possibility to influence traffic flow in the harmonisation area and thus to induce the corresponding benefits.

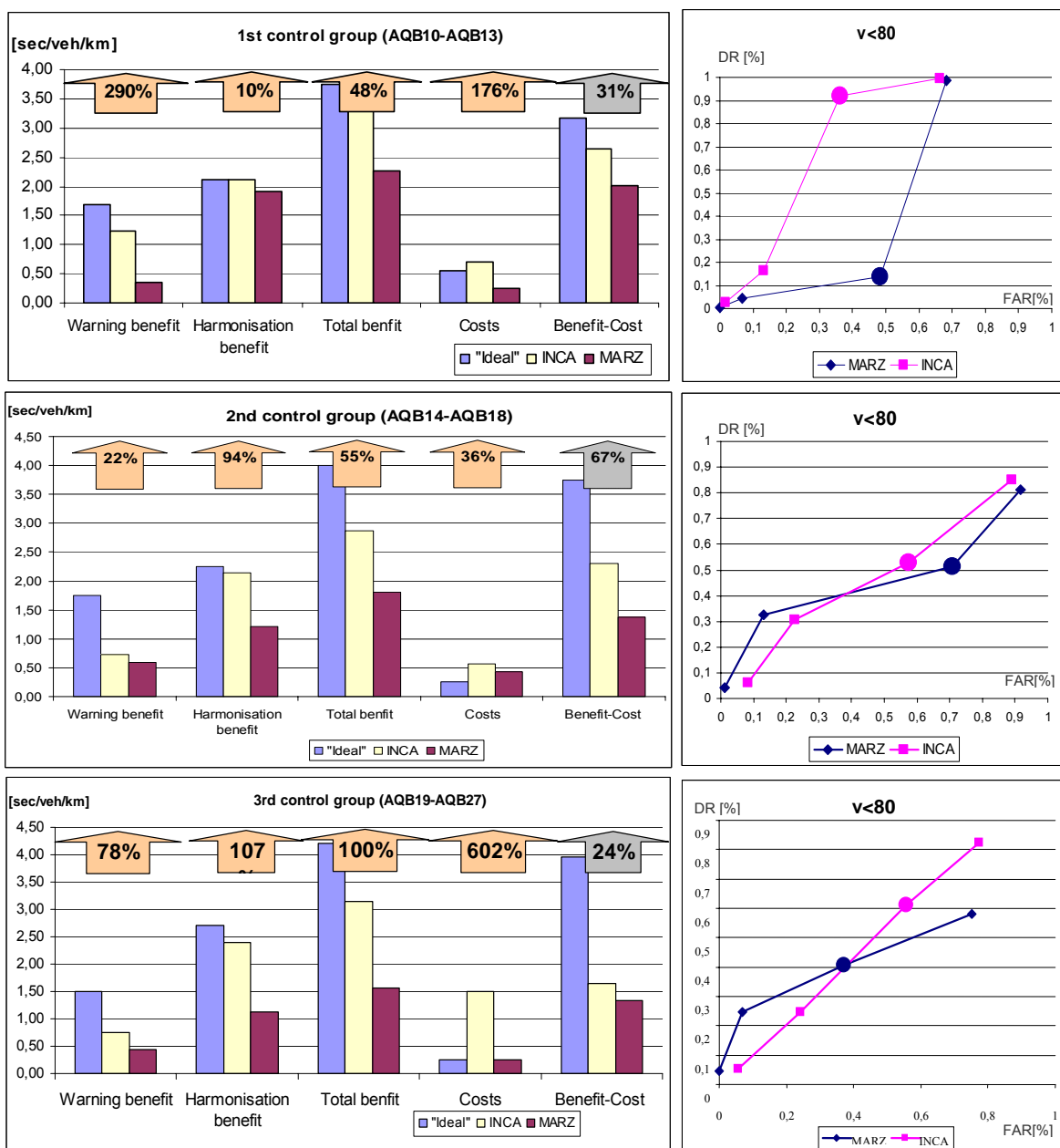


Figure 5-28: Overview of costs and benefits, and DR-FAR curves for each control group. In percentage is shown the increase in benefits/costs between current (MARZ) and INCA control

At first glance, the greatest benefit-cost increase (67%) is achieved in the 2<sup>nd</sup> control group. The major part of this increase is due to the additional harmonisation benefit and only slight

increase in costs. On the other side, the increase in the warning benefit is the lowest of all the control groups and much lower than the benefit of the "ideal" control. This is due to several reasons: in the training data there are not been enough events (incidents) to allow derivation of a more precise control model; low data quality during the test period; *the algorithms used in the control model are not able to detect the specific type of incident dynamics in this control group*, which influences the performance of the control model. The third group is considered critical due to the low data quality. At the location significant increases in warning and harmonisation benefits have been measured but there is also a large increase in costs. This increase in the costs is partly due to the proactive character of INCA and good warning control, but also due to the false alarms usually induced by corrupted data at some AQ stations.

At first glance, it appears that INCA has the same or even higher harmonisation benefit than the ideal control in the second and third group. However, the same figure shows that this "better" harmonisation actually leads to higher costs (even when not all costs can be attributed to the harmonisation control). In other words, INCA activates harmonisation controls even when this will lead to higher costs, while the ideal control finds a better balance between benefits and costs.

DR-FAR curves for speed drops below 80 km/h show that the greatest improvement is in the case of the first group (Irschenberg), with an 80% higher detection rate and a reduction in false alarms by more than 15%. In the case of the second control group for the same situation only a slight increase in the detection (3%) can be observed (which in this case corresponds to the value of the warning benefit of the objective function), but this leads to reduction of the false alarms of more than 15%. The most unstable control decisions are observable by the third control group, where INCA detects more incidents but, at the same time, produces more false alarms.

A more detailed overview of the performances of MARZ, INCA and the ideal control at each AQ station is given in Table 5-8. For better illustration of the potentials of Bootstrap, Ridge regression and division into traffic contexts, in Table 5-9 and Figure 5-29 the performances of controls optimised with and without these mechanisms, measured using the objective function and DR-FAR curves, are shown on examples of AQB11, AQB14 and AQB18. Cases with specific mechanisms are produced in the simulated on-line mode as described in chapter 5.5.1. The control stations in the first control group produced a slight increase in warning benefits, with the exception of AQB13 where significant improvements were noticed. Analysis of the empirical data has confirmed that the majority of the false alarms in this area is due to the control propagation from the downstream control stations at AQB11 and AQB13, which do not propagate all the way up to AQB10 and AQB12. Therefore, the calculated benefits are actually lower than the real one since the control has a proactive purpose whose effects were not observable in the local detector data. In the second group (AQB14-AQB18), INCA outperforms the current MARZ control in harmonisation as well as



warning. The exception is the control for AQB16 where, due to corrupted data, the warning benefits dropped but costs grew. In the case of the third control group (AQB19-AQB29) INCA increases warning benefits for almost all control stations. However, a large increase in costs can be observed at stations AQB23 and AQB24. This is mainly due to data from downstream AQB25 station, which was missing during optimisation and often corrupted during the test period.

Table 5-8: Benefit and costs of INCA, MARZ and ideal control for each VMS

AQB	MARZ					INCA					IDEAL				
	Benefit			Costs	Benefits-Costs	Benefit			Costs	Benefits-Costs	Benefit			Costs	Benefits-Costs
	Warning	Harmonisation	TOTAL			Warning	Harmonisation	TOTAL			Warning	Harmonisation	TOTAL		
AQB10	0.06	2.59	2.65	0.75	1.90	0.05	2.66	2.71	0.98	1.73	0.04	2.61	2.65	0.56	2.09
AQB11	0.50	2.04	2.54	0.06	2.48	0.54	2.29	2.83	0.27	2.56	0.49	2.36	2.85	0.06	2.79
AQB12	0.50	1.84	2.34	0.14	2.20	0.52	2.16	2.68	1.31	1.37	0.50	2.16	2.34	0.14	2.20
AQB13	0.33	1.17	1.51	0.04	1.46	3.67	1.34	5.08	0.24	4.83	5.74	1.31	7.15	1.51	5.64
AQB14	0.49	2.62	3.10	0.78	2.32	0.43	2.34	2.77	0.24	2.53	0.82	2.61	3.43	0.30	3.13
AQB15	0.17	0.89	1.06	0.10	0.96	0.31	2.55	2.86	0.29	2.57	0.54	2.57	3.12	0.21	2.91
AQB16	0.62	1.04	1.67	0.33	1.33	0.34	2.51	2.85	0.72	2.13	1.58	2.62	4.20	0.34	3.86
AQB17	1.05	1.26	2.31	0.32	2.00	1.29	2.11	3.40	0.68	2.72	2.62	1.97	4.59	0.29	4.30
AQB18	0.61	1.08	1.69	0.62	1.07	0.89	1.99	2.87	0.92	1.95	1.91	1.94	3.87	0.16	3.71
AQB19	0.32	0.63	0.96	0.20	0.76	0.27	2.85	3.12	0.36	2.75	0.77	2.52	3.30	0.05	3.25
AQB20															
AQB21	0.67	1.14	1.81	0.43	1.38	0.32	2.74	3.06	0.64	2.42					
AQB22	0.58	1.17	1.76	0.43	1.33	0.80	2.53	3.33	0.93	2.40	1.71	2.16	3.87	0.17	3.70
AQB23	0.59	0.88	1.48	0.28	1.20	0.97	2.40	3.38	2.40	0.98	1.63	2.74	4.37	0.18	4.19
AQB24	0.40	1.36	1.76	0.20	1.56	1.42	1.93	3.35	3.94	-0.59	2.27	2.65	4.91	0.24	4.67
AQB25	0.13	1.00	1.13	0.02	1.11	3.59	2.00	5.59	0.62	4.97	10.38	1.85	12.23	1.34	10.89
AQB27	0.39	1.73	2.12	0.17	1.95	0.60	2.07	2.68	0.75	1.93	1.53	2.57	4.10	0.14	3.96

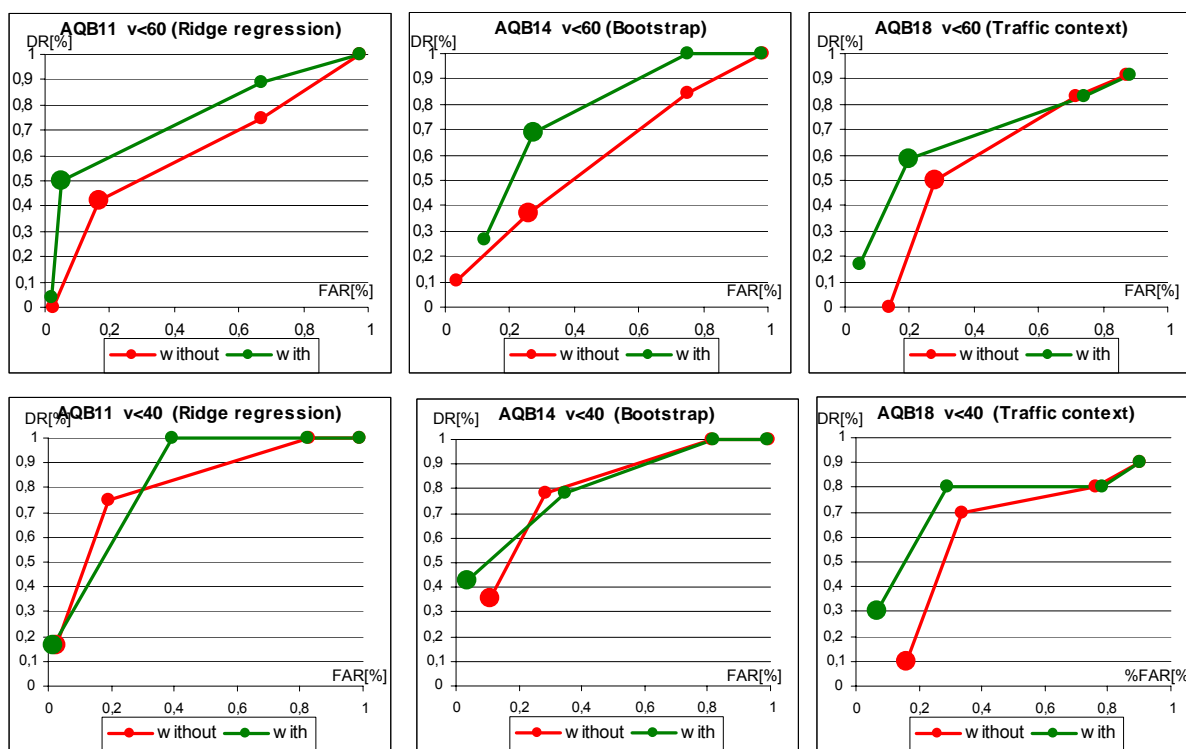


Figure 5-29: Results of simulated on-line controls of the models generated with different optimisation approach (with or without Ridge regression, Bootstrap, consideration of traffic context)

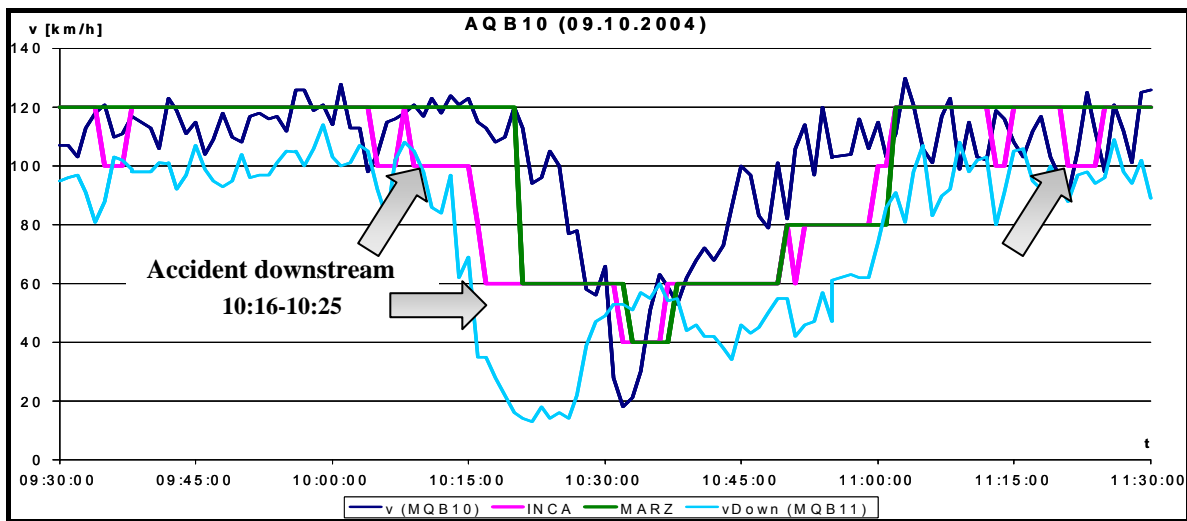
**Table 5-9: Results of the simulated on-line controls of the models generated using different optimisation approaches**

	Warning benefit	Harmonisation benefit	Average benefits	Average costs	Benefit-Costs
<b>Ridge regression</b>	<b>AQB11</b>				
Without	0,54	2,29	2,83	0,27	<b>2,56</b>
With	0,52	2,32	2,84	0,24	<b>2,60</b>
<b>Bootstrap</b>	<b>AQB14</b>				
Without	0,43	2,34	2,77	0,24	<b>2,53</b>
With	0,50	2,47	2,97	0,31	<b>2,66</b>
<b>Traffic context</b>	<b>AQB18</b>				
Without	0,89	1,99	2,87	0,92	<b>1,95</b>
With	0,97	2,11	3,08	1,02	<b>2,06</b>

As expected, the utilisation of Bootstrap, Ridge regression and traffic context consideration, led to more stable control models. Overall, control stations controlled by the model generated using Bootstrap, Ridge regression or considering the traffic context, performed slightly better than other control stations in the same group which used the models generated without these mechanisms (Table 5-8). The application of these techniques shows a tendency to reduce cost of control for all control stations, with the exception of AQB24 (influenced by the problematic AQB25). The application of Ridge regression produced slightly lower warning benefits, while the use of Bootstrap improved detection and resulted in slightly higher warning benefits than the control models produced without Bootstrap. Considering the traffic context under which the control is made shows the tendency of better detection and reduction of the costs.

**5.5.3 Empirical examples**

In this chapter some examples of the INCA and current MARZ controls in different situations (warning of an accident and/or congestion, bad weather, work zones, harmonisation and lower disturbances) are given.



**Figure 5-30: Empirical example of INCA reaction at AQB10 in the case of an accident further downstream.**

Figure 5-30 illustrates INCA control warning and harmonisation features on the example of AQB10. The situations of interests are marked using arrows. Between the hours of 10:16h and 10:25h an *accident* was reported some 3.5 km further downstream at the second control station downstream (AQB12). The effects of accidents are visible at both upstream control stations - MQB11 (vDown) and MQB10 (v). In this situation INCA reacts 7 minutes earlier, reduces surprise and fulfils its warning purpose. The first activation of the 60 km/h control is due to the control propagation rule: at the downstream control station (AQB11) a “congestion” sign is displayed and it triggers the control of 60 km/h at the upstream location. At 10:31h the local control model suggests the lowest possible control and takes over control from the downstream control station. Probably the most *interesting* control is the one shortly before the traffic flow breakdown when INCA posts 100 km/h, while MARZ maintains 120 km/h. At 11:20h, an example of an INCA *harmonisation* control can be seen. Due to the lower speeds at the local but also at the downstream detector, INCA posts the 100 km/h control, while MARZ keeps the 120 km/h control.

Another example of the INCA warning and prevention feature is illustrated on the example of AQB12 in Figure 5-31. The first example of the prevention can be seen at 15:15h when a smaller speed drop has been detected at the local detector station. Very interesting behaviour can be observed at around 17:00h when INCA reacts with the 80 km/h controls even before the real speed drop has been measured at the downstream detector, while MARZ displays the 100 km/h control. After about 10 minutes, disturbances are also detected in the macroscopic data. Here the INCA warning component activated the control timely, some 5 min earlier than MARZ. In the end, the harmonisation feature can be seen at around 18:00h where due to the lower measured local and downstream speed, INCA displays the control 100 km/h while MARZ keeps the 120 km/h control. With this control INCA does not generate additional costs since the driven speeds are around the displayed speed limit, but it provides a harmonisation benefit due to the smoothing of the traffic flow.

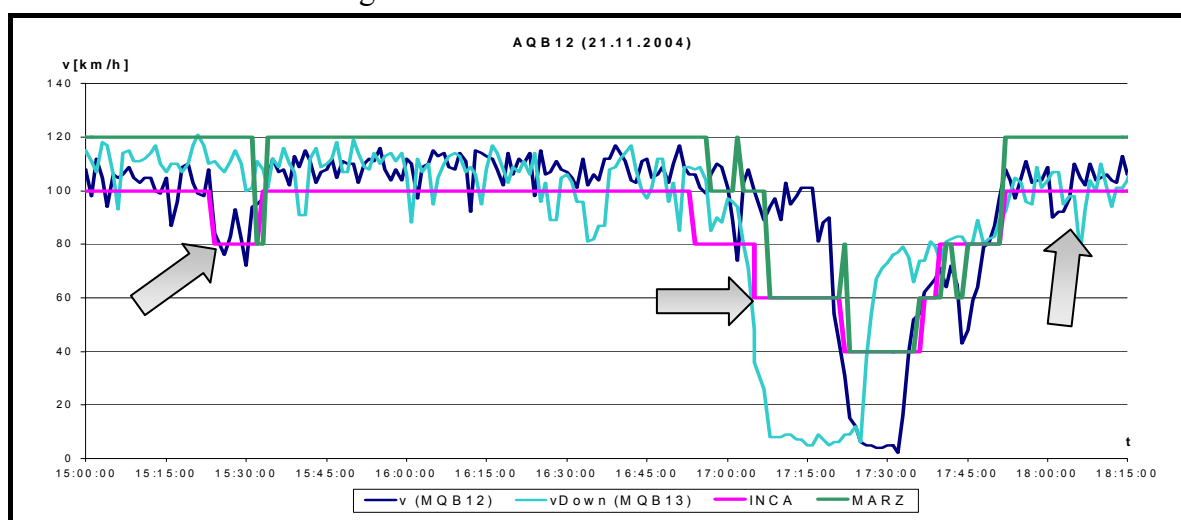


Figure 5-31: Example of INCA-control at AQB12 in case of the congestion further downstream

The reactions of INCA and MARZ under very dense traffic conditions at the local station and higher downstream speeds are shown in Figure 5-32. The INCA warning feature can be observed at 16:00h when INCA warns the driver about the speed drop downstream with the control of 80 km/h (MARZ - 120 km/h). However, the situation at the local station worsens and at the downstream remains the same, which implies that the incident is most likely somewhere in the section between these VMSs. Since INCA includes downstream information into the control decision it allows slightly higher speeds (80 km/h) than MARZ (60 km/h).

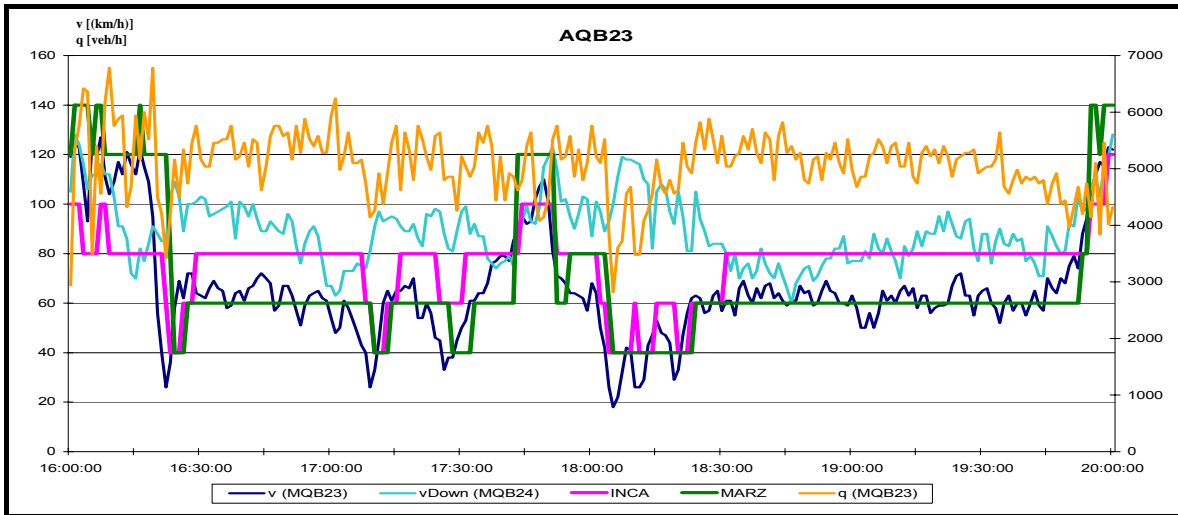


Figure 5-32: Incident in the section between AQB23 and AQB24

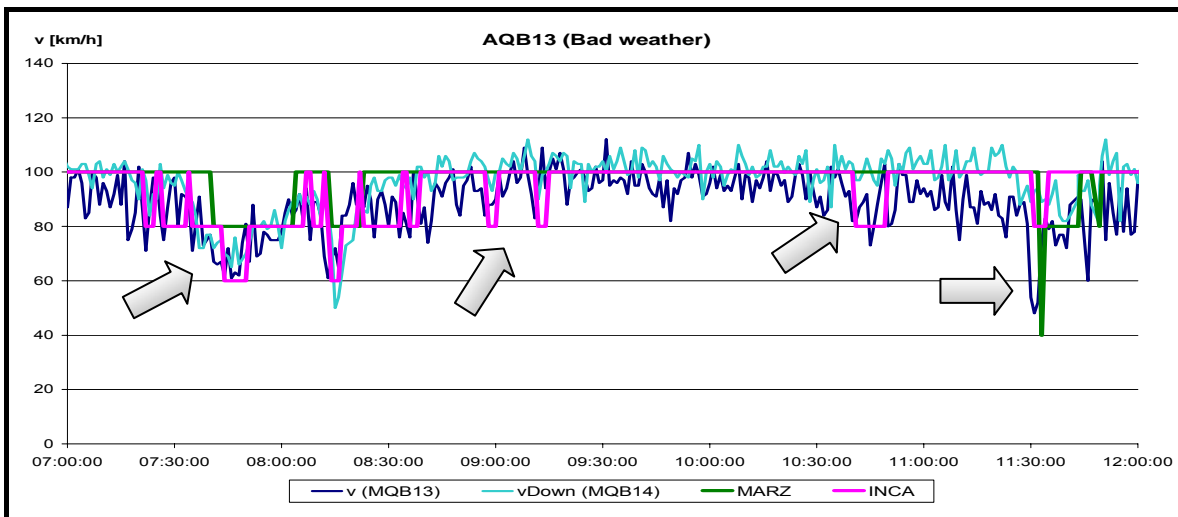
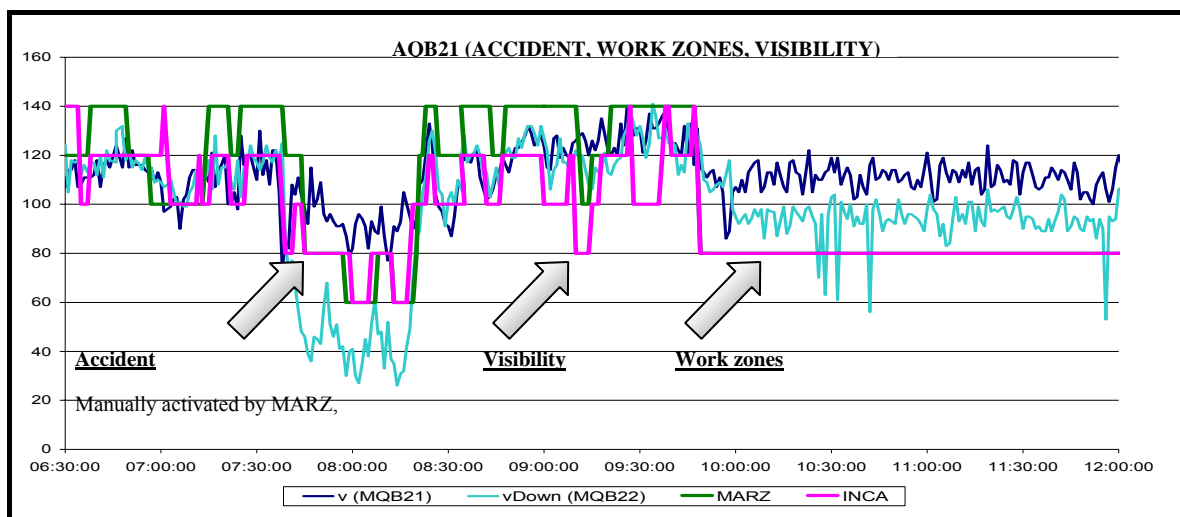


Figure 5-33: INCA control by bad weather conditions on November 8, 2004 (AQB13)

Examples of INCA and MARZ controls under bad weather conditions are shown in Figure 5-33, for November 8, 2004 and AQB13. On November 8, 2004 a considerable amount of rainfall was measured (see Appendix A.4) and from the control data it can be concluded that visibility was bad. Even though INCA does not have special weather detection decision component, it behaves quite good under bad weather conditions. At around 7:30h-8:30h

INCA successfully detected the slight speed drop. Around 09:00h a slightly more restrictive control of 80 km/h was displayed, but as the driven speeds were not much higher and visibility was lower throughout the day this is considered not critical. On the other hand, at around 11:30h MARZ displays sign “congestion” too late, while INCA kept control of 80 km/h.

In Figure 5-34 several simultaneous features of INCA and MARZ (accident case, work zones and bad weather (VISIBILITY) conditions) can be seen. At around 7:30h INCA detects an incident and posts the 80 km/h control, some 8 min earlier than MARZ. Moreover, INCA did not only react earlier but was able to automatically detect an incident, while the current system displayed lower controls and sign ACCIDENT only after manual intervention from the operator. Around 9:30h, INCA suggests a slightly more restrictive control of 80 km/h. At the same time MARZ displays warning sign VISIBILITY and control of 100 km/h. Obviously, adverse weather conditions influenced the traffic flow and caused INCA to post lower speed limits. Finally, after 10:00h the operator manually activates the sign WORK ZONES. Since INCA is configured to take over manual controls, it takes this control as its own.



**Figure 5-34: Accident, work zones; visibility on November 8, 2004 at AQB21.**

INCA behaviour in case of bad data quality is shown on Figure 5-35. At the AQB22 data was sporadically missing during the entire day on November 11, 2004. According to the control protocols the current control model (MARZ) did not activate any control during this day, although the flow was increased and speeds were lower. An exception is the manual control ROAD WORKS and 120 km/h, displayed at around 17:00h. INCA attempted to smooth traffic flow, but due to the missing data and employed correction mechanism more frequent changes between controls of 100 km/h and 120 km/h can be observed. Around 17:00h INCA applied the WORK ZONES sign, but it suggested a more restrictive control (100 km/h) than the one given by the operator (120 km/h).

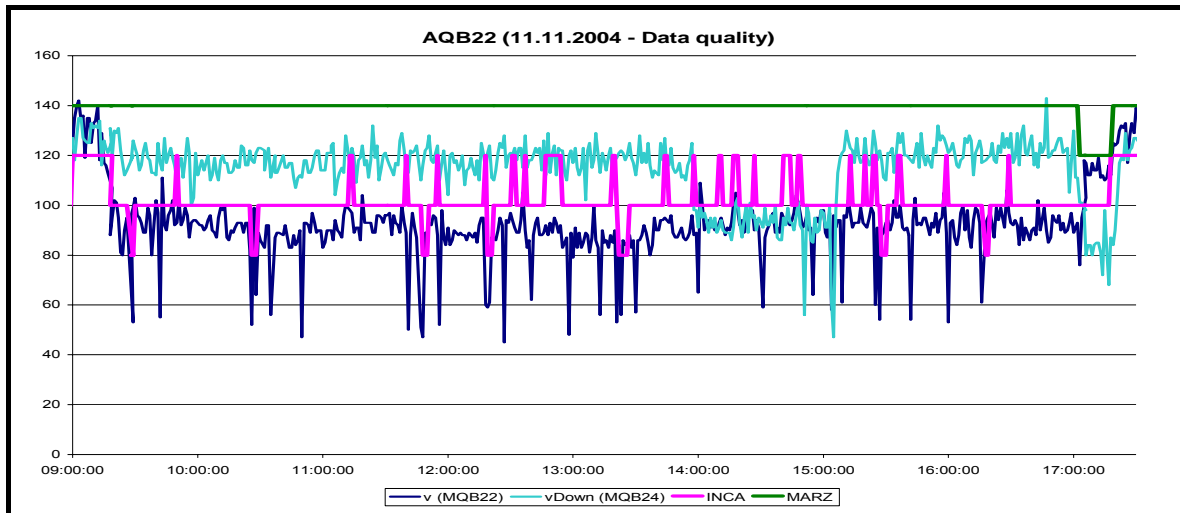


Figure 5-35: Low data quality and INCA control calculations.

## 5.6 Extendibility and flexibility

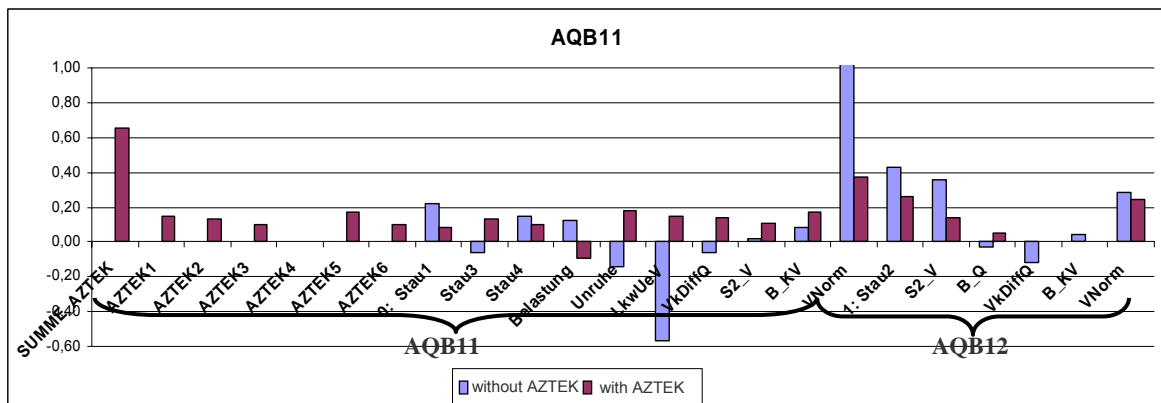
To investigate INCA ability to include new information and automatically determine its “significance” (flexibility), several indicators derived from AZTEK algorithm (HORTER ET AL., 2004) are introduced into the INCA system. The AZTEK algorithm is an advance incident detection algorithm, based on the traffic flow model by Cremer and Kalman filter (BOZIC, 1979). It is believed that it can reconstruct the traffic flow situation between two detector stations and locate the position of the tail of the queue. Thus it has the advantage of recognising incidents earlier and providing INCA with more valuable information. However, its performances strongly depend on the specification of covariance matrix and traffic flow parameters. Inclusion of these indicators in the INCA model also serves to check if advance traffic flow-based algorithms would bring additional benefit to the link control system and incident detection in reality.

A detailed description of the indicators derived from AZTEK and derivation process, is given in KATES ET AL., 2004. Six indicators derived from AZTEK algorithm are included in the system and INCA is re-optimised with this new available data. AZTEK is applied as received from its developer and no additional parameter calibration has been performed. This is due to the desire to maintain “equality” between AZTEK and other algorithms, which have also not been specially calibrated. With the new derived model and available training data, INCA controls are simulated as they would be in reality, i.e. for each second of the simulation, it is checked in the archived data if the data for that second was available. If there was new data available INCA would pass through all the steps as in reality. The control actions produced in such a manner are evaluated using the objective function and DR-FAR curves.

### 5.6.1 Changes in the warning model

With inclusion of AZTEK indicators some previously utilised information is optimised to have lower “weight” or is totally eliminated from the model. Changes in warning parameters

induced by the inclusion of AZTEK indicators, are illustrated in Figure 5-36 on the example of AQB11.



**Figure 5-36: Previous INCA warning module parameters and the new one produced with additionally included AZTEK indicators (at AQB11)**

According to KATES ET AL., 2004, the first three AZTEK indicators are slightly redundant. Each is triggered by one specific speed level, where speed levels include one another (higher includes lower). Therefore their weights are quite similar. The most important, but also the most sensitive, is the AZTEK4 indicator. In the case when the traffic flow model used in AZTEK is not properly calibrated for a specific section (as is the case for AQB11), or when data quality is not satisfactory, INCA assigns the weight 0 to this indicator. AZTEK6 indicator also plays an important role but it is very sensitive to corrupted downstream data. If the quality of the downstream data is low, the AZTEK6 indicator will be optimised to be equal to 0. With the inclusion of AZTEK indicators, the weights of the other algorithms are reduced or even some algorithms are excluded, (e.g. B\_KV and vkDiffQ downstream). In cases when AZTEK4 weight is optimised to be different from 0, the weights of harmonisation algorithms such as Belastung, Unruhe and Lkw-UeV are reduced.

### 5.6.2 Performance assessment

The performance of the new INCA model produced with AZTEK indicators and the old one are evaluated using the objective function and are shown in Figure 5-37. From figure it can be seen that:

- In the case of the first control group (AQB10 - AQB13, Irschenberg) the integration of AZTEK-indicators brings almost no improvement. It is believed that parameters used in the AZTEK flow model were not appropriate for this section, characterised with specific geometry (steep grade).
- In the second control group (AQB14 – AQB18, straight motorway section with only one on-ramp in the area of Weyarn) AZTEK integration resulted in significant improvements of INCA performances. This is also the area where AZTEK4 indicator was optimised to be different from 0, what speaks in favour of its significance. The

15% of corrupt data. The available data for training shows quite different characteristics from the data in the test period. The analysis of the performances of the ideal control shows that objective function appropriately “rewards” and “punishes” specific control strategies, and thus leads the system to the desired state. Furthermore, with the help of adjustable parameters, the objective function can be configured to better correspond to the specific needs in the field and preferences of the authorities. Since it expresses overall effects in time, its interpretation is quite straightforward and more appealing to the user. Furthermore, effects expressed in terms of time (sec/veh/km) can be easily transferred into the monetary measure (e.g. €/veh), which can be then integrated into other, for example a cost-benefit, analysis (e.g. planning purposes).

Following a detailed investigation of the optimisation process and influence of different optimisation settings on the final control model the results of the INCA on-line open loop tests are presented. It has been shown that in most cases INCA significantly outperforms current system in all control groups, especially with respect to speed drops of 80km/h and 40km/h. This, however, leads to higher control costs than of the current MARZ control (which uses prevention potential very little, anyway). According to the DR-FAR curves, further improvements are possible by the control of 60km/h. An increase in harmonisation benefit has also been observed for all control groups. The smallest increase is measured by the first control group where at present usually control of 100km/h or 120km/h is active. The positive effect of the consideration of the traffic state for warning decision, as well as positive influence of Bootstrap and Ridge regression on model stability has been shown. According to the objective function, INCA would increase performance of LCS by some 40% on average (0,7 sec/veh/km more saved than MARZ). During the investigated two-month period, with 40000 vehicles per day, this corresponds to 17,333 hours saved for the drivers, through increased safety and efficiency.

By reasonable data quality INCA detects incidents earlier and displays appropriate prevention controls for which in the current system an operator intervention is usually required. Thus, in addition to greater safety benefits produced by INCA, there is also the potential for reducing the need for manual (operator) control. Dividing the control into the harmonisation and warning components makes the system more tractable. Additionally, the INCA system provides an overview of the “importance” of particular algorithms and their reliability. At each moment it is possible to identify which algorithm or group of algorithms is responsible for control activation. Due to the developed framework, adding and excluding the algorithms from the model can be done in a systematic and tractable manner.

At this stage of development INCA does not have the explicit objective to increase the performances under low data quality. This was, however, present at AQB20 and AQB25 and this influences not only the local control but also the one further upstream. Unstable traffic characteristics during testing and training period speak in favour of a need of longer training periods. Finally, some of the false alarms produced here can be attributed to accidents which have not been detected in the local traffic flow data.



previous INCA model produced 67% more benefit than the current control. With AZTEK further improvement by 35% is achieved, which leads to the total improvement by 102%, compared to the current system. On one side this improvement is due to the good data quality, but it might also be due to the inability of the simple threshold based algorithms to timely detect incidents in the section.

- At the third control group (AQB19 – AQB27, the area AK München-Süd) performances are slightly increased, mainly through reduction in costs due to fewer false alarms.
- In total, the integration of AZTEK (on the entire motorway) brought a slight increase in the benefits, but also reduction in costs due to the fewer false alarms. In conclusion, it could be said that INCA was able to recognize reliable AZTEK indicators capable of detecting incidents with fewer false alarms, and using them instead of the previously used information, which was not so reliable.

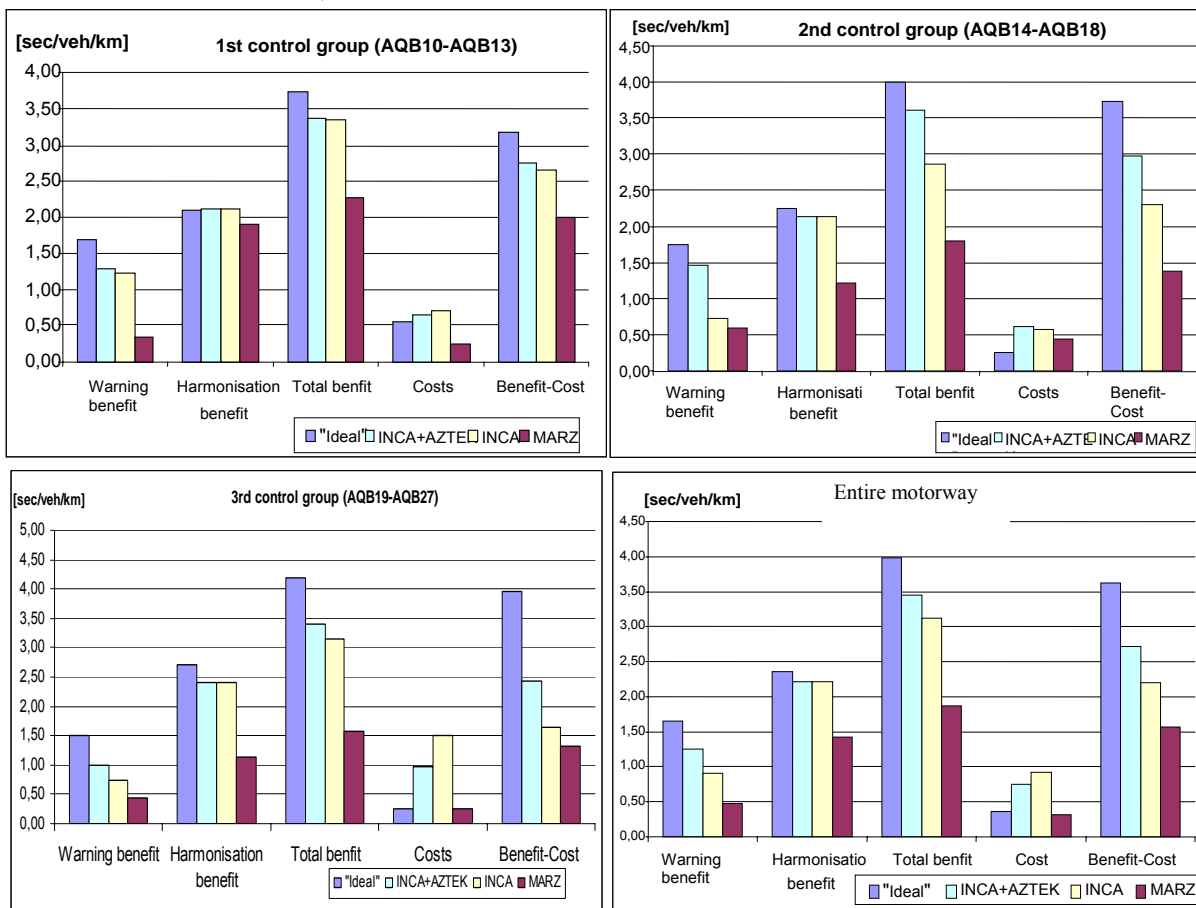


Figure 5-37: Costs and benefits for the three control groups and the entire motorway

### 5.7 Summary

The entire motorway was divided into 3 characteristic control groups, based on the number of speed drops and accidents. This analysis of the test field and data quality shows that data quality for most detectors was acceptable, although on average it contained approximately

Although INCA combines the available data in a flexible manner it is still constrained by performances of the simple mostly threshold based detection algorithms. This can be seen in the performance under unstable dense traffic conditions where the control 60km/h is more or less appropriately activated. To test INCA flexibility and the potential of more advance automatic incident detection algorithms the six AZTEK indicators are integrated into control. It has been shown that INCA is able to automatically recognize their reliability and include them in the system. INCA performance improved on the control sections with normal geometry and not too many on-ramps (2<sup>nd</sup> control group), while it brought very little improvement by other controls groups. It is believed that better calibration of the traffic flow model that lies behind AZTEK, would increase AZTEK and, hence, INCA performances. However, this is exactly the down side of advance algorithms since they require an extensive calibration approach. Finally, the optimisation of the model with more data should also lead to a more stable and more performant control system.

## **6 Conclusion and future research**

Link control systems are probably one of the least investigated areas in the traffic control field. Knowledge about their effects is still limited thus making development of control models quite difficult. The motivation of this dissertation was to provide more insight into the complex link control problem and propose a systematic, theoretically founded and practically applicable, control framework.

### **6.1 Summary and conclusion**

In this thesis the comprehensive, practically applicable, adaptive link control framework called INCA was developed and presented. The crucial components of the system are the cost-benefit based objective function, flexible information fusion model, and (parameter and model) optimisation process. Framework is supported by an additional data-correction and imputation mechanism to increase system robustness. All information needed for the control model is derived and analysed in the knowledge base component.

The objective function quantifies the basic warning and harmonisation benefits against the travel-time losses, by referencing the incident risk probabilities (potential or real) in different situations/controls and considering the information propagation in the upstream direction. Costs of the control are estimated as the time that would have been lost if drivers obeyed a control.

The benefit calculation is based on the calculated average accident density and acute accident densities. The control which reduces the costs is “rewarded” with the sum of cost reduction. Since harmonisation controls are in the higher speed area it is assumed that their positive effect will be in order of the average accident density. Warning benefit is encountered in situations where the control could reduce driver surprise and it is estimated to induce benefit

of an order of magnitude of acute accident density value. Objective function has several configurable parameters which can be adjusted to be in better accordance with specific user needs.

Expressing effects in terms of time savings is not only more appealing to the user but makes it possible to further utilise the measure in other, for example, cost-benefit calculations. Due to the utilised objective function which quantifies speed limit effects, the focus of control model is on the speed control. Control model combines the available information in a flexible manner, which accounts for their varying reliability under different situations. The major challenge was to find a model that is understandable and tractable but still accurate enough. The control decision is made in two steps, first at each control station and then coordinated along the corridor. Each control station is supported with one “intelligent” local controller, responsible for making the best control decision with respect to the situation at the section in front. Since harmonisation and warning are based on the different models and goals, each local controller consists of separate warning and harmonisation control model. Harmonisation decision is made on the basis of the modified MARZ Belastung algorithm. To each identified specific traffic state one warning (logit-based) regression model is assigned. Regression-based model is an appealing solution, especially for applications in practice, since it offers the possibility for information “weighting”, an additional mechanism for determining the appropriate trade-off between model bias and variance can be easily incorporated, optimisation can be done computationally efficient, and it is flexible and understandable even with a huge number of control parameters. Other link controls, such weather-related or truck overtaking prohibitions are included in the next step, as defined by the user. Additionally, the global level is supported by several global rules which should lead to the coordination of control actions (for example control propagation in the upstream direction aimed at smoothing speed gradients).

The goal of the optimisation process is to obtain the best possible parameter estimates with a limited amount of data and in such way to ensure model robustness and transferability. To achieve that, exclusion of the ineffective parameters is accomplished with the help of Ridge regression. Determination of the parameter distribution and thus model variance with limited available training data is done using re-sampling method Bootstrap. The search for the optimal parameter vector is done with the downhill simplex optimisation method. Model parameters are optimised through several iterations and the one showing unstable behaviour are excluded from the model. Generation of an ideal control with the help of genetic algorithm is also part of the optimisation process. The generation of the ideal control is based on the assumption that if we know exactly the situation without control in future and what is proper behaviour - it is also possible to determine the right sequence of the control actions. Through the described steps, the integrated optimisation framework is formulated for the calibration of the new control model applicable in the practice.

To test and evaluate INCA, the proposed framework is implemented in the form of a software called INCAS, and installed at Munich Traffic Management Centre. A 38km long section of BAB A8, Salzburg direction Munich, with 20 VMS panels was chosen as the test site. Two months of data was collected and used for calibration and preliminary model tests.

To investigate objective function validity ideal control is generated and evaluated using DR-FAR curves. For better insight into objective function estimation, different empirical examples were analysed. It was shown that the objective function appropriately “rewards” and “punishes” link controls and moves the system towards the desired behaviour. During the calibration process, the influence of several adjustable optimisation settings on the final parameter estimation was investigated.

INCA was tested in Traffic Management Centre in on-line open-loop modus during a two month period (September 15 - November 25, 2004). Data was acquired and control decisions were made in real time, but they were not sent back to the VMS. This way, all situations that could arise during real-time implementation are included in the model evaluation, ranging from the unsynchronised time stamps of data, communication failures to the data corruption.

The controls produced on-line are evaluated using the objective function and DR-FAR curves for complete motorway and each control segment separately. The results of INCA control were compared with the current system and with the ideal control. Additionally the empirical examples with characteristic traffic situations and calculated controls are shown. The INCA system outperforms the existing control significantly, especially with the control of 80km/h and 40km/h. Detection of the incidents of level 80km/h increased by 60%, and with a reduction in false alarms by approximately 10%. A 16% increase in the detection of incidents of speed level 40km/h was measured, with a slight increase in false alarms (4%), while by incidents of speed level 40km/h increased detection and false alarms were observed. According to the objective function, the overall increase of the performances in comparison with current system was 40%, where an even higher improvement of warning and harmonisation benefits was weakened by the increase in costs. The highest improvement was measured at the 2<sup>nd</sup> control group (Weyarn), where a great part of the benefit is due to the improvement in harmonisation control. Highest warning improvements were achieved in the 1<sup>st</sup> control group with 290% or 1sec/veh/km improvement compared to the existing system. At the first control group due to the present control of 100 km/h or 120 km/h very little improvement of the harmonisation was achieved. Finally, there is additional potential of improving INCA performances in the area of 60km/h. The control model includes different simple detection algorithms, and although it could achieve higher performances than the simple combination, it still depends on the algorithms performances.

The overall increase of the warning performances in comparison to the current system is 0.7 sec/veh/km (of total 2.2 sec/veh/km => ~32% improvement). This means that during 2 months period, for the 38km long motorway, with average daily traffic of 40,000, all together 17,333 hours would be saved. Taking that the average hourly wage is 10€/h and the average

cost of injury accident is 50,000€, the said time saving corresponds to an order of magnitude of 3.5 prevented injury accidents in two month period, or 173,330€. If the trend would be constant for the entire year, then this value would correspond to 21 accidents annually.

The model extendibility has been illustrated on the example of the inclusion of indicators from the traffic-flow model based algorithm AZTEK. Its inclusion also revealed some important issues in relation with usability of the traffic-flow based incident detection algorithms in real time operation. In overall, the AZTEK indicators brought the slight increase of the INCA performances. However, in the Weyarn area, with ideal geometry and only one on- and off-ramp, a significant improvement of the performances has been measured. In other areas, with a more “complicated” geometry (e.g. steep grade) or low data quality, AZTEK brought only minor improvements, which can also be partly contributed to the better fitted model parameters. This speaks in favour of the high potential of AID, based on the traffic flow models, provided that it is properly calibrated and data quality is satisfying. It has been shown that INCA was able to recognize the importance of the new indicators and assign appropriate weights. More sensitive indicators are excluded from the model if the geometry or data quality was not satisfactory. In situations when new indicators have been included, other already existing ones such as Belastung or Unruhe have been excluded from the system.

Since the link control manages long motorway corridors, data quality plays an important role in overall system performance and effect. Corrupted data might lead to the instable controls. Moreover, the corrupted data does not only negatively influence the local control station but also the upstream VMS which could be activated due to the “wrong” control propagation rule. It has been shown that even though INCA has an integrated data correction mechanism, it is still sensitive to the data quality.

By reasonable data quality INCA detects incidents earlier and displays appropriate prevention controls for which in the current system an operator intervention is usually required. Thus, in addition to greater safety benefits produced by INCA, there is also the potential for reducing the need for manual (operator) control. Dividing the control into the harmonisation and warning components makes system more tractable. Besides, INCA system gives an overview of the “importance” of the particular algorithm and its reliability. At each moment it is possible to identify which algorithm or group of algorithms was responsible for control activation. The system’s modular structure allows easy extension, re-configuration and further independent development of each component.

The developed model (INCA) is practical and generic in that it can be applied on any motorway route equipped with a traffic data collection system, given a historical database of measurements is available for calibration and validation. Contrary to the existing approaches, INCA offers a systematic framework for the control of link control systems, where for each particular motorway section a general control model can be derived, using the proposed optimisation procedure and objective function.

## **6.2 Future research**

The analysis of the current link control systems and the design, development and evaluation of the new INCA model suggests several directions for future research: improved knowledge about the link control effects; extension and refinement of certain model aspects; usage of additional sources of information; on-line closed loop tests; model application in other areas.

### **6.2.1 Improved knowledge about the link control effects**

Link control systems influence motorway corridors by displaying a variety of information (e.g. the speed limit). This effects the traffic flow in complex manner where it is often difficult to distinguish control effects from spontaneous changes in the traffic flow itself and even more difficult to quantify them. Moreover, although the link controls usually have an enforcement character, it has been proven that driver attitude plays a crucial role in overall control effects. In the worst case scenario even if the control is objectively valid but drivers do not obey it, LCS can have even more negative effects than not using a control at all.

It has been shown in previous long term studies that warning and harmonisation of the traffic flow should lead to a safer and more efficient driver experience. However the knowledge about how the LCS changes the behaviour of drivers and how this in turn influences overall traffic flow performances is still limited. This is especially the case regarding the effects of harmonisation controls, truck overtaking prohibition or weather conditions signs. For example, it is known that link control can reduce standard deviation of the speed by 9 km/h which increases safety but it is not always clear under which situation such positive effects are to be expected. Therefore for the harmonisation it would be of interest to know exactly under which conditions (traffic flow conditions, weather, etc.) the control would have positive effects, and what are these effects in terms of traffic flow characteristics. Further investigation of the warning controls effects is needed also: for example, for the warning controls of interests it would be to know how far away from the location of the incident drivers should be informed (not too early but not too late) so as for the control to have the optimal safety effect and positive long-term effects (not negatively influencing acceptance).

Link control systems (LCS) have existed since the late 60s but they are still one of the least investigated areas of the traffic control field. Understanding and quantifying their interaction with the traffic flow is a challenge which should be resolved in years to come. Improved knowledge about LCS effects would also make possible development of more accurate model based link control framework.

### **6.2.2 Extension and refinement of certain model aspects**

The INCA modular and extendable link control framework offers possibility for easy improvement of each defined component. Basically, with the proposed data-driven approach, as long as a measure of performance is available, the information used can be enhanced.

The objective function captures the safety and efficiency effects of the link control. The effects were estimated by assessing speed limit effects by assuming surprise in case of speed reduction and prevention through flow smoothing in case of dense traffic flow. However, link control has other means to improve traffic flow (e.g. truck overtaking prohibition) which are currently not modelled. Moreover, inclusion of information about speed deviation could help better model link control prevention effects. Finally, the generic form of the proposed objective function offers an expansion possibility by quantifying the effects of the other controls (e.g. ramp metering). Having such extended objective function, would make it possible to develop an integrated control model.

The control model developed in INCA is a compromise between simplicity and accuracy. It was made under the assumption that with classification into traffic states, and separate models of harmonisation and warning control, most parsimony models can appropriately approximate control decisions. Although the control performance analysis shows significant improvements compared to the existing system, there are several ways in which the control model could be improved. Here only two are analysed: *the data fusion model* is kept the same but the *utilised algorithms* and *traffic state estimation* can be further refined and/or another method can be used for *data fusion*.

Currently INCA utilises simple algorithms and indicators. As it has been shown (by AZTEK inclusion), the inclusion of more advanced incident and/or prognostic algorithms or other sources of information (chapter 6.2.3) could further improve model performance. Information about weather or about data quality could be also included in the control model, either in terms of additional “parameter” or new traffic state class. By keeping the current model structure, but including of these new “traffic states” it could be expected that model will recognize information variability under them, thus improving model performances.

Instead of logit-based regression model used in INCA another more non-linear model could be used for data fusion. The more non-linear method would most probably lead to system that is difficult to interpret, but might provide better insight into the further potentials and points for the improvement of the link control model. Here different techniques can be used such as the abovementioned neural networks and fuzzy system. However, both approaches should be supported with methods for model reduction. For example, for the combination of the available data fuzzy control can be used. Initial algorithm sets can be defined by the expert. The rule base can be automatically created with methods proposed in TEODOROVIC AND EDARA, 2005B by utilising at each minute calculated ideal control and algorithms outputs. The proposed method introduces also mechanisms for rule base reduction from redundant or irrelevant rules. Furthermore, fuzzy sets of so derived model can be further adjusted through optimisation with genetic algorithm or neural networks. Finally, the control model has available information about probability of the model being the hit which is not used in the current system. Their utilisation can give valuable information about control confidence and further improve its performances.

Integration of INCA with other motorway control systems is also possible. Ramp metering, or re-routing controls can be included in the decision model either as an additional parameter (algorithm), or by defining a new traffic state. The key point is that in order to make this inclusion systematic, their effects must somehow be modelled in the objective function. The integrated control is still not a sufficiently investigated area and its huge potentials as well as its complexity could be resolved with a systematic data-driven approach such as INCA.

The optimisation procedure supports derivation of the general control model, with techniques such as Ridge regression and Bootstrap. Especially interesting is Bootstrap which allows us to make estimation of the model variance. However, Bootstrap can be also used in the “other direction”, i.e. for estimation of the sufficient number of events (data) for the stable optimisation by estimation of the confidence level of the model parameters, and/or for generating the validation set from limited available data.

Currently, optimisation is performed with the aim of finding the optimal parameters of the control model. However, as shown on example of the Belastung algorithm (harmonisation model), the objective function could also be used for optimising individual algorithms. The investigation of the usage of the objective function for optimisation of the warning algorithms and their performance before and after optimisation is important for improvement of the currently used MARZ and new INCA system, as well as at existing incident detection and management systems.

### **6.2.3 Additional sources of information**

The INCA framework and its approach for utilisation of different information for producing better control action is currently based on data which comes from loop detectors, commonly present at the German motorways. Today, various additional sources of information such as video detectors, traffic eyes (DGG), floating car data (FCD), etc. are available and their great potentials in the area of traffic monitoring and control are reported. This “spatial” information offers better insight into the situation on the stretch (and not only locally at the detector location). INCA has been designed to allow inclusion of these additional sources of information. It is assumed that the proposed model should be able to identify the importance and reliability of this information even though they have different spatial characteristics and frequency from loop-detector data. The difference in spatial characteristics and frequency should previously be handled by the data management and knowledge base component.

### **6.2.4 On-line close-loop tests**

In this thesis, INCA performances are assessed by on-line open-loop tests. Although this provides a quite realistic estimation of INCA performance in real-life conditions, it is still based on the assumption that the current control does not significantly influence traffic in harmonisation area and that breakdowns cannot be prevented but only postponed if demand remains high for a longer period of time. Therefore, the challenge is to test the closed-loop



performances of INCA, where traffic flow and corresponding data are not influenced by some other system.

As part of the European founded project CORVETTE in the year 2006 INCA is to be applied for the control of the motorway A8 Ost Munich direction Salzburg in on-line closed loop tests. In this way it will be possible to assess the full potential of the newly developed INCA framework: the ability to timely anticipate/detect the disturbances in the traffic flow and apply the proper control action; the ability to harmonise traffic flow by reducing the number of disturbances; robustness with respect to data corruption and communication failures; etc..

INCA application on the new motorway can be done in a systematic and automatic way by applying the proposed optimisation framework. To this end data about road geometry, position of detectors and VMS panels and a sufficiently large database of historical data are needed. To test the ability to control the new environment INCA will be tested in 2006 at the motorway control centre Nord-Bayern on some of the chosen motorways (unfamiliar to INCA).

### **6.2.5 Application of INCA framework in other areas**

Due to its general framework, with an objective function which captures the safety and efficiency control effects and the ability to learn from given data, the INCA model: can be used in other areas of traffic management on motorways; could be applied to other networks (urban or rural); could be used in advance in-car systems (e.g. advance driver assistant systems or automatic cruise control). For example, in-car systems could be supported by the information from the INCA model. This information can be used for timely informing the driver about the unexpected event downstream and in this way increasing safety. The difference from the current model is that the control is not disseminated via VMS but via an in-car system. The INCA model could be used for navigation systems, by calculating the best route driven not only by efficiency measure but also by estimating safety impact of the taken decision. The same extension could be applied to re-routing systems and algorithms. In urban networks, the framework similar to INCA, with a modified objective function and utilised algorithms, could be also applied for estimating the best sequence of the traffic signal control.

The future will show that data-driven approaches, if carefully designed and if appropriately model effects, can be applied in many different fields and that they can substantially improve applications in many domains

## Nomenclature

LCS	Link control system
VMS	Variable message sign
MARZ	Currently used link control algorithm in Germany
DR-FAR	Detection vs. false alarm rate: measure of model ability to detect a critical even
TTS	Total time spent in the system: measure of (traffic) system effectiveness
$K, k$	Density of traffic flow [veh/km]
$Q, q$	Traffic flow volume [veh/h]
$V, v$	Average speed of traffic flow [km/h]
$U, u$	Wave speed in [km/h] or [m/sec]
$s_v$	Deviation of averaged measured speed [km/h]
$occ$	Occupancy [%]
$Q_b$	Forecasted weighted traffic flow volume [veh/h]
$Vcar_f$	Forecasted speed of passenger vehicles [veh/h]
$V(m)$	Speed drop of level $m$ , $m=1, \dots, 5$ (120, 100, 80, 60, 40 km/h, where 40 km/h corresponds to sign “congestion”) [km/h]
$V_s$	Average shockwave propagation speed [m/sec]
$t_p$	The time needed for congestion to propagate in upstream direction between two control stations in [sec] or [min]
$T_{p_j}$	Average time needed for congestion to propagate from $VMS_{j+1}$ to $VMS_j$ [sec]
$\Delta X_j$	Distance between control stations $VMS_j$ and $VMS_{j+1}$ [m]
$n(m)$	Average daily number of speed drops of level $m$ ( $V(m)$ )
$f(m)$	The part of all accidents that can be contributed to the speed drops of the level $m$ ( $V(m)$ ), where $0 \leq f(m) \leq 1$
$FA(m)$	Acuteness of speed drop of severity level $m$ ( $V(m)$ )
$U_0$	Average accident time density [sec/veh/km]

$U_m$	Acute accident time density [sec/veh/km]
$V_{down}$	The minimum speed that the driver would experience during travel along the section [km/h]
$G$	Total effects of the control [sec/veh/km]
$B_T$	Total “benefits” of the control [sec/veh/km]
$B_H$	“Benefits” of the control due to the harmonisation control (due to traffic flow smoothing) [sec/veh/km]
$B_W$	“Benefits” of the control due to warning (reduction of “surprise”) [sec/veh/km]
$C_{TT}$	“Cost” of the control – travel time loss caused by control [sec/veh/km]
$V_C$	Control decision - speed limit [km/h]
$V_w(m)$	Warning speed control $m$ , $m=1,\dots,5$ (120, 100, 80, 60, 40 km/h, where 40km/h corresponds to “congestion” sign)
$V_H(m)$	Harmonisation speed control $m$ , $m=1,\dots,3$ (120, 100, 80 km/h)
$t_0$	Time when congestion is formed at the section (downstream from control station)
$t_n$	Time when congestion reached the (upstream) control station
$\alpha$	Importance of the warning strategy, where $0 \leq \alpha \leq 1$
$Q_{min}$	Critical flow per lane: only if the measured flow per lane is above this value LCS control benefits could be expected [veh/lane/h]
$\gamma$	Warning benefit shrinking factor
$V_N$	Speed threshold used for the normalisation of the warning effects [km/h]
$V_I$	Incident speed threshold value: if the measured speed falls below $V_I$ then LCS (warning) effects will no longer be significant since drivers are already driving very slowly [km/h]
$V_{range}$	Allowed difference between measured and harmonisation control speeds: if the difference between measured and controlled speeds is greater than this value no harmonisation benefit can be expected [km/h]
$V_{bias}$	Allowed bias for $V_{range}$ [km/h]

$V_{H\min}$	Minimal speed for which a harmonisation benefit could be still expected: if the measured speed goes below this value harmonisation benefits should not be expected
$S_1$	Free flow traffic state
$S_2$	Synchronized flow traffic state
$S_3$	Congested flow traffic state
$Z$	Warning indicator: based on this value warning control decision is made
$\beta_k$	“Weight” (importance) of k-th algorithm
$\alpha_m$	Decision point for the warning control m
$B$	Bootstrap subsamples
$\mu$	Regularization term by Ridge regression
$Stau1$	$(\beta_{Stau1})$ : “Weight” of Stau1 algorithm
$Stau2$	$(\beta_{Stau2})$ : “Weight” of Stau2 algorithm
$Stau3$	$(\beta_{Stau3})$ : “Weight” of Stau3 algorithm
$Stau4$	$(\beta_{Stau4})$ : “Weight” of Stau4 algorithm
$Belastung$	$(\beta_{Belastung})$ : “Weight” of Belastung algorithm
$Unruhe$	$(\beta_{Unruhe})$ : “Weight” of Unruhe algorithm
$LkwUeV$	(norm): “Weight” of TruckPercentage algorithm
$S2\_V$	$(\beta_{S2\_V})$ : “Weight” of X(1) indicator
$vkDiffQ$	$(\beta_{vkDiffQ})$ : “Weight” of X(2) indicator
$B\_Q$	$(\beta_{B\_Q})$ : “Weight” of X(3) indicator
$B\_KV$	$(\beta_{B\_KV})$ : “Weight” of X(4) indicator
$V_{norm}$	$(\beta_{V_{norm}})$ : “Weight” of X(5) indicator
$vkDiff$	$(\beta_{vkDiff})$ : “Weight” of X(6) ( $vkDiff$ ) indicator
$vbDiff$	$(\beta_{vbDiff})$ : “Weight” of X(7) ( $vbDiff$ ) indicator

## References

- Adeli, H.; Karim, A. (2003): Fast Automatic Incident Detection on Urban and Rural Motorways Using Wavelet Energy Algorithm. Part of the Journal of Transportation Engineering, Vol. 129, No.1, January 1, 2003
- Autobahndirektion Südbayern (2000): Weiterentwicklung des Steuerungsalgorithmus der Streckenbeeinflussungsanlage A8 Ost im Hinblick auf die Umfelddatenanalyse mittels eines Fuzzy-Logik-Ansatzes. Untersuchungsbericht des Instituts für Straßenwesen, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen
- Baher A.; Ritchie G. S.; Iyer M. (1999): Implementation of Advanced Techniques for Automated Freeway Incident Detection. California PATH Research Report
- Banks, J.H. (1990): Flow processes at a Freeway Bottleneck. Transportation Research Record 1287.
- Barcelo, J.; Ferrer, J.; Montero, L.; Perarnau, J. (2000): Assessment Of Incident Management Strategies Using Aimsun. Universitat Politècnica de Catalunya
- Barcelo, J.; Montero, L.; Perarnau, J. (2001): Automatic Detection and Estimation of Incident Probabilities for Incident Management Purposes. A Case Study in Barcelona, IST Congress in Bilbao
- BAST: Bundesanstalt für Straßenwesen (Hrsg): Merkblatt für die Ausstattung von Verkehrsrechnerzentralen und Unterzentralen (MARZ). Bergisch Gladbach, 1999
- BAST: Bundesanstalt für Straßenwesen (Hrsg): Technische Lieferbedingungen für Streckenstationen (TLS). Ausgabe 2002 Entwurf, Bergisch Gladbach, 2002
- BAST: Bundesanstalt für Straßenwesen (Hrsg): Grundlagen streckenbezogener Unfallanalysen auf Bundesautobahnen. Mensch und Sicherheit Heft M153, Bergisch Gladbach, 2003
- Balz, W.; Zhu, Y. (1994): Nebelwarnsystem A8 Hohenstadt-Riedheim. Wirkungsanalyse im Auftrag des Landesamtes für Straßenwesen Baden Württemberg, Stuttgart
- Balz, W. (1995): Wirkungen kollektiver Verkehrsbeeinflussungsanlagen. Straßenverkehrstechnik, Heft 7/95
- Beale, S. (2002): Traffic Data: Less is More, Road Transport Information and Control. Conference Publication No. 486
- Ben-Akiva, M.(1998): Dynamit, a Simulation-based System for Traffic Prediction and Guidance Generation. in Proceedings of TRISTAN III conference, San Juan, Puerto Rico.
- Bishop, C. M. (1995): Neural Networks for Pattern Recognition, Oxford University Press, United Kingdom.
- BMV: Bundesministerium für Verkehr (Hrsg.): Richtlinien für Wechselverkehrszeichen an Bundesfernstraßen (RWVZ). Bonn, 1997a
- BMV: Bundesministerium für Verkehr (Hrsg.): Richtlinien für Wechselverkehrszeichenanlagen an Bundesfernstraßen (RWVA). Bonn, 1997b

- BMVBW: Bundesministerium für Verkehr, Bau- und Wohnungswesen (Hrsg.): Moderne Verkehrsbeeinflussungsanlagen auf Bundesfernstraßen: Stand der Entwicklung und Zukunftsperspektiven. Bonn, 2002a
- BMVBW: Bundesministerium für Verkehr, Bau- und Wohnungswesen (Hrsg.): Programm zur Verkehrsbeeinflussung auf Bundesautobahnen 2002 bis 2007. Bonn, 2002b
- Bogenberger, K. (2001a), Adaptive Fuzzy Systems for Traffic Responsive and Coordinated Ramp Metering. Ph.D. Thesis at the Fachgebiet Verkehrstechnik und Verkehrsplanung, Munich, 2001.
- Bogenberger, K.; Keller, H.; Vukanovic, S.(2001b): A Neuro-Fuzzy Algorithm for Coordinated, Traffic-Responsive Ramp Metering. IEEE, 4th International Conference on Intelligent Transportation Systems, Oakland, CA, USA, August 26-29, 2001
- Bogenberger, K.; Vukanovic, S.; Keller, H.(2002a): ACCEZZ–Adaptive Fuzzy Algorithms for Traffic Responsive and Coordinated Ramp Metering. 7<sup>th</sup> International conference on Applications of Advance Technologies in Transport, Boston, MA, USA, August 5-7, 2002
- Bogenberger K., (2002b): Mit Fuzzy-Algorithmen gegen den Stau. BMW publication [http://www.bmwgroup.com/news/mail/news\\_d.shtml??8\\_5?&pkey=8922&highlight=14&call=4](http://www.bmwgroup.com/news/mail/news_d.shtml??8_5?&pkey=8922&highlight=14&call=4)
- Bolte, F. (1984): Stauwarneinrichtungen an Autobahnen – Konzeption und Erfahrungen. Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik, Heft 422, Bonn
- Box, G.E.P.; Jenkins, G.M. (1976): Time Series Analysis: Forecasting and Control. Holden Day Inc., San Francisco
- Bozic, S. (1979): Digital and Kalman Filtering. Edward Arnold pub., London
- Breiman, L. (1996): Bagging Predictors. Machine Learning, 24(2):123–140.
- Brilon, W.; Regler, M.; Geistefeldt, J. (2005): Zufallscharakter der Kapazität von Autobahnen und praktische Konsequenzen – Teil 1. Organ der Forschungsgesellschaft für Straßen- und Verkehrswesen, der Bundesvereinigung der Straßen- und Verkehrsingenieure der Österreichischen Forschungsgemeinschaft Straße und Verkehr
- Brilon, W.; Drews, O. (1996): Verkehrliche und ökologische Auswirkungen der Anordnung von Überholverböten für Lkw auf Autobahnen. Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik, Heft 731, Bonn
- Busch, F (1971): Verkehrsbeeinflussung auf Bundesautobahnen - Ein Rahmenplan des Bundesministers für Verkehr, Straße und Autobahn, Heft 9/1971
- Busch, F. (1986): Automatische Störungserkennung auf Schnellverkehrsstraßen – ein Verfahrenvergleich. PhD Thesis at Technical University Karlsruhe, 1986
- Busch, F.; Ghio, A. (1994a): Automatic Incident Detection on Motorways by Fuzzy Logic, Siemens AG, Munich
- Busch, F.; Ghio, A. (1994b): Dynamic Traffic State Estimation on Motorways – A Model-Based Traffic Approach, Siemens AG, Munich
- Cassidy, M.J.; Bertini, R.L. (1999): Some Traffic Features at Freeway Bottlenecks. Transportation Research, 33B(1).

- Cedar, A.; Livneh, L. (1982): Relationship between Road Accidents and Hourly Traffic Flow. *Accident Analysis and Prevention*, 14(1): 19-44.
- Cirillo J. A. (1968): Interstate System Accident Research Study II, Interim Report II, Public Roads, Vol. 35, No. 3
- Connolly, T.; Aberg, L. (1993): Some Contagion Models of Speeding. *Accident Analysis & Prevention*. 25(1), 57-66
- Cremer, M. (1981): Incident Detection on Freeways by Filtering Techniques. 8. International IFAC Congress Kyoto
- Cremer, M.; May, A. (1986): An Extended Traffic Flow Model for Inner Urban Freeways. Preprints 5th IFAC/IFIP/IFORS International conference. on Control in Transportation Systems, Wien.147
- Daganzo, C.F. (1997): Fundamentals of Transportation and Traffic Operations. Elsevier Science Inc., New York.
- Daganzo, C.; Laval, J.; Munoz, J.C. (2002): Some Ideas for Motorway Congestion Mitigation with Advanced Technologies, Institute of Transportation Studies, University of California, Berkeley
- Dailey, D.; Harn, P.; Lin, P.-J. (1996): ITS Data Fusion. Final Research Report, Research Project T9903, Task 9, ATIS/ATMS Regional IVHS Demonstration, Washington state department of transportation
- DG-TREN (2002): TEN-T Workshop on: Variable message Signs: the Users' Experience. Valencia, Spain
- Denaes S., Kates R.(2001): Analyse der implementierten Steuerungsverfahren für die untersuchten Anlagen. CORVETTE report
- Denaes S (2002): Bewertung der implementierten Steuerungsverfahren SBA A8Ost, CORVETTE report
- Denaes S., Scharer R., Vukanovic S., Kates R.(2004): Improvements of LCS in Bavaria. Intelligent Infrastructure for Trans-European Network, I2TREN, 2<sup>nd</sup> Conference on Euro-regional Projects, Wien (Austria)
- Di Febbraro, A.; Parisini, T.; Sacone, S.; Zoppoli, R. (2001): Neural Approximations for Feedback Optimal Control of Freeway Systems. *IEEE Trans. on Veh. Techn.*, 50(1):302.312.
- Dougherty, M.S.; Kirby, H.R. (1993): The Use of Neural Networks to Recognize and Predict Traffic Congestion. *Traffic Engineering and Control*, 34(2).
- Dinkel, A. (2004): Simulation der Effekte von intelligenten Streckenbeeinflussungsanlagen auf den Verkehrsablauf auf Autobahnen. Diplomarbeit, Technical University Munich
- Drew, D. (1968): Traffic Flow Theory and Control. McGraw-Hill Series in Transportation
- Efron, B.; Tibshirani, R. J. (1993): An Introduction to the Bootstrap. New York: Chapman & Hall.
- ETSC: European transportation safety council (1999): Intelligent transportation systems and road safety. Brussels

- ETSC: European transportation safety council (2001): Transport accident and incident investigation in the European Union. Brussels
- Evanco, M. V.(1996): The Impact of Rapid Incident Detection on Freeway Accident Fatalities. Mitretrek Report, McLean, VA
- Everts, K.; Zackor, H., (1976): Untersuchung von Steuerungsmodellen zur Verkehrsstromführung mit Hilfe von Wechselwegweisern. Forschung Straßenbau und Straßenverkeherstechnik, Heft 199/1976
- Färber, B.; Färber, Br. (2000): Akzeptanz von Schildern einer variablen SBA. Zwischenbericht im FE 03.321 (BAST), Universität der Bundeswehr München, Institut für Arbeitswissenschaft, Neubiberg
- FHWA: Federal Highway Administration (Hrsg.): Synthesis of Safety Research Related to – Speed and Speed Management. Publication No. FHWA-RD-98-154, July 1998
- FHWA: Federal Highway Administration (2004a): [http://ops.fhwa.dot.gov/congestion\\_report/](http://ops.fhwa.dot.gov/congestion_report/)
- FHWA: Federal Highway Administration (2004b) Traffic flow theory: A state-of-the-art report: <http://www.tfhrc.gov/its/tft/tft.htm>
- Ferrari, P. (1991): The Control of Motorway Reliability. Transportation Research, Vol. 25A, No. 6, p. 419 - 427
- FGSV: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Hinweise für Steuerungsmodelle von Wechselverkehrszeichenanlagen in Außerortsbereichen. FGSV 359, 1992a
- FGSV: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Hinweise für Steuerungsmodelle von Wechselverkehrszeichenanlagen in Außerortsbereichen. Köln, 1992b
- FGSV: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Empfehlungen für Wirtschaftlichkeitsuntersuchungen an Straßen - EWS - Aktualisierung der RAS-W 86. Köln, 1997
- FGSV: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Hinweise für neue Verfahren zur Verkehrsbeeinflussung auf Außerortsstraßen. Köln, 2000
- FGSV: Forschungsgesellschaft für Strassen- und Verkehrswesen (Hrsg.). Handbuch zur Bemessung von Straßenverkehrsanlagen HBS. 2001
- FGSV: Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.): Merkblatt für die Ermittlung der Wirksamkeit von Verkehrsbeeinflussungsanlagen. draft version, working group on Verkehrsführung und Verkehrssicherung, Ausgabe November 2004
- Fildes, B. N.; Lee, S. J. (1993): The Speed Review: Road Environment, Behaviour, Speed Limits, Enforcement and Crashes. Report No. CR 127, Federal Office of Road Safety, Canberra, Australia.
- Golob T.; Recker W. (2002): Relationships among Urban Freeway Accidents, Traffic Flow, Weather and Lighting Conditions. Institute of transportation studies, University of California, Irvine
- Goolsby, M.E. (1971): Influence of Incidents on Freeway Quality of Service. Highway Research Record, 349.
- Haj-Salem; H; Mangeas, M. (1999): Off-line simulation results. DACCORD, deliverable 6.3.



- Haj-Salem, H.; Lebacque, J. P. (2002): Reconstruction of False and Missing Data with First-order Traffic Flow Model. *Transportation Research Record* 1802, 155-165
- Hall, J.W.; Pendleton, O.J. (1989): Relationship between V/C ratios and Accident Rates. Report FHWA-HPR-NM-88-02, U.S. Department of Transportation, Washington, DC.
- Hall, F.L.; Hall, L.M. (1990): Capacity and Speed-Flow Analysis of the Queen Elizabeth Way in Ontario. *Transportation Research Record* 1287.
- Hall, R. (1999): *Handbook of Transportation Science*, Kluwer Academic Publishers
- Hauer E. (1971): Accidents, Overtaking and Speed Control, *Accident Analysis and Prevention*. Vol. 3, No. 1
- Hauer, E. (1991): Comparison Groups in Road Safety Studies: an Analysis, *Accident Analysis & Prevention*., 23, 609
- Haykin, S. (1999): *Neural Networks – A Comprehensive Foundation*. Prentice Hall International, 1999
- Hensher, D.; Green, W. (2003): *The Mixed Logit Model: The State of Practice*. Kluwer Academic Publishers
- Hegyí, A.; De Schutter, B.; Hellendoorn, J. (2003): MPC-based Optimal Coordination of Variable Speed Limits to Suppress Shock Waves in Freeway Traffic, *Proceedings of the American Control Conference*. Denver, Colorado, pp. 4083-4088, 2003
- Holland, J.H. (1975): *Adaptation in Natural and Artificial Systems*. University of Michigan Press. (Second edition: MIT Press, 1992.)
- Hoops M.; Kates R.; Keller H.: Bewertung von Verfahren zur Erkennung von Störungen im Verkehrsablauf in Theorie, Praxis und Simulation. *Forschung Straßenbau und Straßenverkehrstechnik*, Heft 797/2000
- Horter, S.; Schober; M., Wehlan, H. (2004): Advancing an Automatic Incident Detection Algorithm Regarding Ranges of Traffic Intensity. in *proceedings of the Symposium Networks for Mobility*, Stuttgart
- Isaksen, L.; Payne, H.J. (1973): Freeway Traffic Surveillance and Control. *Proceedings of the IEEE*, Volume 61, No. 5.
- ITSA Fact Sheets (1995): *Development of Advanced Traffic Management Systems in the US by State agencies*. Institute for Traffic Safety Analysis, USA
- Jang, J.-S.; Sun, R. C.-T.; Mizutani, E. (1997): *Neuro-fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Inc., New Jersey.
- Jansson, J.O. (1994): Accident Externality Charges. *Journal of transport Economics and Policy*, 28(1): 31-43.
- Jin, W.; Zhang, M. (2000): *Evaluation of On-Ramp Control Algorithms*. University of California at Davis.
- Jin, X.; Cheu, R.; Srinivasan, D. (2000): Development and Adaptation Of Constructive Probabilistic Neural Network in Freeway Incident Detection, *Transportation Research Part C* 10

- Joksch H. C. (1993): Velocity Change and Fatality Risk in a Crash-A Rule of Thumb. *Accident Analysis and Prevention*, Vol. 25, No. 1.
- Jordan, M.; Jacobs, R. (1995): Modular and Hierarchical Learning Systems. In: M. A. Arbib (Ed.) *The Handbook of Brain Theory and Neural Networks*. THE MIT PRESS, Cambridge: Massachusetts, p. 579-582.
- Kates, R.; Keller, H.; Lerner, G. (1999): Measurement-based Prediction of Safety Performance for a Prototype Traffic Warning System. *Proceedings Traffic Safety on Two Continents*, Malmö
- Kates R. (2001): Stochastic Transfer Model for Assessment of the Effectiveness of Variable Speed Limits. Internal report at Fachgebiet Verkehrstechnik und -planung, Technical University Munich
- Kates, R.; Vukanovic, S.; Denaes, S. (2002): Erweiterung und Optimierung der Steuerung von Streckenbeeinflussungsanlagen: Systemanalyse und Konzeption. *Störungsdetektion und -management*. CORVETTE report
- Kates, R. (2004): Integration des Verfahrens AZTEK in das neue SBA-Steuerungsverfahren. CORVETTE report
- Keller, H. (1972): Kosten-Effektivitäts-Analyse der Maßnahmen zur Soforthilfe für Nothalte auf Bundesautobahnen, Schriftenreihe des Instituts für Verkehrsplanung und Verkehrswesen der Technischen Universität München
- Keller, H; Busch, F.; Everts, K.; Müller, F. (1992): Hinweise für Steuerungsmodelle von Wechselverkehrszeichenanlagen in Außerortsbereichen. FGSV, Köln
- Keller, H.; Kämpf, K. (2001): Wirkungspotentiale der Verkehrstelematik zur Verbesserung der Verkehrsinfrastruktur- und Verkehrsmittelnutzung. Schlussbericht, Auftrag des Bundesministeriums für Verkehr, Bau- und Wohnungswesen, Berlin
- Keller (2002a): Verkehrstechnik – Verkehrsablaufs, Verkehrssicherheit und Bemessungsverfahren im Straßenverkehr. Publication at Fachgebiets Verkehrstechnik und Verkehrsplanung, Technical University Munich
- Keller, H. (2002b): Materialien – Verkehrsleitsysteme im Straßenverkehr – Methodik und Geräte zu Erfassung des Verkehrsablaufs. Publication at Fachgebiets Verkehrstechnik und Verkehrsplanung, Technical University Munich
- Keller, H. (2002c): Materialien – Entscheidungs- und Bewertungsmethoden im Verkehrswesen, Publication at Fachgebiets Verkehrstechnik und Verkehrsplanung, Technical University Munich
- Kerner, B (2002): Empirical Features of Congested Patterns at Highway Bottlenecks, *Transportation Research Record* 1802, Paper No. 02-2918
- Kim, Y.; Keller, H. (2001): Zur Dynamik zwischen Verkehrszuständen im Fundamentaldiagramm. Internal publication at Fachgebiets Verkehrstechnik und Verkehrsplanung, Technical University Munich
- Kim, Y. (2002): Online traffic flow model applying dynamic flow-density relation. Publication at Fachgebiets Verkehrstechnik und Verkehrsplanung, Technical University Munich

- Kockelman K.M.; Ma J. (2004): Freeway Speeds and Speed Variations Preceding Crashes, within and across Lanes. Submitted for presentation at the 2004 Annual Meeting of the Transportation Research Board
- Kosko, B. (1992): Neural Networks and Fuzzy Systems: a Dynamical Systems Approach to Machine Intelligence. Prentice Hall, New Jersey.
- Kühne, R.; Anstett, N.; Nabel, D.; Welsch, M. (1995): Streckenbeeinflussung mit abschnittsbezogenen Anzeigesystemen auf Autobahnen und Ermittlung kritischer Bereiche. Auftrag des Bundesministers für Verkehr, Stuttgart
- Labs, U. (1987): Eignung, Erfassung und Auswahl von psychophysiologischen Funktionsgrößen zur Ermittlung der Belastung von Kraftfahrern in Abhängigkeit von Elementen der Straßenverkehrsanlage. PhD Thesis at Hochschule für Verkehrswesen „Friedrich List“ Dresden
- Lange, D.; Struif, R. (1997): Fahrerverhalten im Bereich von Streckenbeeinflussungsanlagen. Straßenverkehrstechnik, Heft 8/97
- Lesser, V. (1999): Cooperative Multiagent Systems: A Personal View of the State of the Art. IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, January/February
- Leutzbach, W. (1972). Einführung in die Theorie des Verkehrsflusses. Springer Verlag, Berlin-Heidelberg.
- Lighthill, M.J.; Whitham, G.B. (1955): On Cinematic Waves; I: Flow Movement in Long Rivers, II: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society, A229, No. 1178.
- Mahmassani H. S.; Haas C.; Peterman J.; Zhou S. (1999): Evaluation of Incident Detection Methodologies. Project report for Texas department of transportation
- Mamdani, E.H. (1976): Advances in Linguistic Synthesis of Fuzzy Controllers. International Journal of Man-Machine-Studies, Vol. 7.
- Martin, P.; Perrin, J.; Hansen, B. (2001): Incident Detection Algorithm Evaluation. University of Utah, Prepared for Utah Department of Transportation
- May, A.D. (1964): Experimentation with Manual and Automatic Ramp Control. Highway Research Record 59.
- May, A.D.; Keller, H. (1968): Evaluation of Single- and Two-regime Traffic Flow Models. Institute of Transportation and Traffic Engineering, University of California.
- Maycock, G. (1997): Speed, Speed Choice and Accidents. In G.B. Grayson (ed.) Behavioural Research in Road Safety VII. Transport Research Laboratory: Crowthorne.
- McCullagh, P.; Nelder, J.A. (1983). Generalized Linear Models. Second Edition. Chapman and Hall, New York, New York.
- Middelham F. (2002): The Dutch Motorway Control System – 21 Years of Evolution. AVV-transport research centre, Ministry of transport, Public works and water management
- Nelles, O. (1999): Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Systems. Lecture notes.
- Newell, G.F. (1995): Theory of Highway Traffic Flow 1945-1965. Course Notes UCB-ITSCN-95-1, Institute of Transportation Studies, University of California, Berkeley.

- OBB: Oberste Baubehörde im Bayerischen Staatsministerium des Inneren (1993): Das Unfallgeschehen auf den Straßen des überörtlichen Verkehrs in Bayern 1993
- OBB: Oberste Baubehörde im Bayerischen Staatsministerium des Inneren (2000): Verkehr und Unfallgeschehen auf Straßen des üblichen Verkehrs in Bayern. Jahresbericht 1999/2000
- O'Day J.; Flora, J. (1982): Alternative Measures of Restraint System Effectiveness: Interaction with Crash Severity Factors. SAE Technical Paper 820798, Society of Automotive Engineers, Warrendale, PA.
- Papageorgiou, M. (1981): Ein hierarchisches Konzept zur Regelung des Verkehrsablaufs auf Schnellstrassen. PhD Thesis at Technical University Munich
- Papageorgiou, M.; Blosseville, J.-M.; Hadj-Salem, H. (1990): Modelling and real-time Control of Traffic Flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. Part II: Coordinated on-ramp metering. Transportation Research A, Volume 24A, No. 5.
- Papageorgiou, M. (1991): Concise Encyclopedia of Traffic & Transportation Systems. Pergamon Press
- Papageorgiou, M. (2001): Short Course on Dynamic Traffic Flow Modelling and Control, 3-7 May 1999, Chania, Greece
- Parr R.E. (1990): Hierarchical Control and Learning for Markov Decision Processes. PhD thesis at University of California at Berkeley, computer science department
- Payne, H.J. (1971): Models of Freeway Traffic and Control. Simulation Council Proceedings, Vol. 1.
- Peeta S.; Anastassopoulos I. (2002): Automatic Real-time Detection and Correction of Erroneous Detector Data Using Fourier Transforms for On-line Traffic Control Architectures. 81<sup>st</sup> Annual Meeting of the Transportation Research Board
- Persaud, B.; Dzbik, L. (1993): Accident Prediction models for freeways, Transportation Research Record 1401, TRB, NRC, Washington, D.C.
- Petty, K.; Ostland, M.; Kwon, J.; Rice, J.; Bickel, P. (2001): A New Methodology for Evaluation of Incident Detection Algorithms. Transportation Research Part C, 10, pp. 189 - 204
- Ponzlet, M. (1996): Auswirkungen von systematischen und umfeldbedingten Schwankungen des Geschwindigkeitsverhaltens und deren Beschreibung in Verkehrsflussmodellen. Schriftenreihe Lehrstuhl für Verkehrswesen Ruhr-Universität, Heft 16, Bochum
- Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. (2002): Numerical Recipes in C++. Cambridge University Press
- Rämä, P.; Kulmala, R. (2000): Effects of Variable Message Signs for Slippery Road Conditions and Driving Speed and Headways. Transportation Research Part F, Vol. 3, Nr. 2, pp. 85 - 94
- Ritchie, S.; Cheu, R. (1993): Neural Network Models for Automated Detection of Non-Recurring Congestion. California PATH Program, Institute of Transportation Studies, University of California, Berkeley

- Reijmers, H: (2003): Automatic Incident Detection, ET4-024 Traffic Guidance Systems 2003 – 2004, Chapter 12
- Ren, J.; Ou, X.; Zhang, Y.; Song, J.; Hu, D. (2002): SVM-Based Detection of Traffic Incident, Institute of System Engineering. Department of Automation, Tsinghua University, Beijing, China
- Ripley, B. D. (1993): Statistical Aspects of Neural Networks, in O. E. Barndorff-Nielsen, J. L. Jensen and W. S. Kendall, eds, 'Networks and Chaos—Statistical and Probabilistic Aspects', Chapman & Hall, London, pp. 40–123.
- Ripley, B.D. (1996): Pattern Recognition and Neural Networks. Cambridge University Press.
- Roy, P.; Abdulhai, B. (2003): GAID: Genetic Adaptive Incident Detection for Freeways. Department of Civil Engineering, University of Toronto
- Sachse, T.; Keller, H. (1992): Einfluss des Bezugsintervalls im Fundamentaldiagrammen auf die zutreffende Beschreibung der Leistungsfähigkeit von Straßenabschnitten. Forschung Straßenbau und Straßenverkehrstechnik; Heft 614, Bonn
- Sanderson, R.W. (1998): Speed Management and the Roadway Environment. Traffic Safety Summit, Kananaskis, Alberta, October 4-7
- Sarle, W. S. (2001): Frequently Asked Questions (FAQ) of the Neural Network Newsgroup, Newsgroup: comp.ai.neural-nets. <ftp://ftp.sas.com/pub/neural/FAQ.html>
- Sarle, W. S. (1995): Why Statisticians should not FART. <ftp://ftp.sys.com/pub/neural/fart.doc>
- Sastry S.; Bodson M. (1989): Adaptive control: Stability, Convergence, and Robustness, Prantice Hall
- Schick, P. (2002): Einfluss von Streckenbeeinflussungsanlagen auf die Kapazität von Autobahnabschnitten sowie die Stabilität des Verkehrsflusses. PhD Thesis draft, Institut für Straßen- und Verkehrswesen, University Stuttgart
- Schnittger, S. (1991): Einfluss von Sicherheitsanforderungen auf die Leistungsfähigkeit von Schnellstraßen. PhD Thesis at Institut für Verkehrswesen, University Karlsruhe, Schriftenreihe Heft 45/91
- Shell (2000): Pkw-Szenarien. Deutsche Shell Aktiengesellschaft Abt. Energie- und Wirtschaftspolitik. Edition: Analysen und Vorträge.
- Sieber, M (2003): Entwicklung einer neuen Methodik zur Bewertung von Algorithmen zur automatischen Störungserkennung auf Autobahnen unter Berücksichtigung der Verkehrszustände. Diplomarbeit, Technical University Munich.
- Siegener, W.; Träger, K.; Martin, K.; Beck, T. (2000): Unfallgeschehen im Bereich von Streckenbeeinflussungsanlagen unter besonderer Berücksichtigung der Verkehrsbelastung. Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik, Heft 787
- Solomatine; D.P. (2002): Data-driven Modelling: Paradigm, Methods. Experiences. Proc. 5th Int. Conference on Hydroinformatics. Cardiff, UK, July 2002, pp.757-763.
- Smulders, S.A. (1989): Application of Stochastic Control Concepts to a Freeway Traffic Control Problem. Control, Computers, Communications in Transportation. Selected papers from IFAC/IFIP/IFORS Symposium, Paris, France, edited by J.-P. Perrin.

- Steinauer, B.; Bölling, F.; Wienert, A. (1999): Wirksamkeitsanalyse der Streckenbeeinflussungsanlage A8 (Ost) zwischen AK München/Brunnthal und AS Bad Aibling. Bericht im Auftrag der Autobahndirektion Südbayern, Aachen
- Solomon, M. (1991): A Review of Automatic Incident Detection Techniques. ADVANCE Program Technical Report
- Steinhoff C.; Kates R.; Keller H. (2001a): Methods for Analysing Observed Compliance Rates to Mandatory Speed Limits. 8<sup>th</sup> ITS-World Congress, Sydney, Australia
- Steinhoff C; Kates R; Keller H; Färber B; Färber B (2001b): Problematik präventiver Schaltungen von Streckenbeeinflussungsanlagen. Bundesanstalt für Straßenwesen, Forschungsvorhaben FE 03.321/1998/IRB
- Steinhoff C (2002); Online-Bewertung der Akzeptanz und der Wirksamkeit präventiver Maßnahmen durch Streckenbeeinflussungsanlagen. PhD Thesis at Fachgebiet Verkehrstechnik und Verkehrsplanung, Technical University Munich
- Sullivan, E.C. (1990): Estimating Accident Benefits of Reduced Freeway Congestion. Journal of Transportation Engineering, ASCE, 116(2): 167-180.
- Teodorovic, D. (1999): Fuzzy Logic Systems for Transportation Engineering: the state of the art. Transportation Research A, 33.
- Teodorovic, D.; Edara, P. (2005a): Highway Space Inventory Control System. In Transportation and Traffic Theory: Flow, Dynamics and Human Interaction (Editor Hani Mahmassani), Elsevier (in press).
- Teodorovic, D.; Edara, P. (2005b): A Real-Time Road Pricing System: The Case of Two-Link Parallel Network. Computers & Operations Research (accepted).
- TRB: Transportation Research Board (2000a): Special Workshop on Speed Management. TRB, 79th Annual Meeting, January 9, 2000, Washington, D.C.
- TRB: Transportation Research Board (2000b): Highway Capacity Manual (HCM). National Research Council, Washington, D.C.
- TSS Transport Simulation Systems (1998): GETRAM AIMSUN 2 Version 3.2. Users manual, first draft.
- USDOT: United States. Department of Transportation (1996): Safety Management Systems: Good Practices for Development and Implementation'. Available at x, Office of Safety, Room 3407, 400 Seventh St., SW., Washington, DC 20590, or electronically.
- Van Lint, J.W.C. (2004): Reliable Travel Time Prediction for Freeways. Trail Thesis Series no.T2004/3, The Netherlands TRAIL Research School
- Vickrey, W. (1969): Congestion Theory and Transport Investment. American Economic Review, 59(2): 251-260.
- Vukanovic S., Stojkovic O.; Petrovic B.(1999): Applying a Choquee Fuzzy Integral in Reliability Analysis. SINFON, Belgrade, Serbia
- Vukanovic, S.; J. Etz; Bogenberger, K. (2001). A Fuzzy Controller in C++. Internal report at Fachgebiet Verkehrstechnik und Verkehrsplanung, Technical University Munich.

- Vukanovic S.; Kates R.; Daneas S.; Keller H. (2003a): The new methodologies for evaluation of AID and traffic management systems. 10<sup>th</sup> ITS World Conference, Madrid (Spain), November
- Vukanovic S.; Denaes S.; Kates, R. (2003b): Software handbuch: SW-Implementierung des neuentwickelten Verfahrens zur verbesserten SBA-Steuerung für on-line Test im offenen Regelkreis in der VRZ Freimann. Internal CORVETTE report
- Vukanovic S.; Kates R.; Daneas S. (2004): Kalibrierung und Test der neuen Steuerungsmodell INCA im offenen Regelkreis. CORVETTE report
- Vukanovic, S.; Kates, R.; Busch, F. (2005b): Ein intelligentes Modell zur Streckenbeeinflussungsanlagen und ein empirisches Verfahren zur Optimierung im praktischen Einsatz. Heureka Tagungsbericht, Forschungsgesellschaft für Strassen- und Verkehrswesen
- Wall, M. (1996): GALib: A C++ Library of Genetic Algorithm Components, version 2.4. Documentation Revision B, Massachusetts Institute of Technology
- Wang, Y.; Sisiopiku, V. (1998): Review and Evaluation of Incident Detection Methods. Proceedings of the 5th World Congress on Intelligent Transport Systems, Seoul, Korea
- Wang Li-Xin; Mendel J.M. (1992): Generating Fuzzy Rules by Learning from Examples. IEEE transactions on systems, man, and cybernetics, vol. 22., no. 6
- Wardrop, J. (1952): Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers, II/1.
- West, L. B.; Dunn, J. W. (1971): Accidents, Speed Deviation and Speed Limits. Traffic Engineering 11 (41), Washington D.C.
- Yen, J.; Langari, R. (1998): Fuzzy Logic: Intelligence, Control, and Information. Prentice Hall, Inc, New Jersey.
- Zackor, H. (1972): Beurteilung verkehrsabhängiger Geschwindigkeitsbegrenzungen auf Autobahnen. Forschung Straßenbau und Straßenverkehrstechnik, Heft 128, Bonn
- Zackor, H.; Schwenzer, G. (1988): Beurteilung einer situationsabhängigen Geschwindigkeitsbeeinflussung auf Autobahnen. Schriftenreihe Forschung Straßenbau und Straßenverkehrstechnik, Heft 532, Bonn
- Zackor, H.; Keller, H. (1999): Entwurf und Bewertung von Verkehrsinformations- und Leitsystemen unter Nutzung neuer Technologien. Straßenverkehrstechnik, Heft 9/99
- Zadeh, L. (1965): Fuzzy sets. Information and Control 8, 338-353.
- Zadeh, L. (2002): Toward a perception-based Theory of Probabilistic Reasoning with Imprecise Probabilities. Journal of Statistic Planning and Inference 105, 223-264
- Zhang, X.; Busch, F.; Blosseville, J.-M. (1995): Guidelines for Implementation of Automatic Incident Detection Systems. Deliverable No. AC10 – Part2, Commission of the European Communities – R&D Programme Telematics Systems in the Area of Transport (DRIVE II)
- Zografos, K.; Androutsopoulos; K.; Vasilakis, G. (2000): A real-time Decision Support System for Roadway Network Incident Response Logistics. Transportation Research Part C 10

## A. Appendix

### A.1. Automatic incident detection/estimation/prediction algorithms

In recent decades there has been great amount of literature and projects on the subject of automatic incident detection algorithms and their performances. For detail overview of the algorithms, their characteristics, reported effects, etc. please refer to BUSCH, 1986; BAHER ET AL., 1999; ADELI AND KARIM, 2003; JIN ET AL., 2000; WANG AND SISIPIKU, 1998; ZHANG ET AL., 1995, RITCHIE AND CHEU, 1993; etc. AID algorithms could be classified according to:

1. the technology they use: loop detector based, video detection, automatic vehicle identification, floating car data, etc.
2. the manner in which they deal with the spatial characteristics of traffic:
  - a. (local) point-based algorithms that use information from one detector only (such as many comparative or statistical approaches), or
  - b. using information from other detectors along the road (e.g. model-based algorithms)
3. the methodology they use:
  - a. comparative
  - b. statistical-based,
  - c. time series and smoothing/filtering,
  - d. model-based algorithms (traffic-model and theoretical algorithms)
  - e. artificial intelligence algorithms, and
  - f. decision-support algorithms - that combine two or more algorithms from different classes.

Due to the historical development, where in the past only static data collection systems were available, the majority of developed algorithms are intended for work with loop-detector data. The typical problem with macroscopic data is that it also aggregates the effect of incidents. Thus detecting incidents is not only difficult but could also lead to many false alarms. In the meantime, different new technologies have been developed offering good possibilities for detecting an incident in a more microscopic manner. Lately studies have illustrated the great potential of video detection based, or automatic vehicle identification, etc., algorithms. They also report more stable performances compared to those based on static data collection systems. However, here we are interested in the algorithms based on classical data collection systems such as loops, since these are usually present in the existing systems, and will present them based on the methodology they use. It should be noted that basically each algorithm (independent of technology used) delivers an “alarm” that reports whether some kind of incident has been detected or not. In that sense all these algorithms provide the *same final information*. The *difference* however lies in the frequency of information, its reliability and



the level of details it provides (does it report only an incident or some specific kind of incident). Finally a brief discussion about algorithms performances will be given.

*Comparative or Pattern Comparison Algorithms* are based on the assumption that traffic patterns changes when an incident occurs and is different from the normal ones (WANG AND SISIOPIKU, 1998). The core logic of these algorithms is based on comparing observed upstream and/or downstream traffic data, within and between lanes, with pre-established threshold values to declare the occurrence of an incident. These are perhaps incident detection algorithms most widely used in practice and include all California-type algorithms, Pattern Recognition (PATREG) algorithms, the Minnesota algorithm, MONICA, APID, Stau1 and Stau2 (MARZ), etc. Their advantage is that they are fairly simple to operate and understand, the calibration effort is much smaller in comparison with more complex algorithms (though still not simple), and they do not require complex data preparation and sources. Due to their simplicity however, they are not able to capture all the dynamics of traffic flow, and the probability of false alarms is high. Furthermore, due to their fixed threshold values, they are usually able to capture only specific kinds of incidents, and are sensitive to the data quality.

*Statistical algorithms, Time series algorithms and smoothing algorithms* are essentially based on the same logic. They all compare the measured values with the estimated ones, and depend on the difference (that is by statistical approaches determined by confidence interval) to determine whether an incident exists or not. Statistical approaches obtain confidence limits using a representative incident-free database. Instead of using one fixed threshold, this type of algorithm considers the rate of change of the control traffic variables, and keeps updating it based on previous intervals (BAHER ET AL., 1999). The *Standard Normal Deviate algorithm* and *Bayesian algorithm* belong to the category statistical algorithms. *Auto-Regressive Integrated Moving-Average Algorithm (ARIMA)* is a *time series algorithm*, and is used to develop short-term forecasts. Another time series algorithm is the UK *High Occupancy (HIOCC)*, which checks occupancy data from individual loop detectors for the presence of stationary or slow-moving vehicle. MARZ algorithms STAU3 and STAU4 also belong to this group. But this method expands the time of detection. Most *smoothing algorithms* used for incident detection purposes can be expressed as a single or double smoothing function. Representatives are *Dutch MCSS algorithm*, *low-pass filtering*, *(Double) exponential smoothing algorithm*. In general, these algorithms (except smoothing ones) require large historical databases and their results strongly depend on the data in the database. The algorithm performance depends on the prognostic capability. In order to reduce possibility of false alarms due to the fluctuations in traffic variables they usually use persistence checks or tests for compression waves to delay sounding an alarm until the detection thresholds are exceeded for a specified time period. Although more resistant than comparative algorithms they are still sensitive to the data quality.

*Traffic model and theoretical algorithms.* These algorithms use complex traffic-flow theories to describe and predict traffic behaviour during incident conditions. Actual traffic parameters

are compared to those predicted by the model. Kalman filter (model based) algorithms (HORTER ET AL., 2004) assume that in the vicinity of an incident traffic flow behaviour significantly deviates from the one produced by the model. Theoretical algorithms, such as McMaster, are based on the premise that while speed changes sharply when traffic changes from a congested to uncongested state, flow and occupancy change smoothly. Using historical flow-occupancy relationships during changes from congested to uncongested traffic, the algorithm develops a volume - occupancy template. The template is composed of four areas, each corresponding to a particular traffic profile. The algorithm logic performs two comparison tests between the template and actual loop data. The catastrophe theory algorithm (modified McMaster), Low volume incident detection (for detecting incident under low volume conditions), etc. also belong here. These algorithms offer valuable information and can also detect incidents that do not induce reduction of speeds along the motorway. However, they require extensive calibration effort and model-based algorithms require a large amount of input data. Furthermore, missing data will cause model-based algorithms to restart, and wait several minutes to warm-up before they are able to produce output again (VUKANOVIC ET AL., 2004). This can be critical in practical applications where data is often missing or corrupt, and detection time plays an important role for successful warning of drivers.

*Artificial intelligence* approaches try to exploit features of methods from artificial intelligence field (fuzzy logic, neural networks, genetic algorithm, logit model), such as handling of uncertainty, imprecision or vagueness and ability to learn from previous experience. There are several algorithms based on fuzzy logic (Fuzzy Set algorithm, Fuzzy ART and Fuzzy ARTMAP) that are, based on the pre-specified sets of fuzzy variables and fuzzy rules, determining if there is an incident or not (or deliver incident probability). Neural network approaches are based on its possibility to approximate non-linear systems by learning from available data. There are different approaches in literature, the most advanced of which is probabilistic neural networks (PNN) algorithm, which have the ability to incorporate prior probability of occurrence, road conditions, and the cost of misclassification (RITCHIE AND CHEU, 1993). Logit-based models attempt to recognize incident patterns by using incident index. The incident index, which represents the probability of occurring incidents, is estimated by a multinomial logit model. When the two distribution curves of traffic variables (speed), for normal and incident conditions overlap, it gives rise to a large area of uncertainty in that region (BARCELO ET AL., 2000). Artificial intelligence approaches have an advantages in handling uncertainty and imprecision, as well as the possibility to learn from available data. However, they usually require huge training database and a significant calibration effort is needed.

Decision-based incident detection algorithms are actually based on a combination of two or more algorithms from the different classes listed above. It is believed that each algorithm has better performances in a specific situation (e.g. under low flow traffic situation), and by combining them one could exploit their advantages (good detection) and reduce disadvantages

(false alarms). In the simple form they are usually rule-based approaches, where combination of algorithms outputs is achieved with IF-THEN rules. For example, HERMES combines California #9 and extended Kalman filtering techniques for determining if there is an incident. A more complex algorithm is FuzzyAID, which utilises trend factor, speed-density difference and platoon detection and makes decision with the help of fuzzy sets and rules. The output from the system is the probability of an incident (BUSCH AND GHIO, 1994A). There are also techniques utilising Bayesian Belief Networks or multinomial logit models. These algorithms require an extensive calibration process, which often relies on expert knowledge and interpretation. They have the possibility to detect a wider spectrum of incidents but, if not properly calibrated, could also result in more false alarms.

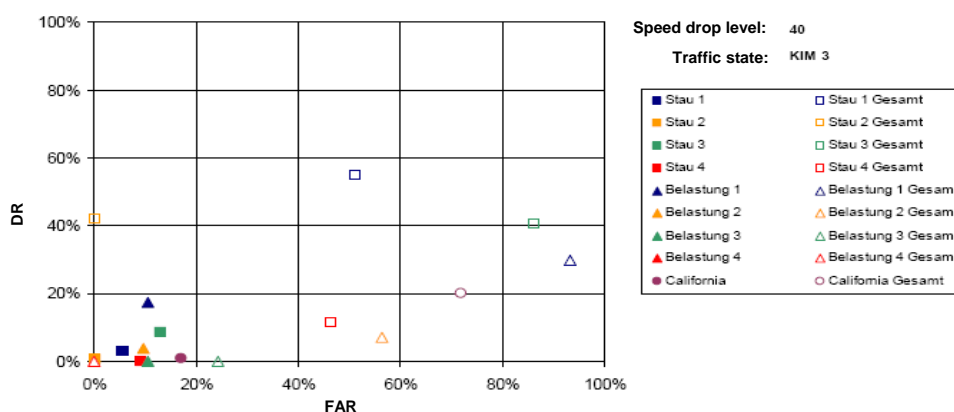
*Estimation algorithms* do not attempt to detect an incident but situations that could lead to an incident. This could be seen as harmonisation algorithms, usually detecting situations in high flow where instabilities could lead to the disturbances. In Germany these are Belastung, Unhrue and truck overtaking prohibition (TruckPercentage) algorithms (chapter 3.5.1). Belastung is based on the density-flow relationship, and when the measured flow and/or density or speed reaches threshold value it triggers an alarm. Unhrue is based on the idea that high speed deviation is indicator of instable traffic conditions and triggers an alarm when it reaches the threshold value. Finally, when percentage of truck traffic reaches the threshold value truck the overtaking prohibition algorithm will be activated.

### **Algorithm performances**

Although there are many researches on the subject of AID and many algorithms have been developed, the systematic comparative evaluation of their performances is still missing. An additional difficulty is the fact that most algorithms are presented alone in the literature and in such a way that meaningful, literature based comparisons between AID's are nearly impossible (PETTY ET AL., 2001). By analysing the qualitative conclusions from these reports it could be concluded that the common problems of the above listed AID algorithms is their high false alarm rate, sensitivity to bad data and/or need for extensive calibration. In general, while less complex algorithms are not flexible enough, more complex algorithms require an extensive tuning process. Furthermore, algorithms (especially comparative) that perform well at one location could, even calibrated, have much worst performances at the other. We will summarize some reasons why the AID algorithms are frequently found too problematic for implementation in practice:

- Sometimes algorithms are developed with sets of data from some specific location, thus inherently representing traffic flow specific characteristics for that location (e.g. composition of traffic, geometry). It can happen that they are not flexible enough and could not be transferred to other location at all. However, even if they can be transfer they need to be recalibrated.

- Even when algorithms are configurable and can be tuned by setting appropriate threshold or parameter values, this process relies on the expert knowledge. Usually used is the multicriteria measure of detection versus false alarm rates (see Appendix A.2) which represents the well-known Pareto optimum problem. Thus, setting parameters can be very difficult in practice, and performance can be very sensitive to these settings. Since poor performance usually translates into an unacceptably high rate of false alarms, many practitioners find AID's too problematic for implementation in practice (PETTY ET AL., 2001).
- Data quality strongly influences detection performances. Depending on how the corrupted data is handled, it could result in false alarms and/or restarting of more complex algorithms (e.g. model based) (HOOPS ET AL., 2000).
- Finally, AID algorithms are usually developed to detect some specific situation – e.g. incident in free-flow or in synchronized flow, etc. Thus, an algorithm might have very good performances under some specific conditions but averaged, over complete spectrum of situations, their performances might be unsatisfactory. In Figure A-1, the example of how performances of algorithms could vary under different situations is given (SIEBER, 2003). From the complete data set, only records that represent synchronized traffic state (situation close to capacity) (KIM, 2002) are taken; then only these records are used to assess algorithm capability to detect severe incidents (causing speed to drop under 40 km/h). As it can be observed, in this situation algorithms such as Stau4 or California produce only false alarms. Intuitively, ignoring their alarms in a synchronized state would lead to an increase in their overall performances.



**Figure A-1: Variation of algorithm performances under different traffic flow conditions. The algorithm's capability to detect severe incidents (resulting in speeds under 40 km/h) while the traffic at detector location is still in synchronized traffic state, is shown in this figure. Unfilled symbols represent algorithm performance in total (over complete evaluation period). Full symbol represent performance under synchronized traffic state only. As it can be seen, under these circumstances, algorithms Stau4, Belastung3 and California produce only false alarms (SIEBER, 2003).**

## A.2. DR-FAR

The importance of two-dimensional (DR-FAR) representation was recognized some time ago (BUSCH, 1986; HOOPS ET AL., 2000) and it still continues to be a useful tool for users and traffic operators. Detection versus false alarm rate curve is widely used for assessing of AID performances. Since they reflect the algorithm's ability to detect, they could be transferred in the link control field and taken to be the warning ability of link control. The detection rate represents this ability. At other side, alarms (or controls) that are activated without the need (there were no safety critical event at the road) are classified as false. False alarms also reflect the control negative impact on traffic efficiency (slowing down of traffic for no reason). Usual<sup>38</sup> definitions of DR and FAR are:

$$DR = \frac{n_d}{N_i} * 100\%, \quad FAR = \frac{n_{far}}{N_a} * 100\%, \quad \text{Eq. 0-1}$$

$n_d$  - number of detected incident cases,

$N_i$  - total number of incident cases,

$n_{far}$  - number of false alarms,

$N_a$  - total number of alarms (decisions).

A link control algorithms generally involves dynamic comparison of a quantity derived from the incoming traffic data with one or more threshold values. A shift of threshold results in a tradeoff between DR and FAR. This tradeoff is related to the well-known statistical concept of receiver operating characteristic as illustrated in the Figure A-2, in which the tradeoff between FAR and DR is expressed by a curve in a two-dimensional representation connecting FAR-DR points for different values of specific control parameters. The nearer the curve is to the top left corner of the graph, the better the intrinsic quality of link control warning capability (VUKANOVIC ET AL., 2003A).

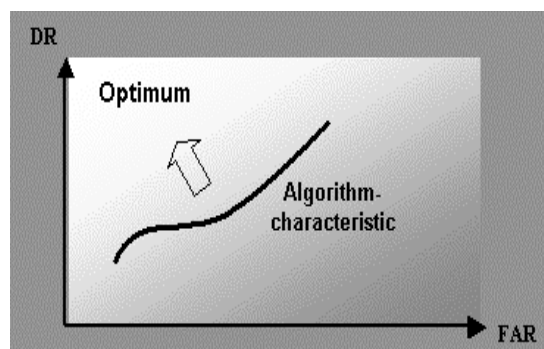


Figure A-2: DR-FAR curve

Due to the importance of timely warning, in addition to DR and FAR, it is necessary to add the measure that would reflect control efficiency. As a measure of control efficiency a total or mean time to detect is usually used (TTD or MTTD respectively). TTD is the difference between the time when the incident was detected (drivers were warned) and the actual time that the incident occurred. MTTD is its averaged value over all incidents. Analogue to the trade-off between DR and FAR there is also the trade-off between FAR, DR and TTD. While a longer detection time permits an algorithm to analyze more data, increasing detection rates and reducing false alarms, a longer detection time also results in a greater impact on traffic.

<sup>38</sup> Please note that these definitions may vary in literature

The importance of a two-dimensional representation has been recognized for some time and continues to be a useful tool for users and traffic operators. However, this approach, as previously applied, has several potential pitfalls that were not addressed until recently (VUKANOVIC ET AL., 2003A):

- First, the calculation of performance requires an a priori judgment by the researcher *as to which of all the reported disturbances constitute real incidents for which detection should be achieved* (BUSCH, 1986; PETTY ET AL., 2001). This subjectivity can lead to separate studies classifying different events as incidents on the same data set. The same interpretation problem occurs when identifying an alarm as false or valid. Consequently, fair comparisons between different algorithms were nearly impossible solely from literature, and evaluation results were always relative to the specific chosen methodology.
- A second potential issue is that it is unreasonable to treat all *incidents with equal importance, i.e.*, it is desirable to distinguish between failing to detect a major incident and failing to detect a low-impact breakdown (BUSCH, 1986; HOOPS ET AL., 2000).
- A third issue is that performance criteria appropriate for harmonization and stabilization strategies typical in complex traffic management systems differ from those appropriate for incident detection algorithms (VUKANOVIC ET AL., 2003A).
- Balancing the DR and FAR (and sometimes MTTD) represents the well known Pareto optimal problem (see PAPAGEORGIOU, 2001), where any point along the characteristic curve (Figure A-2) is a potential operating point of the algorithm. Therefore, the calibration of rule-based algorithms involves testing different parameter values until the point, where the increase in detection rate does not lead to a large increase in false alarm rate, is determined. However, usually it is up to the expert to decide what the large increase in FAR is. Hence, an optimisation routine is necessary when there are more than two parameters that require calibration.

Some of the above mentioned potential pitfalls have been addressed in BUSCH, 1986. The more detail one-dimensional measure is developed that incorporates incident importance (severity) and warning efficiency (time to detect):

$$DI = \frac{\sum_{i=0}^n E_i * G_i}{\sum_{i=0}^n G_i} * 100$$

$n$  – number of event in data

$G_i$  - weight of the reason for congestion ranging from 0,6 for total closure to 1 for only speed drop, in step 0,1.

$E_i$  - “goodness of detection” for one incident =  $S_i * (1 - DT_i)$

$S_i$  – Importance of alarm (0=not detected, 0,5=not significantly detected and 1 = detected)

$DT_i'$  - relative time to detection

$$\begin{cases} 0 - & DT \leq DT \text{ min} \\ 0,5 - & DT \text{ min} < DT \leq DT \text{ max} \\ 1 - & DT > DT \text{ max} \end{cases}$$

DT – Time to detection (s)

DT<sub>min</sub> – minimal time to detection (60s)

DT<sub>max</sub> – maximal time to detection (600s)

However this is a microscopic measure, which is not applicable in operation with macroscopic (aggregated) data that is often the only available data in practice. Inherently, during link control operation we often do not know the reason for the incident, and its real severity.

Therefore, based on the work from BUSCH, 1996 in HOOPS ET AL., 2000 new DR-FAR concept, appropriate for quantifying link control warning capabilities and capable of working with aggregated data, has been introduced. In the function the time to detect is directly included in the DR and FAR calculation. Furthermore, the speed drop levels have been introduced in order to distinguish between incident severities. DR and FAR are calculated using the following formula

$$DR(c, s) = H(c, s) / N(s)$$

$$FAR(c, s) = F(c, s) / A(c)$$

DR(c,m) – Detection rate: percentage of control c that has been followed by speed drops of level m,

FAR(c,m) – False alarm rate: percentage of all controls equal or greater then c that has not been followed by speed drop m,

H(c,m) – number of hits: Control c is considered a hit if it detected speed drop of level (m) in appropriate time interval  $T_{detect}$ ,

$T_{detect}$  – time in which warning about speed drop should be activated in order to be considered successful (5min),

N(m) – number of speed drops of level m, plus reported accidents and road works (if information is available)

$F(c,m)$  - Number of false alarms of the control (c) under speed drop of level m. Control is considered a false alarm if  $T_{before}$  minutes before (usually 5min) and  $T_{after}$  minutes after (usually 10min) control there was no speed drop.

$A(c)$  - Total number of controls c or higher

This measure seems reasonable for assessing of link control warning performances. It includes the systematic classification of incident severities (adjusted to be applicable with available data) and classification of the speed controls. However, with the respect to the link control problem, it still has some **important shortcomings**: it is still the Pareto optimum problem that requires some additional transformations (e.g. weighting) or measures to be included, travel time loss can not be precisely determined, and it does not include link control harmonisation performances.

### A.3. Downhill Simplex

Downhill simplex search requires the construction of an n-dimensional simplex in parameter space. A simplex is the geometrical figure consisting, in N dimensions, of N+1 points (or vertices) and all their interconnecting line segments, polygonal faces, etc. (for example, in two-dimensional space simplex is triangle, in three dimensional tetrahedron, etc.). A simplex is defined through specification of N+1 non-degenerated points, and thus all corresponding lines are linearly independent. In such way an n-dimensional area in parameter space is formed. Each simplex defines a solution in the search space.

Since initial simplex is defined by N+1 points, to start the method this points should be defined. First the initial  $P_0$  point is chosen; then the other N points can be expressed by:

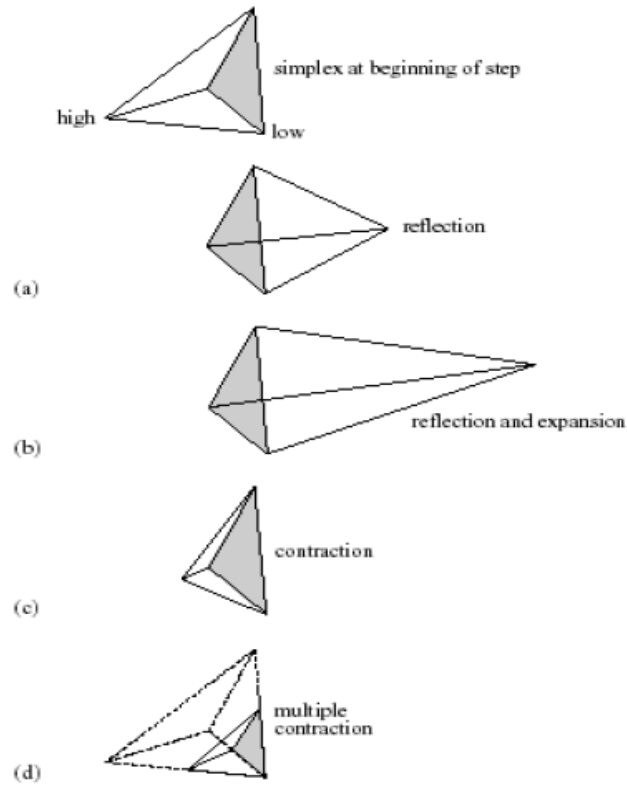
$$P_i = P_0 + a_i * e_i$$

where  $e_i$  are N unit vectors, and  $a_i$  are constants that characterize the length scale for each vector direction. In INCA several  $a_i$  values are used, depending on the optimised variable and their scales. For the normalised algorithms output  $a_i = 0.15$ , for values from armonisation model (chapter 3.6.2.3):  $Q_{bup}(m)$  and  $Vcar_{fup}(m)$  values  $a_i=30$ , and for  $k_{up}(m)$   $a_i=5$ .

The algorithm is then supposed to take its own way downhill until it encounters a (at least a local) minimum - hence the name downhill simplex. The method takes a series of steps, called reflections. In each reflection procedure the simplex is changed with respect on minima/maximal function values found on the corners of the simplex, where the new simplex corner points are determined by linear combinations of selected existing corner points. If the reflection yields a loss function decrease, an expanded reflection is performed; if it yields the objective function increase, a contracted reflection is performed. There are of course several combinations of the above. The series of *random* steps goes as follows. First, it finds the point



where the objective function is highest (high point) and lowest (low point). Then it reflects the simplex around the high point. If the solution is better, it tries an expansion in that direction, otherwise if the solution is worse than the second-highest point it tries an intermediate point. If no improvement is found after a number of steps, the simplex is contracted, and started again (PRESS ET AL., 2002).



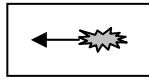
**Figure A-3: Downhill simplex method (PRESS ET AL., 2002)**


In Figure A-3 possible outcomes for a step in the downhill simplex method in a three dimensional case  $f(x_1, x_2, x_3)$  are shown. The simplex at the beginning of the step, here a tetrahedron, is shown at the top of the figure. The simplex at the end of the step can be any one of (a) a reflection away from the high point, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions towards the low point. An appropriate sequence of such steps will always converge to a minimum of the function. According to (NELLES, 1999) this expansion and contraction procedures allows a significant convergence speed up.

The methods could be terminated when the decrease in the function value in the terminating step is fractionally smaller than some tolerance  $f_{tol}$  (usually 0.00001), or when predefined number of iterations is reached. However, the  $f_{tol}$  (or some other) criteria might be fooled by a single anomalous step that, for some reason, failed to get anywhere. Therefore, it is usually recommended to restart a multidimensional minimisation procedure at the point where it claims to have found a minimum.

### A.4. Accidents, road works and weather conditions

Accident and road works reports produced at the VRZ have been extracted and are plotted on the following two diagrams (Figure A-4 and Figure A-5). Road works are marked with orange blocks. Accidents are represented with the symbol:



Where  represents the location where the accident happened (according to the police reports), and the arrow goes upstream up to the location (control station) where accident affected the traffic flow. The symbol for accident is located on the y-axis so to correspond to the reported time of occurrence.

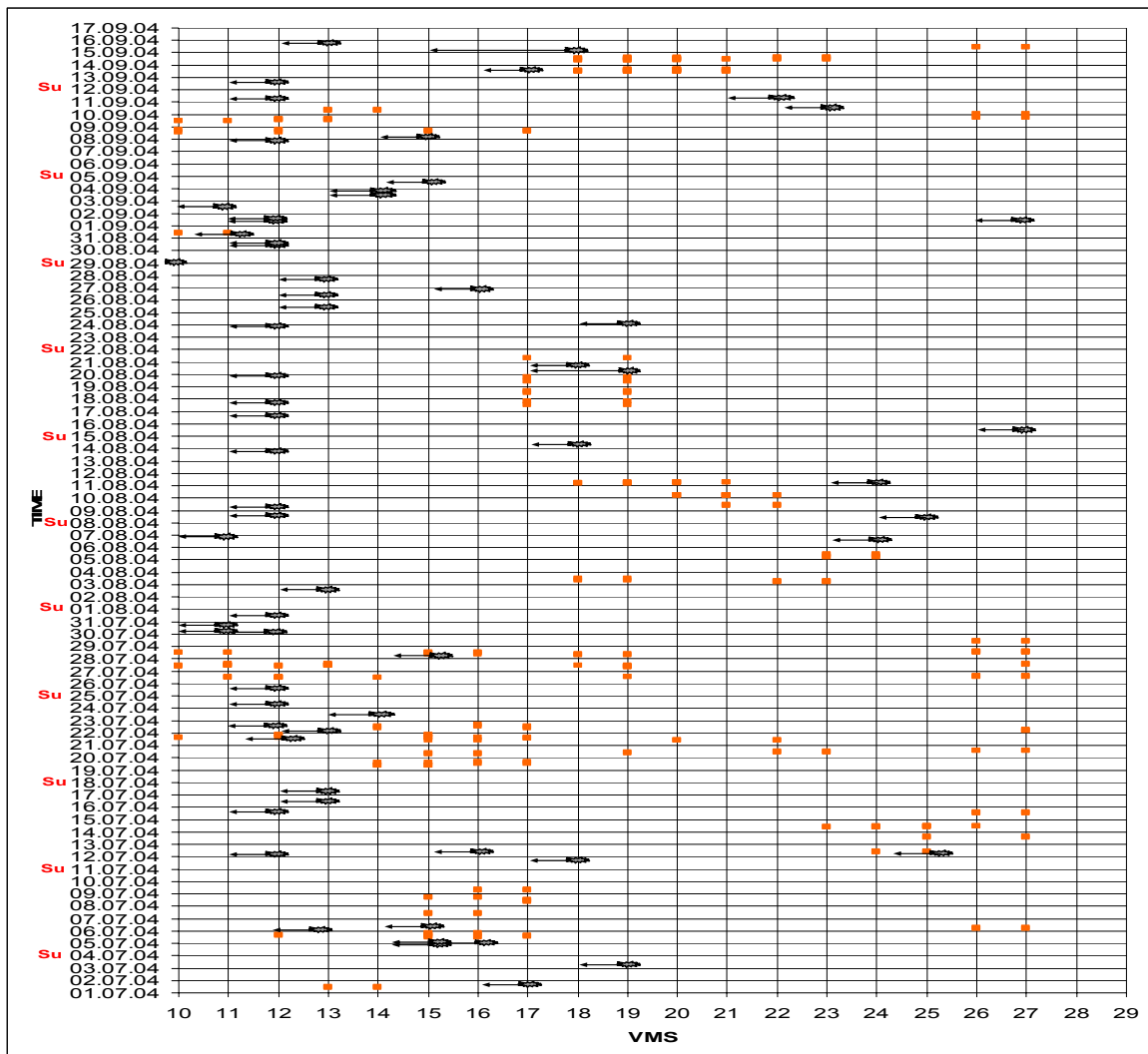


Figure A-4: Accidents and road works between July 1 to and September 17, 2004

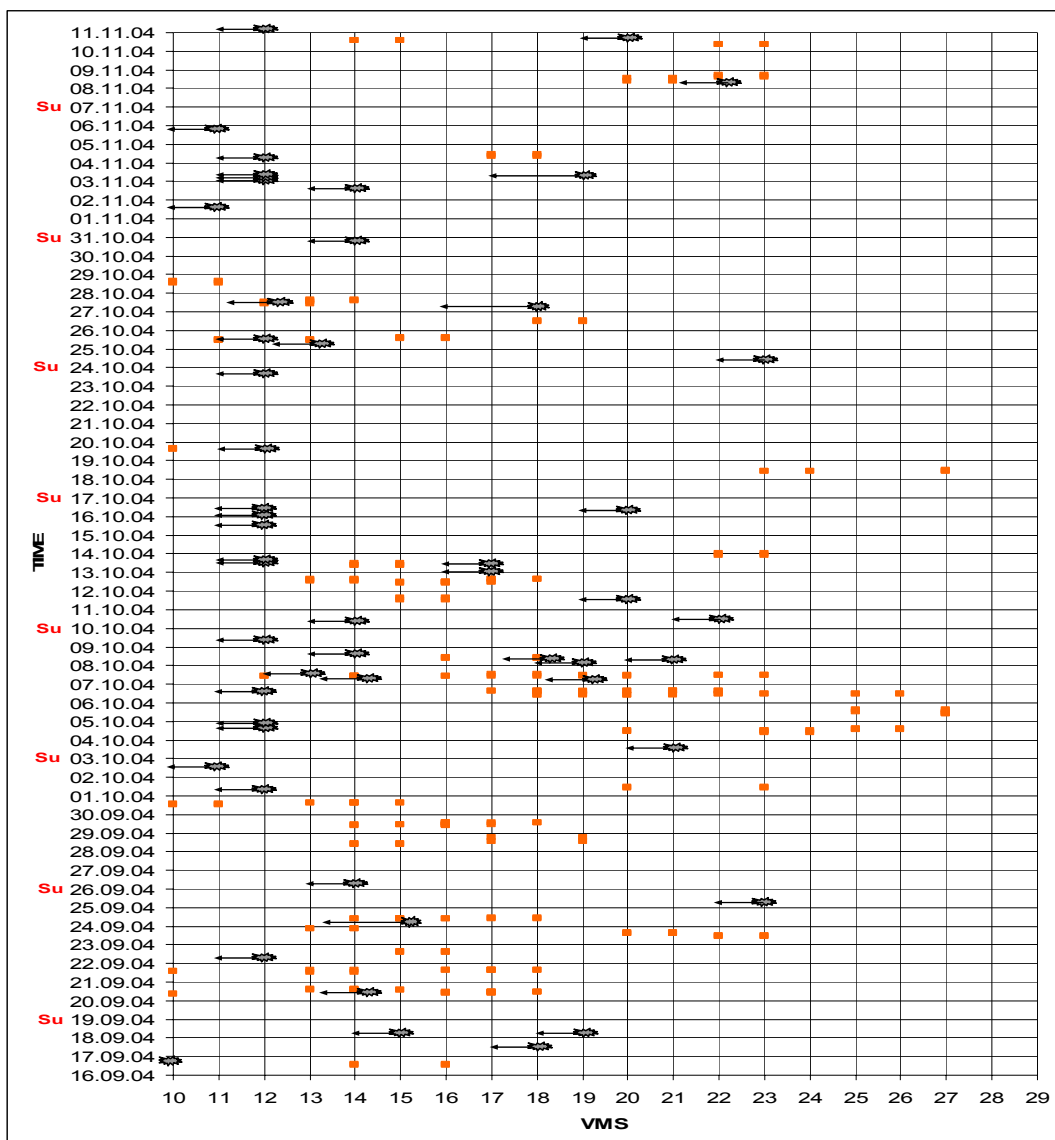


Figure A-5: Accidents and road works between September 16 and November 11, 2004

**Weather**

The available metrological data is taken from the weather station “Karolinenfeld” near A8Ost, Irschenberg. Unfortunately, at the station only data about the amount of rain is collected and therefore is used for weather related comparison. In Figure A-6 and Figure A-7 show measured amount of rain in the training and test period.

The average value for the training set is 2.9 [mm], and for the test set is 3.5 [mm]. However, almost no difference could be noticed in weighted average, where only days with rain were included in the calculation. The weighted average in the training set is 5.63 [mm] and in the test set is 5.65 [mm].

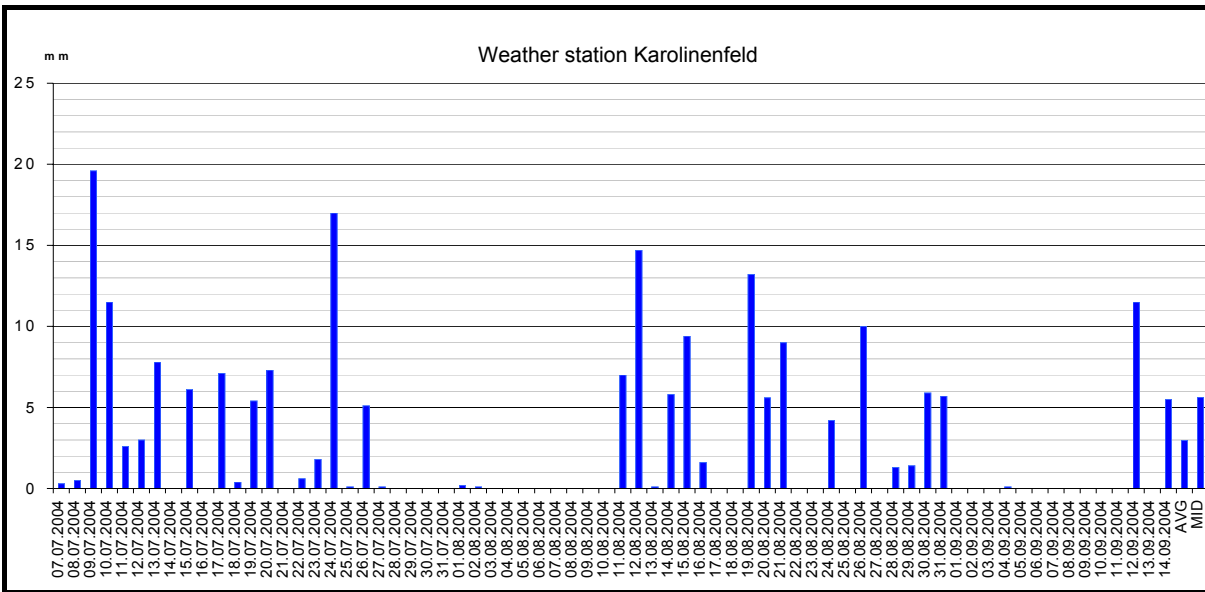


Figure A-6: Amount of rain in the training period

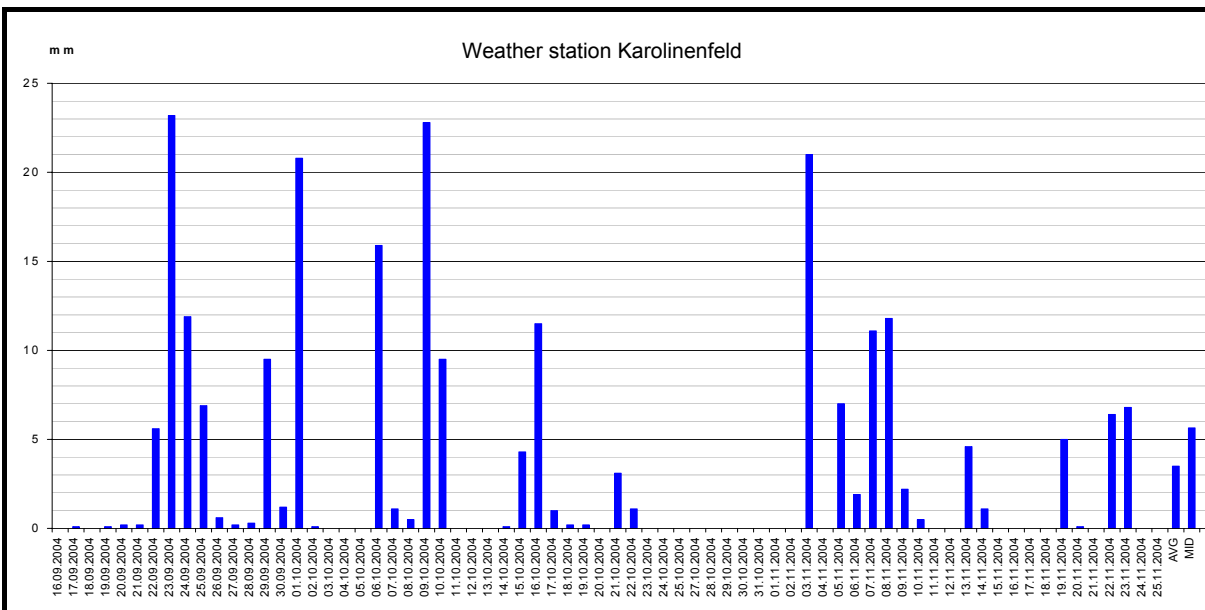


Figure A-7: Amount of rain in the test period

Here it should be noted that INCA does not have an algorithm for weather dependent controls, but takes these from the existing system. Consequently, wrong weather detection would have negative influence on both systems. Taking into account the days with adverse weather conditions, it can be seen if INCA is able to “detect” bad weather situations on the basis of measures of traffic flow characteristics only.