

Crowd-Based Road Surface Monitoring and its Implications on Road Users and Road Authorities

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

M. Sc. Inform.-Wirt. Kevin Laubis

Tag der mündlichen Prüfung: 11.12.2017

Referent: Prof. Dr. Christof Weinhardt

Korreferent: Prof. Dr.-Ing. Hansjörg Fromm

Karlsruhe, 2017

Contents

List of Figures	v
List of Tables	vii
List of Abbreviations	ix
I Foundations	1
1 Introduction	3
1.1 Challenges	4
1.2 Research Questions	8
1.3 Structure of the Thesis	12
2 Road Condition Monitoring	15
2.1 Traditional Approach	15
2.1.1 Road Condition Rating	15
2.1.2 International Roughness Index	19
2.1.3 Quality of Inspection Technologies	19
2.2 Crowdsensing-Based Approaches	21
2.2.1 Crowdsensing	21
2.2.2 Applications in the Vehicular Domain	23
2.2.3 Inertial Measurement Unit	25
2.2.4 Adoption Requirements	26
2.2.5 Related Work	28
2.3 Costs of Road Usage and Maintenance	32
2.3.1 Road Users	32
2.3.2 Road Authorities	33
II Crowdsensing-Based Road Surface Monitoring	37
3 Self-Calibration	39
3.1 Introduction	39
3.2 Regression Methods and Performance Metrics	41

3.3	Research Design	43
3.3.1	Map-Matching	43
3.3.2	Data Alignment	44
3.3.3	Feature Extraction	44
3.3.4	Model Building	45
3.3.5	Scenario Setup	46
3.4	Results	50
3.4.1	Estimation Performance of Single Drives	50
3.5	Conclusion	51
4	Data Reduction	53
4.1	Introduction	53
4.2	Feature Selection and Sensing Frequencies	55
4.3	Research Design	56
4.3.1	Permutation Importance	56
4.3.2	Scenario Setup	58
4.4	Results	60
4.4.1	Sensitivity to Feature Selection	60
4.4.2	Sensitivity to Sensing Frequency	62
4.5	Conclusion	63
5	Combination of Estimations	65
5.1	Introduction	65
5.2	Combination in Road Surface Monitoring	67
5.3	Research Design	68
5.3.1	Weighting	68
5.3.2	Optimal Weights	70
5.3.3	Shrinkage	70
5.3.4	Scenario Setup	71
5.4	Results	72
5.4.1	Weighted Combination	72
5.4.2	Optimal Weights and Regularization	74
5.5	Conclusion	77
6	IT Infrastructure	81
6.1	Introduction	81
6.2	Scaling Technologies	82
6.3	Research Design	84
6.3.1	Scaling Model	84
6.3.2	Scenario Setup	86
6.4	Results	89
6.4.1	Scaling Efficiency	89
6.4.2	Resource Reduction Potential	90

6.5	Conclusion	92
7	Integrated Service	95
7.1	Introduction	95
7.2	Service Systems	96
7.3	Research Design	98
7.3.1	Service Layer	100
7.3.2	Current Situation	101
7.3.3	New Intermediary	101
7.3.4	Hot Spot Analysis	103
7.3.5	Timeliness Metric	104
7.3.6	Scenario Setup	105
7.4	Results	107
7.4.1	Frequent Decision Support	107
7.4.2	Utility to Authorities, Intermediary and Participants	109
7.5	Conclusion	112
III	Economic Assessment	113
8	Road User Side Assessment	115
8.1	Introduction	115
8.2	Path-Finding in Road Networks	117
8.3	Research Design	118
8.3.1	Cost and Trade-Off Matrices	118
8.3.2	Savings Potential	119
8.3.3	Scenario Setup	120
8.3.4	Parametrization	120
8.4	Results	122
8.4.1	Savings Potential	124
8.4.2	Sensitivity to Fuel Price	127
8.5	Conclusion	127
9	Road Authority Side Assessment	129
9.1	Introduction	129
9.2	Markov Models in Maintenance Scheduling	131
9.3	Research Design	132
9.3.1	Partially Observable Markov Decision Processes	132
9.3.2	Value Iteration	132
9.3.3	Scenario Setup	133
9.3.4	Parametrization	133
9.4	Results	137
9.4.1	Savings Potential	137

9.5 Conclusion	138
IV Finale	141
10 Conclusion	143
10.1 Contributions	143
10.2 Outlook	147
10.3 Conclusions	149
V Appendix	151
A Feature Importance	153
B Prediction Errors and Combination Weights	157
C Server Load	161
References	163

List of Figures

1.1	Structure of the Thesis	12
2.1	Schema for Aggregating Characteristics of the Road Surface	18
2.2	Crowdsensing within Crowdsourcing Typology	23
2.3	Illustration of IMU Sensor	26
2.4	Histograms of IRI Measured with Laser Sensors and IMUs	29
2.5	Transportation Costs of Road Users and Road Authorities	33
2.6	Components of RUEs	34
2.7	Development of Road Condition over Time	35
3.1	Outline of Self-Calibration Approach	40
3.2	Sequence IRI Values from the District Road K3535, Germany.	47
3.3	Histogram of IRI from District Road K3535, Germany.	47
4.1	Motivation of Data Reduction Approach	54
4.2	Mean PI of Ten Most Important Features	60
4.3	Mean PI of Ten Least Important Features	61
4.4	Sensitivity of Prediction Performance to Number of Features	62
4.5	Sensitivity of Prediction Performance to Sensor Frequency	63
5.1	Outline of Combining Estimations From Multiple Drives	66
5.2	Weights of R^2 -, RMSE- and SA-Based Combinations	73
5.3	Out-of-Sample R^2 of Aggregated Predictions	74
5.4	Out-of-Sample NRMSE of Aggregated Predictions	75
5.5	Absolute Errors of In-Sample Predictions for Two Drives	76
5.6	Correlation Matrix of In-Sample Prediction Errors	77
5.7	OW in Comparison to SA Weights	77
5.8	Performance of OW Combinations Linearly Shrunk to SA	78
6.1	Resource Allocation with Horizontal Scaling	84
6.2	Resource Allocation with Diagonal Scaling	85
6.3	Outline of Scaling Simulation	87
6.4	Resource Efficiency of Horizontal and Diagonal Scaling	90
6.5	Resource Allocation and QoS Criterion for Load Pattern	91
6.6	Resource Allocation Reduction by Diagonal Scaling	92
7.1	Service Map of Crowd-Based Road Condition Monitoring	99

7.2	Timeliness Metric Determination	106
7.3	Spatial Hotspots—Region of Aizuwakamatsu	108
7.4	Spatial Hotspots—City Center of Aizuwakamatsu	109
7.5	Variation of Hot Spots over Time	110
8.1	Outline of Crowdsensing-Based Vehicle Navigation	116
8.2	Overview of Definitions for Road User Side Assessment	121
8.3	RUEs Considered in the Road User Side Assessment	122
8.4	Trade-Off Matrices for Medium Car	125
8.5	Trade-Off Matrices for Light Truck	126
8.6	Savings Potential Depending on Fuel Price	128
9.1	Outline of Crowdsensing-Based Road Maintenance	130
9.2	RUEs Considered in the Road Authority Side Assessment	137
9.3	Annual Total Costs for Optimal Maintenance Policies	138
A.1	Mean PI of the 1 to 15 Most Important Features	153
A.2	Mean PI of the 16 to 55 Most Important Features	154
A.3	Mean PI of the 56 to 95 Most Important Features	155
B.1	Absolute Errors of In-Sample Predictions for Drives 1 to 3	157
B.2	Absolute Errors of In-Sample Predictions for Drives 4 to 7	158
C.1	Histogram and Boxplot of Server Load	161

List of Tables

2.1	IQL, Corresponding Indicators and Inspection Methods	20
2.2	Summary of Related Literature	31
3.1	Features Extracted From Smartphone Sensors	49
3.2	Out-of-Sample Performance of Single IRI Estimations	51
5.1	Out-of-Sample Performance of \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE}	73
5.2	Out-of-Sample Performance of all Combination Functions	78
6.1	Overview of Scaling Parameters	86
6.2	Overview of Scaling Variables	86
6.3	Values of Scaling Parameters Used for Simulation	88
8.1	VOC per Unit and Vehicle Type	123
8.2	Factors Describing the Effect of the IRI on VOC	123
8.3	VOC per KM and Vehicle Type Depending on the IRI	123
9.1	Differentiation of Markov Models	131
9.2	Road Condition States ST	134
9.3	Transition Probabilities TR	135
9.4	Observation Probabilities OB	136
9.5	Annual Maintenance Costs and RUC	136
B.1	Weights of Combination Methods for Each Drive	159

List of Abbreviations

AWS	Amazon Web Services, Inc.
CPU	Central Processing Unit
CWT	Continuous Wavelet Transformation
GPS	Global Positioning System
HDM	Highway Development and Management
IaaS	Infrastructure as a Service
IMU	Inertial Measurement Unit
IQL	Information Quality Level
IRI	International Roughness Index
MDP	Markov Decision Process
NCHRP	National Cooperative Highway Research Program
NRMSE	Normalized Root Mean Square Error
OSM	Open Street Map
OSRM	Open Street Routing Machine
OW	Optimal Weights
PI	Permutation Importance
POMDP	Partially Observable Markov Decision Process
QoS	Quality of Service
RF	Random Forest
RFR	Random Forest Regression
RMS	Root Mean Square
RMSE	Root Mean Square Error
RQI	Ride Quality Index
RUC	Road User Costs
RUE	Road User Effects
SA	Simple Average
SVM	Support Vector Machine
SVR	Support Vector Regression
VM	Virtual Machine
VOC	Vehicle Operating Costs
WLP	Weighted Longitudinal Profile
ZEB	Program for Evaluating German Federal Highways

Part I
Foundations

Chapter 1

Introduction

Knowledge of the current road condition, such as the road's evenness, is of great relevance for road users and road authorities. The former can align their behavior with regard to an economically efficient, safe and comfortable mobility. The latter require the knowledge to enable efficient planning of maintenance strategies. In order to enable cost-efficient road maintenance, it is necessary to have accurate and up-to-date information on road conditions. The condition of a road can change significantly in a short period of time especially within a winter season in which the road is exposed to several freeze-thaw cycles. In addition, knowledge about the current road condition would be valuable for road users when they are used for advanced driver assistance systems. Today, vehicles equipped with special laser technology and high-resolution cameras are utilized for the assessment of roads. For the federal road network in Germany, for example, the associated costs lead to four-year inspection intervals (Hoppengarten et al., 2006). For state and district roads, there is either no regular or no monitoring at all. For municipal roads, the assessment is often carried out with pen and paper leading to inaccurate and incomplete results.

As smart devices are becoming increasingly popular and are equipped with ever more precise sensors, crowdsensing is becoming more relevant in a wide range of applications (Ma et al., 2014). Crowdsensing can be defined as the primarily passive utilization of mobile devices by the crowd to gather information about their environment (Prpić, 2016). It is not surprising that crowdsensing-based approaches are also being adopted in areas of mobility. This adoption provides the potential for an unprecedented investigation of not only vehicle and driver-specific characteristics, but also of the vehicle's environment and the condition of the road infrastructure. Accordingly, crowdsensing-based approaches, such as those described by Ericsson et al. (2006) or Mohan et al. (2008a), attempt to infer information about the road from the inertial measurement units (IMUs) built into consumer devices, such as smartphones. They provide the foundation of a near real-time assessment of the road condition that could supplement or substitute the today's road condition monitoring.

However, there are challenges when providing a low-cost alternative for a real-time-based and nationwide monitoring of the road network based on the crowd.

The participants have different vehicles with different physical characteristics and different devices, which the system has to take into account. In addition, questions arise about the manageability of the resulting data volumes from several points of view. On the one hand, mobile devices have limited storage space and limited transmission capacity, and on the other hand, a backend of such a crowd-based system has to process a large and fluctuating amount of load sent by the crowd.

The aim of this work is to create a deeper knowledge of the feasibility and utility of a crowdsensing-based solution for a nationwide and timely accurate road condition monitoring. For this, two main perspectives are considered in this work. On the one hand, technical issues are addressed and on the other hand, the economic benefits of such a system are investigated. In particular, the easy integration of new participants, an efficient handling of fluctuating data volumes and the economic utility for both stakeholders, road users and road authorities are examined. The remaining sections in this chapter present the addressed challenges, the covered research questions and the structure of the thesis.

1.1 Challenges

The ever-increasing number of sensors included in modern vehicles and carried by drivers and passengers, such as smartphones and wearables, provides an enormous potential for a crowd-based real-time monitoring of the road condition. However, there are several challenges that need to be resolved. For instance, data provision by automobile manufacturers is probably thought of first. However, even though automobile manufacturers partly intend to share and provide information from their vehicles—reference is made to the company HERE International B. V., which is owned by Audi, BMW, Daimler and four additional companies—it is not at all certain that a widespread provision of crowd-based information from automobile manufacturers can be expected soon. Thus, the utilization of consumer devices, such as smartphones, mobile navigation systems, wearables, etc. that can be carried within any vehicle, is promising as it is not restricted on vehicle models from certain manufacturers. Next to an immediately possible realization, the use of consumer devices would also allow for a rapid increase in the number of participants. Consequently, research was carried out in this field (Eriksson et al., 2008; Mohan et al., 2008a; Tai et al., 2010; Bhoraskar et al., 2012). However, a vast number of the existing work focuses on the detection of individual road anomalies, such as potholes, speed bumps, sunken manhole covers, etc. Even if these individual anomalies are relevant for road users and road authorities, since they represent potential hazards that should be avoided and repaired, other criteria are of greater importance for road authorities. In particular, the longitudinal road profile can be assumed as the most common metric and is considered by almost every road authority (Múčka, 2016).

It is important for all high level metrics, such as safety, comfort, substance, etc. (FGSV, 2001). In comparison to individual road anomalies, the crowd-based determination of a roughness-metric is more difficult. This is mainly because it is not a binary classification—pothole is present or not—but at least a multiple classification or better a regression, which describes the degree of roughness on a ratio scale. There are also research activities focusing on the crowd-based determination of longitudinal roughness. However, these have limitations, which restrict a widespread application in the crowd. Some approaches rely on defining the characteristics of the vehicle through a carefully designed physical model that describes the vehicle's body, the suspension system, the shock absorption by the tires, etc. (Hanson et al., 2014). Despite a careful description of these models, additional manual effort is required since test drives have to be performed for determining the model's quality. Alternatively, promising approaches based on machine learning are pursued (Nitsche et al., 2014). Here, the determination of the complex individual vehicle physics is partly addressed within the training phase of the machine learning process. To train a vehicle-specific model, test drives are also required, in which the vehicles are usually equipped with additional special sensors, such as lasers. In order to avoid this manual effort of equipping the vehicles with additional sensors for an initial calibration, other approaches use only a few single models, which are applied for entire vehicle groups, such as for small, medium sized and large vehicles (Forslöf and Jones, 2015). However, this also ignores the individual vehicle characteristics, which has an inherently negative effect on the estimation quality. The limitations of current approaches can be summarized as follows: either standard models that do not consider individual vehicle characteristics have to be used, or calibrations with manual effort must be carried out by describing physical dependencies or by temporarily attaching special sensors to the participants' vehicles. These limitations are not conducive to a crowdsensing-based approach. An automatic and therefore easy consideration of individual vehicle characteristics of the participants is required. Therefore, a self-calibration approach is presented in this work. In general, this approach is based on the idea of automatically calibrating new vehicles as they drive on roads for which accurate measurements from engineering companies are available, so that they can then enhance and update the database themselves on other roads or for the same road at a later point in time.

The amount of data that can be stored and processed by smart devices is inherently limited. Furthermore, the proposed self-calibration requires full feature sets to be sent to a backend system for the initial training. Even though, mobile phone tariffs have constantly lower rates for data transmission, pay-per-use tariffs or tariffs with limited data volumes are still widely used. Therefore, it is important to address data reduction potentials. In addition to the selection of relevant features that have to be sent to the backend, also the potential for reducing the sampling frequency of a device should be taken into account. On the one hand, this may allow for further reduction of the amount of data and on the other

hand, it may also allow older devices to be integrated into a crowdsensing-based road condition monitoring system, which for technical reasons can only operate at lower sensing frequencies.

One of the greatest strengths of crowdsensing-based systems is the fact that they have the potential to provide a large number of measurements. For the domain of road condition monitoring, this can lead to several measurements for different drives per road segment within a certain time frame. In addition to the fact that a crowd-based approach provides a widespread coverage, it also allows an increase of estimation accuracy per road section by combining multiple measurements. Although an increase in estimation accuracy through multiple measurements is known in other domains, such as financial forecasting (Timmermann, 2006; Blanc and Setzer, 2016), it is important to determine to which extent this applies to the domain of crowdsensing-based road condition monitoring. Even though, in this domain work was done, which takes multiple measurements into account, only aggregations using the simple average (SA) were performed, such as by Hara et al. (2014). Furthermore, it was not investigated how this aggregation affects the accuracy of estimating the road condition. In order to achieve a high benefit from multiple measurements it is important to investigate which aggregation methods are suitable and how the estimation performance depends on the number of measurements.

The proposed approach of crowdsensing-based road condition monitoring requires an IT infrastructure serving as a backend for collecting, processing and storing the data from the crowd. Since the load generated by the crowd is subject to seasonal fluctuations and probably also to a trend component, the resources of the backend should address this fluctuation through an automatic scaling capability. Much research has been carried out on the automatic determination of scaling decisions (Vasić et al., 2012; Urgaonkar et al., 2005; Heinze et al., 2014). Likewise, there is much work that addresses different scaling concepts (Chieu et al., 2009; Dutta et al., 2012). However, it is questionable, how efficient modern scaling approaches—such as so-called diagonal scaling, which integrates different scaling concepts, as described by Han et al. (2012)—are in the case of a fluctuating load. An according examination would allow finding out whether diagonal scaling is a suitable scaling method for the crowd-based approach.

In addition to the consideration of single technical challenges, it is also essential to investigate the applicability and utility of a crowdsensing-based road condition monitoring system as a whole. As described above, a crowd-based service requires not only data from the crowd but also conventional data services from engineering service providers in order to initialize and regularly adapt the models of the participants from the crowd to an accurately captured database. Therefore, in addition to the crowd itself, services from conventional data providers must also be considered in order to offer a crowdsensing-based service. Next to these data provisioning services, services for analyzing the data and business services that can be directly provided to the two main stakeholders are required.

This results in a service system, which has to orchestrate multiple single sub-services for co-creating value to the customer (Unterharnscheidt and Kieninger, 2010; Goldberg et al., 2014). Even though, there is much work done in the field of service science and value-co-creation in general and also some in the field of crowdsensing-based service, such as from Merlino et al. (2016), to the best knowledge there is no work that considers the systems of a crowdsensing-based road condition monitoring service.

Given a crowdsensing-based road condition monitoring service, it is necessary to consider its economic value for road users and road authorities. The potential demand for such a service and thus the necessity of realizing it depends directly on the customers' value. It is apparent that road users benefit from roads that are in an overall good condition. In case of the longitudinal road roughness, a direct connection can be made from roughness to the wear of a vehicle. This can be explained by the physical effects on the vehicle's components (Tan et al., 2012). Thus, if such a service leads to an improvement in the condition of the road network, an indirect benefit for road users is imminent. However, it is questionable whether road users will only benefit indirectly or whether road users will also benefit from the availability of road condition information directly and thus be potential customers of a service that is based on accurate and timely road condition information. Driver assistance systems, for example, are services designed to create direct added value for road users. Although there are driver assistance system approaches that use information about the road condition to warn drivers of potholes, they only refer to individual anomalies rather than to the condition of the entire road (Rode et al., 2009). This is also because approaches, which only refer to individual anomalies do not take into account the condition of the rest of the road and therefore a holistic view of the economic effects on the road user cannot be made. Therefore, it is essential to examine the direct economic benefit for road users of a crowdsensing-based service that provides accurate and timely road condition information. For this purpose, a driver assistance system is proposed in this work, which allows road users for reducing their vehicle's wear by avoiding rough road segments.

In addition to the fact that a crowdsensing-based service should be of benefit to road users, it should also create added value for road authorities. With regard to road maintenance, the fact that the road network is often monitored at very low frequency, leads to a reactive approach, which directs resources for roadwork and improvement to road segments that already reached a critical condition with severe damages. By the described crowdsensing-based monitoring service, it gets possible to provide road authorities with more frequent measurements. Although research was conducted to determine the effect of inspection intervals on maintenance efficiency, such as from Smilowitz and Madanat (2000), it was not applied in the domain of crowdsensing. In this context, it must be taken into account that the inspections carried out by the crowd are less reliable. In order to determine the utility of a crowdsensing-based service for road author-

ities, this trade-off between more frequent but less accurate measurements needs also to be investigated.

1.2 Research Questions

This thesis addresses technical and economic aspects of a crowdsensing-based system to monitor the longitudinal road roughness. From a technical perspective, due to a variety in vehicles and in sensors in a heterogeneous crowd, an automatic calibration is required when a new driver wants to contribute to the system. For this purpose, in Chapter 3 a self-calibration approach is presented and technically evaluated on own test drives. The approach allows new vehicles to be calibrated by just driving on road segments for which information about the road condition is already available. This recent road condition information is either provided by specially equipped vehicles as they are used today, or by other participants, which are already calibrated. The alignment of the participant's sensor data and the ground-truth from former measurements includes a map matching based on the global positioning system (GPS) coordinates and a road segment-based feature extraction. It is unclear whether spatial inaccuracies in these data alignment steps allow for a sufficient prediction accuracy. Accordingly, the first research question, which is addressed in this work, is stated as follows:

Research Question 1 – Self-Calibration

- a) How accurate can participants contribute to a crowd-based road roughness measurement system if their vehicle and smartphone get calibrated automatically while driving on a public road?
- b) How robust are these results with respect to repeated test drives and to different machine learning algorithms?

According to the fact that for such a crowdsensing-based system, aspects of data reduction should be considered, Chapter 4 addresses two different approaches. Since the self-calibration approach is based on machine learning methods, the first attempt for reducing the required amount of data is to reduce the number of features considered. Therefore, an embedded feature selection is used to determine the importance of each feature in order to obtain candidates for elimination by means of the less important features. In a second approach, the effect of a reduction of the sensing frequency and accordingly a reduction of the

accruing data volume on the performance of the estimations is investigated. The research question in this regard is as follows:

Research Question 2 – Data Reduction

- a) Which are the most important features for road roughness prediction that can be collected with a smartphone and how sensitive is the prediction performance to the elimination of less important features for data reduction purposes?
- b) How sensitive is the prediction performance to variations in the sensing frequency and what is the data reduction potential by sensing at a lower frequency?

The effects of having multiple measurements per road section by applying a crowdsensing-based approach and the methods that should be used were insufficiently examined. Therefore, Chapter 5 of this thesis investigates the effect of aggregating multiple estimations per road segment. The data set already considered in Chapter 3 and 4, which resulted from own test drives, consists of seven repeated drives on the same road link. First, in addition to the SA, several naively weighted averages are applied to this data set. In addition, based on the training data optimal weights (OW) are determined for each drive (Bates and Granger, 1969). In order to determine an out-of-sample optimal weighting, these in-sample OW are linearly shrunk towards the SA. Thus, the following research question is addressed:

Research Question 3 – Combination of Estimations

- a) To what extent does the aggregation of crowd-based road roughness measurements from multiple cars increase the model performance?
- b) Can the model performance be increased by applying OW and does a shrinkage of the OW to an unweighted aggregation affect the model performance?

In order to address the determination of a back-end infrastructure that allows for serving the load generated by a crowdsensing approach efficiently, Chapter 6 examines the efficiency of scaling technologies. In infrastructure as a service (IaaS) cloud computing it is commonly distinguished between vertical and horizontal scaling. Diagonal scaling, as a combination of horizontal and vertical scaling, attempts to combine the advantages of both approaches (Han et al., 2012). To determine which scaling technology should be used for the deployment, the corresponding resource efficiencies are evaluated based on a real load pattern.

Accordingly, the research question, addressing efficiency of scaling technologies, is formulated as follows:

Research Question 4 – IT Infrastructure

- a) How efficiently can the challenge of having a fluctuating demand in IT resources be served by horizontally and diagonally scalable IaaS?
- b) What resource reduction can be achieved by using diagonal scaling compared to conventional horizontal scaling?

As mentioned above, there are many parties required to cooperate to provide a crowdsensing-based monitoring service that supports decision-making of road users and road authorities on an accurate and frequent information basis. However, there is no current research that addresses the challenge of describing and orchestrating services for providing a crowdsensing-based road condition monitoring service. Chapter 7 addresses this need for orchestration by proposing a new intermediary that integrates data suppliers—namely the crowd and engineering companies—for serving road authorities and road users with accurate and timely information about the condition of the roads. The resulting service and the sub-services it is composed of are modeled as a service map (Kohlmann et al., 2010). The utility of such an integrated service is inherently important for its acceptance and success. The research question, which addresses this regard, is as follows:

Research Question 5 – Integrated Service

- a) How can existing services be integrated by a new intermediary for providing a crowd-based road condition monitoring service allowing for a frequent and accurate support in decision-making?
- b) What are the utilities to the intermediary, to the crowd and to customers of such an integrated crowd-based road condition monitoring service?

The direct utility of an integrated crowd-based service for road users is immanently of relevance. However, the economic added value of a service that provides accurate and timely road condition information was not quantified from a road user's perspective yet. In order to address this shortcoming, a vehicle navigation is presented in Chapter 8, which uses information about the longitudinal road roughness for finding the most cost efficient path while also considering the vehicle's wear. The international roughness index (IRI) as the most prominent metric describing longitudinal roughness is used for this examination (Sayers et al., 1986). As described by Tan et al. (2012), the wear caused by road conditions

also depends on the vehicle type. In addition, the fuel price is a crucial factor in determining a cost-effective path. Thus, these dependencies must be carefully considered. The research question in regard to quantify the monetary surplus of using road roughness information for path planning is stated as follows:

Research Question 6 – Road User Side Assessment

- a) What are the potential individual cost savings per road user that result from adapted vehicle routing based on the improvement of the overall IRI score per year and for which road conditions is the rerouting monetarily feasible?
- b) How sensitive are the results to different vehicle types and different fuel prices?

In addition to the utility of a crowdsensing-based road condition monitoring service for road users, the road authority's benefit of such a service needs to be assessed likewise. Currently, the road condition of main roads, such as highways, is assessed with quite accurate cameras and lasers, but at a very low frequency—up to four year intervals in the case of the federal road network in Germany—and at relatively high cost. Chapter 9 examines whether crowdsensing-based road condition inspections, which provide more frequent but less accurate information, have the potential to complement or substitute conventional inspection methods by reducing the overall maintenance costs. Partially observable Markov decision processes (POMDPs) are used to determine the costs of optimal maintenance strategies for different inspection scenarios. Hereby, the following research question is addressed:

Research Question 7 – Road Authority Side Assessment

- a) Given crowd-based road condition inspections compared to accurate laser-based inspections, what are the effects of different inspection accuracies and inspection frequencies on an optimal maintenance policy?
- b) To what extent can maintenance costs and road user costs (RUCs) be reduced, when combining crowd-based and laser-based inspections?

Summarizing, this work addresses questions that focus on both the technical feasibility and the economic utility of a crowdsensing-based road condition monitoring system. The research questions in mind, they have not or only partly been covered by related work yet.

1.3 Structure of the Thesis

This section outlines the structure of the thesis, as depicted in Figure 1.1. To address the research questions presented in the former section, the thesis is organized in four parts.

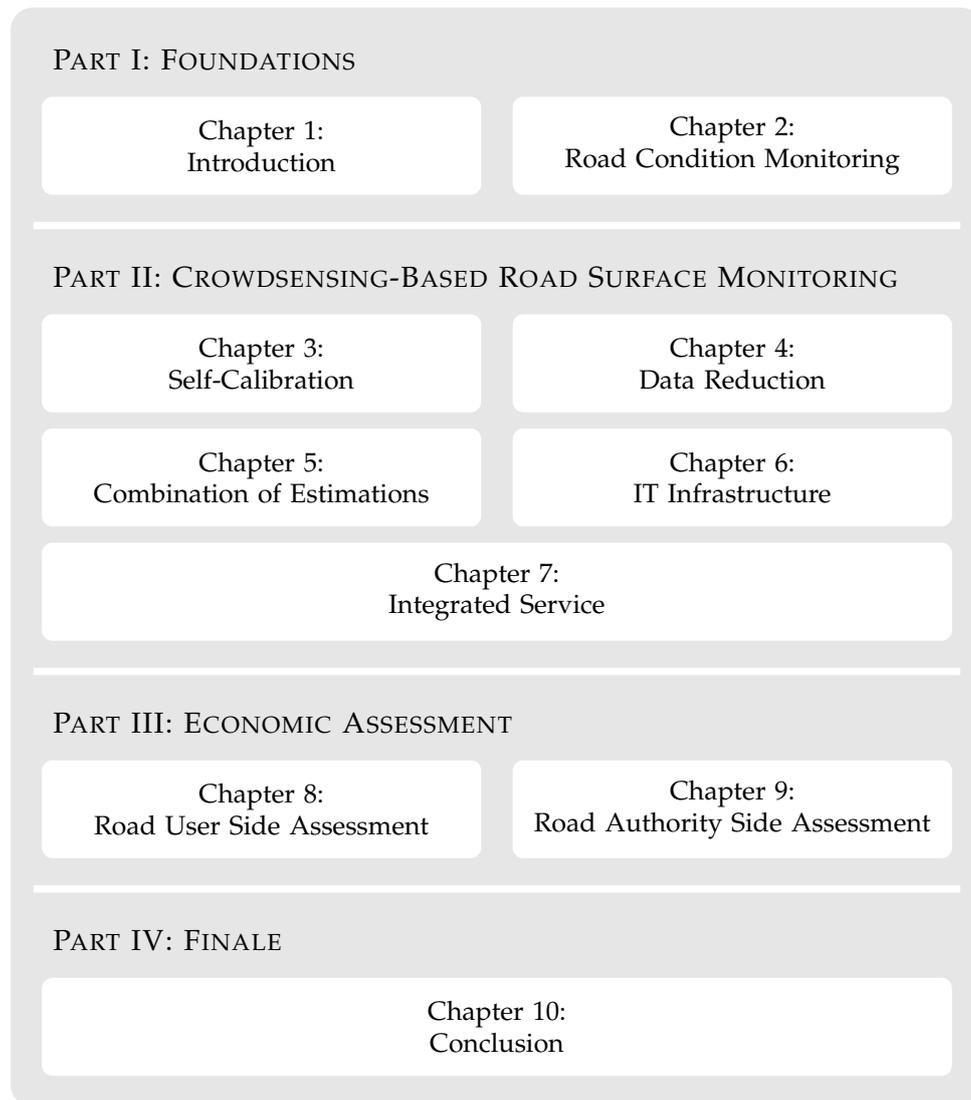


Figure 1.1: The thesis is organized in four parts. An introduction, a problem description and the foundations the thesis is based on are provided in Part I. Part II addresses the feasibility and possible improvements of a crowdsensing-based monitoring of the road surface from a technical and a system science perspective. This approach is economically assessed in Part III. The drawn conclusions and an outlook on further research are provided in Part IV.

Part I provides the foundations for the thesis. Chapter 1 introduces the approach of crowd-based monitoring road conditions. Next to a general motivation, the research questions addressed within this thesis are derived. Chapter 2 covers both, the road authorities' current approach of monitoring the road network manually or with specially equipped vehicles and the existing emerging crowd-based road condition monitoring approaches. For both, relevant technical and economic aspects are discussed.

Part II focuses on the proposed crowdsensing-based approach of monitoring the road surface condition by exploiting the sensors build in smart devices, such as smartphones. In Chapter 3 the central problem of easily integrating new participants to such a system is addressed. It is investigated how models for certain vehicles and smart devices can be determined without the need of attaching additional sensors to the vehicle for calibration reasons. The data sets utilized for evaluating this self-calibration are described. These data sets are also used for evaluation purposes for most of the remaining chapters. This self-calibration requires the transmission of training data to a backend system for each new participant. The potential of reducing the amount of data that has to be sent to the backend is covered by Chapter 4. Therefore, the approach's sensitivity to reducing the number of features extracted from the sensors and its sensitivity to lowering the sensing frequency are investigated. Applying the self-calibration enables an easy integration of new participants. Having many participants allows for the collection of a vast amount of single road condition estimations. Chapter 5 assesses the potential of spatially aggregating multiple measurements. Weighted and unweighted combinations are compared. An IT infrastructure is required to serve as a backend system that allows the collecting, storing, processing and distributing the data gathered by the crowd. In Chapter 6 the requirements of such a backend system are determined. An IT infrastructure architecture is designed accordingly. This is followed by a sound selection of technologies implementing the architecture's components. A simulation is performed to evaluate the system's elasticity while processing a daily fluctuating load pattern. Applying a crowdsensing-based monitoring approach encompasses the crowd as a data provider and the road users and road authorities as information consumers. However, a third party is required for deriving meaningful road condition information from the raw sensor data and providing decision support. Thus, Chapter 7 presents an integrated service that introduces an intermediary between the crowd as a data provider and road users and road authorities as service customers. At the end of this chapter, a hot spot analysis illustrates the applicability of the integrated service by exemplary providing a spatio-temporally accurate decision support.

Part III assesses the utility of the crowdsensing-based road condition monitoring to the two main stakeholders, road users and road authorities. Chapter 8 investigates the utility to the road user. A smoother road is expected to cause less fuel expenses and less vehicle wear (Tan et al., 2012). Therefore, a scenario of considering crowdsensing-based road condition information for vehicle nav-

igation is investigated. The road user's potential cost savings of being rerouted to a smother road are determined for different vehicle types and different fuel prices. Chapter 9 assesses the effects of using crowdsensing-based road condition inspection for scheduling road maintenance actions. Hereby, the system's utility to road authorities is determined. A crowdsensing-based road condition information system allows for more frequent inspections compared to a conventional approach. However, the measurements are less reliable. Both, frequency and accuracy of road condition inspects affect the efficiency of road maintenance policies. Different inspection and maintenance scenarios are investigated to determine which of the two effects—the higher frequency or the lower accuracy—predominates.

Part IV finally, summarizes the conclusions drawn in this thesis and discusses them critically in Chapter 10. Limitations of the work are addressed and a proposition for future research is provided. This thesis is partially based on and extends former publications from Laubis, Simko, and Schuller (2016a,b,c); Laubis, Simko, and Weinhardt (2016); Laubis, Simko, Schuller, and Weinhardt (2017) and Laubis, Konstantinov, Simko, Gröschel, and Weinhardt (2018).

Chapter 2

Road Condition Monitoring

This chapter introduces and discusses foundations of road condition monitoring. After the current monitoring and assessment of road condition is described, approaches of crowdsensing-based collecting road condition information are presented. The chapter concludes by describing the costs associated with road condition for road users and road authorities.

2.1 Traditional Approach

Roads are assigned to public and private road authorities responsible for maintaining them. The part of the public road network in Germany is again assigned to different road authorities. In principle, a distinction is made between federal roads, state roads, district roads and municipal roads. For the federal roads regular inspection are performed (Hoppengarten et al., 2006). For state, district and municipal roads there is either no regular or no monitoring at all.

2.1.1 Road Condition Rating

There are multiple road characteristics which can be used to describe different road condition facets. Basically, a distinction can be made between surface characteristics and those characteristics of deeper pavement layers. For the investigation of characteristics of the deeper road layers, stationary and partly also invasive techniques are used, such as the so-called falling weight deflectometer, for the determination of the pavement strength (Bennett and Paterson, 2000). Since these methods are stationary, they require a temporary closure of the road segment that has to be inspected and are therefore only carried out sporadically. In this section, characteristics of the road surface are briefly explained, since in the further course of the work the focus will be on the longitudinal road roughness, which is one of the road surface characteristics. The selection of the following characteristics was made on the basis of the program for recording and evaluation of the state of federal highways in Germany (ZEB) (Hoppengarten et al., 2006).

Longitudinal roughness: The longitudinal roughness can be expected as one of the most important characteristics to describe road condition. Roughness in the longitudinal direction, such as bumps or the general waviness of the road, have an effect on the vehicle, the vehicle's passengers and the road itself (Sayers and Karamihas, 1998). The forces acting on the vehicle, directly cause it to vibrate. This causes an increased vehicle wear and fuel consumption. Additionally they have an effect on safety, as the braking distance may be extended (De Weille, 1966; Chatti and Zaabar, 2012). Through the vibration of the vehicle, the passengers' ride comfort is indirectly affected. In addition, the road itself is exposed to a higher stress. One of the most prominent metrics for describing the longitudinal road roughness is the international roughness index (IRI). It was defined by a multinational research consortium in 1986, which was funded by the World Bank (Sayers et al., 1986). Basically, it describes the ratio between the accumulated vertical displacement of a reference car and the distance traveled. Next to the IRI, there are other metrics, such as the weighted longitudinal profile (WLP) (Ueckermann and Steinauer, 2008; Maerschalk et al., 2011) or the Michigan ride quality index (Janisch, 2015). A comprehensive survey through different metrics of the longitudinal road roughness is provided by Múčka (2016).

Rutting and standing water: Ruts are depressions in the road along the driving direction. They usually occur in the wheel paths (Miller and Bellinger, 2014). Similar to the longitudinal roughness, ruts are a deformation of the road surface. Thus, they reflect the deterioration of the road substance. The extent of rutting is defined by the depth of the rut. There are different definitions regarding the lateral reference and the distances in which the depth is measured repeatedly. The measurement with a 1.2 m straight edge every 15.25 m should be mentioned here as an example (Simpson, 1999). Ruts lead to an increased risk of aquaplaning as they can accumulate water. Thus, they do not just reflect the road's substance, but also safety aspects due to standing water. Therefore, the mentioned metrics are extended by taking the lateral slope of the road into account for determining the maximum water depth.

Friction: "Pavement friction is the force that resists the relative motion between a vehicle tire and a pavement surface" (Hall et al., 2009). It has been shown, that the friction of a road is a crucial safety aspect since it is directly related to the risk of accidents (Henry, 2000). This is especially true when the road surface is wet and no sufficient deceleration can be achieved or if the friction is too low for resisting the lateral forces in a curve. Common metrics used by road authorities currently for describing the friction condition of a road are the coefficient of friction and the side-force coefficient (Hall et al., 2009). The first coefficient describes the resistance of the road surface when brak-

ing on a straight road section. The second coefficient describes the road's resistance in curves.

Cracking: Several types of cracks are differentiated, such as fatigue cracking, block cracking, edge cracking, longitudinal cracking, reflection cracking at joints and transverse cracking (Miller and Bellinger, 2014). If the cracks are of a small extent, they have no effect on driving comfort or safety aspects. However, a first appearance of cracks leads to an accelerated deterioration of the road. On the one hand, this is because the pavement is less resilient to traffic and on the other hand because the penetration of water into the pavement during the frost periods often leads to more severe damage, such as potholes. A common metric is the number of square meters affected by cracks. This allows to determine the ratio between the number of affected square meters and the total number of square meters for a certain road section.

Patching: A patch can be described as a part of the pavement of at least 0.1 m^2 , which has been repaired by either removing the worn surface and replacing it or by applying additional material on top (Miller and Bellinger, 2014). Patches usually do not directly affect the road usage. However, they reflect the degree of deterioration of the road. Furthermore, patches tend to be subject to further damages. For example, deformations of its surface, such as ruts, can occur. Similar to the cracks, a common metric is the number of affected square meters. Alternatively, the number of actual patches can also be used.

A common approach of an integrated consideration of the mentioned road condition features is to apply a weighted combination, as depicted in Figure 2.1. The bottom of the figure shows the road characteristics considered within the assessment of the federal highways in Germany. Therefore, each characteristic described above in this section is represented by one box at the bottom. The metrics of the individual characteristics—also described above—are normalized to be comparable. Thereafter, the normalized metrics are aggregated to an usage rating and a substance rating. In simple terms, the usage rating is of interest to road users and the substance rating is of primary interest to road authorities. A more thorough differentiation and an explanation of the interaction of the costs for road users and road authorities related to the road's condition are provided in Section 2.3. The aggregation is weighted according to the labels at the edges. Depending on whether the longitudinal roughness or the rut depth has a worse normalized value, it is decided which of the two corresponding metrics—the worse one—is considered for contributing with a weight of 25 % to each of the intermediate ratings, usage and substance. This is also illustrated in Figure 2.1. The other characteristics, friction, water depth, cracks and patches contribute with 25 % or 50 % to either one of the intermediate ratings. Similar to the consideration of the

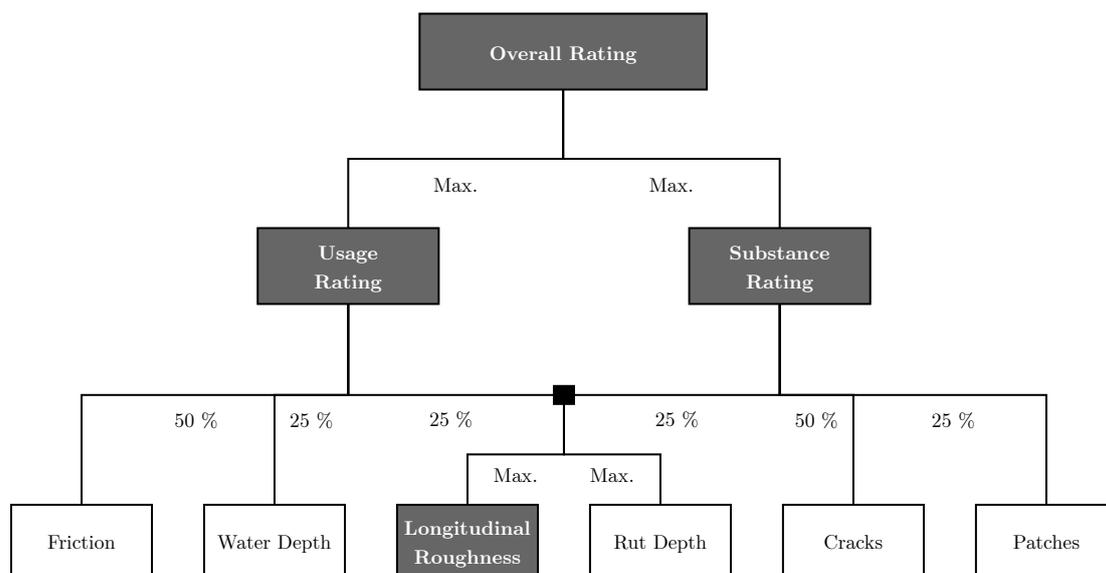


Figure 2.1: Schema for aggregating characteristics describing the surface condition of asphalt roads (FGSV, 2001). The boxes in the bottom row represent actual road condition characteristics. Each characteristic can be described by a certain metric, such as the IRI for the longitudinal roughness. The metrics are then normalized across all characteristics. The labels on the edges indicate the weightings with which the normalized characteristics are aggregated to an usage and substance rating. The worse scored rating is considered as the rating for indicating the overall road's condition. The highlighted components indicate that the longitudinal roughness affects both, the usage and the substance rating.

characteristics, longitudinal roughness and rut depth, the worse rated intermediate rating constitutes the overall rating. In the case of the assessment of German federal highways, the metrics and ratings are generally determined for 100 m segments. The exemplary weighting schema is for the case of asphalt roads. For concrete roads there is another aggregation schema that is slightly different for the substance rating (FGSV, 2001). This is because there are no patches on concrete roads. In contrast, however, especially edges and corners of concrete plates can be subject to additional wear. This is taken into account in the aggregation schema for concrete roads.

The longitudinal roughness is important for both types of road. This is the case as it is included in all ratings, in the two intermediate ratings and in the overall rating, at least if it is in a relatively bad condition. The influence of the longitudinal roughness is indicated by the highlighted boxes in Figure 2.1. Since the longitudinal roughness has a huge importance in the assessment of the condition and is relevant for both, road users and road authorities, it was chosen as the considered characteristic in this work. In addition, the IRI—as the most prominent metric to determine the longitudinal roughness—is primarily used as the metric

addressed in this work. Subsequently, it is described in more detail in the section below.

2.1.2 International Roughness Index

The longitudinal road roughness can be quantified by several metrics (Múčka, 2016). One of the most important metrics is the IRI. It indicates whether the road is rough and bumpy, contains anomalies, is in an overall wavy condition or can be considered as being smooth. It captures the longitudinal profile of a road and acts as a proxy for several road characteristics, such as the comfort, safety, deterioration and the additional stress imposed on every vehicle driving on the road. It was developed and defined in the course of the road roughness experiment by a multinational research consortium in 1986 (Sayers et al., 1986). This research was funded by the World Bank. The IRI was originally developed for emerging countries but is actually considered in most of today's road condition surveys commissioned by road authorities. It is defined as the ratio of the accumulated vertical movement of a suspension system and the hereby driven distance. The ratio can be given in the unit m/km or inches/miles. For convenience, the unit is omitted sometimes. The values are always given in m/km. In particular, since this displacement of the suspension system depends on the physics of the entire vehicle, a single reference car was determined. In addition to the dependence on the vehicle's physics, the movement of the suspension system depends on the driven speed. Therefore, a fixed speed of 80 km/h has been specified. It was supposed to be representative for the widest speed range. However, the IRI is most sensitive to wavelengths between 2.4 m and 15 m (Sayers and Karamihas, 1998). Although it is defined by the suspension movement it is actually determined by measuring the road's profile and afterwards simulating the suspension system's movement using a mathematical representation of the reference car—defined by a mathematical quarter-car-model (Jazar, 2014). Thus, it is possible to calculate the IRI, independently of the inspection method, solely on the basis of the road profile. As part of the work for the thesis at hand, an R-package `rroad` (<https://CRAN.R-project.org/package=rroad>) was developed which allows to calculate the IRI based on a road profile. It is published in "The Comprehensive R Archive Network" repository (Simko and Laubis, 2017). Since the IRI is widely adopted and determined by most road authorities, it is the metric employed in this thesis.

2.1.3 Quality of Inspection Technologies

Depending on the inspection methods and tools required to measure the IRI, an information quality level (IQL) can be assigned to the resulting measurements (Paterson and Scullion, 1990). IQL 1 comprises measurement methods that re-

quire sophisticated instrumentation and provide highly accurate quantitative output. IQL 2 represents the information detail level usually required for decision making at a project level. IQL 3 can be used for decisions relating to a road network. For this purpose, the measurements may be subject to minor inaccuracies, which are unlikely to influence a decision at an aggregated level. IQL 4 is primarily used for reporting purposes based on key attributes describing the overall pavement condition. This even higher aggregation level allows to have even less restrictions on the inspection method. IQL 5 represents the highest aggregation level of road condition information. Key performance indicators that represent an aggregate of IQL 4 key attributes and road network performance related measurements—such as structural appropriateness and traffic influences—are to be assigned to the IQL 5. Table 2.1 summarizes these five IQL instances. Examples of indicators and inspection methods are presented on the basis of the IRI.

Table 2.1: IQLs, corresponding indicators and inspection methods based on the example of the IRI metric (Paterson and Scullion, 1990; Bennett et al., 2006). To determine the indicators at different IQL, inspection methods of different accuracy are required.

IQL	Indicator	Inspection method
I	Profile-based roughness (e.g. IRI metric in m/km)	Accurate lasers
II	Roughness classes (e.g. longitudinal roughness characteristic)	Profilometer
III	Ride quality classes (e.g. usage rating and substance rating)	Correlation-based
IV	Pavement condition (e.g. overall pavement condition rating)	Subjective rating
V	Road Performance (e.g. combination of condition and traffic effects)	—

Most accurate official surveys to determine the IRI mainly rely on laser profilometers to measure the profile and calculate the IRI afterwards (Paterson and Scullion, 1990). Keeping the IRI as an example, the IQL 2 can be determined by deriving roughness classes. If only these classes are of interest, it is not necessary to use high-precision laser technology, but simpler profilometer approaches can be followed. Crowdsensing-based approaches, such as smartphone-based measurements, correspond to the IQL 3, since they often do not measure the metric itself, but a correlated one. They are explained in more detail in the following section. If only an overall rating in terms of the IQL 4 is to be determined, a further aggregation of IQL 3 ratings can be performed, as outlined in Figure 2.1. Road performance indicators of the IQL 5 consider the pavement condition in the context

of, for example, maintenance task scheduling and can therefore not be measured by a single inspection method.

2.2 Crowdsensing-Based Approaches

Smart devices have recently become a great source of sensor data from included accelerometers, gyroscopes and global positioning system (GPS) units. This established the crowdsensing-based road condition monitoring as an interesting research field. In the following sections, in addition to a definition of crowdsensing, fields of application in the vehicular domain are described. Thereafter, the inertial measurement unit (IMU) is described as a sensor built in modern smartphones and employed in many crowdsensing-based approaches. Furthermore, requirements for a crowdsensing-based monitoring of road roughness are derived to discuss related work based on these criteria.

2.2.1 Crowdsensing

Crowdsourcing was initially described by Howe (2006) as the application of information technology in order to leverage the crowd for the fulfillment of tasks. Later, Estellés-Arolas and González-Ladrón-de Guevara (2012) defined crowdsourcing more detailed as a "type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task". The technological development lead to ever more applications of crowdsourcing. With new crowdsourcing-based applications made possible by today's technologies, Prpić (2016) took up the concept of crowdsourcing and describes four new fields of application for it. Therefore, based on an exploratory analysis on the development of crowdsourcing, Prpić (2016) coins the term of "next generation crowdsourcing", in which he describes the four new fields of crowdsourcing, which have emerged from the availability of new technologies. These new fields are situated crowdsourcing, spatial crowdsourcing, wearables crowdsourcing and crowdsensing. They are briefly describe in the following:

Situated crowdsourcing: Situated crowdsourcing refers to hardware installed on site, which aims at exploiting human skills directly at the location of interest. These may be tablets or other input-enabled devices, for example. Here, the participant is assumed to be actively involved within the process.

Spatial crowdsourcing: Spatial crowdsourcing, in contrast to situated crowdsourcing, refers to the environment of the participant and not to the individual itself. This type of crowdsourcing became possible as mobile devices and mobile networks became increasingly prevalent.

Wearables crowdsourcing: Wearables crowdsourcing is about using mobile devices that are worn by the participants to gather information about themselves. Since even the activation of respective devices is today widely automatized, no or almost no contribution is required by the participant.

Crowdsensing: Crowdsensing refers to passive approaches—similar to wearables crowdsourcing. However, in this case, mobile devices are leveraged for gathering information about the environment rather than the participant itself, which in turn is similar to the situated crowdsourcing approaches.

In addition to these new fields of crowdsourcing a typology is derived for classifying them. Therefore, two dimensions are considered for a differentiation, namely the type of human participation and on the type of intelligence as it is depicted in Figure 2.2. The first dimension distinguishes between whether the participant must always actively contribute to the completion of the task or whether mainly contributes passively. Both, situated and spatial crowdsourcing require an active participation of the individuals. Wearables crowdsourcing and crowdsensing, however, require only little activity of the participants. The second dimension addresses differences with respect to about what intelligence should be derived. Wearables and situated crowdsourcing focus on the participants themselves, whereas spatial crowdsourcing and crowdsensing address the participant's environment.

Since crowdsensing is a passive way of acquiring information about the environment it is located to the upper-right corner in Figure 2.2. Once more referencing Prpić (2016), a definition of crowdsensing, which is also often referred to as participatory sensing or social sensing, is provided as follows:

"Crowdsensing [...] leverages the built-in sensors in smartphone devices to gather environmental data such as location, temperature, and acceleration, as a result of human mobility, smartphone portability and wifi/mobile networking."

This definition is consistent with and is itself based on the works of Malatras and Beslay (2015), Sun et al. (2016) and Zenonos et al. (2016). There are many existing fields of application for crowdsensing and there is a constantly increasing number of new opportunities as discussed by Ganti et al. (2011) and Ma et al. (2014).

According to the presented typology and definition, the approach of this work—to use smartphone sensor technology to investigate the road condition without having the participants performing active tasks—is covered by the term crowdsensing. In order to refocus on the domain of road condition monitoring, the following section provides an overview of crowdsensing-based approaches in the domain of vehicles, which is also referred to as vehicular sensing (Zhao et al., 2015; Strazdins and Mednis, 2011).

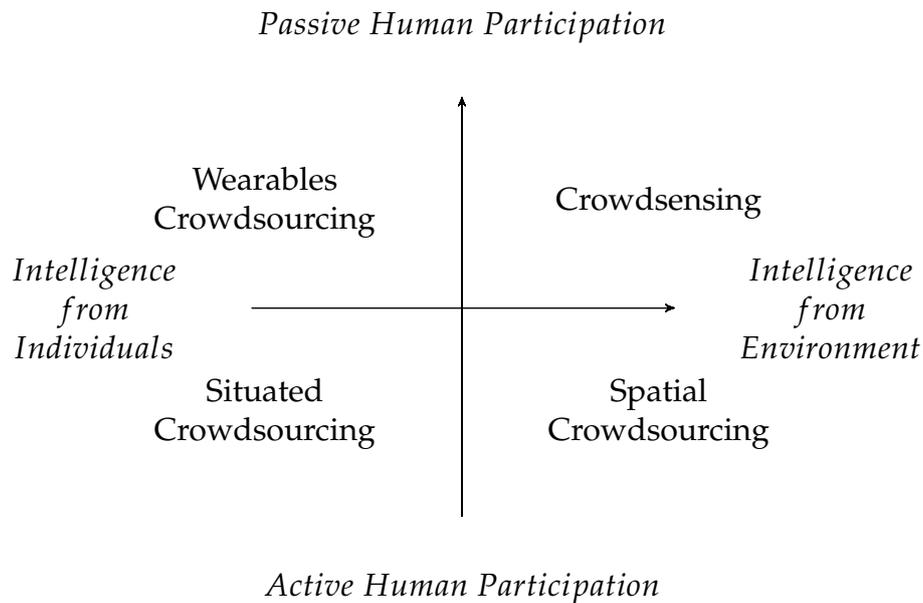


Figure 2.2: Crowdsensing within the crowdsourcing typology defined by Prpić (2016) allows for a classification in two dimensions. First, with regard to the type of human participation it is possible to distinguish between an active and passive participation. On the other hand, it is possible to differentiate according to about what intelligence should be derived. Concrete instances in this second dimension are the individual or its environment. Since crowdsensing is a passive way of acquiring information about the environment it is assigned to the upper-right corner.

2.2.2 Applications in the Vehicular Domain

Zhao et al. (2015) describe the term vehicular sensing, as the concept of making use of sensors integrated in vehicles or smartphones for gathering environmental information. Thus, it can be considered as a specialization of crowdsensing, which is aimed at the environment of the vehicle. This is also apparent, since Zhao et al. (2015) describe challenges in vehicular sensing, which are analogously discussed by Ma et al. (2014) for the crowdsensing domain.

Crowdsensing applications for vehicles and road users are predominantly found in the area of traffic. In particular, the measurement of traffic and the estimation of travel times are prominent vehicular sensing applications (Caudel et al., 2008; Thiagarajan et al., 2009). By a GPS-based determination of vehicle positions, speeds and routes are derived, which in turn facilitate the estimation of average travel times and unforeseen delays in real-time. Mohan et al. (2008a), for example, follow an interesting approach by additionally considering braking behavior and honking detected via the smartphone's microphone to monitor traffic conditions in a city. Next to traffic related applications, there are approaches targeting air pollution, such as the quantification of the CO₂ emissions (Hu et al.,

2009; Ortenzi and Costagliola, 2010; Hasenfratz et al., 2012; Yi et al., 2015). These approaches require either access to vehicle sensors, such as the lambda sensor, or the connection of additional sensors to the mobile device. Approaches such as from Sahlholm and Henrik Johansson (2010) try to estimate the longitudinal road grade. They integrate information from the engine, the gear box, the brake system and GPS from multiple drives for estimating the road grade. Providing this information to following road users allows for a fuel efficient look-ahead vehicle control. For example, this can be achieved if the cruise control of a vehicle automatically accelerates if there is an ascending road ahead or automatically decelerates shortly before reaching the top of a hill to take advantage of the downhill road segment. A further application is the detection of weather-related road conditions. For example, Jokela et al. (2009) investigate the possibility of determining black ice, snow and water coverage of the road surface. The detection of so-called black ice is crucial. It is a thin and transparent ice-layer on the road surface, which because of its transparency allows to see the asphalt, which is predominantly black in the USA. Because of its minimal visual contrast to the asphalt color, it is difficult to recognize. Hainen et al. (2012) attempt to indirectly determine the snow coverage of roads by analyzing changes in traffic flow. Many of these crowdsensing approaches in the vehicular domain have in common that they attempt to leverage the potential of low-cost sensors in consumer products, which are already present, by intelligently combining many, albeit inaccurate, measurements to obtain results as accurate or more accurate as those obtained from high-precision sensors.

In order to refocus on the subject of this work, reference is now made to application areas, which are concerned with the road's surface condition in terms of its degradation. In addition to approaches focusing on the longitudinal road roughness—which are discussed in more detail in Section 2.2.5—there are studies addressing the determination of other characteristics of the road surface. There is a vast number of work focusing on the detection of single road anomalies, such as potholes, speed bumps, sunken manhole covers, etc. (Eriksson et al., 2008; Mohan et al., 2008a; Tai et al., 2010; Koch and Brilakis, 2011; Mednis et al., 2011; Bhoraskar et al., 2012; González-Gurrola et al., 2013; Seraj et al., 2014; Mohamed et al., 2015; Savera et al., 2016). A prominent approach called Pothole Patrol was described by Eriksson et al. (2008). They equipped taxis in Boston, USA with smartphones for detecting potholes in that area. To detect actual anomalies, they implemented several machine learning-based filters for detecting expansion joints, rail crossings, speed bumps, door slamming, etc. Mohan et al. (2008a) developed a system called Nericell or TrafficSense (Mohan et al., 2008b) which also detects potholes and bumps. Their experiments were carried out in Bangalore, India. In contrast to the former approach, they deployed an algorithm for reorienting the smartphone axes dynamically to avoid limitations of manual alignment. Mednis et al. (2011) describe a similar machine learning-based approach for detecting road anomalies in Riga, Latvia. However, using their algorithm, they were able to

distinguish between different anomaly types, such as large potholes, small potholes, pothole clusters, gaps and drain pits. Bhoraskar et al. (2012) developed a system called Wolverine. They applied a classification approach by distinguishing road segments in terms of the existence of a speed bump. The evaluation of their approach is based on experiments conducted on the roads of Mumbai, India. Tedeschi and Benedetto (2017) focus on utilizing smartphones to identify different types of cracks in the road surface, such as longitudinal-transversal cracks and fatigue cracks. Chugh et al. (2014) and Gonzalez et al. (2017) provide an overview of crowdsensing-based approaches to detect single road anomalies, such as potholes.

The presented approaches are provided for putting the estimation of the longitudinal road roughness into perspective. Although these individual anomalies are of relevance, since they represent potential hazards, the longitudinal road profile can be assumed to be the most common metric taken into account by road authority (Můčka, 2016). The IMU is described in more detail in the following section. It is a sensor built into modern smartphones and is used for many crowdsensing-based applications.

2.2.3 Inertial Measurement Unit

Modern smartphones already include a variety of sensors that can be exploited for crowdsensing purposes. Ambient light, proximity, dual cameras, dual microphones, GPS, accelerometer, gyroscope—rotation rate sensors—and compass are a few of them (Lane et al., 2010). In addition, it can be assumed that the sensor density in mobile devices will continue to increase—air quality sensors are an example (Ganti et al., 2011). In addition to sensors in smart devices, vehicles themselves are a further source of sensors, which can provide information about the vehicle's environment. To name a few, accelerometer, cameras, radar, lidar, infrared, rain, tire pressure sensors, etc. are mentioned (Fleming, 2008). With regard to road roughness, vehicle vibrations are an important indicator. Therefore, the acceleration and gyroscope sensors—both inertial sensors—are of particular interest.

An IMU is a combination of multiple inertial sensors (Titterton and Weston, 2004). It usually consists of 3 accelerometers and 3 gyroscopes, which are positioned orthogonally to each other. Thus, the accelerations and rotation rates for all spatial dimensions can be determined. An IMU, which comprises twice three individual sensors, is referred to as a 6-axis IMU. Figure 2.3 shows the usual terms used to describe the measurements in relation to a vehicle. The acceleration is measured for the axes x , y and z . The acceleration is measured in m/s^2 and of the rotation speed in rad/s , in which $2\pi \cdot \text{rad} = 360^\circ$ of rotation. The rotations around these axes are referred to as roll, pitch and yaw. Some IMUs also include a 3-axis magnetometer, which is accordingly denoted as a 9-axis IMU. This mag-

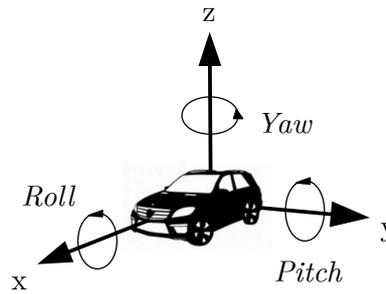


Figure 2.3: Illustration of accelerometer (x , y and z) and gyroscope (roll, pitch and yaw) forces relative to a car. Both, the acceleration in each dimension and the rotation speed for each dimension can be determined with a 6-axis IMU. The unit of acceleration is m/s^2 and of the rotation speed is rad/s , in which $2\pi \cdot \text{rad} = 360^\circ$ of rotation.

netic field sensors or compass enable the determination of the vehicle's absolute position.

2.2.4 Adoption Requirements

For a thorough review of literature directly related to the present work, criteria are identified in the following, which are necessary to facilitate a broad adoption of a crowdsensing-based monitoring system. The related literature will then be discussed on these criteria.

Calibration: A crowdsensing-based road roughness monitoring system is considered to be capable of being calibrated if it is able to handle the heterogeneity of vehicles and sensors in the crowd by taking into account their differences. A common approach for inferring from the behavior of a vehicle—for example from its body movement—on the road condition is to thoroughly describe the vehicle and its dynamic behavior by a physical model. Thus, most existing approaches explicitly or implicitly rely on a so-called quarter-car-model, which is a physical model representing a car and its suspension system (Jazar, 2014). For the application in a crowd-based system, in which many different vehicles and sensors should contribute, appropriate model parameters have to be determined for each participant. These parameters can be estimated based on empirical data using machine learning methods. Accordingly, the first criterion is fulfilled if the crowd-based system provides empirical calibration facilities, such as by machine learning methods.

Automatic reference: In order to obtain a scalable system, which allows to encompass many participants, the effort necessary for the calibration described in the previous criterion must be manageable, even if the number of

participants is very large. The use of supervised learning algorithms for calibration purposes requires a training data set, which is a combination of the participant's sensor data and the actual values of the target variable—the latter one is called ground truth. Accordingly, in the case of sensing road roughness, information about the roughness from former measurements are required to be referenced with the sensor data. Thus, the "automatic reference" requirement is considered to be fulfilled if the system facilitates an automatic alignment of the participant's sensor data to the ground truth. A calibration capability in combination with an automatic referencing can be referred to as a self-calibration capability of the system.

Consumer device: In the vehicular sensing domain, as described in Section 2.2.2, a basic distinction can be made between the use of vehicle sensors and sensors carried by passengers by means of consumer devices, such as smartphones. The use of the sensors included in consumer devices has its advantages. There is no need for the cooperation with automobile manufacturers. Apart from the consent of the individual participants, no additional allowance from companies is required. The use of consumer devices therefore enables an immediate realization and a rapid increase in the number of participants.

Combination: A substantial advantage of crowd-based applications is the opportunity to have access to a large number of individual contributions. In the context of crowd sensing, this can result in being provided with many single measurements. In order to make use of this aspect, a combination of several measurements should be performed. Besides the fact that customers of a crowdsensing-based system may only be interested in the aggregates, a temporal combination can lead to more robust results (Ma et al., 2014). Depending on whether an aggregation of individual measurements is carried out to increase robustness and performance, the prerequisite for a combination is considered to be fulfilled.

Infrastructure: The operation of a crowdsensing-based system requires an IT infrastructure, serving as a backend. In terms of resource efficiency, the backend system should be able to flexibly adapt its capacity to fluctuations in the load generated by the crowd. It is to be expected that the load is subject to seasonal fluctuations—such as daily and weekly—and that there will be a trend component particularly when the system is launched. Therefore, the IT infrastructure should be able to utilize appropriate scaling technologies. If scaling capability is addressed, the criterion infrastructure is considered to be fulfilled.

Utility: The long-term viability of a crowdsensing-based system depends on its utility to stakeholders. As already indicated, road users can benefit from the

consideration of road condition information to reduce their vehicle wear. In addition, road authorities can improve their maintenance strategies if they are provided with a spatio-temporally accurate information basis. Therefore, when developing a crowdsensing-based road condition monitoring system, the utility of both potential customers should be examined. If so, the utility criterion is met.

Criteria relevant to realize and operate a system for crowdsensing-based monitoring road roughness were outlined. In the following section related literature is discussed based on these criteria.

2.2.5 Related Work

This section discusses prominent research that is directly related to the crowdsensing-based road roughness monitoring described in the thesis at hand. It complements the more general literature presented in Section 2.2.2. Besides a brief description of each reference, a critical discussion is provided based on the requirements listed in the previous section.

Forslöf and Jones (2015) describe an approach to use IMU sensors from smartphones to determine the IRI. First, a single physical model for all participants was applied. To target the requirement of providing a calibration, an ability to linearly adjust the physical models is considered by multiplying the estimations with a factor, which can be specified individually for each participant. However, this factor is manually determined by an operator, while comparing the model's results with the actual road roughness, which is measured with a laser-based approach. Even if the fact that this is a rather subjective approach is neglected, such a manual intervention does not allow for a widespread adoption. For addressing this limitation, next to the purely physical models, machine learning-based approaches to estimate the IRI are considered additionally. In particular, linear regressions are applied. Since the parameters have to be determined for each vehicle individually, just a few representative vehicles are considered for training. This results in relying on a few pre-calibrated models—for small, medium sized and large cars—that should serve for all participants. This comes hand in hand with the burden of an increased inaccuracy. Forslöf and Jones (2015) regularly conduct surveys with the Automobile Association "Motormannen" (<http://motormannen.se>) in Sweden. Within this cooperation, for example, 92 000 km of roads were surveyed in Sweden in 2014. The Volvo V40 inspection cars equipped with Android smartphones were used. The histograms depicted in Figure 2.4 emphasize the difference between distributions of the official and crowd-based IRI measurements from a survey in Gävleborg, Sweden. The quantiles of the actual IRI values are: 0.33 (0 % quantile), 1.17 (25 % quantile), 1.80 (50 % quantile), 2.92 (75 % quantile) and 24.62 (100 % quantile). They claim to

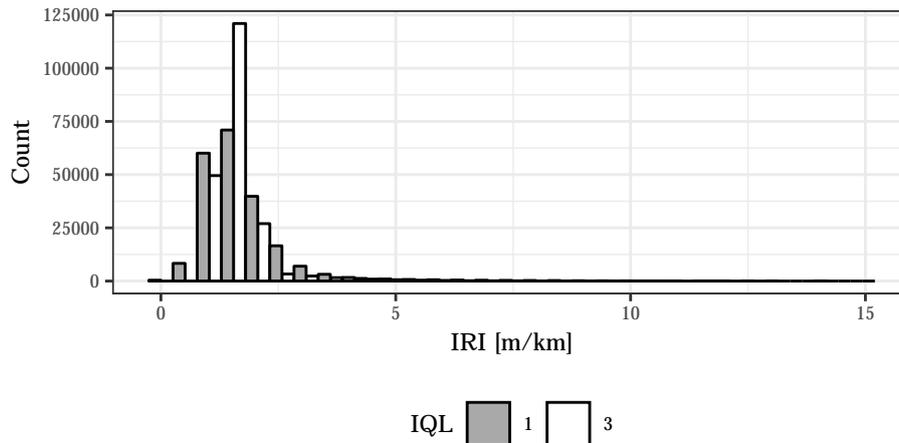


Figure 2.4: Histograms of the IRI measured with laser sensors and IMUs of national and county road segments in Gävleborg county, Sweden in 2014 (Forslöf and Jones, 2015). The dark grey histogram represents the laser measurements of the IRI (IQL 1). The white histogram shows the measurements with the IMU sensors (IQL 3).

achieve a performance up to a coefficient of determination R^2 of 51.54 %. The difference between crowdsensing-based and actual measurements can be assumed to be caused by just considering linear dependencies between the features from the IMU sensors and the actual IRI values and by the usage of a few standard models that are applied to all vehicles.

Yagi (2014) describes the usage of physical models—which model the movement of the vehicle chassis depending on the absorption characteristics of the suspension system and the absorption behavior of the tires—for determining the IRI crowd-based. Analogous to the previous approach, the data is collected by using IMU sensors included in smartphones, which itself are attached to the dashboard of the vehicle. However, neither a manual nor an automatic calibration is considered. Accordingly, this absence of an individual calibration to the vehicle’s and sensor’s characteristics has to be regarded as a limitation. Since an implementation of the system exists, it can be assumed that questions regarding infrastructure and utility were addressed. However, the report does neither include any information on the selected IT infrastructure and its scaling capability, nor does it provide a thorough investigation of the benefits of different potential customers.

Nitsche et al. (2014) apply support vector machines (SVMs) for distinguishing classes based on the WLP. As described in Section 2.1.2, the WLP is another metric besides the IRI to describe road roughness. The WLP was defined by Ueckermann and Steinauer (2008). On the one hand, Nitsche et al. (2014) determine the actual road roughness by attaching additional laser sensors to the probe vehicle for calibration reasons. Hereby, they are able to define roughness

classes serving as the ground truth. On the other hand, the accelerometer measurements which reflect the actual participants' sensor data contributions and the corresponding laser measurements are collected in parallel with the same vehicle. Thus, there is no need to additionally align probe vehicle measurements and laser measurements—no automatic referencing was required. However, attaching additional sensors to newly participating vehicles for calibration reasons—and thus avoiding an alignment of sensor data and ground truth—is not feasible in a crowdsensing system on a large scale. Furthermore, they use additional special sensors and no consumer devices for the acceleration measurement. The participants would not only have to include them during the calibration phase but for all drives. In addition to acceleration measurements, they also access information from the vehicle, such as the steering wheel position, wheel speed, pedal position and engine rotation.

Douangphachanh and Oneyama (2013) also apply a supervised machine learning approach for estimating the road roughness based on accelerometer sensor data. Unique in this work is that they have investigated the positioning of the smartphone in different places—near the gearshift and in the driver's pocket. The sensor measurements for each dimension and for the aggregate from all dimensions of the IMU were transferred to the frequency domain. The resulting magnitudes were used as features in a linear regression to estimate the actual IRI. This way, at least linear relationships can be determined empirically analogous to the first approach of Forslöf and Jones (2015) and thus a basic calibration for different vehicle types can be realized. Additional sensors in the form of an external GPS sensor are used to determine the position of the smartphone. The results are only representative for a productive system if such external GPS sensors were also used there. However, they are not nearly as widespread as smartphones, which is a limitation of the approach. A further limitation is that the ground truth is also based on readings from another single IMU. Since the calibration is performed on IQL 3, the trained models may not attempt to estimate the actual IRI, but an incorrect one. This can be assumed by considering the comparison of IQL 1 and IQL 3 readings shown in Figure 2.4. Accordingly, the model performances stated by Douangphachanh and Oneyama (2013) are to be interpreted in light of this fact. In addition, the sensor data from the smartphones and the data used to determine the ground truth were gathered simultaneously with the same vehicle, which is similar to the approach from Nitsche et al. (2014). Although different GPS receivers were used for the different recording methods, combining the data to generate a training data set was not subject to lateral deviations. No explanation has been given of how this referencing was carried out.

Jang et al. (2017) set a slightly different focus than the approaches described above. First, no estimation of a roughness metric, such as the IRI, is performed, but a classification into the classes smooth, rough and impulses. The latter describes the presence of single road anomalies. Therefore, the work is not of interest from an automatic calibration perspective. But an interesting approach

for the aggregation of multiple measurements is described. For this purpose, a clustering algorithm is used, which aggregates driving segments based on their location and direction. The aggregation of the individual classification results is then carried out using a voting mechanism that cannot directly be transferred to a regression approach. In addition, there is no investigation of whether or how the aggregation affects the accuracy of the classification. Jang et al. (2017) also provide a description of a client-server-infrastructure—at least on a high level. However, the low detail level does not allow for a direct implementation. Furthermore, no aspects of scalability are addressed.

All of the aforementioned approaches have limitations in their applicability in a real-world crowdsensing-based system: they are not feasible for a crowd-based approach because of the inability to automatically provide individually calibrated models. Thus, they lack easy adaptable models for new participants. Other approaches rely on additional sensors, which is also not feasible for a crowd-based approach. Requirements regarding an IT infrastructure and the system’s utility to customers are partially discussed in some approaches. A summary of the literature discussed and its fulfillment of the requirements described in Section 2.2.4 is provided in Table 2.2. As the resources for road maintenance

Table 2.2: Summary of related literature. Next to the discussed literature the present work is listed in the bottom row. The degree to which the requirements identified in Section 2.2.4 are fulfilled is indicated for each reference (●: fulfilled, ◐: partially fulfilled, ○: not fulfilled or not stated).

Reference	Criterion					
	Calibration	Automatic reference	Consumer device	Combination	Infrastructure	Utility
Forsl�f and Jones (2015)	◐	○	●	○	◐	◐
Yagi (2014)	○	○	●	○	◐	◐
Nitsche et al. (2014)	●	○	○	○	○	○
Douangphachanh and Oneyama (2013)	◐	◐	◐	○	○	○
Jang et al. (2017)	○	○	○	◐	◐	○
This work	◐	●	●	●	●	◐

are still strictly limited and as the potential of actually utilizing the crowd for performing road condition inspections increases, there is a lack of research about how road condition inspections can be performed by a heterogeneous crowd, how a backend infrastructure for a practical application has to look like, and how

a crowd-based service can be facilitated to provide benefits to road users and road authorities. This information demand is addressed by the research questions itemized in Section 1.2 which are investigated in the remaining parts of this thesis.

2.3 Costs of Road Usage and Maintenance

When considering costs related to road condition monitoring, again the two parties road users and road authorities have to be considered. While road users incur fuel costs, vehicle wear and other expenses through the use of the road infrastructure, road authorities incur costs, such as for the initial construction of roads and for recurring maintenance tasks. These costs of both are directly and indirectly interdependent. For example, road maintenance tasks, such as road works, can cause congestion and thus directly lead to higher travel time costs for road users. Not performing maintenance tasks, however, let the road deteriorate and a resulting bad road condition can indirectly increase the road user costs (RUCs) due to an increased vehicle wear.

The World Bank developed a collection of Highway Development and Management (HDM) models to assist road authorities in planning of efficient construction and maintenance strategies (Watanatada et al., 1987; Kerali et al., 2006). The models are based on the study from De Weille (1966), which was also funded by the World Bank. The currently latest version of the model is HDM-4. Bein (1993) gives an overview of the historical model development. The HDM models are intended for planning efficient infrastructure investments and therefore consider road construction, maintenance costs and road usage costs. In terms of maximizing the welfare, the total transportation costs should be minimized, as depicted in Figure 2.5 (Watanatada et al., 1987; Kerali et al., 2006). The respective cost components are explained in the following two sections.

2.3.1 Road Users

Regarding the road usage the HDM models consider road user effects (RUEs). Figure 2.6 gives an overview of the components of RUEs. RUEs comprises the effects of road transportation on the road user and its environment. In accordance to Figure 2.6, this includes, among other components, fuel consumption, travel time, accidents and emissions. When an RUE becomes apparent, costs directly or indirectly caused by the RUE can be assigned to it. For instance, fuel consumption can cause fuel costs, travel time leads to opportunity costs—since no alternative activities can be performed—, accidents cause damage costs and emissions result in health costs. All these costs caused by RUE components are referred to as RUCs. Often only subsets of the RUCs are considered. These are also often referred to as vehicle operating costs (VOCs). According to Bennett

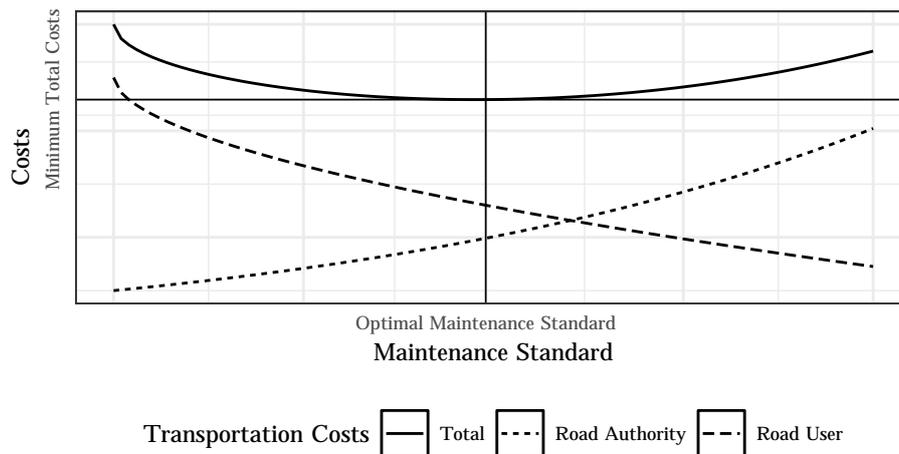


Figure 2.5: Interplay of transportation costs of road users and road authorities. Providing a higher maintenance standard leads to higher maintenance costs for the road authorities. For road users, on the other hand, a higher maintenance standard has a favorable effect on their costs associated with transportation. The minimum total costs should be aimed, even if the resulting optimal maintenance standard is not cost minimizing from the point of view of the road authorities.

and Greenwood (2003), the VOCs comprise costs resulting from fuel, tire, oil, lubricant consumptions, repairs, financing and overheads.

The HDM models claim a relationship between road conditions, such as the IRI, and the VOCs. However, the model calibrations were based on data from roads in developing countries—Kenya, India and in the Caribbean within 1971 and 1986 and in Brazil within 1975 and 1984. Chatti and Zaabar (2012) addressed these limitations to developing countries by conducting a survey on US-roads in 2011. This work was performed within the National Cooperative Highway Research Program (NCHRP). They calibrated the latest HDM fuel consumption and tire wear models and developed an improved repair and maintenance model. Accordingly, they investigated a subset of the VOC. In particular, they showed for different vehicle types that—and to what extent—fuel consumption, tire consumption and repair rates are higher when the vehicles are driven on roads with an overall higher IRI, that means on rougher roads.

2.3.2 Road Authorities

The condition of roads exposed to traffic, weather, material fatigue, etc. deteriorates over time (Kerali et al., 2006). Road authorities are required to take cost efficient maintenance actions to ensure a road condition that is appropriate for road users. Figure 2.7 schematically depicts the road condition development—in terms of the IRI—over time. A distinction is made between the two scenarios,

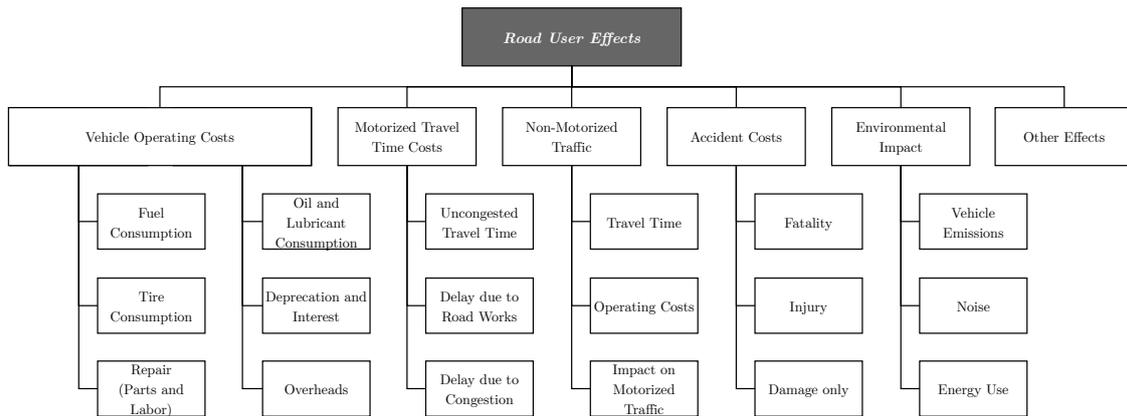


Figure 2.6: Components of RUEs (Bennett and Greenwood, 2003). RUEs are the effects caused by road transportation on road users and the environment. The RUEs can cause costs which are referred to as RUCs.

with and without maintenance. Decisions that have to be made—which have an influence on the road condition—after an initial construction at t_0 are basically, when and which maintenance action, in this case at t_1 and t_2 have to be carried out for rehabilitating the road condition. Additionally, a differentiation has to be made with regard to different maintenance actions, such as an early resurfacing versus a late total reconstruction. Since the decisions depend on the stochastic development of the road condition, it is also necessary for road authorities to decide whether, when and which inspections should be carried out in order to obtain information on the actual road condition.

Extensive work has been done in the modeling of road deterioration and maintenance task scheduling through Markov decision processes (MDPs). Gao and Zhang (2013) developed an MDP model for a road maintenance optimization problem including RUC and computed an optimal maintenance policy. Smilowitz and Madanat (2000) extended MDP approaches by considering uncertainty within the inspection methods. They call this approach latent MDP. Similar approaches are also known as partially observable Markov decision processes (POMDPs). These approaches consider discrete condition states. Schöbi and Chatzi (2016) describe the implementation of a continuous state POMDP that considers inspection uncertainty within maintenance planning. Their results are compared to a discrete model. Han et al. (2014) use a Bayesian Markov hazard model and time series data for highway deterioration forecasting as an extension to a conventional MDP. Other approaches to determine optimal maintenance strategies also take into account imperfections in maintenance and inspections (Pham and Wang, 1996; Le and Tan, 2013). Even though there has been research done in the field of optimal scheduling of road maintenance tasks based on road conditions, a combination with crowdsensing-based road condition inspections has not been investigated.

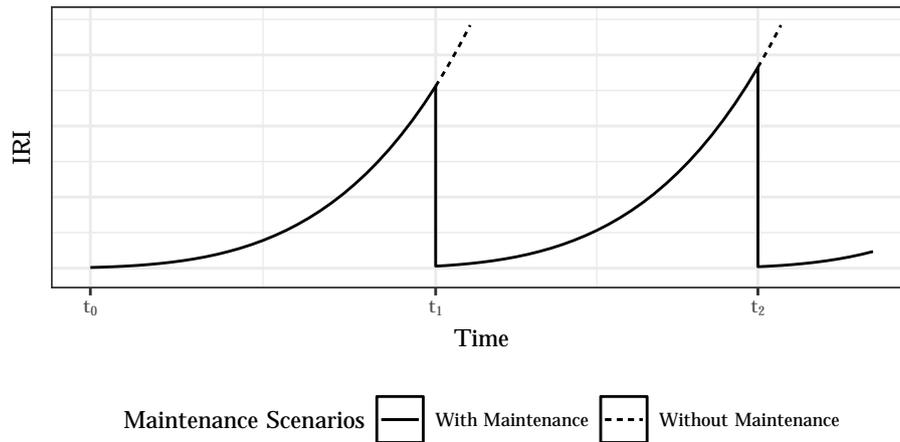


Figure 2.7: Development of road condition over time (Kerali et al., 2006). The two scenarios, with and without maintenance by a road authority, are distinguished. t_1 and t_2 represent rehabilitation actions after initial construction of the road at t_0 .

Besides considering maintenance costs, road authorities should consider RUCs related to the road condition (Ouyang and Madanat, 2004). As explained in the previous section, the RUCs are partially dependent on road conditions. For instance, higher fuel consumption, wear and longer travel times may have to be taken into account if the road condition is poor or if maintenance work hinders traffic flow (Bennett and Greenwood, 2003). However, it is also the case that the final decision as to where and when, a particular maintenance action is to be carried out, is made by human decision makers. This applies to both public roads and privately operated roads. This is why there may be a tendency towards a reduction in the costs of the road authorities and not that of both, road authorities and road users.

Part II

Crowdsensing-Based Road Surface Monitoring

Chapter 3

Self-Calibration

The widespread adoption of smart devices has the potential for an unprecedented real-time assessment of road conditions. One of the challenges, however, is the heterogeneous nature of measurements from different sensor and vehicle types that contribute to a crowdsensing-based road monitoring. In this chapter, a self-calibration approach is proposed that utilizes statistical models trained individually for each vehicle, which in turn can automatically be integrated into a crowdsensing-based system. Since the international roughness index (IRI) is an important road profile quality indicator well suited for a crowd-based sensing approach, it is the considered metric in this and the following chapters. The approach is evaluated on a data set collected from seven independent drives with a total distance of 32 km with a smartphone-equipped car. The data set contains time, global positioning system (GPS) and inertial measurement unit (IMU) measurements. The results show that the self-calibration approach is applicable and can reach a mean R^2 of 0.68 and a normalized root mean square error (NRMSE) of 9% in mean over all single car predictions. Parts of this chapter are adapted from the publication: Laubis, K., V. Simko, and A. Schuller (2016c). Road Condition Measurement and Assessment: A Crowd Based Sensing Approach. In *Proceedings of the 37th AIS International Conference on Information Systems*, pp. 1–10.

3.1 Introduction

The usage of crowdsensing-based measurements from, for example, the IMU in smart devices, such as smartphones, potentially allows for a near real-time assessment of road conditions for given standard quality indicators, such as the IRI. Crowd-based road condition sensing currently faces the challenge to utilize measurements from different vehicles and sensors. The main reason for this is that vehicles have versatile physical behaviors, such as different suspension systems, and that smartphones are placed at different locations within the vehicle. Thus, facilitating a system with many participants for robust real-world application is difficult and requires both, an easy integration of new participants and the ability to adapt to changes for existing participants.

The focus of this chapter is to develop and evaluate the ability to utilize measurements from heterogeneous vehicles and sensors by automatically build participant-specific models in a self-calibration manner. The possibility of individual calibration for the vehicles and measurement devices in order to determine the IRI for the evaluation of current road conditions is investigated. Accordingly, Research Question 1 is addressed in this chapter by the following two sub-questions: (a) How accurate can participants contribute to a crowd-based road roughness measurement system if their vehicle and smartphone get calibrated automatically while driving on a public road? (b) How robust are these results with respect to repeated test drives and to different machine learning algorithms?

Therefore, an approach to include individual vehicle characteristics in the assessment and the interpretation of sensor data is proposed. Given a global road condition database, which initially just contains high-quality measurements, an alignment of this road condition information to the measurements from the crowd is performed in an alignment module. This is a challenging task, since not only GPS measurements from the participants can be faulty, but also the ground truth data set regarding the alignment to the road network. The main idea of the research scenario is depicted in Figure 3.1.

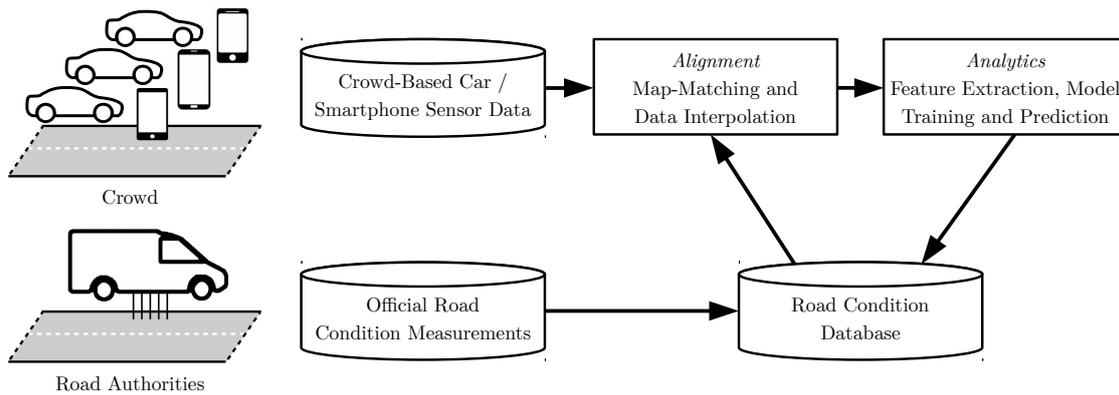


Figure 3.1: Outline of self-calibration approach for an effortless adoption of new participants in a crowdsensing-based road surface monitoring system. Raw sensor data from a new participant is sent to the backend system. There, the data is aligned to actual road condition information serving as a ground truth. This aligned data set is used for extracting features and the training of an individual model for the new participant.

The chapter is structured as follows: in the next section, research related to the self-calibration approach is briefly discussed with a focus on the applied machine learning methods and foundations of regression methods are provided. The following section describes the theoretical basis of the self-calibration and a scenario for its evaluation. Then the results based on a real-world data set from drives with a smartphone-equipped vehicle are presented and discussed before the chapter finally concludes.

3.2 Regression Methods and Performance Metrics

Within the proposed self-calibration approach, regression methods are used to determine participant-specific models. These allow the determination of the road roughness on a ratio scale type rather than, only on a nominal or ordinal scale type—as it is the case for many of the approaches represented in the literature, which, for example, only distinguish whether an anomaly is present or not (Eriksson et al., 2008; Tai et al., 2010; Mohan et al., 2008b). A regression is necessary because the IRI represents a ratio scale type and not just a classification. Random forest regressions (RFRs) and support vector regressions (SVRs) are briefly described below as exemplary regression methods. In addition to their methodology, their characteristics are discussed. Furthermore, performance metrics for regression methods, which will be used in the further course of this thesis, are briefly introduced in this section.

RFRs, which represent the regression variant of random forests (RFs), are a method for training prediction models, which increase in popularity (Breiman, 2001). They consist of multiple single regression trees (Breiman et al., 1984). The final prediction is determined by the average of the single trees' predictions. Each tree is trained on a random subset of the whole feature set. The number of considered features to train each tree is a parameter to choose. At each split, a single feature and a corresponding value are chosen for splitting, which maximizes the change of an impurity measure, such as the Gini impurity, between the parent node and the child nodes. RFRs are especially feasible when handling problems with a small number of samples and a high number of dimensions without tending to overfitting. Furthermore, they scale very well with an increased number of dimensions since the number of considered features per tree is inherently limited.

SVR is a further common method to build prediction models, which is based on support vector machines (SVMs) (Vapnik et al., 1997; Schölkopf et al., 2002). Both approaches are related to reproducing kernel Hilbert spaces (Cristianini and Shawe-Taylor, 2000). The training data—represented as data points—is mapped into a higher-dimensional feature space to make non-linear problems separable by a so-called hyperplane. The hyperplane within the feature space is defined to maximize its distances to its closest data points, which are called support vectors. To avoid overfitting, single data points can be allowed to be misclassified in the training phase. Therefore, a cost parameter trading error penalty for stability can be determined for adjustment. A large cost parameter leads to low bias and high variance models, and vice-versa. For computing efficiency, a so-called kernel trick can be applied so that by choosing a suitable transformation function, no actual transformation has to be performed, but the optimization problem can be solved in the original space itself.

Whereas accuracy, precision and recall are common metrics to determine the performance of a classification, for regressions there are different goodness of fit metrics. The metrics used in this thesis are briefly introduced in the following.

The $R^2 \in \mathbb{R}$ indicates how much of the variance in the ground truth data is explained by the prediction model. It is defined in Equation 3.1

$$R^2 = \frac{\sum_{i=1}^n (\bar{y} - \hat{y}_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (3.1)$$

where

- n is the number of samples in the considered data set,
- $y_i \in \mathbb{R}^+$ is an actual value of the data set, such as a laser-based measured IRI value,
- $\hat{y}_i \in \mathbb{R}$ is a corresponding predicted value, such as an IRI estimation based on crowdsensing,
- $\bar{y} \in \mathbb{R}^+$ is the mean of the actual values y .

The numerator represents the variance explained by the model and the denominator represents the total variance in the data set. It is a widely used metric for determining the goodness of fit of a regression model and its reliability. Since it is a relative measure, it is easy to interpret and can be compared easily between models. As stated above, the metric, which can be interpreted as a percentage, can become negative when the chosen model fits the data very poorly. This is the case if the model explains less variance than a naive estimate with the mean value \bar{y} .

Next to the R^2 performance metric, the root mean square error (RMSE) $\in \mathbb{R}^+$ is a common absolute error metric for describing the model's performance. It is defined in Equation 3.2

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (3.2)$$

The squaring weighs larger estimation errors higher, thus they have a much larger influence on the metric. The square root allows for an easy interpretation, as the metric has the same unit as the actual values.

Let $y_{max} \in \mathbb{R}^+$ and $y_{min} \in \mathbb{R}^+$ be the maximum and minimum of the actual values, such as the laser-based measured IRI values, of the considered road. The normalization of the RMSE to the spread of the actual values is defined in Equation 3.3

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}. \quad (3.3)$$

The NRMSE was chosen for the analyses in addition to the R^2 and RMSE for better comparability of the results with regard to its implications.

3.3 Research Design

The self-calibration utilizes GPS fixes, IMU measurements and a road network. The first step is to apply a map matching to the GPS fixes, followed by sensor data alignment through multiple interpolations. Thereafter a feature extraction is performed. The last step is the model training and the prediction. These steps are formally described in this section and a scenario to evaluate the approach is presented.

3.3.1 Map-Matching

First, the GPS fixes delivered from a vehicle are mapped onto a road network, which is necessary because GPS fixes are subject to measurement inaccuracy—especially when GPS sensors of consumer devices are employed. Therefore, the first step is to address the set of GPS fixes G . Every GPS fix $g \in G$ is a tuple as defined in Equation 3.4

$$g = (lat, lon, ts), \quad (3.4)$$

where

- lat, lon are GPS coordinates—latitude and longitude (e. g. $lat = 56.78901$),
- ts is a timestamp in milliseconds. For convenience, the notation ts_g is used to denote the timestamp of a particular GPS fix g .

A map-matching algorithm *mapmatch* is used, which considers road network information $Rnet$. The corresponding function is defined in Equation 3.5

$$mapmatch_{Rnet}(G) \mapsto G', \quad (3.5)$$

where G' is a new set of GPS coordinates that are matched to the actual road network $Rnet$. Both sets are of the same size, i. e., $\|G\| = \|G'\|$. Timestamps do not change during map-matching, as described in Equation 3.6

$$\forall i : g_i \in G, g'_i \in G', g_i = (lat, lon, ts), g'_i = (lat', lon', ts). \quad (3.6)$$

A function *interpolate* is used to convert matched GPS fixes G' into virtual GPS fixes VG that are equidistantly placed on the road (with distance ed). The size of VG is much larger than G' , i. e. $\|VG\| \gg \|G'\|$. Thus, Equation 3.7 defines the interpolation

$$\begin{aligned} & interpolate_{Rnet}(G', ed) \mapsto VG, \\ & \forall g' \in G' : \forall vg \in vgnext(g', ed) : \\ & ts_{vg} = ts_{g'} + \frac{ts_{next(g')} - ts_{g'}}{\|next(g') - g'\|} \cdot \|vg - g'\|, \end{aligned} \quad (3.7)$$

where

- function $next(g')$ finds the next matched GPS coordinate based on timestamp from g' ,
- function $vgnext(g', ed)$ returns all virtual GPS fixes $vg \in VG$ such that it holds: $ts_{g'} < ts_{vg} < ts_{next(g')}$.

3.3.2 Data Alignment

Following the determination of virtual GPS fixes VG , sensor readings are assigned to them based on the timestamps. For every virtual GPS fix vg , there can be multiple sensor readings available—multiple ones per vehicle. This is when the sensor readings are gathered at a higher frequency than the GPS readings. Therefore, an aggregation function, such as the mean, is applied. A continuous approximation, such as linear or cubic spline, as given in Equation 3.8 to the sensor data S , is also performed

$$interpolate(S) \mapsto S'. \quad (3.8)$$

Thus, the virtual GPS fixes and the approximated sensor data can be joined through their timestamps, as described in Equation 3.9

$$timejoin(VG, S') \mapsto \{s \in S' : \exists g \in VG, ts_g = ts_s\}. \quad (3.9)$$

As a result, sensor data mapped to equidistant samples in space are obtained that do not have the time component anymore. This allows to aggregate samples from different sensors and simplifies predictions and further analyses.

3.3.3 Feature Extraction

The data set can be assumed as a matrix M with n rows and k columns. Columns represent features f_1, \dots, f_k . Rows represent samples $\vec{y}_1, \dots, \vec{y}_n$. Each sample belongs to a slot of length ed . As Equation 3.10 denotes, each sample with index $j = 1, \dots, n$ is defined as

$$\vec{y}_j = (f_1(x_j), \dots, f_k(x_j)). \quad (3.10)$$

The matrix representing the whole data set is then defined as follows by Equation 3.11

$$M = \begin{bmatrix} \vec{y}_1 \\ \vdots \\ \vec{y}_n \end{bmatrix} = \begin{bmatrix} f_1(x_1) & \dots & f_k(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \dots & f_k(x_n) \end{bmatrix}. \quad (3.11)$$

Furthermore, a function $ctxsize$ is defined in Equation 3.12, which assigns a natural number (context size) to each feature f_i

$$ctxsize(f_i) \mapsto \mathbb{N}. \quad (3.12)$$

This number represents how many neighboring samples contribute to the computation of a single sample. For example, the $ctxsize$ of a moving average feature would be its window size. Knowing this context size allows for a meaningful sub-sampling and for a precisely separation of the data sets in a training and testing data set. The context size can be defined for the whole matrix M in Equation 3.13

$$ctxsize(M) = \max_{i=1\dots k} (ctxsize(f_i)). \quad (3.13)$$

One of the features, usually the last feature f_k , represents the outcome variable—such as the actual IRI value of the corresponding road slot. It is used during training and evaluation.

3.3.4 Model Building

Let M be the $n \times k$ data matrix defined in Equation 3.11. The matrix is split row-wise into two matrices M^{test} with size $t \times k$ and M^{train} with size $(n - t) \times k$. A common ratio is $\frac{t}{n} = 0.2$, i. e., 20% of the samples will be used for testing and 80% for training. A prediction model P is trained using the training samples M^{train} as denoted in Equation 3.14

$$P = train(M^{train}), \quad (3.14)$$

where the $train$ function can be a regression method, as described in Section 3.2. The resulting model P is itself a function that can be used for making predictions about input vectors \vec{x} . In other words, it assigns a predicted outcome $\hat{y} \in \mathbb{R}$ to a given input \vec{x} , according to $P : \mathbb{R}^{k-1} \mapsto \mathbb{R}$.

Having trained a prediction model P , the out-of-sample prediction performance E can be evaluated as follows. From the test set M^{test} , a projected matrix X is created by removing the ground truth column f_k (representing the actual outcome values y). This reduction is represented by Equation 3.15

$$X = (f_1, \dots, f_{k-1}) = \begin{bmatrix} f_1(x_1) & \dots & f_{k-1}(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_t) & \dots & f_{k-1}(x_t) \end{bmatrix}. \quad (3.15)$$

The predictor P can be applied to each row of the input matrix X . The outcome can be compared to the ground truth value y . This way, the performance measure

E —such as the R^2 , RMSE or similar metrics, as described in Section 3.2—can be computed following Equation 3.16

$$E(P(X), f_k) = E \left(\begin{bmatrix} P(f_1(x_1), \dots, f_{k-1}(x_1)) \\ \vdots \\ P(f_1(x_t), \dots, f_{k-1}(x_t)) \end{bmatrix}, \begin{bmatrix} f_k(x_1) \\ \vdots \\ f_k(x_t) \end{bmatrix} \right). \quad (3.16)$$

The selection of M^{test} and M^{train} can be done systematically multiple times in a multiple fold cross validation. Each fold gives a different E_{fold} value. Thus, a robust estimate of the out-of-sample performance of a model can be determined based on multiple folds. In the following section, the described self-calibration approach is applied for evaluating its feasibility.

3.3.5 Scenario Setup

This section evaluates the self-calibration approach described above. As outlined in Section 2.1.2, the IRI represents one of the most important road condition indicators, thus it is used for the evaluation of the self-calibration approach (Múčka, 2016). First, the empirical data basis in terms of the ground truth and the data collected from vehicles is described. This is followed by an explanation of how the steps of the self-calibration approach, map matching, data alignment, feature extraction and model building, are applied. The results are given in the later section.

Ground Truth

The evaluation of the self-calibration approach is performed by investigating the longitudinal road roughness for a recently paved 2.28 km road link on the district road K3535 in Germany. For this road link, information about the road profile measured by a laser-equipped vehicle was provided by the Institute of Highway and Railroad Engineering (ISE) at the Karlsruhe Institute of Technology (KIT).

Given the profile in both driving directions, the IRI is calculated according to Section 2.1.2. The IRI was calculated for 100 m segments with an overlap of 80 m. This results in overall 220 samples for 4.56 km. For this IRI calculation, the R-package `rroad` (<https://CRAN.R-project.org/package=rroad>) that was developed in the course of this work, was used (Simko and Laubis, 2017). The resulting IRI values for the road segments are considered as the ground truth for the model building. Figure 3.2 shows a subset of these actual IRI values. Between the segments 190 and 215, roughness caused by two bridges is indicated by an increased IRI. The IRI of all considered segments ranges from 0.8 m/km to 2.94 m/km with a median of 1.2 m/km and a variance of 0.147. Figure 3.3 provides a histogram and a boxplot summarizing these actual IRI values.

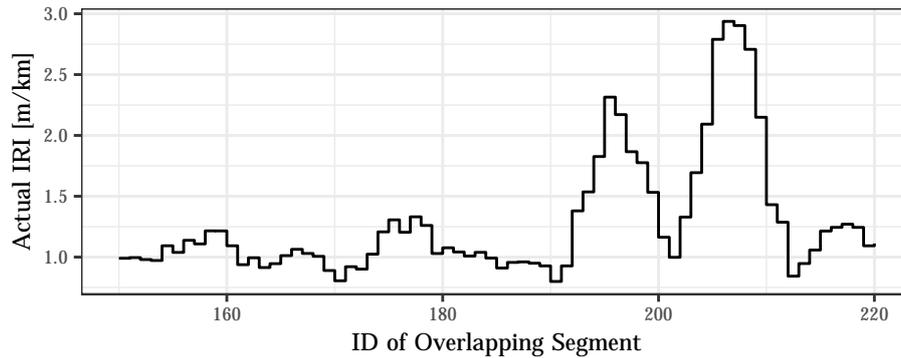


Figure 3.2: Subset of actual IRI values from a section of the district road K3535, Germany. Two peaks between the segments 190 and 215 indicate bridges at these road sections.

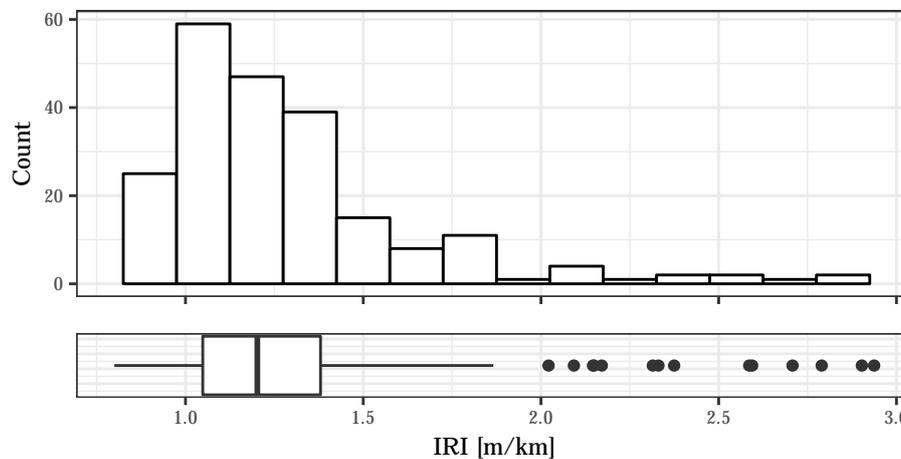


Figure 3.3: Histogram and boxplot of actual IRI values from a section of the district road K3535, Germany. Each sample represents a 100 m segment of the 4.56 km road link. In total 220 overlapping segments are considered.

Data Gathering

For this work, seven drives with a smartphone-equipped passenger car are performed on the K3535 road link to collect values from GPS and IMU sensors. It should be noted that the car drives and the measurements from the laser profiler are performed separately and on different days. Furthermore, the car is not equipped with additional sensors except those from the smartphone. Thus, the car can be assumed as an example for a new participant to a crowdsensing-based road condition monitoring system. The smartphone used is a Nexus 4 and it is attached to the middle of the dashboard with a car mount holder. The car used is a Renault Twingo. Next to the GPS coordinates, the accelerometer and gyroscope sensor values are recorded for all three axes. A new GPS fix is determined at a

frequency of 0.9952 Hz on average and depends on the smartphones' central processing unit (CPU) utilization. The frequency at which the accelerometer and gyroscope sensors record, also depend on the smartphone utilization. On average, a frequency of around 50 Hz is encountered for both sources. The speed was kept nearly constant at 75 km/h on average on all seven drives. Thus, a new GPS fix is received approximately every 20 m and sensing accelerometer and gyroscope values takes place roughly every 0.4 m. The following steps of the self-calibration approach are performed for each drive separately.

Map-Matching and Data Interpolation

Map-matching GPS fixes to a road network, common to all drives and to the ground truth information, is the first step of making measurements from multiple drives comparable and aligning the vehicle's features and the ground truth. Open Street Map (OSM) is used as the road network common for the laser-based measurements and the passenger car's.

The mapping to the road network is achieved by applying a hidden Markov model-based map-matching that considers inverse distance weighting between GPS fixes and the road positions (Newson and Krumm, 2009). The map-matching approach makes use of the Viterbi algorithm for maximizing the product of measurement probabilities and transition probabilities to determine the most likely route. An open source implementation of this algorithm from the project Open Street Routing Machine (OSRM) is employed in this thesis (Luxen and Vetter, 2011).

To have common slots on the road network for further analyses and to drill down the coarse spacial granularity at which the cars' positions are known, the road link is subdivided into 10 cm slots. This distance is chosen to consider sensor frequencies up to 200 Hz without information loss, assuming a speed of at least 72 km/h. The time at which the car passes each slot is determined by linear interpolation between two consecutive mapped GPS fixes. The next step is to align all sensor values to the global road slots based on their timestamps, which is performed by a second linear interpolation. Applying this interpolation leads to sampling of all sensors—accelerometer and gyroscope—from all cars with a common fine-grained spatial representation.

Since different sampling distances and road segment lengths are considered in this chapter, these are summarized for clarification. The ground truth IRI values are provided for overlapping 100 m road segments with a 20 m offset. The relevant road link is split in equidistant slots with a distance of 10 cm. The sensor measurements are aligned to these slots by map-matching and interpolation. For extracting features continuously, each slot is considered as one sample. For the model training and testing a subset of this resulting set of samples is chosen in a way that there is one sample kept for each corresponding ground truth IRI value.

Feature Extraction and Model Training

The unified IMU measurements are used for extracting features. Based on these spatio-temporal samples, the speed is derived and considered as an additional feature next to the accelerometer and gyroscope measurements. Next to the absolute readings from the accelerometer sensor, the relative linear acceleration excluding the gravity is considered. For considering the frequency content of these acceleration and gyroscope features, a continuous wavelet transformation (CWT) is performed (Torrence and Compo, 1998). The biwavelet R-package (<https://CRAN.R-project.org/package=biwavelet>) was used for this (Gouhier et al., 2016). From the bias-corrected wavelet power spectrum, the wavelengths 0.4 m (highest frequency), 0.8 m (high frequency), 2.26 m (medium frequency), 9.05 m (low frequency) and 51.21 m (lowest frequency)—with a likewise contextual information—are selected as additional features. Even though, most likely the features with a smaller contextual information are important, also the larger ones are extracted for determining their importance.

Since in today's road maintenance the IRI is mainly measured for consecutive road segments of 100 m, features with a likewise contextual information are derived from the speed and the scaled accelerometer and gyroscope samples. The aggregation is also performed in a continuous manner for getting additional values for each slot. Aggregation functions used are the mean, range from minimum to maximum, standard deviation, variance and root mean square (RMS). Table 3.1 summarizes the resulting 95 features, which are all z-score normalized.

Table 3.1: Features extracted from smartphone sensors. For each 100 m road segment summary statistics for the GPS velocity and for the IMU readings are determined. This results in 95 features.

Sensor	Aggregation function	Number of features
GPS velocity	Mean, range, std. dev., var., RMS	5
Accelerometer (3-axis)	Mean, range, std. dev., var., RMS, CWT for 5 bands	30
Linear accelerometer (3-axis)	Mean, range, std. dev., var., RMS, CWT for 5 bands	30
Gyroscope (3-axis)	Mean, range, std. dev., var., RMS, CWT for 5 bands	30

To reduce the resulting data set to a manageable size, every 200th segment is considered as one sample for model training and testing. With regard to the features with contextual information, this can be assumed as having 100 m samples with an offset of 20 m. In addition, this reduction is performed to achieve the same amount of samples as the ground truth segments. From the resulting 220

samples per drive—110 for each direction—20 % of the samples are kept back at the early beginning for evaluating the out-of-sample model performance. Within this data splitting, it is made sure to not put samples to the testing set, which overlap with samples in the training set.

An RFR and an SVR are applied to the training set of each drive separately for determining the IRI based on the extracted features. Each drive is treated independently—for both, training and testing—so that the models can be assumed as resulting from a different car. To reduce overfitting, a time slicing-based cross validation is performed within each training phase. The RF models are tuned by the number of considered features at each tree node. For the SVR a default Gaussian radial basis function kernel is chosen. The SVM models are tuned by the cost parameter and a radial basis kernel specific parameter gamma, for lowering the variance within the Gaussian function. The metric considered for cross validated tuning is the coefficient of determination R^2 .

3.4 Results

The following section provides the evaluation results of the self-calibration approach. Research Question 1 is answered before the chapter concludes with a summary and critical discussion.

3.4.1 Estimation Performance of Single Drives

The results of the evaluation give insight to the potential of the self-calibration approach. For evaluating the performance of the approach and likewise, for answering the Research Question 1 a), how accurate participants can contribute to a crowd-based road roughness measurement system if their vehicle and smartphone are calibrated automatically while driving on a public road, the R^2 , the RMSE and its normalized metric the NRMSE are determined. To answer the Research Question 1 b), how robust these results are with respect to repeated test drives and to different machine learning algorithms, the results for all seven drives and for both an RFR and an SVR are listed in Table 3.2. They are generated by cross-validation. The last four rows provide summary statistics of the performance metrics.

Except for the second drive, the RFR outperforms the SVR models. Thus, mainly the results for RFR are discussed. The worst model achieves an out-of-sample R^2 of 0.5968, which means that 59.68 % of the variance in the out-of-sample data is explained by the car's model. Likewise, the best model explains 76.79 % of the actual variance. Considering that the road is very smooth and thus, has not much variance at all, these results are remarkable. The range of the RMSE is from 0.2144 m/km of the second drive down to 0.1627 m/km of the fifth drive. If this error measure can be kept even for roads with higher variance in its profile

Table 3.2: Out-of-sample performance of crowdsensing-based IRI estimations by single drives. The performance is provided for RFR and SVR methods. For each of the seven drives the R^2 , the RMSE and the NRMSE are determined. The bottom rows summarize the performances over all drives.

Drive	RFR			SVR		
	R^2	RMSE	NRMSE	R^2	RMSE	NRMSE
1st drive	0.6319	0.2049	0.0958	0.5440	0.2280	0.1067
2nd drive	0.5968	0.2144	0.1003	0.6416	0.2022	0.0946
3rd drive	0.6207	0.2080	0.0973	0.5899	0.2163	0.1012
4th drive	0.7395	0.1724	0.0806	0.4649	0.2470	0.1156
5th drive	0.7679	0.1627	0.0761	0.7601	0.1654	0.0774
6th drive	0.7115	0.1814	0.0849	0.6598	0.1970	0.0922
7th drive	0.6799	0.1911	0.0894	0.3985	0.2619	0.1225
Max.	0.7679	0.2144	0.1003	0.7601	0.2619	0.1225
Mean	0.6783	0.1907	0.0892	0.5798	0.2168	0.1015
Min.	0.5968	0.1627	0.0761	0.3985	0.1654	0.0774

roughness, the crowd sensed IRI measurements can be assumed as a very meaningful indicator for road condition monitoring and road maintenance. This is because rougher conditions are expected to be more easily detected by ordinary consumer devices.

3.5 Conclusion

In this chapter, a generic approach for allowing an integration of new participants to a crowd-based road condition measurement system is described. The self-calibration approach is based on the use of machine learning methods and automatic referencing of data from the vehicles with ground truth. Thus, the approach does neither have a need of attaching additional sensors to the cars for calibration reasons, nor does it come hand in hand with the burden of inaccurate models. Furthermore, the approach is not limited to specific road characteristics or to specific sensor types. For this an individual calibration for each participant is proposed by training a prediction model based on features extracted from the smartphones' sensor values and on ground truth information of the road condition from accurate laser-based measurements.

With respect to the need for aligning the cars' measurements with the ground truth, the self-calibration approach consists of the two main modules alignment and analytics. Both are described in detail and evaluated by applying them to determine the IRI with a smartphone-equipped car. The steps performed within the evaluation, especially the feature extraction and model training are highly

depending on the road condition metric, being targeted. However, it has to be mentioned that this causes no loss of generalizability of the self-calibration approach. The results show that RFRs outperform SVRs. Applying RFRs, in mean 67.83 % of the variance of the road link's actual IRI is explained by single drive predictions. According to Research Question 1, it is shown how a self-calibration approach can and should be implemented with the aim of a high and accurate coverage.

The underlying research in this chapter has limitations in terms of the amount and variance of the investigated road condition information. Since the road section on the K3535 used for the evaluation was recently paved, the variance of the actual IRI is relatively low. It would be interesting to extend the evaluation to roads which are rougher or have a different construction type, such as concrete roads. Next to evaluating the self-calibration on rougher roads, the effects of using different cars and sensors characteristics—such as the sensitivity of the results to different sensing frequencies—can be examined. The latter aspect is especially important from the background, as different devices with different sensing frequencies would be used in a crowd-based approach. In addition, by taking into account the effects of the sensing frequency, it could be determined whether a reduction of the frequency should be performed in order to reduce the data volume. This allows to address the limitations in storage of mobile devices and in the amount of data that can be transmitted between the devices and the backend system. Therefore, these data reduction aspects are addressed in the next chapter.

Chapter 4

Data Reduction

Road condition estimation based on crowd-sensed data from smart devices placed within vehicles allows for determining given quality indicators, such as the international roughness index (IRI). As shown in the previous chapter, the challenge of utilizing measurements from heterogeneous sources in a crowd-based approach can be faced by individually calibrating statistical models for each participant. This self-calibration requires the training of the individual models to be performed in a backend system. This is because of the required computational power and the need of having information on the actual road condition in place for applying the supervised training. Therefore, the training data gathered by the participants has to be sent to the backend and stored temporarily on the smart devices beforehand. Since mobile data transmission and storage is expensive, in this chapter it is examined how a reduction of the amount of data sent to the backend system for training purposes affects the model's performance. Especially a reduction of the number of considered features and a reduction in sensing frequencies are assessed. Results show that reducing the number of features by approximately 50 % does not reduce the performance of the models. Likewise, it is observed that the approach can handle sensing frequencies down to 25 Hz without a performance reduction compared to the baseline scenario with 50 Hz. This chapter is based on an earlier publication: Laubis, K., V. Simko, A. Schuller, and C. Weinhardt (2017). Road Condition Estimation Based on Heterogeneous Extended Floating Car Data. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1582–1591.

4.1 Introduction

The self-calibration approach proposed in the previous chapter relies on measurements from a smartphone-equipped vehicle and is once more outlined in Figure 4.1. In this chapter, the focus is on the location where the single processing steps take place—in particular, the vehicle and a backend system—and which data streams run between them. The overall goal is to leverage the potential of vehicles for road condition estimation by allowing a seamless integration of new participants to the system. The self-calibration addresses the challenge to handle

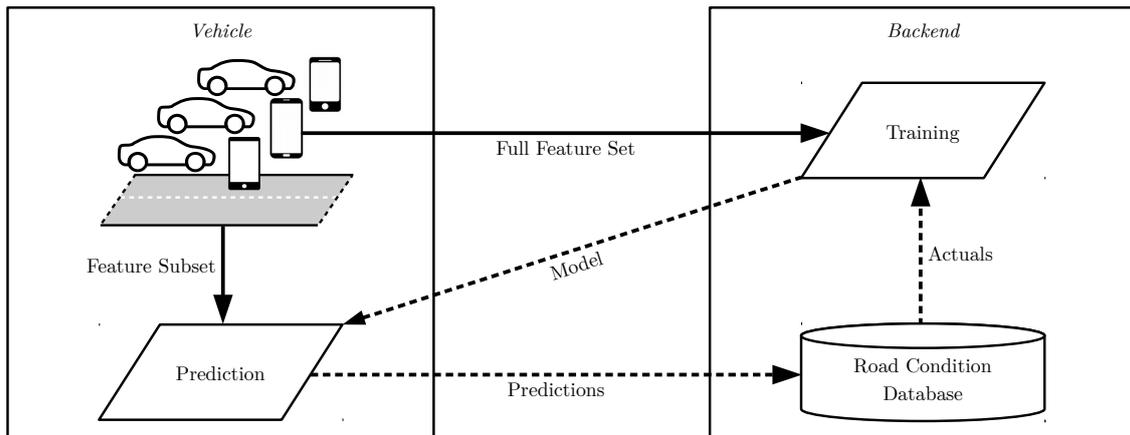


Figure 4.1: Motivation of data reduction approach. For the training phases, the full feature sets gathered by the vehicles have to be stored in the mobile device and sent to the backend system. After the new participant is provided with a unique model, just the single predictions have to be sent. However, depending on the number and the size of features and on the number of road segments, the size of the initial full data set can potentially be very large.

the heterogeneity of the contributing vehicles and smart devices—such as varying maximum sensing frequencies. Since the participants' vehicles and devices can vary strongly, it is not possible to treat all sensor measurements with one single model. This requires fitting a unique model for each participant individually.

As the (re-) training of unique models is computationally expensive and requires information about the road's actual condition, it cannot be performed on the smart device itself, but has to be performed in a backend system. For these training phases, the features gathered by the cars have to be buffered in the mobile device and sent to the backend system as it is depicted in Figure 4.1. Depending on the number and the size of features and on the number of road segments, this amount of data can potentially be very large. Thus, a further challenge of the self-calibration approach is to keep this amount of data small.

This chapter, addresses the aforementioned challenge of handling the limitations in the amount of buffered and transmitted data in accordance with Research Question 2: (a) Which are the most important features for road roughness prediction that can be collected with a smartphone and how sensitive is the prediction performance to the elimination of less important features for data reduction purposes? (b) How sensitive is the prediction performance to variations in the sensing frequency and what is the data reduction potential by sensing at a lower frequency?

Prediction models were built for determining the accuracy with which participants can contribute to a crowdsensing-based road roughness measurement

system. However, it was not considered that there is a limited amount of data that can be transmitted between the cars and the backend system. Furthermore, it was not investigated, how the approach behaves with different sensor frequencies, which is crucial, since the sensors and its sensing frequencies can vary between different devices and even within one device over time.

To answer Research Question 2 a), the most important features according to the permutation importance (PI) criterion are selected. Then new models are build using different feature subsets and their prediction performance is evaluated. Research Question 2 b) is answered by varying the frequencies of the smartphone's sensors. Likewise, to Research Question 2 a), the impact of this variation on the model's performance is determined.

The remainder of the chapter is organized as follows: first, the related literature is discussed with regard to the use of data reduction approaches and general feature selection approaches are differentiated. This is followed by the explanation of the methodological steps for data reduction. The results of the data reduction approaches and their effects on the model performance are presented before the chapter concludes.

4.2 Feature Selection and Sensing Frequencies

The review of the related work in Section 2.2.5 shows that the studies of Forslöf and Jones (2015), Nitsche et al. (2014) and Douangphachanh and Oneyama (2013) apply linear or non-linear machine learning methods.

Forslöf and Jones (2015) and Douangphachanh and Oneyama (2013), however, only consider a single feature—next to the vehicle speed—derived from the acceleration sensors, which they fit to the actual IRI using a linear regression. This approach does not allow the identification of the vehicle information relevant to estimate the IRI. In addition to features from vehicle speed and acceleration, Nitsche et al. (2014) considers additional features, such as the steering wheel and pedal positions. Meanwhile, there is no investigation which of these features are important to estimate the longitudinal road roughness. Thus, there is no possibility for the elimination of less relevant features for data reduction purposes.

With regard to the sensor frequencies, in all three studies a single frequency is considered. Accordingly, no conclusions can be drawn from these studies about how the frequency influences the estimation performance. Therefore, it is not possible to determine up to which degree older smart devices can contribute or whether the frequency can be artificially reduced to achieve a data reduction without a loss of performance.

Thus, in contrast to this chapter, none of the known approaches consider explicit feature selection mechanisms for data reduction reasons nor investigate the model's performance sensitivity to variations in the sensing frequency.

Focusing on the first research question, the following briefly describes different feature selection approaches. Feature selection can be divided into three types: filter methods, wrapper methods and embedded methods (Guyon and Elisseeff, 2003).

Filter methods are performed as a preprocessing step before the actual model building and are mainly based on univariate or multivariate statistics, e. g. methods based on the mutual information criteria, such as the minimum redundancy feature selection algorithm (Auffarth et al., 2010). They are usually fast but do not make use of the machine learning model itself.

Wrapper methods make use of a certain machine learning algorithm by training models for different feature subsets and determining their relevance by comparing the prediction performances of models (Kohavi and John, 1997). Even though, they perform best on the chosen algorithm if they are applied exhaustively, they are computationally expensive since the problem complexity is exponential (Amaldi and Kann, 1998).

Embedded methods are inherently connected to a specific machine learning algorithm since the feature selection is performed within the training phase itself. Since they are making use of the learning algorithm without the need of building multiple models, they demand no or only minor additional computing effort. For this reason they are used in combination with a wrapper-based approach in the further course of this chapter. In particular, the PI for random forests (RFs) is applied as an embedded method (Breiman, 2001).

4.3 Research Design

For the data reduction, two approaches are conducted. First, the feature set is reduced to the important features. The importance of the features is determined by the PI. Second, lower sensing frequencies are employed to reduce the amount of data. In this section, the determination of the PI is formally described and procedures for determining the model's performance sensitivity to the two data reduction approaches are outlined.

4.3.1 Permutation Importance

The PI is a method for feature importance determination that is embedded in the RF training algorithm and likewise in the random forest regression (RFR) (Breiman, 2001). The method for determining the PI is described in this section. Let P be an RFR prediction model that consists of multiple trees $t \in \{1, \dots, ntree\}$.

During the training phase, each tree is evaluated on its own so-called out-of-bag sample set $B^{(t)}$ of size $p \times (r - 1)$ as specified in Equation 4.1

$$B^{(t)} = (f_1, \dots, f_{r-1}) = \begin{bmatrix} f_1(x_1) & \dots & f_{r-1}(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_p) & \dots & f_{r-1}(x_p) \end{bmatrix}, \quad (4.1)$$

where

- p is the number of samples in the out-of-bag sample set for the tree t (for convenience, the notation p is used instead of $p^{(t)}$),
- $r - 1$ is the number of features considered by this tree (for convenience, the notation $r - 1$ is used instead of $r^{(t)} - 1$).

To determine the PI, the values of an individual feature f_i , for which the importance is to be determined, are permuted in all trees, which consider f_i . This permutation for the feature f_i is described for a single tree t in Equation 4.2 by the function PE_i

$$PE_i(B^{(t)}) = \begin{bmatrix} f_1(x_1) & \dots & f_{i-1}(x_1) & f_i(x_{\pi_i(1)}) & f_{i+1}(x_1) & \dots & f_{r-1}(x_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_1(x_p) & \dots & f_{i-1}(x_p) & f_i(x_{\pi_i(p)}) & f_{i+1}(x_p) & \dots & f_{r-1}(x_p) \end{bmatrix}, \quad (4.2)$$

where the function π_i permutes the values $1, \dots, p$.

Additionally considering Equations 3.14 and 3.16, the PI for the feature f_i for a single tree t with is denoted as $PI^{(t)}$ can be determined according to Equation 4.3

$$PI^{(t)}(f_i) = E(P(B^{(t)}), f_i) - E(P(PE_i(B^{(t)})), f_i), \quad (4.3)$$

where

- f_l represents the corresponding actual outcomes of the out-of-bag sample $B^{(t)}$ (for convenience, the notation f_l is used instead of $f_l^{(t)}$),
- $E(P(B^{(t)}), f_l)$ represents the tree's out-of-bag performance without permutation,
- $E(P(PE_i(B^{(t)})), f_l)$ represents the tree's out-of-bag performance with permuted values of the feature f_i .

The overall PI between all trees denoted as PI for feature f_i is determined by the mean $PI^{(t)}$ over all trees $ntree$ as defined in Equation 4.4

$$PI(f_i) = \frac{1}{ntree} \sum_{t=1}^{ntree} PI^{(t)}(f_i). \quad (4.4)$$

The resulting PIs are z-score normalized by dividing them with the standard deviation over all features $1, \dots, k - 1$. Thus, for every feature the mean performance decrease is determined as the difference between the performance with and without permuting the feature's values. For features that are unimportant because they do not have a relation to the outcome or because there is multicollinearity in the feature set, the permutation does not result in a large performance decrease. However, for features that are of importance, the performance decreases when the values are permuted.

4.3.2 Scenario Setup

The procedure for determining the PI described in the previous section is carried out on each of the seven drives, already introduced in the former section. The objective is to determine the importance of the features to remove unimportant ones for data reduction reasons. This scenario is described in more detail in the following section. Thereafter, the sensitivity of the models to variations in the sensor frequency is examined. Both data reduction potentials are investigations using RFRs because they outperformed support vector regressions (SVRs), as identified in the previous chapter. The results are given in the later section.

Importance-Based Feature Selection

To reduce the amount of data that has to be stored on the smartphone and that has to be sent to the backend system for model training purposes, the most important features are determined by computing the PI, as described above. Since the PI determines the feature importance in an embedded manner within the RFR training phase, it allows to determine the feature importance efficiently with nearly no additional effort and to be specially adapted to the machine learning algorithm. Therefore, the PI is deployed in this chapter instead of filter methods, which do not make use of the model specific characteristic.

There exist different approaches for selecting feature subsets based on the features' importance. A naive approach is the Best-Subset Feature Selection, which is a brute force approach by simply creating and testing all possible feature combinations Kohavi and John (1997). This approach becomes very complex with a larger number of features. Common alternative methods are the so-called forward selection and backward elimination. The forward selection method starts

with an empty feature subset and successively adds features to the model. Conversely, the backward elimination works by initially using all features and iteratively removing the least relevant features. These feature selection methods are wrapper methods. Since this chapter focuses on the data reduction potential starting from the full feature set, a variant of the backward elimination is applied—in addition to the embedded PI method. As this chapter uses a combination of embedded permutation importance and backward elimination, it can be considered as a hybrid feature selection method.

The steps carried out to conduct the hybrid feature selection approach are as follows. At the beginning, the importance of each feature is determined using the PI. The features are ranked according to these importance, which allows a first interpretation on which are the more relevant and the less relevant sensors. Thereafter, a backward elimination is applied to remove half of the features—the less important ones. This is followed by calculating new RFR prediction models for all seven drives. The performance of each model is determined. This reduction in the number of features based on the PI and the subsequent determination of the models' performance is carried out repeatedly. This allows to determine the effect of reducing the number of features on the model's performance. Thus, it is determined how much the number of features and correspondingly the data can be reduced.

Down Sampling of Sensing Frequencies

Since it cannot be assured that all smart devices contribute with the same frequencies and since an individual feature extraction and model building is performed to allow contributions from a heterogeneous set of smart devices, the effect of variations in the sensing frequency on the model performance is investigated. Even though, the same Nexus 4 smartphone is used for all test drives, a variation in the sensor's frequency is achieved by subsampling the gathered data. Thus, it is possible to determine whether the findings hold for sensor types and smartphones with lower sensing frequencies as well. The subsampling is done by reducing the number of readings in the sensor data according to different sampling rates. The maximum frequency considered is the actually recorded one with approximately 50 Hz. Frequencies of 25 Hz, 15 Hz and 5 Hz are additionally considered. The upper bound is chosen since 25 Hz approximately relates to the empirically determined Android sensor delay type or sampling rate "User Interface" of the considered Nexus 4. The lowest chosen frequency relates to the default Android sensor delay type "Normal".

As with the investigation of the feature sensitivity, new RFR models and their performance are determined for all three trips. More precisely, for each new subsample, the continuous approximation (Equation 3.8), the alignment with the virtual global positioning system (GPS) fixes (Equation 3.9), the generation of the

feature matrix (Equation 3.11) and the model training (Equation 3.14) and testing (Equations 3.16 and 4.4) are performed again.

4.4 Results

In the previous chapter, it was determined that individual vehicles can contribute to a crowdsensing-based road roughness monitoring system with an average R^2 of 67.83 %—as shown in Table 3.2. For the conducted seven test drives the R^2 ranges from 59.68 % to 76.79 %. These performances are considered as the baseline for the feature selection analysis and for the sensitivity analysis in terms of the sensing frequencies. The results of these two analyses are provided below to answer Research Question 2. A summary and discussion concludes the chapter thereafter.

4.4.1 Sensitivity to Feature Selection

To answer Research Question 2 a), first, the most important features for road roughness prediction—that can be collected with smartphones—are determined by computing the PI for each feature and ranking the features accordingly. Second, the less important features are interactively removed for investigating the prediction performance’s sensitive to this elimination of less important features for data reduction reasons.

Figure 4.2 shows the mean PI over all seven drives for the ten most important features. The labels at the individual PI bars are composed of the aggregation

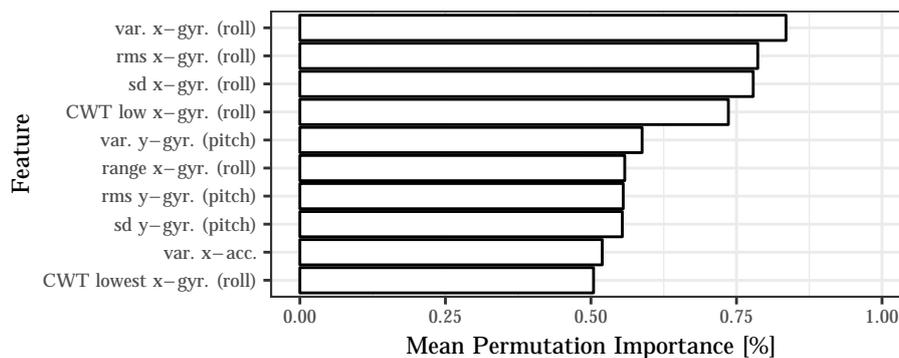


Figure 4.2: Mean PI of ten most important features. It can be observed that the variance features of roll and pitch have a high explanatory value for estimating the IRI. For the roll behavior also the CWT bands with a contextual information of 9.05 m and 51.21 m are of importance.

function (mean, range, std. dev., var., root mean square (RMS) and CWT) and the sensor (speed and axis of inertial measurement unit (IMU) sensors) as defined in

Section 3.3.5. The bias-corrected wavelet power spectrum CWT aggregation function is additionally differentiated into the five frequencies highest, high, medium, low and lowest. The wavelengths described by these frequencies are 0.4 m, 0.8 m, 2.26 m, 9.05 m and 51.21 m. The features extracted from the x-axis gyroscope are prevalent in the ten most important features. Thus, the information whether the vehicle is rolling around the x-axis—as depicted in Figure 2.3—does have a high explanatory value for estimating the IRI. Next to these features, the y-axis gyroscope (pitch) is also important for the estimation, since three of its features are present among the top ten. For both, roll and pitch, the variance seems to be the best aggregation function. The CWT features with the two lowest frequency bands extracted from the x-axis gyroscope appear in the top ten. Thus, for the roll behavior of the vehicle, the bands with a contextual information of 9.05 m and 51.21 m are more important than those with a smaller contextual information. From the accelerometer sensor just the variance of the x-axis is one of the ten most important features.

Figure 4.3 indicates the mean PI for the ten least important features. It is shown that seven out of these ten features are extracted from acceleration sensors. Four out of these are extracted from the absolute and linear y-axes acceleration. Although the pitch and roll sensors are dominant in the ten most important features, the mean aggregation function of both sensor values is unimportant for the prediction. This indicates that the variation in the sensor’s measurements is more important than the absolute value. The mean PI of all 95 features are provided in Figures A.1, A.2 and A.3 in the Appendix.

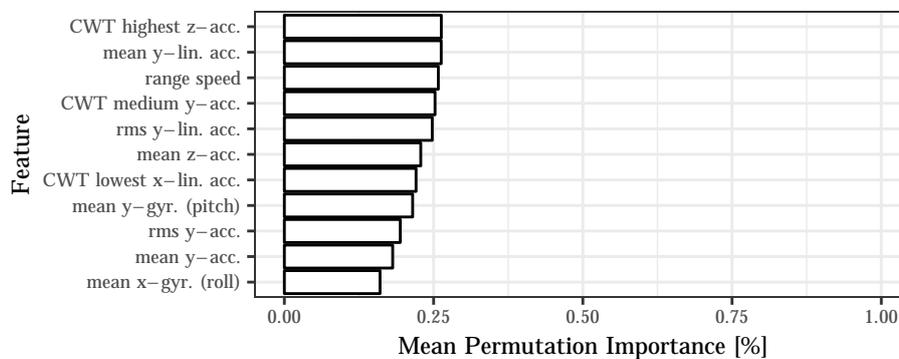


Figure 4.3: Mean PI of ten least important features. Seven out of the bottom ten features are derived from the accelerations. Four out of these features reflect the y-axes accelerations. Although the range, standard deviation, variance, RMS and CWT aggregations of pitch and roll belong to the ten most important features, the mean aggregations of both sensor values are unimportant. This indicates that the variation in the sensor’s measurements is more important than the absolute value.

A sensitivity analysis is performed for evaluating the effect of the PI-based feature elimination on the models’ performance. The baseline is the models’ perfor-

mance, which considered the full set of $k - 1$ features. Compared to this baseline the performances for models with the $\frac{k-1}{2^1}$, $\frac{k-1}{2^2}$, $\frac{k-1}{2^3}$, $\frac{k-1}{2^4}$ and $\frac{k-1}{2^5}$ most important features are determined. Considering initial 95 features, the reductions correspond to 48, 24, 12, 6 and 3 features. The result of the sensitivity analysis regarding the model performance in relation to the number of features is shown in Figure 4.4. It is shown, that reducing the number of features from 95 to 48—which

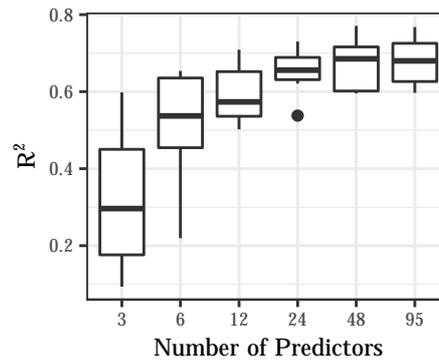


Figure 4.4: Sensitivity of prediction performance to number of features. The boxplot with 95 features serves as the baseline. It is shown, that reducing the number of features from 95 to 48 does not lead to a decrease in the median R^2 . Further reducing the number of features leads to a meaningful performance decrease. Considering 12 features still allows for an out of sample R^2 of more than 50 % for each model trained.

likewise results in a data reduction of nearly 50 %—does not lead to a decrease in the median R^2 over all drives. Further reducing the number of features leads to a performance decrease. However, only considering twelve features still allows for an out-of-sample R^2 of more than 50 % for each model trained. Reducing the number of considered features to 6 leads to a performance reduction to an R^2 of 21 %.

4.4.2 Sensitivity to Sensing Frequency

To answer Research Question 2 b), how sensitive the prediction performance within a crowdsensing-based road roughness monitoring system is to variations in the sensor’s frequency, the results of the sensitivity analysis addressing the variation in the sensing frequencies are depicted in Figure 4.5.

The 50 Hz boxplot serves as the baseline and is equal to the corresponding baseline boxplot for 95 features in Figure 4.4. Reducing the sensor’s frequency to 25 Hz causes a minor increase in the median performance. However, the first and third quartiles are both lower than those from the 50 Hz boxplot. Furthermore, the performances of the single models are more spread and thus, the predictions are not as reliable as those from the baseline models. A further reduction of the sensor’s frequency to 15 Hz leads to a reduction of the median performance from

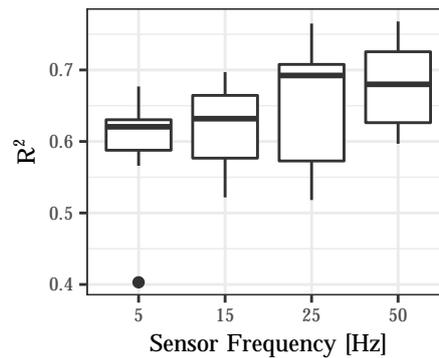


Figure 4.5: Sensitivity of prediction performance to sensor frequency. The 50 Hz box-plot serves as the baseline. However, the performances of the single models are more varied and thus, single predictions are less reliable. A further reduction of the sensor’s frequency to 15 Hz leads to a moderate reduction of the median performance from an R^2 of 0.6923 to 0.6319. Considering a frequency of 5 Hz further reduces the performance to an R^2 of 0.6205.

an R^2 of 0.6923 to 0.6319. Considering a frequency of 5 Hz further reduces the performance to an R^2 of 0.6205. Next to this moderate overall performance decrease it is worth mentioning, that for the 5 Hz frequency there is a single model with a low performance of R^2 0.4031. Although this is only a single model, it may be interpreted as an indication that low frequencies are likely to be subject to outliers, and therefore to less robust estimations.

4.5 Conclusion

This chapter examines the data reduction potential of the crowdsensing-based road condition monitoring system, which was introduced in Chapter 3. For this purpose, reduction potentials were investigated by reducing the number of features considered for the model training and by reducing the frequency at which the sensors gather data. The effects on the model performance are examined. The single prediction models determined in the former chapter have an out-of-sample R^2 of 0.6783 on average. Thus, they explain 67.83 % of the ground truth’s variance. The performance ranges from an R^2 of 0.5968 at minimum to 0.7679 at maximum. These model performances are considered as a baseline.

The PI was determined for all features over all test drives. It is shown that features extracted from the x-axis gyroscope readings are more important and the y-axis accelerometer readings are less important for the prediction models. Reducing the feature set by keeping the 50 % more important features—from 95 features to 48 features—does not lead to a reduction in the median performance of the models. Conversely, since RFRs are robust against unnecessary and corre-

lating features, it was not expected that eliminating unimportant features would increase the model performance. While a reduction of 50 % does not have any significant effect on the performance, further reducing the feature set leads to a drop in the median R^2 .

Analyzing the model's sensitivity to different sensing frequencies shows that a reduction from 50 Hz to 25 Hz does not cause a reduction in the median out-of-sample R^2 . Even though, the R^2 of these measurements has a higher variance, it is shown that devices with a lower measuring frequency can also participate in the crowdsensing-based monitoring approach. The result that a reduction in frequency does not have a negative effect on the estimation of the IRI until the frequency falls below 25 Hz seems to be consistent with the sampling theorem of Nyquist (1928). The sampling theorem states that a signal limited to certain frequency can be exactly reconstructed if it is sampled at a frequency greater than twice of the original frequency. Thus, at a sampling frequency of 25 Hz, frequencies of 12.5 Hz can be perfectly reconstructed. At a speed of 75 km/h, 12.5 Hz represents a wavelength of approximately 1.7 m. Since the IRI, as described in Section 2.1.2, is sensitive to wavelengths greater than 2.4 m, it is reasonable that a sampling frequency of 25 Hz is sufficient at the speed considered.

In both data reduction approaches—feature and frequency reduction—it is shown that a reduction of at least 50 % can be achieved without having losses in the model performance. Considering the results of this chapter, less data needs to be stored on smart devices and likewise less data needs to be sent to the backend system.

A question that remains unanswered is how a combination of both data reduction approaches would affect the performance. For example, it could be investigated whether a combination can achieve a further reduction potential of more than 50 % without a loss in performance. In addition, it is worth addressing the question of whether a crowd-based approach cannot compensate for a lower performance of individual vehicles through making use of multiple measurements. The results of the sensitivity analyses in this chapter describe the model performance of single vehicles. Both in the analysis for the reduction of characteristics as well as in the frequency reduction analysis, an increased spread of the estimation performance and thus a lower robustness of the estimation can be observed. It can be assumed that this robustness loss can be compensated by considering measurements from multiple drives. Therefore, the potential of considering measurements from multiple drives is investigated in the following chapter.

Chapter 5

Combination of Estimations

Applying a crowdsensing-based road surface monitoring approach allows measurements from multiple participants. Each participant contributes given its individual vehicle characteristics, sensor types, driving behavior, etc. Depending on these differences, the estimation accuracy varies between the participants. It seems obvious that a combination of estimations from different participants driven at same road segments could be exploited for resulting in an overall higher estimation accuracy. This chapter focuses on determining to what extent such a combination can increase the estimation accuracy and how a combination should be performed. Therefore, different strategies for aggregating multiple measurements—unweighted and weighted ones—are evaluated. The results confirm that an aggregation of estimations from single drives leads to a higher model performance. This has been expected and confirms the intuition. The overall R^2 could be increased from 0.68 to 0.75 on average and the normalized root mean square error (NRMSE) could be decreased from 9% to 8% on average using the unweighted mean for aggregation. However, contrary to the intuition, the results show that a weighted aggregation should be avoided, which is consistent with findings in other domains, such as in financial forecasting. Parts of this chapter are adapted from a former publication: Laubis, K., V. Simko, and C. Weinhardt (2016). Weighted Aggregation in the Domain of Crowd-Based Road Condition Monitoring. In *INFORMATIK 2016*, pp. 385–393. Bonn: Gesellschaft für Informatik (GI).

5.1 Introduction

As described in the previous two Chapters 3 and 4, smart devices from drivers and passengers can be used to measure and analyze vehicle vibrations and thus estimate the road surface's roughness. However, the low accuracy of sensors in consumer devices, versatile suspension systems, different placements of smart devices in the car and other factors lead to a lower prediction accuracy compared to well-calibrated laser-based measurements. See also Section 2.1.3, in which the different information quality levels (IQLs) are described. To overcome this decline in accuracy, this chapter compares different approaches for aggregating

measurements from multiple vehicles. The outline of this idea is depicted in Figure 5.1.

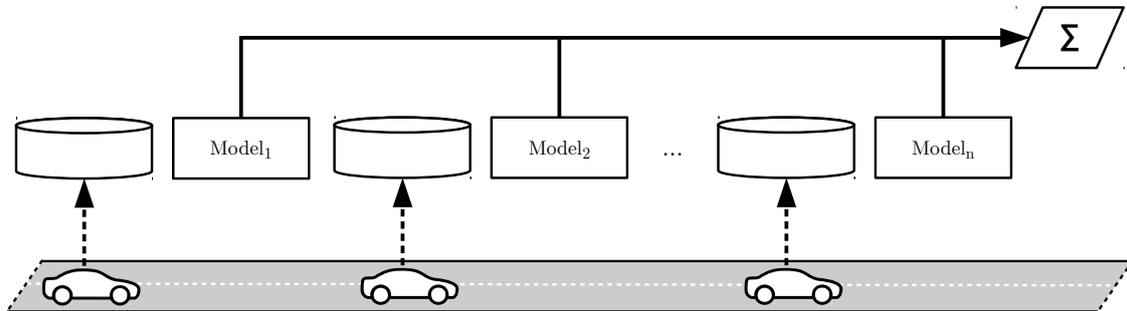


Figure 5.1: Outline of combining estimations from multiple drives. Given crowdsensing-based inspections of the road roughness from multiple drives, a combination of the single estimations is expected to be beneficial. Next to getting more robust overall estimations a systematic outperforming of the best single estimation could be possible. The potential of combining measurements strongly depend on the error structure of the single estimations and on the aggregation method itself.

Assuming uncorrelated prediction errors of the single vehicle's predictions, the unweighted mean is expected to reduce the variance component of the errors and thus, increase the prediction accuracy. However, it is not clear to what extent the accuracy can be increased. Although it may seem plausible to use weighting schemes based on the model performance instead of a simple arithmetic mean when aggregating the results, it has been shown that the simple arithmetic mean often exceeds a more sophisticated weighting (Smith and Wallis, 2009). This is true, since a weighted aggregation could increase the prediction error because of an increase in the variance component of the error. On the other hand, weighting could reduce the bias component of the prediction error. Thus, it has to be investigated empirically, how weighted aggregation functions perform against the unweighted mean.

To determine, to what extent aggregation of single car predictions can increase the model's accuracy and whether weighting of the single car's prediction is worth an implementation, the focus of this chapter is to develop and evaluate the extent to increase the performance of crowd-based road roughness estimations by aggregating estimations from multiple cars. For this purpose, unweighted and weighted aggregation methods are applied to the estimates derived in Chapter 3 to answer the Research Question 3: (a) To what extent does the aggregation of crowd-based road roughness measurements from multiple vehicles increase the model performance when using the simple average (SA) and means weighted by the performance of the single models? (b) Can the model performance be increased by applying optimal weights (OW) and how does a shrinkage of the OW to an unweighted aggregation affect the model performance?

The remainder of this chapter is structured as follows: the next section summarizes the related work with regard to combination approaches in crowd-based road condition monitoring. This is followed by a description of the research design, presenting different weighting concepts, including OW, and a regularization approach. In order to answer the Research Question 3, the results of this chapter are then presented. Finally, a conclusion is provided.

5.2 Combination in Road Surface Monitoring

Within the related works discussed in Section 2.2.5, only one of the papers follows an approach to integrate multiple measurements. This approach by Jang et al. (2017) relies on a clustering algorithm in which the trajectories of individual vehicle drives are split into segments, followed by a consideration of the position and direction of the segment as criteria for the clustering algorithm. Although the approach takes the distances between the single drives into account while aggregating them, the effects of the aggregation itself is not examined, nor is the aggregation mechanism—which is a voting-based mechanism—varied in order to investigate its effectiveness in comparison to other aggregation mechanism. Moreover, the approach is intended for classifications rather than regressions, therefore the approach cannot be applied to the estimation of a continuous international roughness index (IRI) metric—as followed in the thesis at hand.

Another approach to combine estimations from multiple participants is described by Eriksson et al. (2008). They outline a machine learning approach to detect single road anomalies, such as potholes or speed bumps, with a fleet of smartphone-equipped taxis. For getting robust results, the pothole candidates provided by single taxis are geospatially clustered. However, the performance increase by applying this aggregation was not investigated. In addition, it should be noted that this work is intended for binary classification, such as the distinction between the existence and non-existence of a road anomaly. Therefore, the approach is even less applicable than that of Jang et al. (2017) to the estimation of a continuous IRI metric.

Both studies demonstrate that a geospatially alignment between the measurements from single drives is required to combine them. For the combination in the course of this chapter, this alignment is carried out by means of the map-matching described in Section 3.3.1. Here, road segments are defined globally to which the measurements are assigned. Thus, measurements assigned to a segment—and also measurements from neighboring segments—can be combined to a single estimate. While such a geospatially alignment can be found in both studies, none of them investigates the effect of considering measurements from multiple vehicles. That means that the effect on the overall model performance compared to single measurements is not examined. Since the effect of multiple measurements is not determined, the performance of different aggrega-

tion functions is also not determined in the domain of crowdsensing-based road surface monitoring. However, a thorough investigation of the effects of aggregation functions on the estimation performance is essential. On the one hand, this allows to determine whether an aggregation creates a significant performance increase at all. On the other hand, the combination should take into account the characteristics of the combination methods and the underlying data on which the combination is to be applied, in order to increase the performance.

5.3 Research Design

There are several ways to combine multiple drives. In addition to a general formal description of an unweighted and weighted combination, the extreme methods SA and OW are described in more detail in this section. Furthermore, a regularization of the OW is described. Finally, the scenario is outlined how these different combination methods are evaluated in the domain of crowdsensing-based monitoring of the road roughness.

5.3.1 Weighting

Let x be a road slot for which predictions from d vehicles were obtained using their prediction models P_l with $l \in \{1, \dots, d\}$. As defined by Equation 3.14, each model P_l is a function that maps a feature vector of $k - 1$ dimension $\vec{x} \in \mathbb{R}^{k-1}$ to a real outcome, which means $P_l : \mathbb{R}^{k-1} \mapsto \mathbb{R}$. For brevity, $P_l(x) = P_l(f_1(x), \dots, f_{k-1}(x))$ is denoted as the prediction at slot x using the model P_l which uses features $f_1(x), \dots, f_{k-1}(x)$ —these predictions are also often denoted as \hat{y} . It should be noted that each prediction model might use a different feature set. Given the predictions $P_l(x)$ from multiple drives, a combination of these single predictions is defined by Equation 5.1

$$IRI(x) = comb(P_1(x), \dots, P_d(x)), \quad (5.1)$$

where $comb$ is a function $\mathbb{R}^d \mapsto \mathbb{R}$, such as the SA.

The combination of multiple predictions can be performed in different ways. The aggregation functions basically differ in the weights assigned to each prediction $P_l(x)$. The predictions $P_l(x) = \hat{y}_l, \forall l \in \{1, \dots, d\}$ can be combined using a simple arithmetic mean—which is the SA—as described by Equation 5.2

$$\bar{P}(x) = \frac{\sum_{l=1}^d \hat{y}_l}{d}. \quad (5.2)$$

A weighted combination can be defined accordingly by the Equation 5.3

$$\bar{P}_W(x) = \sum_{l=1}^d \hat{y}_l w_l, \quad (5.3)$$

where $W = \{w_1, \dots, w_d\}$ are the weights corresponding to the prediction models P_1, \dots, P_d . Equation 5.2 is a specialization of Equation 5.3, where all weights are equal and sum up to 1, so that Equation 5.4 applies

$$w_l^S = \frac{1}{d}, \forall l \in \{1, \dots, d\}. \quad (5.4)$$

In addition to the SA weighting, the weights can be chosen relatively to the performance of the individual models. Accordingly, an obvious approach is to choose the weighting analogous to performance metrics, such as the R^2 or the root mean square error (RMSE), what can be denoted as \bar{P}_{R^2} and \bar{P}_{RMSE} . In the case of a performance metric, which shows a higher value the higher the performance of the model is—such as the R^2 —, Equation 5.5

$$w_l^G = \frac{G_l}{\sum_{m=1}^d G_m}, \forall l \in \{1, \dots, d\} \quad (5.5)$$

can be used to determine the corresponding weights, where G_l indicates the performance of the l -th model, such as the R^2 . Error metrics, on the other hand, can be considered for the determination of weights using Equation 5.6

$$w_l^H = \frac{H_l^{-1}}{\sum_{m=1}^d H_m^{-1}}, \forall l \in \{1, \dots, d\}. \quad (5.6)$$

In this case, H_l represents an error metric of the l -th model, such as the RMSE.

An extreme alternative for considering the performance of the single models is that only the m -th model with the highest performance is considered, what can be denoted as \bar{P}_{Best} . According to Equation 5.7, all weights are 0 except that of the best model, which is weighted with 1

$$w_l^B = \begin{cases} 1, & \text{if } l = m \\ 0, & \text{otherwise} \end{cases}, \forall l \in \{1, \dots, d\}. \quad (5.7)$$

In addition to these apparent options for determining the weights, the training data set can be used to calculate the weights which minimize the in-sample error. The determination of these OW and a regularization approach for addressing overfitting are described in the following two sections.

5.3.2 Optimal Weights

The OW approach of Bates and Granger (1969) aims at minimizing the error variance of the combined estimate. This minimization addresses the errors within the training set. An aggregation function that weights the individual estimates according to the OW is referred to as \bar{P}_{OW} . In order to estimate the weights \hat{w}^O that minimize the in-sample combined error variance, the error covariances between the individual models are employed. There is no other weighting than the OW, which further reduces the in-sample error variance. Therefore, they are termed optimal.

The OW \hat{w}^O can be determined by applying a least square regression of the actuals on the predictions where the sum of the regression coefficients is restricted to 1 and the intercept term is suppressed. According to Granger and Ramanathan (1984) this is equivalent to regressing the error of the n -th model E_n on the differences between the other errors, $\forall l \in \{1, \dots, n-1, n+1, \dots, d\} : E_l$ and E_n as defined by Equation 5.8

$$E_n = \hat{w}_d^O(E_n - E_1) + \dots + \hat{w}_d^O(E_n - E_{n-1}) + \hat{w}_d^O(E_n - E_{n-2}) + \dots + \hat{w}_d^O(E_n - E_d) + \epsilon, \quad (5.8)$$

where the intercept is suppressed.

The weight \hat{w}_n^O for the n -th model itself can be determined by Equation 5.9 using the restriction that all weights must sum up to 1

$$\hat{w}_n^O = 1 - \sum_{l \in \{1, \dots, n-1, n+1, \dots, d\}} \hat{w}_l^O. \quad (5.9)$$

As this approach for determining the OW shows, the weights are itself estimated based on the training data set. This can result in an additional estimation error caused by an overfitting to the training data set. In such a case, applying the OW to a new test data set—out-of-sample—would result in a lower performance than on the training data set—in-sample. In many areas, the OW have even a lower out-of-sample performance compared to a non-estimation-based approach, such as the SA (Stock and Watson, 2004). To address the risk of overfitting when using OW, a regularization can be applied, as described in the following section.

5.3.3 Shrinkage

To reduce the problem of overfitting when using OW, a regularization in the form of a shrinkage can be performed. More precisely, the OW can be shrunk linearly by the factor $\lambda \in [0, 1]$ towards the SA weights, what can be referred to as \bar{P}_λ . Thus, a trade-off can be found between the overfitting prone OW and the SA,

which does not add an additional estimation error since it is composed instead of estimated. The resulting weights can be described by Equation 5.10, depending on λ

$$\hat{w}_l^\lambda = (\lambda - 1)\hat{w}_l^O + \lambda w_l^S, \forall l \in \{1, \dots, d\}. \quad (5.10)$$

The optimal value for λ depends on the size of the training set, on changes in error correlation and on error variances (Blanc and Setzer, 2016). Empirically, a suitable λ can be determined following an iterative approach. An out-of-sample error metric of the combined estimates can be calculated for a set of λ values. Given an appropriate selection and a sufficiently large number of values for λ , that value resulting in the lowest error metric can be regarded as close to optimal.

5.3.4 Scenario Setup

To determine the effect of combining estimates from multiple drives and to compare the performance of the different aggregation functions empirically, they are applied to the models' estimates resulting from the drives described in Section 3.3.5. The following section first describes the setup for composing the seven drives without an additional estimation of weights, then the setup for combining the drives with OW and shrinkage is given.

Composition of Multiple Drives

In order to compose multiple drives, the combination functions \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE} are applied according to Equations 5.2 and 5.3. The weights are determined according to Equations 5.4, 5.5 and 5.6. None of these methods require a training to determine the weights, so there is no risk of an additional overfitting caused by the combination. First, the weights for the combination of the whole set of all seven drives are presented for interpretation reasons. Secondly, on the basis of a total of seven $n = 7$ single drives, several subsets are considered for further possible drive combinations. In particular, the combination of $k \in \{2, \dots, 7\}$ drives are possible. According to the binomial coefficient, this leads to sample sizes of $\binom{n}{k} = \{21, 35, 35, 21, 7, 1\}$. For all these possible combinations, the functions \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE} are applied again and the resulting performance metrics R^2 and NRMSE are calculated by cross-validation.

Shrinking Optimal Weights to Simple Average

In addition to the composition of multiple drives by using the SA and by using performance metrics for a weighted combination, the OW are estimated and used for the combination. Thus, the function \bar{P}_{OW} is applied according to the Equations 5.3, 5.8 and 5.9. A preceding descriptive analysis of the estimation errors within the training data set is used to assess whether the application of the

OW-based combination appears promising or not. For this purpose, the correlations of the estimation errors between the single drives are determined. Aware of the fact that OW are prone to overfitting, the performance of the combined estimates in terms of the NRMSE is determined both in-sample and out-of-sample. In order to address a potential lower out-of-sample performance when employing OW, a regularized combination function \bar{P}_λ is applied whereby the OW are gradually shrunk towards the SA according to Equation 5.10.

Finally, a comparison of the out-of-sample performances of all considered combination methods is provided. In particular, these are the methods \bar{P} , \bar{P}_{R^2} , \bar{P}_{RMSE} , \bar{P}_{Best} , \bar{P}_{OW} and \bar{P}_λ , which were introduced in the previous sections. This overall result and the intermediate findings themselves—which are presented in the next section—allow a sound understanding of the potentials of combining multiple drives in a crowdsensing-based road monitoring system.

5.4 Results

To answer Research Question 3, the results of the evaluation of the combination methods are provided in this section. The results are presented differentiated according to the sub-Research Questions 3 a) and 3 b).

5.4.1 Weighted Combination

The resulting weights of the combination functions when considering all seven drives \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE} are provided in Figure 5.2. The performances of the individual drives do not differ much—see Table 3.2—, so the weights for \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE} are also similar. The exact values of the weights can be found in Table B.1 in the Appendix. Thus, it can be assumed that the performances of the combined estimates also tend to just differ slightly.

In order to answer the Research Question 3 a) and investigate how both the combination functions and the number of combined drives affect performance, Figure 5.3 shows boxplots describing the distribution of the combined R^2 for each of the combination functions considered and for the different numbers of combined drives. The median R^2 of single drive predictions—see Table 3.2—is indicated by a horizontal line. Even for combining two drives, all combination functions achieve a significantly better median performance compared to its single drive baseline at a significance level of at least 5%. Increasing the number of considered drives further increases the R^2 constantly for each aggregation function. Comparing the performance of the combination functions shows that there is just a minor difference between them. However, while considering four and more drives, the unweighted mean combination \bar{P} outperforms the weighted ones \bar{P}_{R^2} and \bar{P}_{RMSE} .

A comparison regarding the NRMSE is given in Figure 5.4. Similar to the R^2

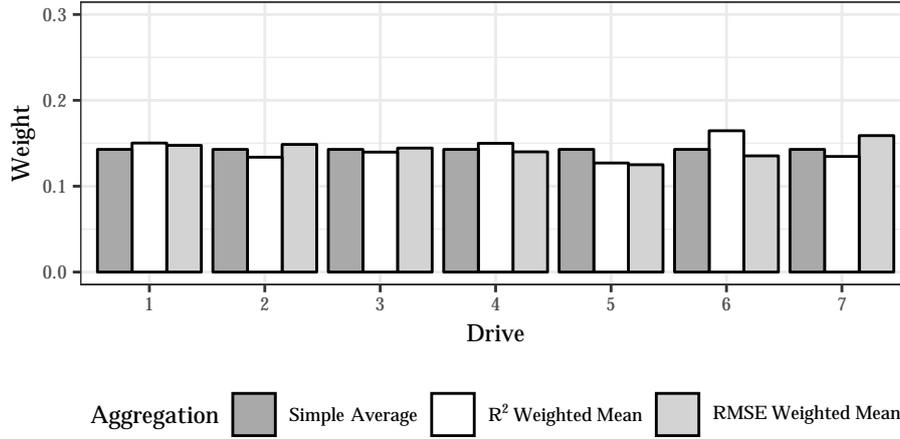


Figure 5.2: Weights of R^2 - and RMSE-based combination in comparison to SA weights. Both, the R^2 and the RMSE of the single models do not differ much. Accordingly, the weights derived from these performance metrics are very similar for all models. Due to the constraint of summing all weights up to 1, the R^2 and the RMSE based weights are also very close to the SA weights.

scenario, a constant decrease of the NRMSE is achieved by considering more drives. Furthermore, the unweighted mean aggregation has a lower median NRMSE than the weighted aggregations except for the case of two drives.

The exact values of the mean R^2 and mean NRMSE over all possible combinations for the considered number of drives are given in Table 5.1. If there

Table 5.1: Out-of-sample mean R^2 and NRMSE of combined prediction functions \bar{P} , \bar{P}_{R^2} and \bar{P}_{RMSE} . Significant performance decreases compared to \bar{P} are indicated.

Drives	\bar{P}		\bar{P}_{R^2}		\bar{P}_{RMSE}	
	R^2	NRMSE	R^2	NRMSE	R^2	NRMSE
1	0.6783	0.0893	0.6783	0.0893	0.6783	0.0893
2	0.7215	0.0832	0.7214	0.0832	0.7196 ***	0.0835 ***
3	0.7359	0.0811	0.7353	0.0812	0.7333 ***	0.0815 ***
4	0.7431	0.0801	0.7422 *	0.0802 *	0.7401 ***	0.0805 ***
5	0.7474	0.0794	0.7463 **	0.0796 **	0.7441 ***	0.0799 ***
6	0.7503	0.0790	0.7491 *	0.0792 *	0.7468 ***	0.0795 ***
7	0.7523	0.0787	0.7510	0.0789	0.7488	0.0793

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

is a significant performance decrease (decrease in mean R^2 or increase in mean NRMSE) of using a weighted mean combination instead of using the unweighted mean aggregation, it is indicated at the corresponding mean performance of the

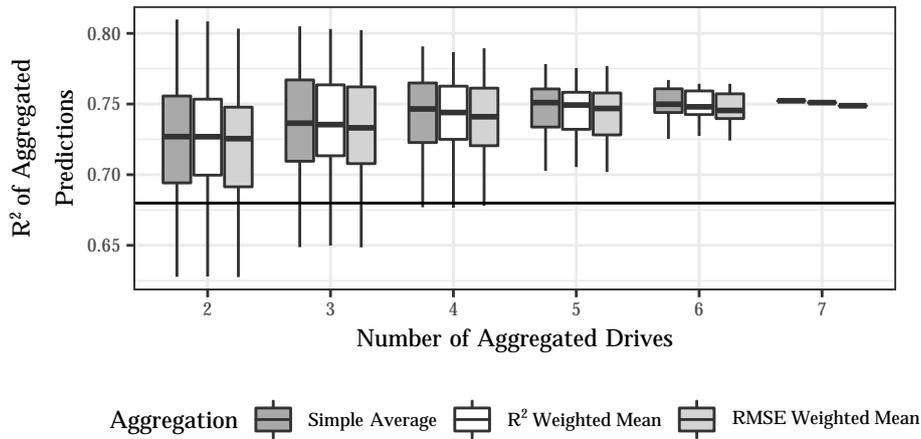


Figure 5.3: Out-of-sample R^2 of aggregated predictions for SA, mean weighted by R^2 and mean weighted by RMSE. Each boxplot represents the performances of all possible drive combinations given the corresponding total number of considered drives. The median R^2 of all single drive predictions is provided as a horizontal line. For each combination size the median R^2 of the SA method is higher than the R^2 of both weighted average methods.

weighted aggregation. Since the first row indicates the baseline with no aggregation, there is no difference between the aggregation functions. Likewise, referring to the number of possible combinations, the seventh row does not contain tests on significant differences in performance. Except for combinations of drives less or equal to 3, the mean performance of the R^2 weighted aggregations are worse than the unweighted aggregations. Regarding the mean performance of the NRMSE the unweighted aggregation outperforms the weighted ones for all considered numbers of drives. Even though, the absolute differences are minor, the decreases compared to the unweighted mean function are mostly statistically significant even for the small sample sizes. This indicates that applying a weighted aggregation increases the variance error component to a higher extent than decreasing the bias error component. A vertical comparison of the performance metrics provided in Table 5.1 was discussed based on Figure 5.3 and Figure 5.4.

5.4.2 Optimal Weights and Regularization

To allow a first assumption in answering Research Question 3 b), Figure 5.5 shows the absolute estimation errors for the first two drives. The first two drives are chosen arbitrarily and for exemplary reasons. The errors of all seven drives are provided in Figures B.1 and B.2 in the Appendix. All of them are the in-sample errors—meaning those in the training data set with the segment numbers $\{1, \dots, 60\} \cup \{80, \dots, 170\} \cup \{190, \dots, 220\}$. The absolute errors shown indicate

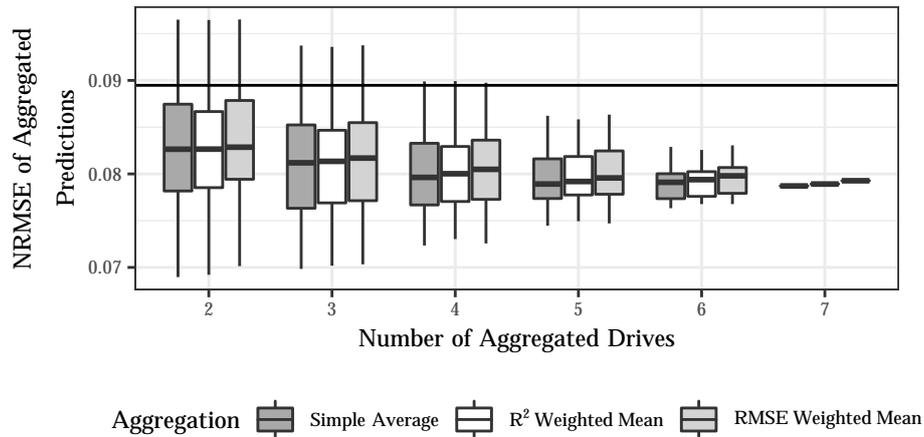


Figure 5.4: Out-of-sample NRMSE of aggregated predictions for SA, mean weighted by R^2 and mean weighted by RMSE. Each boxplot represents the performances of all possible drive combinations given the corresponding total number of considered drives. The median NRMSE of all single drive predictions is provided as a horizontal line. Vice versa to the R^2 , for each combination size the median NRMSE of the SA method is lower than the NRMSE of both weighted average methods.

that similar error structures exist for the different drives. This is confirmed by the error covariances between drives, as shown in the correlation matrix in Figure 5.6. These in-sample errors are exploited in the sense of Equations 5.8 and 5.9 to estimate the OW. The in-sample prediction errors of all drives, however, have a high positive correlation. The lowest correlation is found between the drives 1 and 6 and still has a correlation coefficient of 0.75. Thus, an increase in performance by combining estimates from multiple drives will probably only be possible to a limited extent.

The OW for the combination of all seven drives are shown beside the SA weights in Figure 5.7. The exact values of the weights are provided in Table B.1 in the Appendix. The estimates of four of the seven drives are to be weighted positively and three negatively to obtain an optimal combined estimate within the training set. The estimate of the sixth drive should be given the highest positive weighting. This is also worth mentioning, as the sixth drive is just the third best performing when considering the single drive estimates.

When applying the OW for combination \bar{P}_{OW} within the training data set, an NRMSE of 0.0671 is obtained. This is a lower error compared to a combination \bar{P} using SA weighting, which results in an NRMSE of 0.0761. The fact that in-sample \bar{P}_{OW} performs at least as well than \bar{P} is according to the definition, as described in Section 5.3.2, and was expected.

Since \bar{P} is robust against overfitting, it also meets the expectation that the out-of-sample NRMSE of 0.0787 differs only marginally from the in-sample NRMSE.

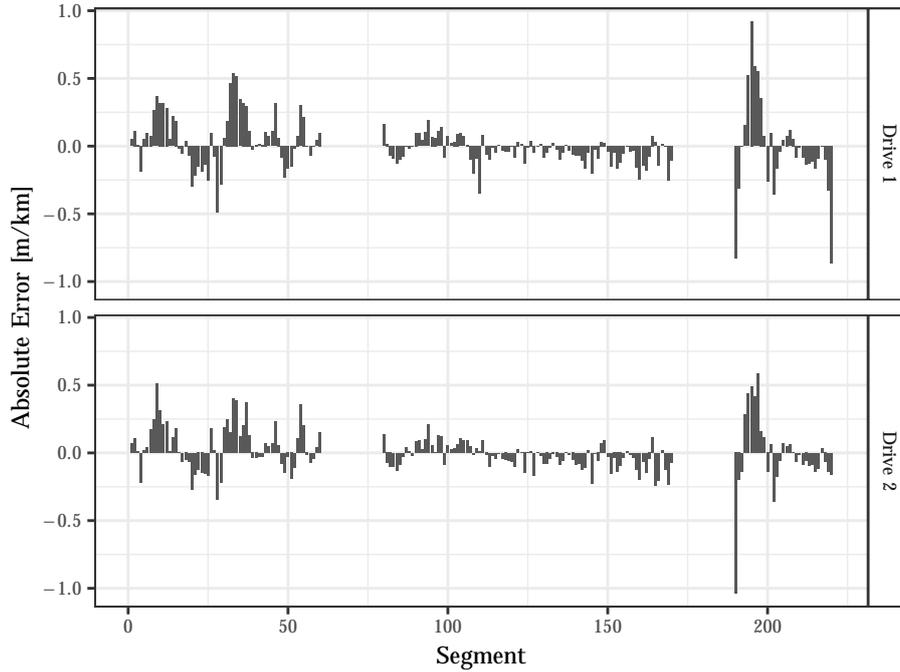


Figure 5.5: Absolute errors of in-sample predictions for drive 1 and 2. Drive 1 and 2 are chosen for illustrating the similarity of their absolute prediction errors for the training segments ($\{1, \dots, 60\} \cup \{80, \dots, 170\} \cup \{190, \dots, 220\}$). The similarity of the errors in turn suggests that they are highly correlated. Plots of all drives are provided in Figures B.1 and B.2 in the Appendix.

However, for the performance of \bar{P}_{OW} , the out-of-sample NRMSE 0.0963 is much higher than the in-sample NRMSE 0.0671. Accordingly, using OW to combine multiple drives cannot be recommended.

To answer the Research Question 3 b) entirely, Figure 5.8 shows the results of the regularized combination \bar{P}_λ , in which shrinkage of the OW to the unweighted combination is carried out. While Figure 5.8a represents the in-sample performance over λ , Figure 5.8b represents the corresponding out-of-sample performance. It is apparent that in the out-of-sample case there is no λ for which the regularized combination \bar{P}_λ has a lower NRMSE than the SA-based combination \bar{P} . Thus, the regularized weighting-based \bar{P}_λ corresponds to the SA-based weighting \bar{P} . Therefore, neither the weighting based on OW nor the regularized weighting—with a $\lambda < 1$ —can be recommended for combining multiple drives.

Table 5.2 summarizes the results by listing the out-of-sample NRMSE for all combination functions considered. It is shown that an SA-based combination is not outperformed by any other combination function in the given scenario.

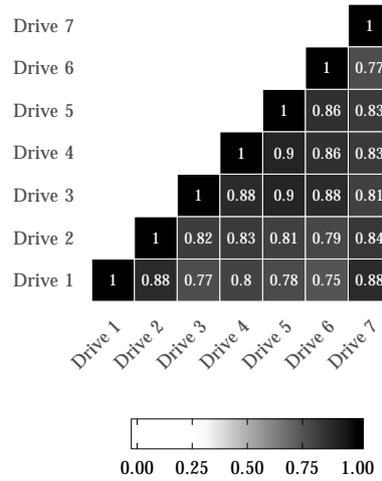


Figure 5.6: Correlation matrix of in-sample prediction errors. The in-sample prediction errors of all drives have a high positive correlation. Therefore, a performance increase through combination of estimations from multiple drives is expected to be limited. This is especially true for combinations that aim to minimize the error within the training set, such as OW.

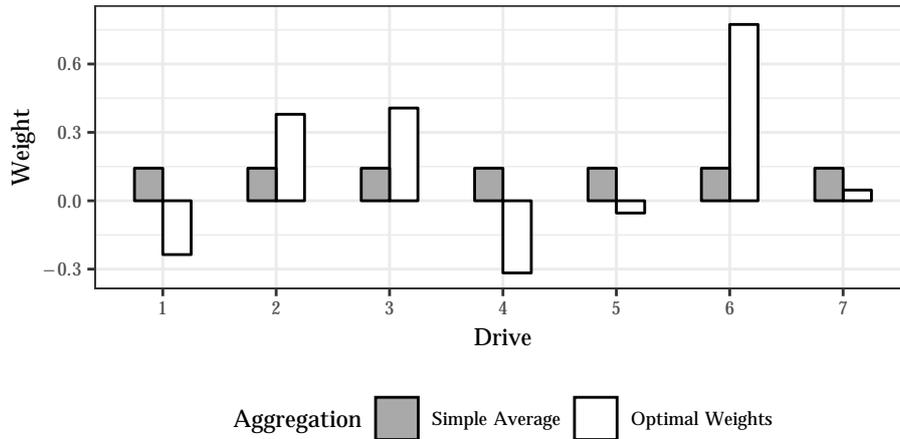


Figure 5.7: OW in comparison to SA weights. The OW of drive 1, 4 and 5 are negative. Combining multiple estimations given these OW allows for minimizing the combined in-sample error. There is no linear weighting method that further reduces the in-sample error.

5.5 Conclusion

This chapter examines how a combination of single crowdsensing-based road condition estimates affects the estimation performance. For this, different combi-

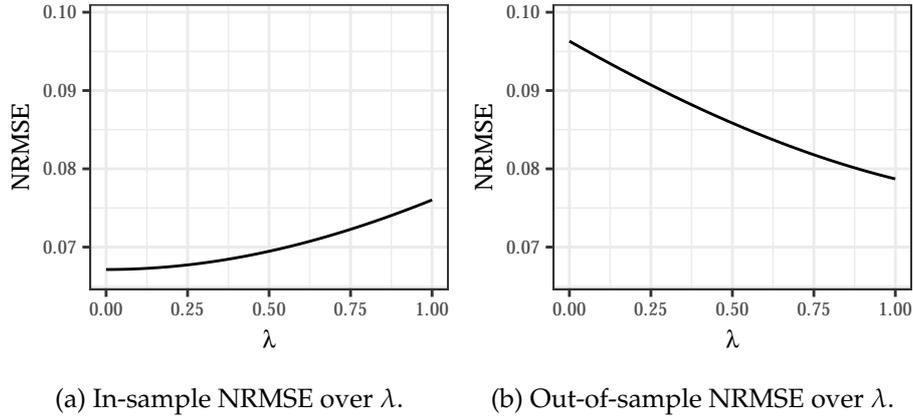


Figure 5.8: Performance of OW combinations linearly shrunk to SA using a shrinkage factor $\lambda \in [0, 1]$. Choosing a $\lambda = 0$ means applying OW and $\lambda = 1$ means that the OW were fully shrunk to the SA. In-sample and out-of-sample NRMSE over λ are provided in Figure 5.8a and Figure 5.8b respectively. Increasing λ leads to a monotonic increase of the in-sample NRMSE. However, the out-of-sample NRMSE monotonically decreases while increasing λ . Thus, shrinking OW totally to the SA outperforms every other degree of shrinkage.

Table 5.2: Out-of-sample NRMSE of prediction functions \bar{P} , \bar{P}_{R^2} , \bar{P}_{RMSE} , \bar{P}_{Best} , \bar{P}_{OW} and \bar{P}_λ . The SA-based combination \bar{P} outperforms every other combination function. Accordingly, when the OW are regularized in terms of \bar{P}_λ , they are completely shrunk to the SA.

Combination function	NRMSE
\bar{P}	0.0787
\bar{P}_{R^2}	0.0789
\bar{P}_{RMSE}	0.0793
\bar{P}_{Best}	0.0849
\bar{P}_{OW}	0.0963
\bar{P}_λ	0.0787

nation functions (\bar{P} , \bar{P}_{R^2} , \bar{P}_{RMSE} , \bar{P}_{Best} , \bar{P}_{OW} and \bar{P}_λ) are applied to the estimates from the seven drives described in Chapter 3.

It is shown that given road roughness estimations from multiple drives, an increase in the estimation performance can be achieved by applying the SA-based combination \bar{P} . This has been expected and confirms the intuition. Thus, the R^2 can be increased from 0.68 to 0.75 on average and the NRMSE can be decreased from 9% to 8% on average. In other words, real-time predictive road maintenance gets better with an increasing number of participants.

Contrary to the intuition, the results also show that weighting aggregations of single predictions should be avoided. This is consistent with the results of the

study of Smith and Wallis (2009), which describes similar findings in the financial forecasting domain. From a technical perspective, this allows a simpler and also more efficient implementation.

It has to be mentioned that all drives were with the same car, so the prediction errors are strongly correlated. A weighted combination could possibly outperform SA-based weighting if predictions from different vehicles are considered. The analyses in this chapter are based on the estimates described in Chapter 3. Accordingly, the estimates were derived by considering the full feature set. Therefore, it might also be possible to investigate how the combination of multiple measurements affects the combined estimation performance when the single estimates are derived from a reduced feature set, as described in Chapter 4.

The procedures described in this and the previous chapters, for road condition estimation, data reduction, estimation combination, etc. require an IT infrastructure on which these processes are performed. A corresponding infrastructure should be able to address fluctuations in the demand and at the same time be resource efficient. These challenges are addressed in the following chapter.

Chapter 6

IT Infrastructure

Collecting, storing, processing and distributing data gathered by the crowd requires a backend IT infrastructure. Since road traffic fluctuates following daily, weekly and yearly patterns, the load generated by the crowd likewise fluctuates. Next to this seasonality in load, a trend component can be expected when rolling out a crowdsensing-based system. This is true if new participants can enter or leave the system over time. Thus, the backend system is required to be scalable to dynamically adapt its resources to the current load. This chapter focuses on determining the potential of different scaling technologies. A simulation-based investigation of the resource efficiency of different scaling technologies is conducted. In addition, the resource reduction potential of diagonal scaling is investigated in comparison to conventional horizontal approaches. Given an empirical load pattern with daily fluctuations, a central processing unit (CPU) allocation reduction potential between 4.16 % and 8.05 % compared to a horizontally scalable service can be achieved. This chapter also builds on and extends a former publication—parts of it are described for comprehensibility and self-containing reasons: Laubis, K., V. Simko, and A. Schuller (2016a). Cloud Adoption by Fine-Grained Resource Adaptation: Price Determination of Diagonally Scalable IaaS. In *Advances in Service-Oriented and Cloud Computing*, pp. 249–257. Cham: Springer International Publishing.

6.1 Introduction

The approach of crowdsensing-based road condition monitoring presented in the former chapters requires an IT infrastructure serving as the backend of the system. Concerning the question of whether the computing resources required for the backend are to be deployed on-premise or on-demand—by means of infrastructure as a service (IaaS) cloud computing—, there are a number of reasons for an on-demand infrastructure.

In addition to the fact that the deployment of cloud resources does not impose an entry barrier through initial costs, there is no need to decide the extent of the deployment prior to operation. This is an advantage especially for small and medium-sized companies, which are unable to fund a deployment across

the company. In addition, the infrastructure costs are directly related to the outcome of the crowdsensing-based system. The more participants participate, the more computing resources are required, but also a higher return on investment can be created by more frequent, accurate and robust results. Therefore, through the use of scalable cloud resources, the risk of investment is kept low, which is particularly advantageous with regard to innovative solutions, such as the crowdsensing-based road condition monitoring presented in the thesis at hand.

Since flexible resource adaptation to fluctuating computing demand is one of the main benefits of cloud usage and likewise a main reason for IaaS cloud adoption (Andrade et al., 2015; Chebrolu, 2011), cloud computing is an omnipresent concept that is still far from its envisioned potential and can be implemented in various technological configurations (Leavitt, 2009; Raza et al., 2015). A decision regarding the technological configuration to be taken is the choice of the scaling technology used. Resource scaling in cloud computing can be performed horizontally, vertically or diagonally. The former, currently the most common approach (Chieu et al., 2009; Urgaonkar et al., 2005), adjusts the resource capacity by adding or removing whole virtual machines (VMs) to or from the deployment. Vertical scaling, instead, adjusts the capacity within a VM (Dutta et al., 2012). Diagonal scaling is a combination of both (Han et al., 2012).

The main contribution of this chapter is thus the evaluation of these scaling technologies regarding their resource efficiency. Based on a real load pattern, which represents requests to a backend system, the resource consumption of horizontal and diagonal scaling is determined by simulations. This allows to answer Research Question 4: (a) How efficiently can the challenge of having a fluctuating demand in IT resources be served by horizontally and diagonally scalable IaaS? (b) What resource reduction can be achieved by using diagonal scaling compared to conventional horizontal scaling?

The remainder of this chapter is as follows: after the introduction a review of scaling technologies is provided. This is followed by the description of a generic scaling model. The scaling model is parametrized—once to describe horizontal scaling and once for diagonal scaling. Given a real load pattern, simulations are performed to determine the resource allocation for both scaling scenarios. After providing the simulation results to answer the Research Question 4, the chapter concludes.

6.2 Scaling Technologies

Resource scaling in cloud computing can be performed horizontally, vertically or diagonally. When scaling resources horizontally, entire VMs are added or removed to or from the deployment. The incoming load is distributed to these VMs (Chieu et al., 2009; Urgaonkar et al., 2005). Adding and removing VMs is also referred to as scale out and scale in. Vertical scaling steps, that is, adjusting the

resources within a VM, are known as scale up and scale down (Dutta et al., 2012). Since diagonal scaling combines horizontal and vertical scaling, all four scaling steps—out, in, up and down—can be performed (Han et al., 2012).

Obviously, vertical scaling can be more fine-grained and usually performs much faster than horizontal scaling (Mao and Humphrey, 2012; Yazdanov and Fetzer, 2012). A scale up can be performed within one second, whereas the provisioning of another VM in terms of a scale out can require several minutes. However, vertical scaling has severe economic and physical limitations compared to its horizontal counterpart. This is because the amount of resources a VM can consist of is physically limited. On the other hand, VMs with an overall higher performance are more expensive than small VMs. This is why a scale-up is either physically or economically unfeasible up from a certain size. Diagonal scaling, as a combination of both approaches, mitigates the disadvantages of only one scaling dimension (Han et al., 2012). It is capable of increasing the adaptability of IaaS deployments and therefore has the potential to reduce resource allocation under fluctuating loads.

The flexibility in adapting cloud computing resources allows for a fine-grained pricing model. This leads to an usage-based pricing, which is an important determinant of cloud adoption (Jamshidi et al., 2014; Lin et al., 2009). In contrast to coarse granular pricing models for horizontal scaling, which often have a minimum contract duration, such as one hour, real usage-based pricing models are possible by diagonal scaling. To achieve a sufficient granularity in terms of time, a pricing model with a resolution of one second is possible (Berndt and Maier, 2013). Since the pricing model based on diagonal scaling techniques allows for being more usage-based than IaaS solutions based on horizontal scaling, the pricing curve fits better to the resource consumption curve. This is illustrated by the Figures 6.1 and 6.2. In both figures, the grey area indicates overprovisioning. The smaller this area is, the more efficient is the scaling technology. The use of horizontal scaling technology enables an adaptation to the load. Smaller scale steps and shorter provisioning durations allow a further adjustment of the resource allocation to the load when scaling diagonally.

To drive the allocation of VMs and associated virtual CPUs in accordance with the number of requests for the system, an automatic decision must be made regarding the scaling steps. Besides a threshold-based scaling, as presented by Han et al. (2012), there are other alternatives based on reinforcement learning (Vasić et al., 2012), queuing theory (Urgaonkar et al., 2005), control theory (Jamshidi et al., 2014; Kalyvianaki et al., 2009) or time-series analysis (Dutta et al., 2012; Heinze et al., 2014). Since threshold-based scaling is very popular, as pointed out by Heinze et al. (2014), and supported by most main IaaS providers, such as Amazon Web Services, Inc. (AWS) (<http://aws.amazon.com/autoscaling>) or Rackspace Hosting, Inc. (<http://rackspace.com/cloud/auto-scale>) this approach is used within this chapter.

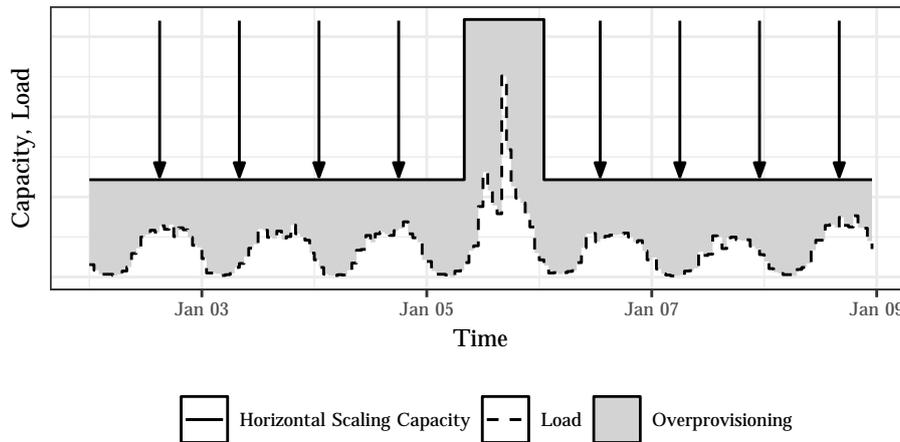


Figure 6.1: Schematic illustration of resource allocation with horizontal scaling. The arrows represent the resource reduction potential of horizontal scaling compared to an on-premise solution.

6.3 Research Design

To determine the scaling efficiencies of horizontal and diagonal scaling and to calculate the resource reduction potential of diagonal scaling, a generic scaling model is provided in this section. This generic model is then used to simulate horizontal and diagonal scaling by setting corresponding parameters. The parameter selection, the steps of the simulation process and the load pattern that has to be processed within the simulation are described.

6.3.1 Scaling Model

To compare horizontal and diagonal scaling alternatives a generic scaling model based on threshold-based auto-scaling is developed. The parameters and variables considered for scaling are provided in Tables 6.1 and 6.2. While different scaling scenarios—such as horizontal scaling and diagonal scaling—can be specified by selecting the parameter values, the variables represent volatile measures of the system, such as the incoming load and the system’s utilization. In the following, a brief discussion is given of both the parameters and the variables through which the generic scaling model is specified.

The CPU demand of a backend service for making threshold-based scaling decisions is considered (Iqbal et al., 2011). An almost linear relation ϕ is assumed between the number of requests in a defined period of time n_{req} and the number of CPUs n_{cpu_dem} required for processing them within a given time (Sedaghat

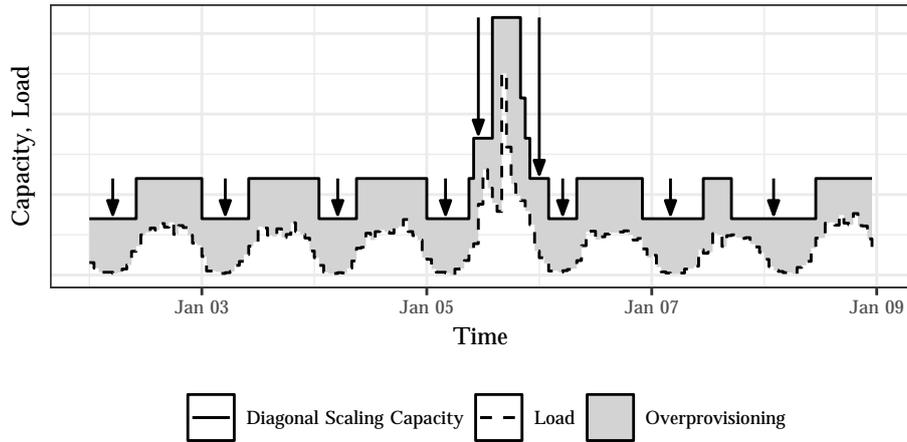


Figure 6.2: Schematic illustration of resource allocation with diagonal scaling. The arrows represent the additional resource reduction potential of diagonal scaling compared to horizontal scaling.

et al., 2013). By applying this constant performance, the CPU demand at all times of the investigated load pattern can be determined by Equation 6.1

$$n_{cpu_dem}(t) = \left[n_{req}(t) \cdot \frac{1}{\phi} \right]. \quad (6.1)$$

CPU is considered as the main determinant of the VM capacity, thus other aspects, such as memory, are neglected. To simulate the actual resource allocation n_{cpu_alloc} and n_{vm_alloc} for a previously determined CPU demand pattern, the scaling model provides continuous monitoring of the CPU utilization u . A fully utilized CPU is able to serve exactly the amount of requests determined by ϕ while meeting the defined response time for each request. The average CPU utilization of a sliding monitoring window w_h and w_v serves as criterion for the scaling thresholds θ_{h_out} , θ_{h_in} , θ_{v_up} and θ_{v_down} and is calculated each second. In both techniques, scaling steps always involve a single CPU. Thus, for vertical scenario the unit is a single CPU, while for horizontal scenario the unit is a VM with a single CPU. A bundled pricing model is assumed for acceptance and simplicity reasons as discussed by El Kihal et al. (2012). Accordingly, a bundle consists of a CPU and further corresponding resources, such as memory, storage, network, etc. For distinguishing between horizontal and vertical scaling steps within the diagonal scaling scenario, a maximum number of CPUs per VM n_{cpu_max} is taken into account. As mentioned in the previous section, the resource reduction potential of diagonal scaling techniques is mainly achieved by shorter provisioning durations d_{prov} and also by shorter contract periods d_{cont} , reflected by the corresponding parameters. To allow a comparison between scaling technologies, a

Table 6.1: Overview of scaling parameters. Selecting the parameter values of the generic scaling model allows the definition of different scaling scenarios. They are specified once initially and remain constant throughout one simulation run.

Parameter	Unit	Description
ϕ	$\text{req} \cdot \text{s}^{-1} \cdot \text{cpu}^{-1}$	Served requests per CPU per second
q	%	Common QoS criterion
w_h	s	Decision time window for horizontal scaling
w_v	s	Decision time window for vertical scaling
θ_{h_out}	%	CPU utilization threshold for scale out
θ_{h_in}	%	CPU utilization threshold for scale in
θ_{v_up}	%	CPU utilization threshold for scale up
θ_{v_down}	%	CPU utilization threshold for scale down
n_{cpu_max}	cpu	Maximum number of CPUs per VM
d_{prov}	s	Provision duration for scaling up
d_{cont}	s	Minimum contract period

Table 6.2: Overview of scaling variables. The values of the variables vary throughout the simulation and therefore represent volatile measures of the system.

Variable	Unit	Description
n_{req}	$\text{req} \cdot \text{s}^{-1}$	Number of incoming requests per second
n_{cpu_dem}	cpu	Number of CPUs required to serve requests
n_{cpu_alloc}	cpu	Number of actually allocated CPUs
n_{vm_alloc}	vm	Number of allocated VMs
u	%	CPU utilization

common quality of service (QoS) criterion q is defined. Thus, both scenarios have to process requests within a given response time (Menasce, 2002).

6.3.2 Scenario Setup

This section first describes how the scaling model is applied to determine the resource efficiency of a threshold-based horizontally and diagonally scalable IaaS through simulations. Based on these efficiencies, the resource reduction potential of diagonal scaling compared to horizontal scaling is evaluated. The overall evaluation process is outlined in Figure 6.3.

The evaluation is performed by comparing the scenarios, given an eleven-day load pattern—with an one-day heat-up phase—of Gloveler GmbH (<http://gloveler.de>). Gloveler is a German web application provider for offering and booking private accommodations. The load pattern is also provided in Figure 6.3. Although this load pattern does not originate directly from a crowd-based appli-

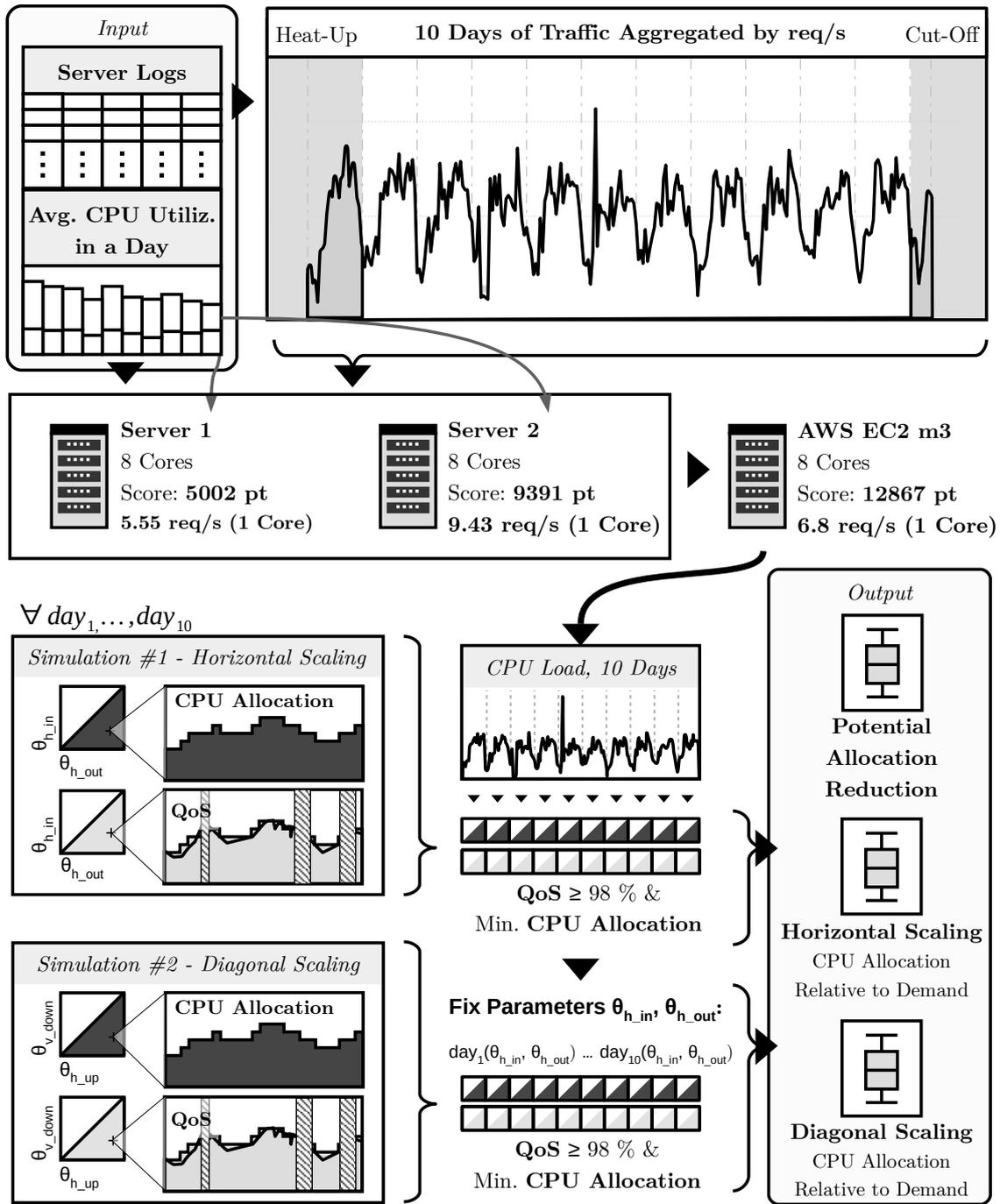


Figure 6.3: Outline of simulation for the determination of scaling efficiencies. Given a load pattern, a CPU demand pattern is derived, which is attempted to be met by simulations based on horizontal and vertical scaling technologies. In each scenario, the minimum resource allocation is determined under the condition that a common QoS criterion is met.

cation and therefore does not represent a load generated by a vehicle fleet, the load pattern has the relevant characteristic of a daily fluctuation. This fluctuation is expected to be similar for a national crowdsensing-based system as during the day more contributions from participants are expected than at night.

Table 6.3 provides the parameter values considered for the evaluation of the horizontal and diagonal scaling scenarios. The value of ϕ represents the perfor-

Table 6.3: Values of scaling parameters used for simulation. The selected parameters allow to determine the resource allocation of a horizontally scalable infrastructure on the one hand and the allocation of a diagonally scalable infrastructure on the other hand by performing simulations.

Parameter	Value horizontal	Value diagonal	
ϕ	$6.8152 \text{ req} \cdot \text{s}^{-1} \cdot \text{cpu}^{-1}$	$6.8152 \text{ req} \cdot \text{s}^{-1} \cdot \text{cpu}^{-1}$	
q	0.98	0.98	
n_{cpu_max}	1 cpu	8 cpu	
d_{cont}	1 min, 1 hour	1 s	
		Horizontal	Vertical
w	600 s	600 s	30 s
d_{prov}	97 s	97 s	1 s

mance of an AWS EC2 m3 general purpose instance, which encompasses exactly one CPU. AWS EC2 instances are selected for comparability reasons. The performance ϕ is stated in the number of requests of the load pattern that can be processed by the instance per second. To determine ϕ , server logs describing the load pattern processed by two physical servers of Gloveler and the CPU utilization of these servers are considered. The maximum number of requests that can be processed by the Gloveler CPUs is calculated and mapped to a CPU of the AWS instances. This mapping is based on CPU benchmarks from PassMark Software Pty., Ltd. ([http://cpubenchmark.net/compare.php?cmp\[\]=834&cmp\[\]=896&cmp\[\]=1220](http://cpubenchmark.net/compare.php?cmp[]=834&cmp[]=896&cmp[]=1220)).

The CPU performance of the AWS instance was chosen for the calculation of the resource demand, which is stated as CPU load in Figure 6.3. Following an example of the AWS user guide the monitoring window w is 600 seconds for horizontal scaling and 30 seconds for vertical scaling (AWS, 2010). To remain comparable in terms of scaling units the maximum number of CPUs per VM n_{max_cpu} for horizontal scaling is one, which is possible with AWS EC2. For diagonal scaling the maximum number is equal to the largest AWS EC2 m3 general purpose instances, which currently is 8 (AWS, 2017). The provisioning duration d_{prov} for horizontal scaling is determined according to an AWS EC2 Linux VM (Mao and Humphrey, 2012) and according to Yazdanov and Fetzer (2012) for diagonal scaling. The minimum contract duration d_{cont} for the horizontal scenario is determined according to the contract duration of AWS EC2 on-demand instances and

for a more conservative calculation a duration of one minute is chosen additionally. For the diagonal scaling scenario the minimum contract duration of one second is defined since vertical scaling allows this.

For both scenarios—horizontal and diagonal scaling—and for each day of the load pattern multiple simulation runs are performed for different thresholds θ of 5 %, 10 %, 15 %, ..., 95 %. The thereby determined resource allocations and QoS are used for selecting the threshold combinations with the lowest resource consumption while meeting the common QoS criterion q of 98 % for each day (Menasce, 2002). This procedure is carried out by first determining the scaling thresholds θ_{h_out} and θ_{h_in} for horizontal scaling, which are used for both scenarios. After determining the thresholds for horizontal scaling, the vertical scaling thresholds θ_{v_up} and θ_{v_down} are determined, which are only applied for the diagonal scaling scenario. Thus, an accurate comparison of the scenarios' resource allocations is possible.

6.4 Results

The simulation results are discussed in the following section. Accordingly, the efficiency of horizontal and diagonal scaling is described as well as the resource reduction potential of diagonal scaling compared to horizontal scaling. This answers Research Question 4.

6.4.1 Scaling Efficiency

To answer the Research Question 4 a), how efficiently the challenge of having a fluctuating demand in IT resources can be served by a horizontally and diagonally scalable IaaS cloud, Figure 6.4 shows the average CPU allocation relative to the demand of all ten days for both scaling scenarios. The worst resource efficiency is obtained when horizontal scaling is applied and the minimum contract duration d_{cont} is one hour. In this case, 183.05 % more CPUs are allocated than necessary for processing. If a contract duration of one minute is assumed, 174.92 % more resources must be allocated. The diagonal scaling scenarios achieve a better resource efficiency with 159.74 % for the one hour contract duration and 163.34 % for the one minute case respectively. These overall high resource allocations, in order to guarantee a 98 % QoS, suggest that a significant reduction potential can be achieved by choosing the appropriate scaling technology. The results of the investigation of this reduction potential are therefore presented below.

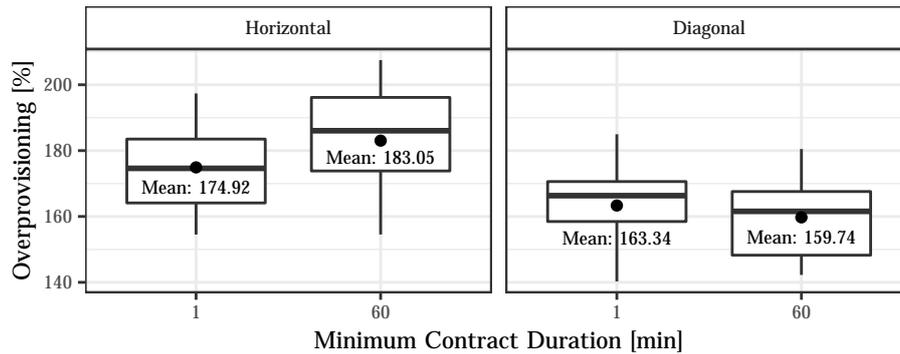


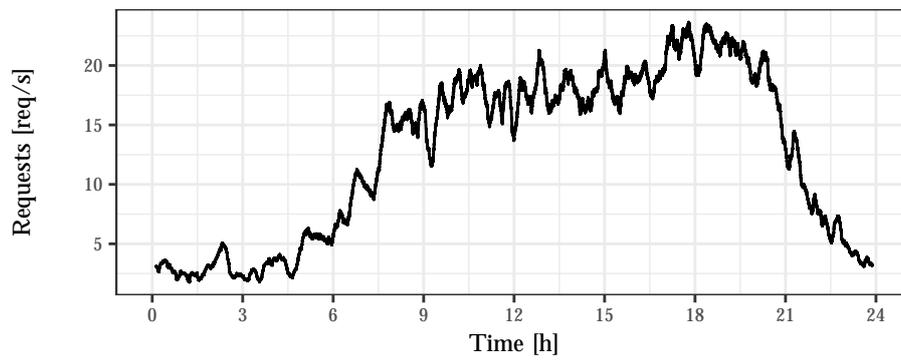
Figure 6.4: Resource efficiency of horizontal and diagonal scaling. For both scaling scenarios, horizontal and diagonal scaling, the overprovisioning is provided. Thus, the difference between the allocated resources and the resources required is deducted. For horizontal scaling—which is also part of diagonal scaling—a distinction is made between two minimum contract durations.

6.4.2 Resource Reduction Potential

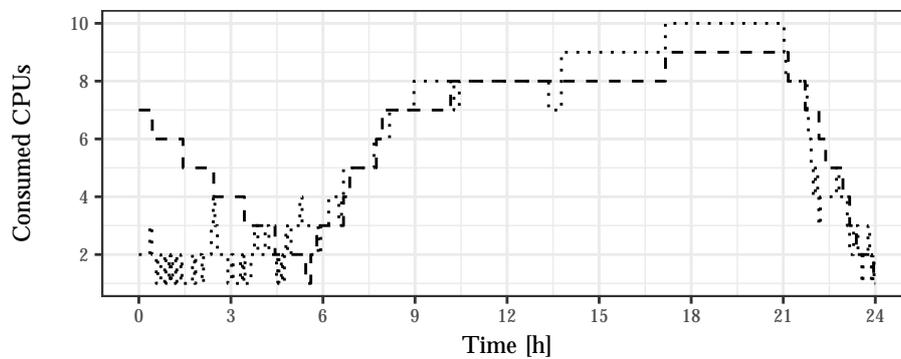
Figure 6.5a shows the one-day load pattern of the 24.07.2014 for exemplary reasons. Additionally, a histogram and boxplot summarizing the number of requests per 15 minutes for the entire load pattern are provided in Figure C.1 in the Appendix. Figure 6.5b and 6.5c show the corresponding resource allocation and the number of requests that were not served in time for both scenarios by using the determined thresholds ϕ for that day and considering a common contract duration d_{cont} of one hour for the horizontal scaling scenario.

Comparing the allocations—in terms of the actually consumed CPUs—, a quicker response to load fluctuation for the diagonal scenario can be observed. At the beginning of the graph—approximately between 12 and 4 a.m.—, the adaptation of the actually allocated resources to the resource demand is delayed in the horizontal scaling scenario. This is mainly because of the longer minimum contract durations. Also during about 5 and 7 a.m. the resource adaptation is faster with diagonal scaling. In this case, however, it is the other way round so that the required CPU bundles are added faster, allowing more requests to be processed in time. This reduced violation of the QoS criterion makes it possible to scale down faster in the diagonal scaling scenario between approximately 9 and 11 p.m. and still meet the overall QoS criterion of processing 98 % of the requests in time.

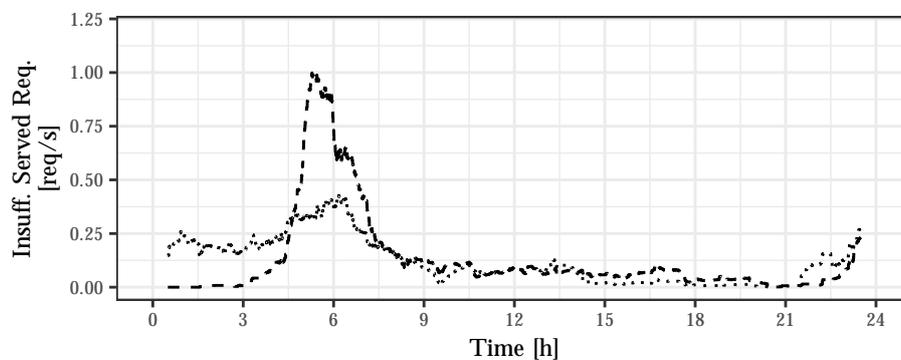
Comparing the resource allocation between horizontal and diagonal scaling for the entire load pattern, the resource allocation reductions shown in Figure 6.6 are observed. These results present, the Research Question 4b), what resource reduction gets possible by using diagonal scaling compared to conventional horizontal scaling, can be answered. A mean CPU reduction of 8.05 % is achieved by



(a) One-day load pattern of the 24.07.2014 (15 minutes average).



(b) Resource allocation for one-day load pattern.



Scaling Diagonal --- Horizontal

(c) Insufficiently served requests for one-day load pattern (1 hour average).

Figure 6.5: Resource allocation and fulfillment of the QoS criterion of horizontal and diagonal scaling for one-day load pattern. The 24.07.2014 was chosen by chance from the load pattern of Gloveler. It can be seen that the diagonal scaling can, for example, cover the required resources more quickly in the morning through a higher elasticity.

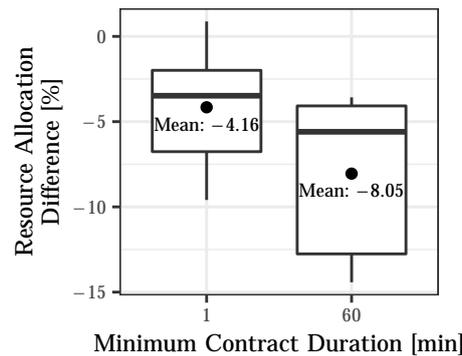


Figure 6.6: Resource allocation reduction by diagonal scaling. Based on the scaling simulations, the resource reduction potential of diagonal scaling compared to horizontal scaling is shown for all of the ten days of the load pattern. For horizontal scaling, a distinction is made between two minimum contract durations.

the diagonal scenario when considering a common contract duration d_{cont} of one hour for the horizontal scaling scenario. If, on the other hand, a contract duration of one minute is assumed, the resource allocation can still be reduced by 4.16 %.

6.5 Conclusion

In this chapter, an automated generic resource scaling simulation model is build. A fine-grained CPU-bundled pricing regime is considered for the simulation. Based on an exemplary real life load pattern of a German web application provider, the resource consumption of a common horizontal scaling scenario is compared with a diagonal one. For both scaling scenarios, the efficiency in terms of required and actually allocated resources is determined. While providing a higher granularity for possible scaling steps, an average saving in resource consumption between 4.16 % and 8.05 % is achieved depending on the assumed minimum contract duration.

Using the exemplary load pattern with daily fluctuations—as it is expected for a back-end IT infrastructure of a crowdsensing-based road condition monitoring system—a significant resource reduction potential is determined by the use of diagonal scaling technologies compared to horizontal scaling. By means of this result it can be recommended to use diagonal scaling for the approach presented in the thesis at hand to keep the actual resource allocation to a minimum while addressing fluctuations in the load.

Limitations in the investigation presented in this chapter are also mentioned here. So far, a rather short load pattern is covered disregarding long-term trends. Although long-term trends in the crowdsensing-based road condition monitoring can be expected when the number of participants increases over time, such a

relatively slow increase in the resource demand should not be critical when using cloud infrastructure. However, the advantage of diagonal scaling compared to horizontal scaling may change with a higher overall load. Additionally, it should be taken into account that the requests are drawn from a web application and different types of requests are neglected. Next to the technical challenges regarding the feasibility of a crowdsensing-based road condition monitoring system, it is also essential to consider the actors involved and examine their interaction in order to enable a corresponding crowdsensing-based service. This integration of actors to provide a service is covered in the following chapter.

Chapter 7

Integrated Service

A frequent monitoring of road conditions is beneficial for road users and road authorities. Applying such an approach, there are basically two parties—the crowd as data provider and the road users and road authorities as data consumers. The crowd faces the challenge of marketing its data, since a single participant provides too few and too inaccurate data that it would be of interest to other road users or road authorities. The road users and road authorities on the other hand require decision support based on frequent and accurate road condition information. These challenges can be overcome by an intermediary, which integrates existing data sources for providing a customizable decision support service. Considering the need in such an integration service this chapter addresses the question, how a new intermediary (service integrator) allows for serving the individual information demands—on a nearly real-time and sufficiently accurate basis—of road users and road authorities by orchestrating data services from multiple single data providers. Bringing together the findings from the former chapters, a smart crowd-based road condition monitoring service is developed that introduces an intermediary between the crowd as a data provider and road authorities and road users as service customers. Next to providing accurate and frequent road condition information to customers, participants in the crowd are hereby enabled to monetize their collected data. This chapter is based on a manuscript currently conditionally accepted for publication in the special issue "Smart Services: The Move to Customer-Oriented" of the *Electronic Markets Journal*: Laubis, K., M. Konstantinov, V. Simko, A. Gröschel, and C. Weinhardt (2018). Enabling Crowdsensing-Based Road Condition Monitoring Service by Intermediary. *Electronic Markets*, 1—16. (forthcoming).

7.1 Introduction

The continuous monitoring of road conditions is crucial for the safety and comfort of road users and for the efficient maintenance of the road network. As the previous chapters describe, the ever-increasing number of sensors in smart devices, which are carried in the vehicle, allow for supplementing or superseding the traditional road monitoring by a crowdsensing-based approach.

However, an unsolved challenge faced by the crowdsensing-based approach is the fact that a single participant is not able to utilize the collected data, for example by selling it directly to road authorities or other road users. This is caused by different reasons. First, the raw sensor data does not directly provide insights into the road's characteristics, such as the road roughness. These characteristics have to be estimated based on sensors in commodity devices. As Chapter 3 shows, the reliability of the information is lower compared to the information provided by engineering companies. Second, the data that a single participant can contribute is limited in space and time. Such limited data sets are not of interest to road authorities and road users as they require a holistic view of their road network or at least of a freely determinable subset of the network.

On the one hand, there is the huge potential of providing road condition information from the crowd, whose participants are not able to market this data, and on the other hand there are the road authorities and road users, who require accurate and frequent road condition information. This results in the need for a new intermediary, which integrates existing data sources for providing a decision support service. In the further course of the thesis, such a service is called smart crowd-based road condition monitoring service.

This chapter focuses on answering the corresponding Research Question 5 by addressing the following two subquestions: (a) How can existing services be integrated by a new intermediary for providing a crowd-based road condition monitoring service allowing for a frequent and accurate support in decision-making? (b) What are the utilities to the intermediary, to the crowd and to customers of such an integrated crowd-based road condition monitoring service?

The remainder of this chapter is as follows: first, related literature from the field of service science is discussed. Second, the applied research design is presented. This also includes the introduction of the road condition monitoring. Evaluation results are then provided before the chapter concludes.

7.2 Service Systems

Road infrastructure monitoring is a service system in which road users, road authorities and service providers interact and mutually create value. Road authorities require service providers to monitor the road condition. They provide the service of an adequate and safe road infrastructure to taxpayers, the road users. Service providers, for example engineering companies, only create value if the infrastructure is used by the road users. Maglio et al. (2009) identify this "co-creation of value" by "a configuration of people, technologies [and] organizations" in the system as a primary characteristic of service systems. In road condition monitoring however, the road user, as an integral part of the service system, does neither participate in the road condition monitoring nor has access to up-to-date road condition information.

Barile and Polese (2010) define smart service systems as systems that are designed for self-management and self-reconfiguration to ensure the provision of a satisfactory service to the participants. As they raise participant's service satisfaction, service systems including smart services are being introduced in various domains, such as electricity grids, home automation and smart city architectures (Farhangi, 2010; Anttiroiko et al., 2014; Byun and Park, 2011). Allmendinger and Lombreglia (2005) argue that a key element of smart services is the introduction of intelligence into the service landscape, which facilitates higher customer engagement for existing services and enables new services. In today's system, communication between road users and road authorities is difficult and customer engagement is low. Consequently, the service system of road condition monitoring shows deficiencies with regards to the satisfaction of road users' and road authorities' needs.

In order to provide a smart service for road authorities and road users, road users have to be integrated into the monitoring process. As this requires the integration of road condition information sources of multiple service providers, multi-sourcing service integration is found to be relevant. Multi-sourcing service integration covers different management approaches and business processes to integrate various external service providers and their interdependent services into an existing organization (Goldberg et al., 2014). Research about multi-sourcing integration originates from the domain of IT outsourcing and experienced a rise in scientific interest in recent years, as multi-sourcing strategies become increasingly important for companies (Herz et al., 2010; Bapna et al., 2010). Originally the retained organization, meaning the part of the IT that is kept in-house and is not outsourced integrates services of multiple service providers (Dibbern et al., 2004; Goldberg et al., 2014). As the integration of crowd-sensed road condition information requires complex analytical processing, road authorities lack the ability to integrate all relevant data sources. Therefore, this direct integration is not applicable in the case of crowdsensing-based road condition monitoring. A different integration model is necessary.

Unterharnscheidt and Kieninger (2010) identify the management of multiple providers as a challenge for companies seeking to adapt a multi-sourcing service strategy. In the case of interdependent services, management of multiple service providers is difficult (Goldberg et al., 2015). Gallivan and Oh (1999) classify outsourcing relationships in a service network. According to the authors' taxonomy, road condition monitoring can be classified as a complex outsourcing relationship because multiple service providers—various engineering companies with accurate measurement ability—can provide their service to multiple customers, such as road authorities and road users. Making the crowd part of the system, this relationship becomes even more complex as a great number of new service providers with different quality levels enter the market. This underlines the necessity for a separate service integrator role in the domain of road condition monitoring.

Rajamaki and Vuorinen (2013) provide a framework for multi-sourcing service provider management. Existing methods are adapted and applied to multi-sourcing service management in public protection and disaster relief organizations. However, they describe a higher-level framework and cover a different domain, which is why their findings are not applicable in the case of crowd-based road condition monitoring.

Goldberg et al. (2014) identify five concepts of managing the integration of several service providers in a service network. In addition to the traditional concept of integration in the retained organization, the role of integrator can also be fulfilled by a prime provider—one of the service providers—, which can be a separate integration entity, or it can be distributed between the stakeholders in the value network. The concepts of a prime provider and a separate integration entity are both applicable to the case of road condition monitoring, as they take the task of complex data processing away from road authorities. However, the authors provide a generic framework for multi-sourcing service integration, whereas this chapter focuses on providing an applicable framework for the domain of road condition monitoring.

Since providing a smart road condition information service is essential for both, road authorities and road users, and has not been addressed in research so far, a general framework for a smart service integrating road condition information from various sources is introduced in this chapter.

7.3 Research Design

To address the research question, a smart crowd-based road condition monitoring service is designed and modeled as a service map. By considering a new intermediary, it solves the problem of bringing both together, the crowd as a provider of raw data and the road authorities and road users as customers requiring customized road condition information. The smart service is evaluated by a hotspot analysis on crowdsensing-based condition information and a descriptive evaluation of the stakeholders' utility.

In this section, an overview of the methods applied is provided. The solution to the identified problem is to overcome the difference between the goal state and the current state. Therefore, both is identified, the current and the ideal state of a road condition monitoring service. Comparing both concrete requirements are defined in terms of the accuracy and granularity that the target service should fulfill. Also a metric is defined to quantify the ability to meet one of the requirements. According to the defined objectives of the solution, a smart service for road condition monitoring is designed. It is different to current solutions as it considers a new intermediary between the information providers and the information customers. For summative evaluations, analytical methods are chosen to determine the degree the service fulfills the timely provisioning of road condi-

tion inspections. The service’s value to road authorities is also addressed in the further narrowed down scenario of scheduling road maintenance actions. This results in an analytically and descriptively evaluated artifact.

Figure 7.1 outlines the smart crowd-based road condition monitoring service and the services it is composed of. The model is based on the concept of service

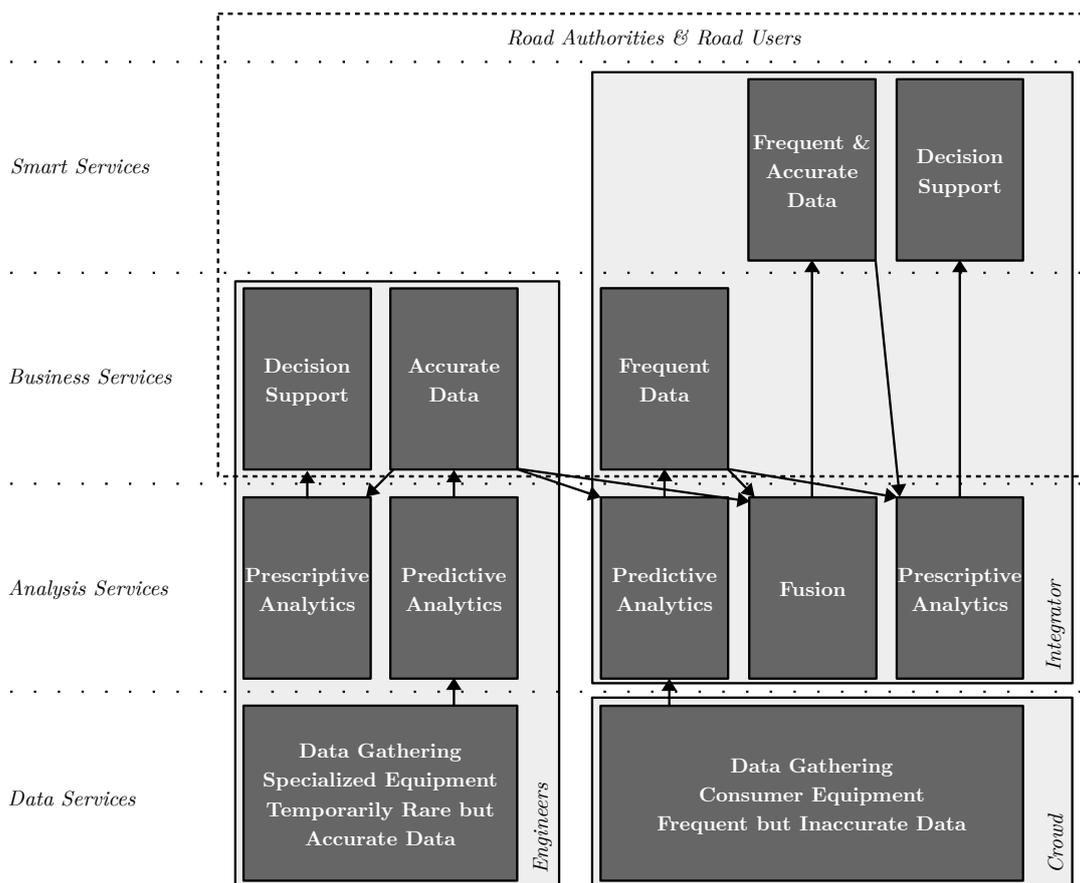


Figure 7.1: Service map of smart crowd-based road condition monitoring service. The service is enabled by a new intermediary that integrates multiple data sources—accurate measurements from engineering companies and frequent measurements from the crowd. The smart service is composed of several single services that are assigned to different service layers. The top two service layers—smart services and business services—provide value for road authorities and road users.

maps, which are adapted to the specific requirements of road condition monitoring. A service map is a visualization of relationships between services and the involvement of stakeholders in these services. The status quo in road condition monitoring and the change through the introduction of a new intermediary are explained based on the concept of service maps in the following sections.

7.3.1 Service Layer

The single services are visualized as small rectangles that are assigned to service layers. Based on the service typology developed by Kohlmann et al. (2010) the model is divided into four service layers. The data and analysis service layers represent technical services. The separation into data and analysis services has been introduced by Demirkan and Delen (2013). Additionally, the business services are further divided into regular business services and smart services to emphasize the need for an intermediary to integrate the crowd into the market and provide crowd-based road monitoring services. Data services are basic data collection services in the model. They do not provide any value to a potential customer by themselves and can be differentiated regarding the data source, the data frequency and the data quality. Analysis services encompass data processing services, needed to derive meaningful information out of raw data and for drawing conclusions based on the derived information. Analysis services are the base for business services as only processed data provides value to potential customers. Business services represent services that are of value to certain customers and thus, can be marketed to them. Smart services are services that combine business services or other smart services itself to derive additional information and offer new services. This layer has similarity to the service cluster layer, as described by Kohlmann et al. (2010), in using the outcome of underlying business services. Yet, in contrast to a service cluster, the outcome of existing business services is not only clustered, but first redirected to the analysis layer, where further analysis steps provide the information that is the basis for the resulting smart service. In addition, it is a necessary condition for a smart service in the crowdsensing-based road monitoring domain to be enabled by a fusion of data from engineering companies and the crowd, as described in Chapter 3. Thus, a smart service meets the different demands of customers regarding space, time and accuracy of road condition information. They also differ from business services as they integrate single business services and provide an automatic tailoring to the customer's needs. The frequent data service, for example, remains a business service, as accurate data from engineering companies is required for the calibration of crowd measurements, but this accurate data itself is not directly included in the services offered to the customer.

Only the two upper layers, namely business services and smart services, can be used by the two stakeholders road authorities and road users. This is indicated in Figure 7.1 by the surrounding dashed rectangle. Next to assigning the services to service layers, the services are grouped by the type of player who provides them. Relevant players are the engineering companies, the crowd and the new intermediary, called the service integrator. This grouping is indicated by solid rectangles surrounding the individual services. Services depend on each other. One service for example requires the output of one or multiple other services. This is indicated by arrows. The arrow is directed from the supplying service to

the consuming service. These dependencies can cross layers and can exist within the boundaries of one type of service provider or they can cross these boundaries.

7.3.2 Current Situation

Nowadays, if a standardized and technology-based road condition monitoring approach is applied at all, it is mainly based on gathering raw data with special purpose vehicles. This data gathering is conducted by engineering companies. The collected raw data, such as global positioning system (GPS) coordinates, laser measured road profiles, vibration patterns, high-resolution images, etc., is of high-quality due to the specialized and sound calibrated sensors. This data gathering service is located in the bottom layer of the service map and is connected to the predictive analytics service. The predictive analytics service includes all tasks from using the raw sensor data to derive information and metrics that are meaningful for road authorities and road users. For instance, services that derive meaningful metrics, such as the international roughness index (IRI), are to be located here. Since this results in information, which is relevant for road authorities and road users and is directly applicable, the data can be provisioned to them. Thus, the accurate data provisioning service of engineering companies is a business service and accordingly located in the service map. Considering that the customers do have to come up with decisions based on this accurate data, prescriptive analytics tasks have to be performed subsequently. Road authorities for example have to decide when to perform which maintenance action at which road link and road users for example have to decide, whether to take a detour based on rough road conditions ahead. Applying such prescriptive analysis services by the engineering company allows for providing decision support services with direct managerial implications to the two main customers, road authorities and road user. As already mentioned, these services rely on highly accurate sensor readings but the relatively high costs and scarce resources prohibit frequent measurements. Thus, the business services provided by the engineering companies are accurate and reliable but of low frequency.

7.3.3 New Intermediary

The tasks of the new intermediary as an enabler of a smart crowd-based road condition monitoring service is to manage both, the demand of the customers, namely the road authorities and road users, and the integration of existing and new crowd-based data suppliers for serving these demands.

The current approach of performing road condition monitoring without a service integrator does not allow the crowd to market their data. This is indicated in Figure 7.1 since the crowd by itself does not have any services in the business or smart service layer. A single participant gathering a certain amount of inaccurate

raw data requires a player that performs certain analysis steps for making the data valuable for road authorities and other road users. Therefore, the intermediary has to come up with a payment regime reflecting the quality of the data provided by the individual service suppliers. Given this service integration model, road users, which contribute by sensing the road's condition, directly benefit by being compensated for the data they contribute. This is possible only because the amount of participants necessary to produce robust and reliable crowd-sensed information can be aggregated by the service integrator. As with the engineering companies, meaningful road condition metrics have to be derived from this raw data. This can be achieved by applying supervised machine learning algorithms to the raw data for calibration, which is represented by the predictive analytics service of the integrator. However, these supervised algorithms require information on the actual road condition for training purposes. Thus, the integrator has to commission accurate road condition measurements for these calibration. The integrator has to decide when and for which road links a procurement of the accurate ground truth is beneficial. Hereby, it has to be considered when and where a calibration of new participants is required. Next to this, the procurement of accurate data is necessary for road links that are driven by those participants with inaccurate models and thus require a recalibration due to changes in the car, the sensors, the driving behavior, etc. This should be done automatically to enable an efficient and easy integration of new participants and an easy recalibration for already existing participants. This self-calibration approach is described in Chapter 3. Even though, combining multiple crowd-based road condition measurements leads to a more robust road condition prediction—as shown in Chapter 5—the provided data tends to be less accurate. However, this way the integrator can provide a frequent data service to road authorities and road users. The information demand of customers especially road authorities differs depending on the decision that has to be supported. Decisions that require specific information are for instance maintenance scheduling, budget planning, estimating the road asset value, technical acceptance of new constructions or resurfacing works, driver navigation, hazard warnings, etc. Furthermore, the service has to be customizable depending on the considered road network. It has to be distinguished whether a nationwide network, a certain city or a single road link should be inspected. To meet these customization needs, a fusion of highly accurate data from engineering companies and frequent but less accurate data from the crowd can be performed. Thus, in addition to the decision of purchasing ground truth data, the intermediary also has to decide when to purchase and when to perform which service for most efficiently serving the demand of customers. Within such a combination, the different data sources have to be weighted according to their reliability, which means both, the sensing accuracy and their timeliness.

The intermediary can compose the provided service depending on the individual information demands of the customer. In the case of road authority customers, for example, roads that are old and likely to degrade in an unforeseen

way, require more frequent measurements than roads that are quite new. Likewise, road segments, which are made of concrete should be monitored more frequently on hot summer days since they tend to suffer from so called "blow ups", which are unforeseen bucklings of the concrete elements. In contrast, newly constructed or newly paved roads do not require a temporally fine-grained monitoring, but a few very accurate final inspections for the acceptance of the work. Thus, an integrated and customized smart crowd-based road condition monitoring service based on frequent and accurate data becomes possible. Next to providing the road authorities and road users with customized road condition information, the intermediary is able to provide a decision support service based on applying prescriptive analyses. Road authorities and road users are interested in identifying these road links and road segments that differ significantly from others. Also clusters of bad road condition segments can be of high relevance. Such segments, called hotspots, can reveal essential insights and for example allow for a prioritization of maintenance tasks.

7.3.4 Hot Spot Analysis

Hotspot analysis is a tool for determining patterns of spatial or spatio-temporal autocorrelation in a geographical area (O'Sullivan and Unwin, 2010). It can be used to provide decision support. Thus, it is regularly applied in various fields, such as criminology, epidemiology, traffic safety, etc. (Ratcliffe et al., 2011; Goovaerts and Jacquez, 2005; Sugumaran et al., 2009; Steenberghen et al., 2004). Hotspots provide a robust insight into the local environment of a measured instance. Regarding the domain of road condition monitoring a hotspot can be defined as a geographical cluster showing a concentration of bad road conditions. Providing a hotspot analysis as a smart service enables road authorities to focus their maintenance activities on the most relevant areas. Different statistical metrics can be applied for revealing hotspots. The Getis Ord G^* metric is chosen in this chapter as it enables the detection of spatial associations of a geographical region with adjacent regions in a selected distance a . Additionally, G^* measures are z-score normalized. This inherently allows for determining statistically significant hotspots (Ord and Getis, 1995). Other applicable metrics are Local Geary's C and Local Moran's I (Anselin, 1995). The latter is especially suited for detecting local outliers, such as potholes. The G^* statistic for the cell i is defined by Equation 7.1

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} y_j - \bar{Y} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}}, \quad (7.1)$$

where y_j is the value of the j -th cell, S is defined by Equation 7.2

$$S = \sqrt{\frac{\sum_{j=1}^n y_j^2}{n} - \bar{Y}^2} \quad (7.2)$$

and \bar{Y} is defined by Equation 7.3

$$\bar{Y} = \frac{\sum_{j=1}^n y_j}{n}. \quad (7.3)$$

The elements $w_{i,j}$ of the spatial weight matrix W are defined according to Equation 7.4

$$w_{i,j} = \begin{cases} r(a_{i,j}), & \text{if } a_{i,j} \leq a^\theta \\ 0, & \text{otherwise} \end{cases}. \quad (7.4)$$

G_i^* is calculated for every cell i in the spatial grid, considering the IRI value y_i of the cell itself and the values y_j of adjacent cells within the convolution distance a . The adjacent cells contribute to G_i^* depending on the spatial weight $w_{i,j}$ that is attributed according to their distances $a_{i,j}$ to cell i . For example, the distance a can be the Chebyshev distance ($a_{i,j} = \max\{\|r_i^{(1)} - r_j^{(1)}\|, \|r_i^{(2)} - r_j^{(2)}\|\}$, with $r_l^{(k)}$ depicting spatial grid coordinates of the regarded cells) and the Manhattan distance ($a_{i,j} = \|r_i^{(1)} - r_j^{(1)}\| + \|r_i^{(2)} - r_j^{(2)}\|$) for further customization (Cha, 2007). A road authority in charge of maintaining a highway for instance might consider a greater distance a and spatial weights of inverse distance $r(a_{i,j}) = 1/a_{i,j}$ if maintenance actions affect larger road segments. For road authorities which cover the road grid of a city and are interested in local anomalies, a smaller distance a and spatial weights of inverse square distance $r(a_{i,j}) = 1/a_{i,j}^2$ would be more appropriate. Thus, the y_j of a close adjacent cell can have a stronger influence because of a higher spatial weight $w_{i,j}$ than that of an adjacent cell that is located further away. In addition, a^θ can be used to determine a threshold up to which distance neighboring cells should be considered. The cell's deviation from the expected value is determined and standardized to provide the z-score.

The hotspot analysis is the basis of a prescriptive analytics service in the service map. It can be customized to the specific needs of a single stakeholder depending on the parameterization of the G^* statistic.

7.3.5 Timeliness Metric

For evaluation a metric describing the timeliness of hotspot-detections with different temporal granularities is defined and calculated. Hotspots can be determined at different temporal granularities, such as monthly and yearly. It is obvious that hotspots reported monthly can provide information sooner than yearly

reports as outlined in Figure 7.2a. This can be formally defined as a timeliness metric T , which allows the comparison of two methods that use different time granularities. Let F be a method, which uses the finer-grained granularity—in the latter evaluation a monthly reporting period. Let C be a method, which uses the coarser-grained granularity—in the latter evaluation a reporting period of ten months. Each method delivers a tensor of measurements, which has two spatial dimensions and one time dimension. A single cell in the tensor F is denoted as $F_{x,y}^{(t)}$, where t is the time coordinate, x and y are space coordinates. In the evaluation is $t \in \{0, -1, \dots, -9\}$. The tensor from C is a simple matrix since it represents only a single time point. Therefore, the notation $C_{x,y}$ is used for a single cell. The timeliness metric T is computed according to Equation 7.5

$$\begin{aligned} T(C, F^{(0)}, F^{(-1)}, F^{(-2)}, \dots) &= \frac{\text{Score}}{\text{Norm}} \\ &= \frac{\sum_{\forall x,y} S(C_{x,y}, F_{x,y}^{(0)}, F_{x,y}^{(-1)}, F_{x,y}^{(-2)}, \dots)}{\sum_{\forall x,y} A(C_{x,y}, F_{x,y}^{(0)}, F_{x,y}^{(-1)}, F_{x,y}^{(-2)}, \dots)}, \end{aligned} \quad (7.5)$$

as a normalized sum of penalty scores S from all pixels where the coarse grained model agrees with the majority of the fine-grained models. See also Figures 7.2b and 7.2c. The score S and the agreement factor A for a single pixel are defined by Equations 7.6 and 7.7

$$S(c, f^{(0)}, f^{(-1)}, f^{(-2)}, \dots) = \begin{cases} \min\{i : f^{(i)} = m\}, & \text{if } c = m \\ 0, & \text{otherwise} \end{cases}, \quad (7.6)$$

$$A(c, f^{(0)}, f^{(-1)}, f^{(-2)}, \dots) = \begin{cases} 1, & \text{if } c = m \\ 0, & \text{otherwise} \end{cases}, \quad (7.7)$$

where $m = \text{median}(f^{(0)}, f^{(-1)}, f^{(-2)}, \dots)$.

7.3.6 Scenario Setup

A twofold approach is followed regarding the evaluation of the designed smart service. On the one hand, analytical methods are selected to determine whether the specified requirement of an increase in the temporal resolution are fulfilled. On the other hand, a descriptive evaluation is performed for investigating the services' value to road authorities, the intermediary and the crowd.

Input Data and Preprocessing

For the analytical evaluation to determine whether frequent decision support can be provided, a hotspot analysis is applied to a crowd-sensed road condition data

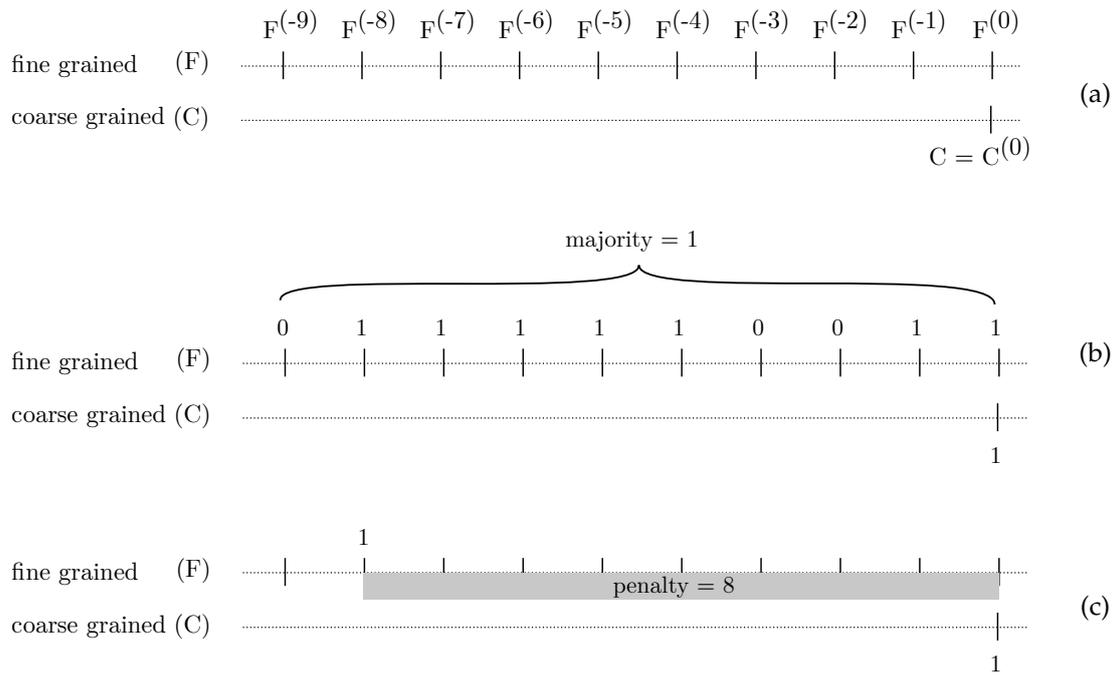


Figure 7.2: Overview of the determination of the timeliness metric T based on a comparison between hotspots on temporally fine and coarse granular data. Figure 7.2a shows the different temporal inspection granularities. Figure 7.2b depicts how hotspots based on temporal fine-granular inspections are aggregated by majority voting. Figure 7.2c illustrates a later detection of the hotspot when coarse grained inspections are performed. A penalty is taken into account when determining the timeliness metric T .

set obtained from BumpRecorder Co., Ltd. (<http://www.bumprecorder.com/en>) covering the geographical region of Aizuwakamatsu in Japan (Yagi, 2014). The data was collected using accelerometer sensors of smartphones, which were attached to the dashboard of the car. The data set contains already derived IRI values processed through predictive analytics. Even though the process applied for the IRI calculation is an intellectual property of BumpRecorder it can be expected to lead to similar results to those of the former chapters. However, it is not to be assumed that automatic calibration has been carried out or that measurements from multiple drives have been combined. The data set is chosen since it covers a large region and includes multiple measurements. It also covers several months of the year 2016, which allows for the investigation of the temporal evolution of hotspots.

The data is preprocessed by assigning each single measurement to a cell of a spatial grid with an edge length of 22 m. A measured instance is characterized by longitude, latitude and time at which the road condition was measured. According to the temporal dimension a period from January to October 2016 is covered. The data set consists of 1 443 632 single measurements. On average 2.35 instances

per grid cell were measured each week. The most frequented roads—third quartile of aggregated instances per cell—were measured 4.68 times a week.

Getis Ord G^*

The IRI estimations are aggregated over the whole time span and a hotspot analysis based on the G^* statistic is applied. The Manhattan distance is chosen to calculate $a_{i,j}$ for adjacent cells and the convolution distance is set to $a^\theta = 2$. The spatial weights are set to $w_{i,j} = 1/a_{i,j}$ for $i \neq j$ and $w_{i,i} = 2$.

Next, the data is aggregated on a temporal basis. To do so, a monthly level is selected as it provides both, high data consistency and sufficient measurements per time period. The amount of measurements in the whole period examined was stable. To demonstrate the feasibility of determining changing patterns in the hotspot distribution, three different time periods are analyzed additionally. In a first step, the monthly subsamples are filtered for similar cells in order to have a consistent data base. Then, identified hotspots are compared between the different time slots. January, May and September of 2016 are selected as they are equally distributed on a temporal axis. Considering the union of the hotspots from all three subsamples results in $N = 20\,184$ grid cells, which are considered for further analysis. In terms of their intersection, 819 common hotspots were identified, which accounts for 4 % of the considered road condition cells.

7.4 Results

This section answers Research Question 5. The results of the hotspot analyses at different temporal granularities are shown. In addition, the result of the timeliness metric applied to the hotspots is provided and interpreted. Furthermore, the utility for the involved parties is demonstrated by a descriptive evaluation.

7.4.1 Frequent Decision Support

Figure 7.3 provides the resulting map of identified spatial hotspots. Longitudinal and latitudinal coordinates define the geographical location of the hotspots. Dark grey indicates a higher G^* value and thus, a cluster of rough road segments. The number of unique cells in the spatial grid with road condition data available is $N = 69\,689$, from which 6 999 cells, or 10.04 % of the considered road network, are identified as hotspots on a 95 % confidence level. Considering a 99 % confidence level, 3 893 cells, which account for 5.59 % of the examined road network, are identified as hotspots. Figure 7.4 provides a more detailed insight into the result by displaying the hotspots in the city center of Aikuwakamatsu.

The spatial distribution of the common hotspots over all three time slots is presented in Figure 7.5a. Hotspots that are exclusively detected in a specific time

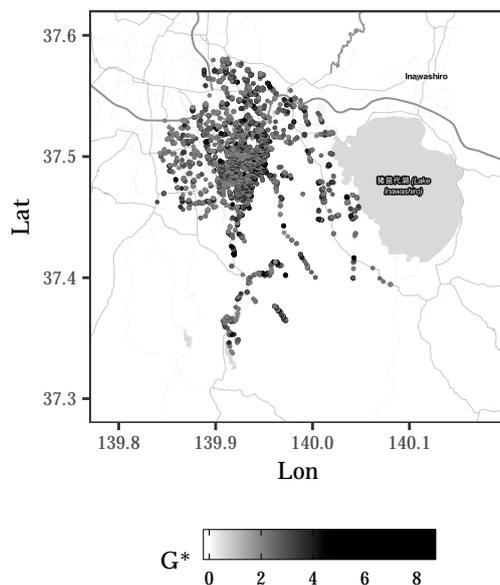


Figure 7.3: Spatial distribution of hotspots on a 95% confidence level—region of Aizuwakamatsu, Japan. Dark grey points indicate a higher G^* value and thus, a cluster of rough road segments.

slots are provided in the Figures 7.5b, 7.5c and 7.5d. It can be observed that in the month of May the amount of additionally identified hotspots is lower ($N = 131$) than in January ($N = 378$) and September ($N = 378$). Thus, the results show that it is possible to determine changing patterns in road condition on a monthly base. This fulfills the need for frequent road condition monitoring of road authorities and road users. Additionally, with an increasing amount of crowd-sensed data, the same method can be applied to provide significant road condition hotspots even more frequently leading to a real-time road condition monitoring.

Since, the fine grained data correspond to monthly aggregates, and the coarse grained data corresponds to the aggregation over ten months, the highest score for a pixel according to Equation 7.6 is 9. Applying Equation 7.5 to the data set results in $T = 8.37$. Thus, more frequent measurements result in hotspots being discovered on average 8.37 months earlier than with coarse granular measurement intervals. This provides a quantitative answer to Research Question 5 a), how existing services can be integrated by a new intermediary for providing a crowd-based road condition monitoring service allowing for a frequent and accurate support in decision making. As for 69.3% of all hotspots the hotspot assessment from coarse and fine granular aggregation coincide, high consistency of hotspot detection can be stated.

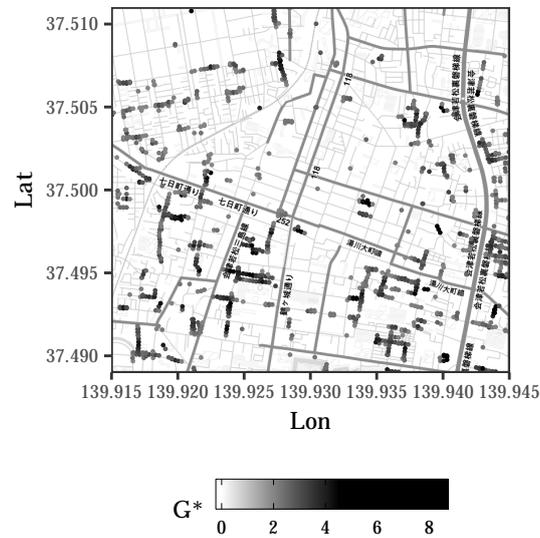
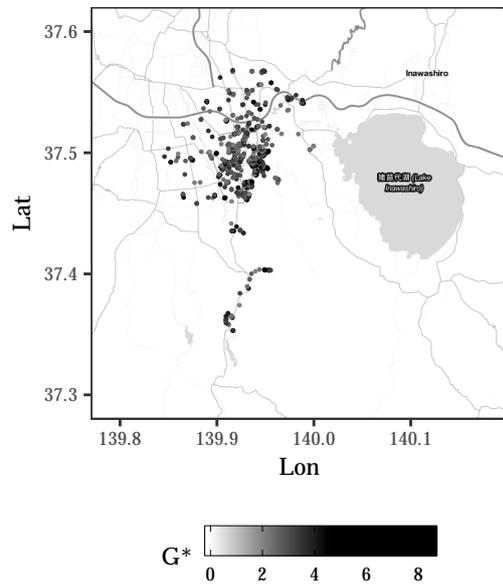


Figure 7.4: Spatial distribution of hotspots on a 95 % confidence level—city center of Aizuwakamatsu, Japan. Dark grey points indicate a higher G^* value and thus, a cluster of rough road segments.

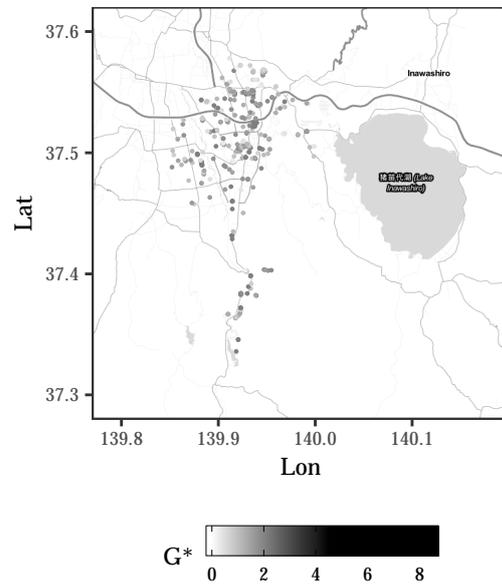
7.4.2 Utility to Authorities, Intermediary and Participants

The stakeholders of the smart service and its utility to them is inherently of importance (Allmendinger and Lombreglia, 2005). In order to answer the Research Question 5 b), beside the analytical evaluation, the service's utility is demonstrated considering the scheduling of road maintenance tasks as a concrete scenario. By expounding the utility for road authorities, the necessity of the smart road condition monitoring service and likewise the monetary potential for an intermediary providing this service is demonstrated. Finally, the utility for the crowd is presented as an additional enabler of the service.

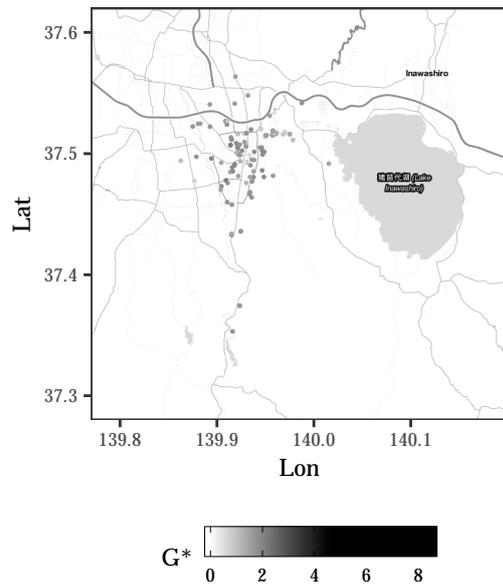
As mentioned, road authorities require accurate information about the road condition for efficiently scheduling maintenance actions. The goal is to find an optimal maintenance strategy for minimizing both, the costs for operation and maintenance actions and the costs for gathering information about the road condition (Watanatada et al., 1987). It is at the discretion of the authorities to decide when to perform which maintenance action and when to perform which type of road inspection—respectively when to purchase which type of road condition information. Spending more money for road inspections allows for a more efficient scheduling of maintenance actions and thus saves maintenance costs and vice versa. Formally written, this is a problem of finding an optimal policy in an accessible, stochastic environment with a known transition model. The envi-



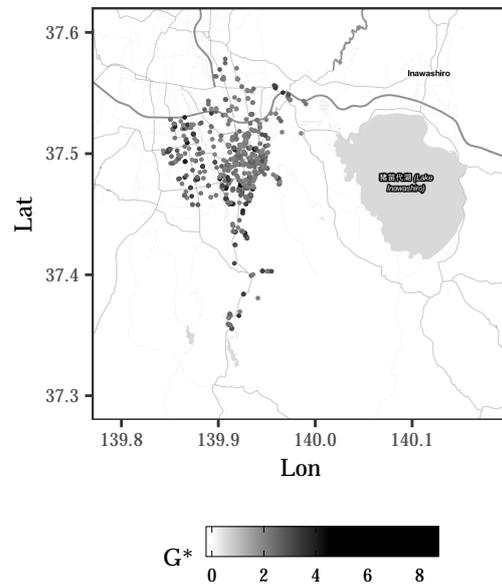
(a) Distribution of common hotspots.



(b) Distribution of exclusive hotspots 01/2016.



(c) Distribution of exclusive hotspots 05/2016.



(d) Distribution of exclusive hotspots 09/2016.

Figure 7.5: Variation of hotspots over time. In addition to showing the hotspots, which are present in all three time slots in Figure 7.5a, the hotspots that exclusively occur in a single time slot are shown in Figures 7.5b, 7.5c and 7.5d.

ronment is accessible since it can be observed by inspections and influenced by maintenance actions. It is stochastic since both, the deterioration over a time period and the rehabilitation through a maintenance action, is not always the same. Such an optimization problem can be described as a Markov decision process (MDP) (Puterman, 1994; Gao and Zhang, 2013). Solving this MDP addresses the trade-off between maintenance costs and inspection costs. The frequency of performing highly accurate road condition inspections is constrained by high costs. However, being provided with the smart service designed in this chapter, the MDP would be subject to less constraints since road condition information could be purchased at nearly arbitrary time intervals. Even though, this also incurs costs, it definitely results in a higher degree of freedom to act. This likely results in an MDP for which a more efficient and thus, cheaper policy can be found. The demonstrated cost reduction potential is a clear incentive for road authorities to consume the smart road condition monitoring service. At the same time, this encourages the provision of such a crowdsensing-based service by a new intermediary.

From the intermediary's point of view, a way to monetize the service would be to charge for spatio-temporal information packages. Thus, the road authority can for example purchase information about the condition of rural roads in their administrated road network for the last year on a monthly granularity. The utility of the service depends on the intervals the road authority would purchase road condition information. The pricing regime should reflect these different utilities to the customers by providing discounts if information packages are purchased more frequently. Assuming an appropriate spatio-temporal information coverage, such a value-based pricing regime is feasible since information is an intangible asset (Hand and Lev, 2003).

In addition to the road authorities and the intermediary, utility for the crowd must be ensured, as the crowd as a data provider is of fundamental importance for the crowdsensing-based road condition monitoring service. Creating utility for the crowd and thus achieving the willingness to install the requisite software and to provide the gathered data can be supported by different concepts. Zhang et al. (2016) distinguish between entertainment, service, and money as possible incentives for mobile crowdsensing. Entertainment as an incentive aims at ensuring that the sensing itself is not understood as a mere task, but that the participants enjoy taking part. In the sense of gamification, a common approach would be to award points corresponding to the distance traveled, enabling a competition among the participants. Using a service as an incentive would mean that the participants gain access to a service in return for collecting and providing data—most likely the same service made possible by the crowdsensing. Since the crowd consists of road users in this case, the participants are potential customers of the smart service anyway. Thus, there is no need for providing further examples for service-based incentives. Finally, another way to motivate the crowd to partici-

pate is through monetary incentives. Accordingly, payments can be made for the recorded kilometers or for the recording performance.

7.5 Conclusion

A smart road condition monitoring service based on crowd-sensed data was designed, introducing a new service provider into the value-added network. To do this, the needs of road authorities and road users with regard to timeliness of the road condition information were taken into account. The new service provider acts as an intermediary between the crowd and the road engineering companies as service providers and the road authorities and road users as service customers. The smart service is described in form of a service map. The approach was evaluated by implementing spatio-temporal hotspot analyses on crowd-sensed data for exemplifying the prescriptive analytics service. Thus, the timeliness of the service was demonstrated by an analytical evaluation. The results show that, given the new intermediary, road authorities can be served with both, accurate and frequent road condition information. The benefits for the road authorities, the intermediary and the crowd are demonstrated by a descriptive evaluation.

Despite demonstrating the utility of the smart service through evaluations, there are also limitations. The designed framework is proposed to be generalizable. However, deploying other sensors, such as cameras and microphones, and considering other road condition metrics than the IRI has to be evaluated with regard to its feasibility. In this chapter, the hotspot analyses were all performed on the same area. However, if hotspot analyses are to be compared across different areas, the instability of the G^* statistic must be taken into account, as described by Bruns and Simko (2017). Another issue concerns the evaluation of the utility. Even though, the utility of the integrated service was demonstrated by a descriptive evaluation, an extended summative evaluation that quantifies the utility is necessary. Thus, the degree of utility especially for the customers should be determined. Therefore, the following Part III of the thesis at hand focuses on these economic utilities. Concrete scenarios are defined to assess the utility for road users and road authorities.

Part III

Economic Assessment

Chapter 8

Road User Side Assessment

The quality of the road has an impact on road users. With respect to the longitudinal road roughness, this is mainly because of the physical effect of the surface's waviness to the vehicle and to its components. Thereby, vehicle operating costs (VOC), such as expenses for fuel and wear, depend on the road roughness (Tan et al., 2012). Thus, being provided with accurate and frequent information about the road surface condition—as it is possible with a crowdsensing-based monitoring—is beneficial for road users. Given a certain origin and destination, the road roughness depending VOC can be considered when choosing eligible road links from the complete road network. To assess the economic value of a crowdsensing-based road monitoring from an user side perspective this chapter quantifies the monetary impact of using road roughness information for path planning. Using a crowd-based data source and vehicle cost models a sensitivity analysis is performed to investigate the monetary implications on vehicle owners. The results are presented as a collection of trade-off matrices showing potential yearly cost savings for different vehicle types and road roughness levels. Moreover, the dependency between fuel price and overall cost savings is investigated. Although the cost savings depend on the vehicle type and the fuel costs, the results show that the main factor is the average baseline international roughness index (IRI) of the considered road network. In particular, vehicle owners benefit from rerouting to a smoother road in regions with a baseline road roughness at least $IRI \sim 4 \text{ m/km}$. Parts of this chapter are adapted from a former publication: Laubis, K., V. Simko, and A. Schuller (2016b). Crowd Sensing of Road Conditions and its Monetary Implications on Vehicle Navigation. In *Proceedings of the IEEE International Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress*, pp. 833–840.

8.1 Introduction

Due to ever increasing traffic flows, the quality of roads in urban and suburban areas is decreasing. Commuters are exposed to higher costs of individual transportation either due to the additional time spent in traffic jams or due to in-

tensified utilization of their vehicles. Factors contributing to the overall costs of individual transportation are for example fuel costs, vehicles maintenance costs and travel time costs. Hereby, the road condition and its influence on the vehicle performance and efficiency plays an important role. As described in Section 2.3, the VOCs are affected by the road roughness. Driving on rougher roads causes a higher wear of the vehicle and also leads to a higher fuel consumption. The longitudinal road roughness directly effects both, the frequency of repairs and the fuel consumption and tire wear. Accordingly, the costs for the repair of broken parts and for fuel and tires depend on the road condition. The presence of accurate and up-to-date road condition information—as it becomes possible by a crowdsensing-based approach as presented in Part II—can play an important role in crowd-based applications, such as considering the road condition while routing vehicles. Such a crowd-based routing approach has the potential to reduce transportation costs of the vehicle owners. The outline of this crowdsensing-based vehicle navigation approach is illustrated in Figure 8.1.

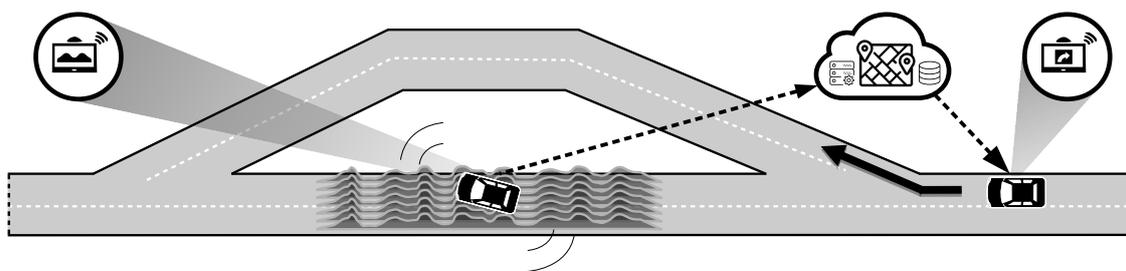


Figure 8.1: Outline of the crowdsensing-based vehicle navigation approach. For reducing VOCs, the following vehicle is redirected because of rough road segments ahead. Current road condition information can be provided applying the crowdsensing-based monitoring approach described in Part II.

The road condition of a certain road link is determined by a single vehicle or as an aggregate of the measurements from multiple vehicles. According to the procedure described in the previous part a roughness metric, such as the IRI, can be determined in nearly real-time. Following vehicles are allowed to use this information for choosing a preferable route. Even though, monetary metrics do not directly reflect the utility or disutility to persons, VOCs are chosen to reflect the preference of a route. This allows for an evaluation for the road users' side without the need for deriving an utility function. The driver of the following vehicle is thus assumed to reduce its VOCs.

The decision, whether a detour should be taken based on additional information, depends on both, the distances and the road roughness of the alternative routes. Both affect the VOCs. This results in a trade-off between the additional distance that has to be driven for thus being able to benefit from a smoother road. Assuming road users have the objective to minimize their VOCs, it is of interest,

what the actual cost savings are when following such an approach. These potential savings are likely to depend on the road network's overall condition, the vehicle characteristics and the fuel costs.

To determine the cost savings potential, a trade-off analysis is performed in this chapter. Both effects are taken into account, the additional distance travelled and the aspect of being exposed to less rough roads, which affects the cost of fuel consumption, tire consumption and repairs. Thus, it is possible to determine under which circumstances a detour is economically profitable for a road user and how much costs can be saved. To determine the road condition depending costs the VOC models introduced in Section 2.3 are considered. Since all the considered costs components vary depending on the vehicle type the analyses are performed for two different vehicle types—cars and light trucks. For each vehicle type an individual cost model is considered. Next to the vehicle type, external factors, such as the fuel price, are likely to vary for different regions and in time. Strong fluctuations in the fuel price and ongoing technological changes, such as the increased adoption of electric vehicles, stress the need to reflect this in the analyses. Therefore, a sensitivity analysis is described in this chapter for addressing the uncertainty regarding the fuel price. These evaluations provide answers to Research Question 6: (a) What are the potential individual cost savings per road user that result from adapted vehicle routing based on the improvement of the overall IRI score per year and for which road conditions is the rerouting monetarily feasible? (b) How sensitive are the results to different vehicle types and different fuel prices?

The remainder of this chapter is structured as follows: first, the work related to vehicle navigation is summarized. The core part of the chapter follows, where all the important assumptions, parameters, concepts and experimental results are explained. The results are then presented before the chapter concludes by summarizing the results, discussing the limitations and providing an outlook.

8.2 Path-Finding in Road Networks

There is a substantial body of work in the area of routing algorithms for finding the shortest path between two nodes in a graph. To name the most prominent, Dijkstra (Dijkstra, 1959) and A* (Hart et al., 1968) are mentioned. In automotive navigation systems, fast path searching requires pre-computation of the road network (Jun et al., 2008). Beside static approaches, there are also dynamic approaches for considering real-time information, such as disturbances in the traffic flow (Zhao, 1997; Deflorio, 2003). Next to just considering a single criterion, there are also approaches for considering multiple objectives in parallel, such as finding a pareto-optimal path regarding distance and time (Fujimura, 1996). An example of an open source routing application is given in (Luxen and Vetter, 2011). It is a state-of-the-art routing engine for determining the shortest path in road

networks. It can determine solutions that span across continents within milliseconds. It is designed to use the Open Street Map (OSM) data.

In navigation systems, it is common to reduce the cumulative weights of the edges. The weights could represent distance, time, safety, energy consumption, greenhouse gas emissions, etc. related to road segments (Ben-Akiva et al., 1991; Ericsson et al., 2006). However, there are no approaches that consider road user costs (RUCs) caused by road roughness for finding a cost efficient path. In addition, there are no approaches considering crowdsensing-based information of the road roughness for determining the path with the lowest VOCs.

8.3 Research Design

This section defines concepts that are crucial to examine the trade-off between additional distances travelled and the savings that result therefrom when driving on smoother roads. First, theoretical models are formally defined before scenarios are described to parametrize the models.

8.3.1 Cost and Trade-Off Matrices

First, a VOC model VC^c is defined. It is specific to a vehicle type c and can be described by the tuple in Equation 8.1

$$VC^c = (IRI, Tire, Repair, Fuel), \quad (8.1)$$

where

- $IRI \subseteq \mathbb{R}^+$ is a set representing IRI values in units m/km. It is the subset of those values for which the corresponding VOC are determined.
- $Tire : (IRI, \mathbb{R}^+) \mapsto \mathbb{R}^+$ is a function that assigns tire costs—for all wheels—of a vehicle for a given pair (iri, d) ; $iri \in IRI$ and $d \in \mathbb{R}^+$ is the distance traveled in km. For convenience, the notation $Tire(iri)$ is used for $Tire(iri, 1)$.
- $Repair : (IRI, \mathbb{R}^+) \mapsto \mathbb{R}^+$ is a function that assigns repair costs—parts and labor—of a vehicle for a given pair (iri, d) ; $iri \in IRI$ and $d \in \mathbb{R}^+$ is the distance traveled in km. The notation $Repair(iri)$ is used for $Repair(iri, 1)$.
- $Fuel : (IRI, \mathbb{R}^+) \mapsto \mathbb{R}^+$ is a function that assigns fuel costs of a vehicle for a given pair (iri, d) ; $iri \in IRI$ and $d \in \mathbb{R}^+$ is the distance traveled in km. The notation $Fuel(iri)$ is used for $Fuel(iri, 1)$.

Based on the VOC models VC^c absolute cost matrices AC^c can be determined. This is again done specifically for a vehicle type c . The matrix AC^c is defined in Equation 8.2

$$AC^c = \begin{bmatrix} h(d, 1) & \dots & h(d, iri) \\ h(d+1, 1) & \dots & h(d+1, iri) \\ \vdots & \ddots & \vdots \\ h(d'-1, 1) & \dots & h(d'-1, iri) \\ h(d', 1) & \dots & h(d', iri) \end{bmatrix}, \quad (8.2)$$

where

- $d, d' \in \mathbb{R}^+$, $d < d'$ be parameters representing distanced traveled by the vehicle and
- $iri \in IRI$ is a specific upper baseline.
- The function h is a helper function defined in Equation 8.3

$$h(a, b) = a \cdot (Tire(b) + Repair(b) + Fuel(b)). \quad (8.3)$$

It should be noted that rows of AC^c represent a range of distances (d, \dots, d') traveled by the vehicle. Columns represent IRI values from 1 up to a specified upper bound. Thus, it holds $i \in IRI : 1 \leq i \leq iri$.

Given the absolute costs matrix AC^c with parameters d , d' and iri as defined above, the trade-off between the additional distance travelled and the thus possible avoidance of rough roads can be quantified through the cost differences of the alternatives. $AC_{1,iri}^c$ is a specific cell from matrix AC^c at *row* = 1 and *column* = iri . Then, a trade off matrix TO can be defined by Equation 8.4

$$TO^c = AC_{1,iri}^c - AC^c. \quad (8.4)$$

The trade off matrix TO^c represents cost savings comparing to the situation when the vehicle traveled distance d on a road with IRI value of iri . It should be noted that the row index 1 in the matrix AC^c corresponds to distance d —the beginning of the given interval (d, \dots, d') .

8.3.2 Savings Potential

The trade-off matrix TO^c contains cost savings for every combination of distance and IRI differences. Each cell $TO_{i,j}^c$ of the matrix represents the cost difference between two alternative routes. This allows a detailed investigation of the trade-off. In order to allow a more compact representation of the entire savings potential and to enable a direct comparison between multiple trade-off matrices, a corre-

sponding metric is derived below. Let TO^{c+} be a matrix of all positive cells in the trade-off matrix TO^c as defined in Equation 8.5

$$TO_{ij}^{c+} = \begin{cases} TO_{ij}^c & \text{if } TO_{ij}^c > 0 \\ 0, & \text{otherwise} \end{cases}. \quad (8.5)$$

The savings potential $SP^c \in \mathbb{R}^+$ is then defined as a sum of all positive cells in the trade-off matrix TO^c as Equation 8.6 shows

$$SP^c = \sum_{i,j} TO_{ij}^{c+}. \quad (8.6)$$

This way, an useful metric is defined that allows to compare multiple trade-off matrices. Only the positive values are considered, because they represent feasible cost savings.

8.3.3 Scenario Setup

The particular instances of the definitions described above depend on the parameters selected. Especially the fuel price is an integral parameter and is therefore explicitly addressed in Research Question 6. For this, scenarios SC are defined in Equation 8.7 to investigate the sensitivity of the savings potential to the fuel price

$$SC : (c, iri, fm) \mapsto (VC^s, AC^s, TO^s, SP^s), \quad (8.7)$$

where

- $fm \in \mathbb{R}^+$ is a multiplier for fuel costs,
- $VC^s = (IRI, Tire, Repair, fm \cdot Fuel)$ is a similar model than VC^c with the *Fuel* component re-scaled using the fm multiplier,
- AC^s , TO^s and SP^s are similar to AC^c , TO^c and SP^c , but these are derived from VC^s instead of VC^c .

The parameters c , iri and $VC^c = (IRI, Tire, Repair, Fuel)$ are, as already defined, the vehicle type, the IRI baseline—which is attempted to be reduced by a detour—and the VOC model. It should be noted that $fm = 0$ represents an extreme scenario, in which fuel costs do not play any role in the total costs computation. Vice versa it should be noted that $fm = 1$ implies $VC^c = VC^s$. Figure 8.2 depicts how the definitions contribute to the analysis and results.

8.3.4 Parametrization

This section explains all assumptions and input parameters important for the analysis.

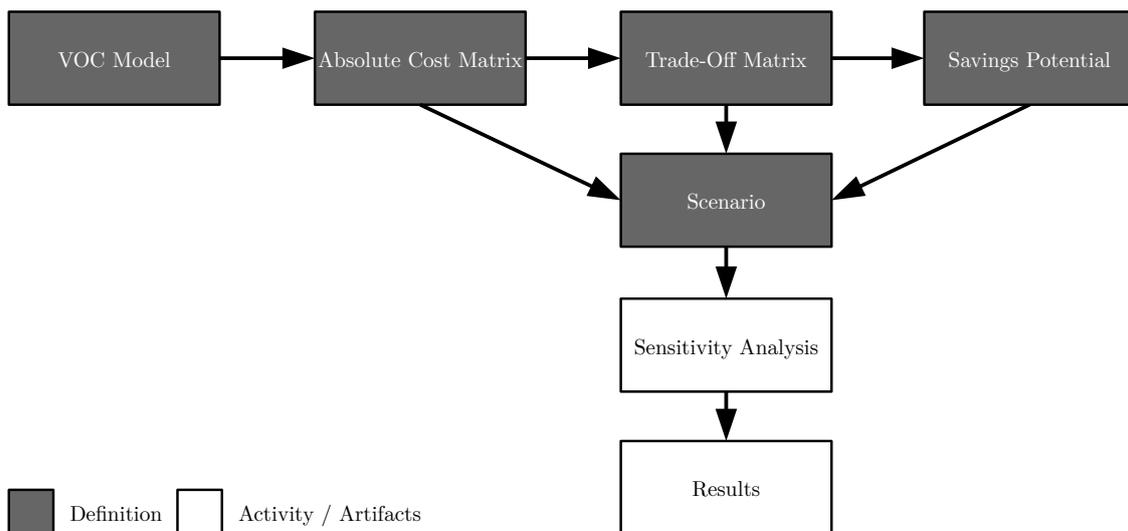


Figure 8.2: Overview of the definitions for the road user side assessment. Conceptual model of the sensitivity analysis showing dependencies between definitions. A differentiation into definitions and activities / artifacts is given.

Vehicle Operating Cost Model

As described in Section 2.3, the World Bank developed Highway Development and Management (HDM) models to provide decision support of road authorities. These models also quantify the relationship between the road roughness—in the sense of the IRI—and the VOCs. For the analysis in this chapter, the National Cooperative Highway Research Program (NCHRP) cost model provided by Chatti and Zaabar (2012) is applied. This model is chosen because it is based on the latest version of the well established HDM-4 model (Kerali et al., 2006) and provides up-to-date calibrations for the USA as a developed country. The NCHRP model considers the VOC components fuel consumption, tire wear and reparation costs. Figure 8.3 gives once more an overview of all effects that are caused by road usage—the road user effects (RUE). Some of these are not considered in the NCHRP model and likewise not in this analysis. The fuel costs and tire wear costs are significantly affected by the road roughness due to the rolling resistance. The reparation costs are determined empirically and likewise depend on road roughness.

Parameter Selection

The basic idea of the trade-off analysis is to compare baseline paths with multiple corresponding alternative paths. The alternative paths—that can be chosen by an alternative routing—vary in road roughness and travel distance. Thus, there are two dimensions to vary in the analysis—the decrease in IRI and the additional

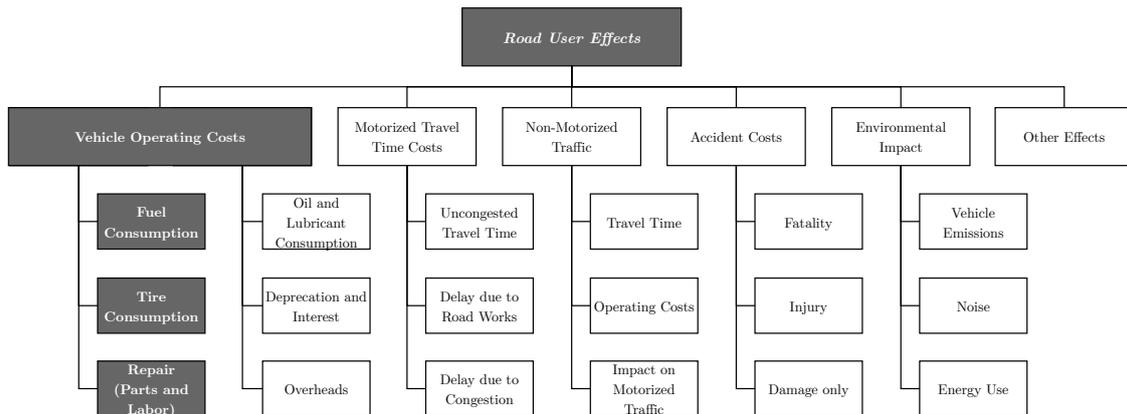


Figure 8.3: Components of RUEs (Bennett and Greenwood, 2003), which are included in the road user side assessment. The highlighted components are considered in the NCHRP model and likewise in the road user side assessment conducted in this chapter.

distance traveled due to rerouting. For the baseline paths, IRI values from 3 up to 6 are assumed. The lower bound is chosen by considering the 75 % quantile from the actual IRI values in Gaevleborg, Sweden shown in Figure 2.4. The upper limit is chosen by considering the upper IRI values in the survey of the NCHRP model (Chatti and Zaabar, 2012). For each of these baseline IRI values, the difference in VOCs are examined by choosing an alternative path with a roughness reduction of the IRI values down to 1.

For the analysis, the 88 km/h speed level and two of the vehicle types—medium car and light truck are chosen from the NCHRP model. The medium car is a "Mitsubishi Galant" with gas engine from 2008 with a tare weight of 1.46 t and a tire diameter of 0.38 m. The light truck is a "GMC W4500" with gas engine from 2006 with a tare weight of 3.7 t, a maximum allowable load of 2.9 t, a GVW of 6.6 t, a load weight of 2.8 t and a tire diameter of 0.4 m. Table 8.1 contains the unit costs for the considered VOC components for both vehicle types. The values were chosen according to the NCHRP survey. Table 8.2 shows the factors given by the NCHRP survey of VOC for both vehicle types and the three cost components relative to a smooth road—given an IRI of 1. Combining the unit values from Table 8.1 and the factors from Table 8.2 Table 8.3 was calculated. It shows the VOCs per km, which serve as the basis for the analysis.

8.4 Results

Using the definitions above, all the scenarios can be specified and their results can be presented. First, the cost saving potentials for the two vehicle types medium car and light truck are presented. Then the results of the sensitivity analysis regarding the fuel price follow.

Table 8.1: VOC per unit and vehicle type. Repair costs are exemplary for an IRI of 1. The costs are from Chatti and Zaabar (2012) and are translated from US Dollar into Euro at the closing exchange rate 0.90086 on 02.11.2016.

	Fuel costs [€/L]	Tire costs [€/tire]	Repair costs [€/km]
Medium car	0.86	90	0.017
Light truck	0.86	158	0.026

Table 8.2: Factors describing the effect of the IRI on VOC. The factors are provided for two vehicle types and are to be considered relative to a smooth road with an IRI of 1 (Chatti and Zaabar, 2012).

		IRI [m/km]										
		1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Car	Fuel	1.00	1.01	1.03	1.04	1.05	1.05	1.08	1.09	1.10	1.11	1.13
	Tires	1.00	1.00	1.01	1.01	1.02	1.03	1.03	1.03	1.04	1.04	1.05
	Repair	1.00	1.00	1.00	1.00	1.00	1.05	1.10	1.25	1.40	1.55	1.70
Truck	Fuel	1.00	1.00	1.01	1.01	1.02	1.02	1.03	1.03	1.04	1.04	1.05
	Tires	1.00	1.00	1.01	1.01	1.02	1.03	1.04	1.04	1.05	1.05	1.06
	Repair	1.00	1.00	1.00	1.00	1.00	1.10	1.20	1.45	1.70	1.95	2.20

Table 8.3: VOCs [cent/km] for two vehicle types and different road roughness. The costs are from Chatti and Zaabar (2012) and are translated from US Dollar into Euro at the closing exchange rate 0.90086 on 02.11.2016.

		IRI [m/km]										
		1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Car	Fuel	7.21	7.31	7.42	7.50	7.57	7.60	7.78	7.86	7.93	8.04	8.15
	Tires	0.50	0.50	0.51	0.51	0.51	0.52	0.52	0.52	0.52	0.53	0.53
	Repair	1.71	1.71	1.71	1.71	1.71	1.79	1.88	2.14	2.40	2.65	2.91
Truck	Fuel	15.58	15.66	15.74	15.82	15.89	15.97	16.05	16.13	16.21	16.29	16.36
	Tires	1.14	1.14	1.14	1.15	1.16	1.17	1.18	1.19	1.19	1.20	1.21
	Repair	2.61	2.61	2.61	2.61	2.61	2.87	3.13	3.79	4.44	5.10	5.75

8.4.1 Savings Potential

The scenarios start from the baseline $iri = 3$ —approximately the 75 % quantile—up to $iri = 6$ —approximately the 95 % quantile—, which is also the upper limit of the NCHRP model. For the vehicle type medium car mc the following scenarios are considered:

$$SC(mc,3,1), SC(mc,4,1), SC(mc,5,1), SC(mc,6,1) \\ SC(mc,3,0), SC(mc,4,0), SC(mc,5,0), SC(mc,6,0)$$

The result for the vehicle type mc are presented in form of several trade-off matrices in Figure 8.4. Similarly, for the vehicle type light truck lt the following scenarios are considered:

$$SC(lt,3,1), SC(lt,4,1), SC(lt,5,1), SC(lt,6,1) \\ SC(lt,3,0), SC(lt,4,0), SC(lt,5,0), SC(lt,6,0)$$

Again, the result are presented in form of several trade-off matrices, which are shown in Figure 8.5. With regard to Research Question 6 a), it is shown that individual cost savings per users can be achieved by considering road roughness in vehicle routing. However, higher overall savings were expected. For a medium car and a baseline $iri = 6$, absolute savings up to €518.63 per year can be achieved by being rerouted to smoother roads with the same overall distance and with an overall IRI of 1. This represents a quite optimistic scenario. Choosing a more realistic baseline $iri = 3$ an increase in distance up to 3 % is acceptable if hereby a reduction to an IRI of 1 is achieved. Vice versa, no routes with an additional distance should be chosen if just an IRI reduction to 2 would be achieved. The savings vary depending on the vehicle type. The trade-off matrices for light trucks show higher savings compared to the savings for medium cars in most cases. Just for the case of $iri = 3$ the medium car savings are higher. However, it has to be considered that the absolute VOC for light trucks are higher than for medium cars. The relative savings for light trucks may therefore not be higher than those for medium cars. Basically, the rerouting is most feasible for road networks which are in a bad overall condition. Thus, a reduction from $iri = 6$ has much higher effects than from a lower IRI baseline. In particular, the results indicate that there is a relevant cost saving potential for road networks with an IRI of 4 or higher, which may be the case of developing countries.

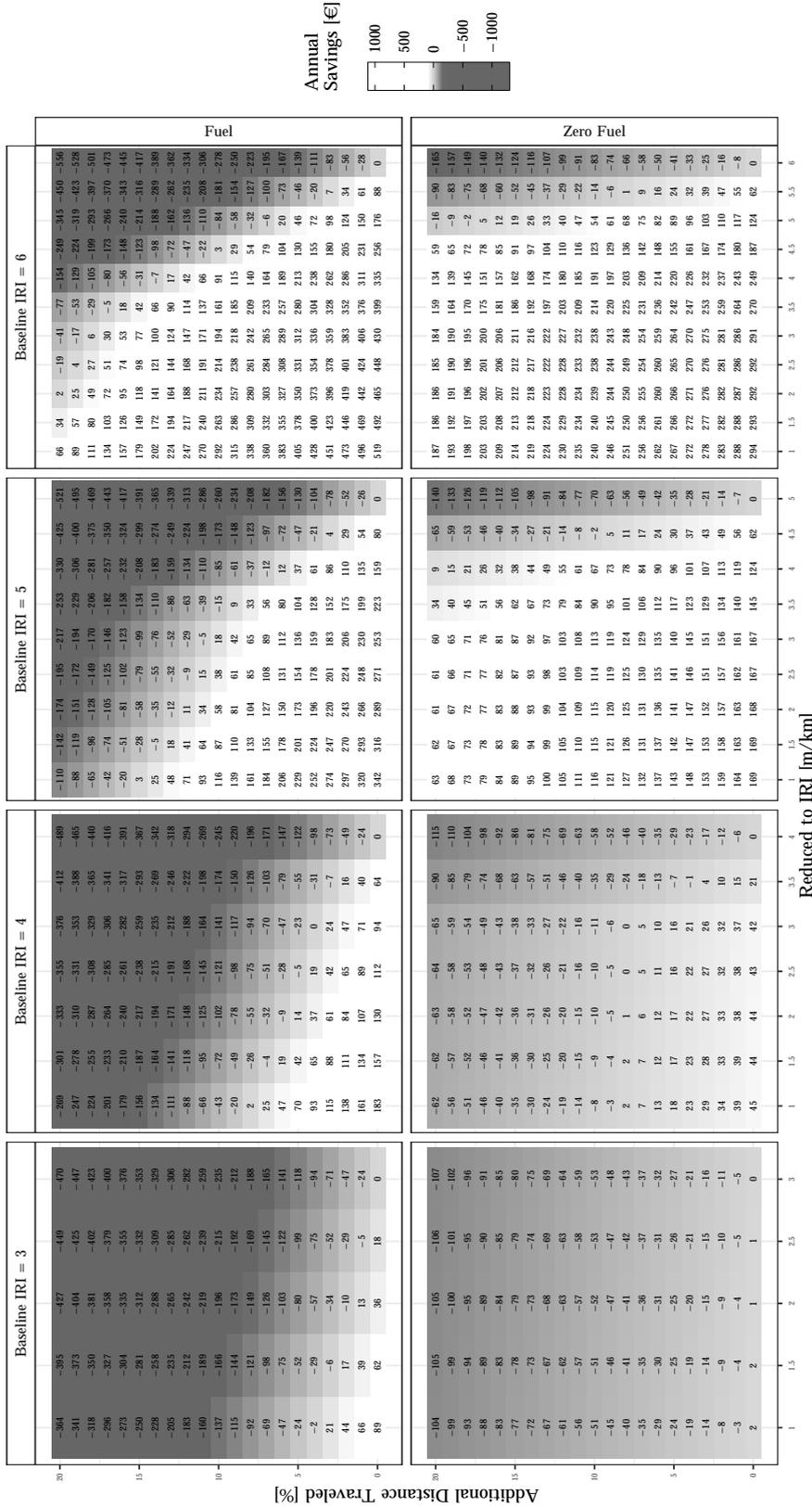


Figure 8.4: Trade-off matrices normalized to a year showing cost savings in Euro, computed for $IRI \in \{3, 4, 5, 6\}$ for a medium car driving at 88 km/h. An average distance of 24 000 km traveled per year is assumed. Additionally, trade-off matrices for a medium car without fuel consumption are shown.

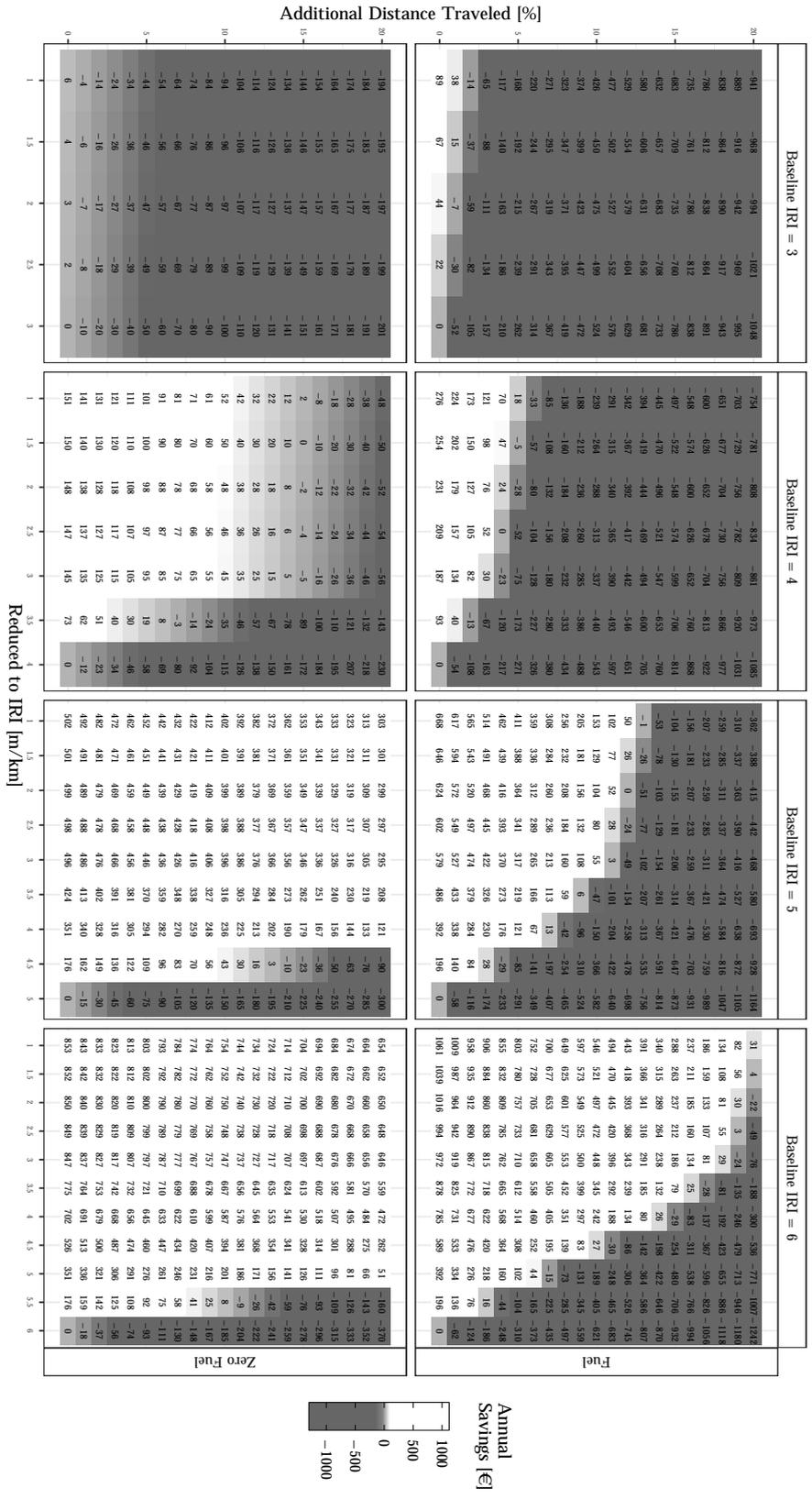


Figure 8.5: Trade-off matrices normalized to a year showing cost savings in Euro, computed for $IRI \in \{3, 4, 5, 6\}$ for a light truck driving at 88 km/h. An average distance of 24 000 km traveled per year is assumed. Additionally, trade-off matrices for a light truck without fuel consumption are shown.

8.4.2 Sensitivity to Fuel Price

In order to investigate the effects of the fuel price, the following scenarios are examined for the vehicle type medium car *mc*:

$$\begin{array}{lll}
 SC(mc, 6, 0), & SC(mc, 5.5, 0), & \dots SC(mc, 4, 0) \\
 SC(mc, 6, 0.1), & SC(mc, 5.5, 0.1), & \dots SC(mc, 4, 0.1) \\
 \vdots & \vdots & \ddots \vdots \\
 SC(mc, 6, 1.9), & SC(mc, 5.5, 1.9), & \dots SC(mc, 4, 1.9) \\
 SC(mc, 6, 2), & SC(mc, 5.5, 2), & \dots SC(mc, 4, 2)
 \end{array}$$

These additional scenarios are examined in order to take the vehicle type light truck *lt* into account:

$$\begin{array}{lll}
 SC(lt, 6, 0), & SC(lt, 5.5, 0), & \dots SC(lt, 4, 0) \\
 SC(lt, 6, 0.1), & SC(lt, 5.5, 0.1), & \dots SC(lt, 4, 0.1) \\
 \vdots & \vdots & \ddots \vdots \\
 SC(lt, 6, 1.9), & SC(lt, 5.5, 1.9), & \dots SC(lt, 4, 1.9) \\
 SC(lt, 6, 2), & SC(lt, 5.5, 2), & \dots SC(lt, 4, 2)
 \end{array}$$

The resulting saving potentials for both vehicle types are provided in Figure 8.6.

The purpose of these two diagrams is to show how the savings potential varies based on fuel price. Given these sensitivity analyses, Research Question 6 b) can be answered. The savings vary depending on the fuel price. The lower the fuel is, the higher are the savings potentials. In addition, the savings potentials differentiate more strongly between the different IRI baselines if the fuel price is low. This is an interesting finding for scenarios in which the driver does not have to pay for the fuel at all.

8.5 Conclusion

Given the annual savings for medium cars depicted in detail in Figure 8.4 and for light trucks in Figure 8.5 as well as the aggregated saving potentials relative to the fuel price in Figure 8.6, the Research Question 6 was answered. However, the procedure in this chapter is also subject to limitations. Different baseline IRI values are assumed for the analyses. Even though, they are based on real-world measurement from Gävleborg county in Sweden—and are therefore assumed to be realistic—a simulation-based investigation given an actual road network instead would lead to a higher external validity. Thus, it could be determined to what extent the avoidance of rough road segments is possible at all and which of the chosen baseline scenarios is most realistic. Furthermore, determining the cost savings potential by rerouting to roads with an IRI < 1 would be meaningful

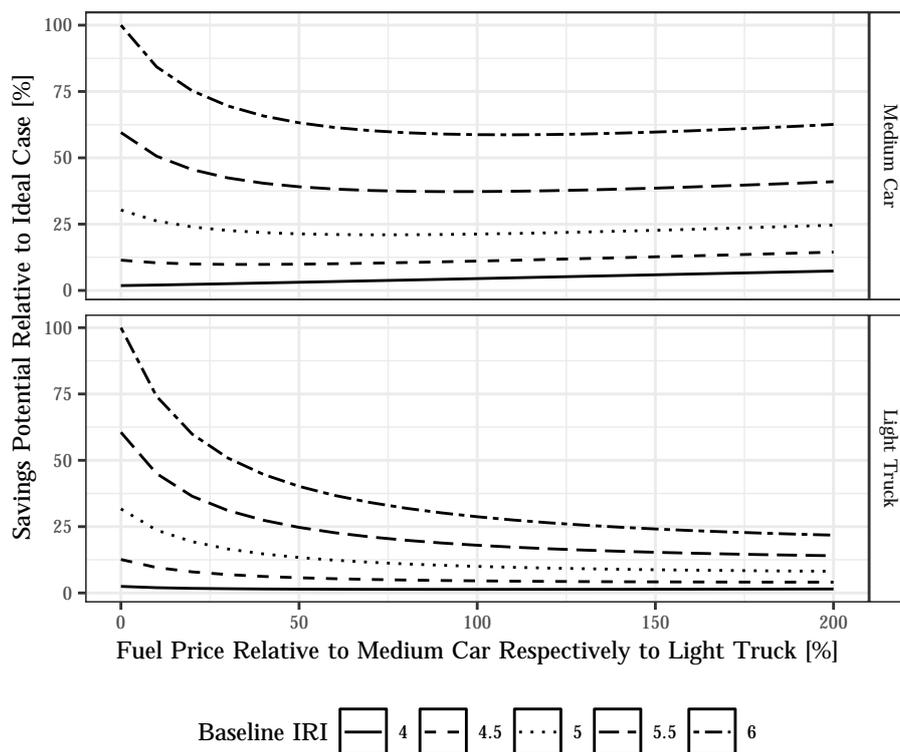


Figure 8.6: Savings potential depending on fuel price for medium cars and light trucks. Comparison of saving potentials from different baseline IRI and for different fuel prices. The savings potentials are shown relative to the ideal scenario. The ideal scenario is for both vehicle types if the maximum IRI is taken as a baseline and no costs for fuel are assumed.

as well. However, the survey performed within the NCHRP just reports costs related to $IRI \geq 1$. Next to this the time spent in addition because of the rerouting, different speed levels and further vehicle types can be considered to extend the investigation. The benefit of information on the current condition of roads—as it gets possible with a crowdsensing-based monitoring—was examined for road users in a concrete use case. However, the question remains what monetary benefit road authorities can derive from a crowd-based condition monitoring system. This question is referred to in the following chapter.

Chapter 9

Road Authority Side Assessment

Road authorities are responsible for maintaining a certain road network. From their perspective, knowing the current road condition is essential to schedule maintenance actions in an efficient and sustainable manner. For the federal road network in Germany, relatively expensive laser-based road inspections are scheduled every four years for serving this information demand. In future, they could be extended or even completely replaced with a crowdsensing-based approach. Crowdsensing-based road condition measurements are less accurate than laser-based ones. However, in comparison, a productive crowd-based system allows for much more frequent measurements. From a road authority's perspective the question, whether the lower accuracy of a crowd gathered data is redeemed by its timely data updates, has to be answered. Therefore, this chapter addresses the managerial implications of substituting or supplementing laser-based road condition monitoring by a crowdsensing approach. Road authorities have to establish strategies for maintaining their road network that not just minimizes the inspection and maintenance costs, but also the road user costs (RUC). Thus, the question is, how the inspection accuracy influences the long-term overall costs. These are comprised of the maintenance and user costs, such as vehicle's wear and traffic jams due to construction sites. Scenarios, which differ in the type of inspection and the frequencies these inspections are performed are defined in this chapter. For each scenario, a partially observable Markov decision process (POMDP) is applied for determining a cost minimal maintenance policy. A comparison of these resulting policies show that crowd-based inspections aggregated once per year can decrease total maintenance and RUC by 5.9% compared to laser-based inspection on a four-year basis. An approach, which combines both inspection types, reduces the costs by 6.98%.

9.1 Introduction

The road network is regarded as a valuable asset. Next to the decisions "when", "where" and "how" to construct new road links for expanding the network, also the existing infrastructure has to be maintained by road authorities. Besides considering maintenance costs, road authorities have to consider RUCs related to the

road condition (Ouyang and Madanat, 2004). A sustainable road maintenance requires information of the actual roads' conditions for coming up with efficient decisions about maintenance actions.

The crowdsensing-based monitoring presented in this thesis enables to supplement or even replace the current laser-based, highly accurate road condition inspections. The main potential of such a crowdsensing approach is the spatio-temporal coverage. Gathering and analyzing data in a centralized backend system would constitute only a small additional effort. Even though the measurements performed by individual cars are not as accurate as those from official inspection drives are, they are performed by many more cars and at a much higher rate. Figure 9.1 outlines the crowdsensing-based road maintenance, which is investigated in this chapter. It depicts the conventional approach of considering

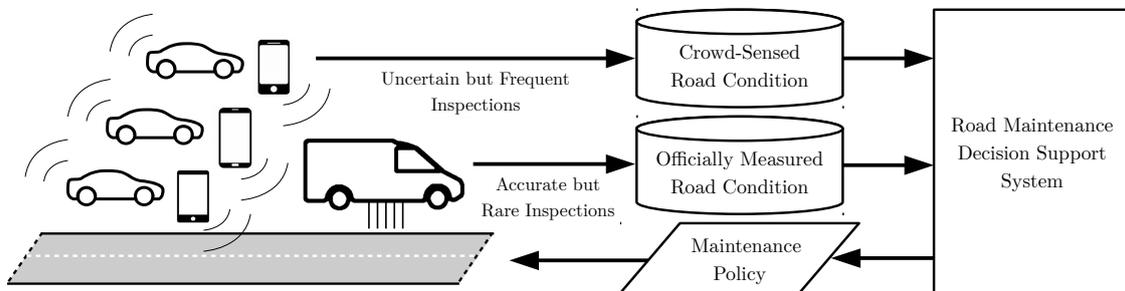


Figure 9.1: Outline of crowdsensing-based road maintenance. Road condition monitoring for supporting maintenance decisions based on official measurements with specially equipped inspection vehicles and a crowdsensing-based approach as substitute or supplement.

laser-based road condition inspections within a road maintenance decision support system and it indicates inspections based on smart technologies as a substitute or supplement. However, it has to be investigated whether crowd-sensed inspections alone or combined with current inspection cycles are beneficial from a managerial perspective. Thus, this chapter quantifies the monetary effect of a crowd-based road condition monitoring on the performance of maintenance policies. Accordingly, this chapter answers Research Question 7: (a) Given crowd-based road condition inspections compared to accurate laser-based inspections, what are the effects of different inspection accuracies and inspection frequencies on an optimal maintenance policy? (b) To what extent can maintenance costs and RUC be reduced, when combining crowd-based and laser-based inspections?

The chapter is structured as follows: in the next section, POMDPs and related Markov models are presented. This is followed by a description of the research design pursued in this chapter. Finally, the results are provided before the chapter concludes with a summary.

9.2 Markov Models in Maintenance Scheduling

Markov chains have been proved useful for modeling of stochastic processes (Revuz, 1984). The key characteristic of Markovian chains is the Markov property, which describes the fact that Markov chains are "memoryless". Thus, the conditional probability distribution of a system's state at time step $t + 1$ only depends on the present state at time t . This property allows for computing probabilities for state transitions and thus related costs. The basic model can be extended by including uncertainty about the system's state, what is called a hidden Markov model. Or it can be extended by allowing to influence the state transition probabilities by the agent's actions, what is called a Markov decision process (MDP). Furthermore, applying a POMDP, both can be modeled, uncertainty and control over state transitions. Table 9.1 provides an overview of the characteristics of the mentioned Markov models. A verbal differentiation is given below.

Table 9.1: Differentiation of Markov models. Differentiation is made along two criteria, the ability to control transitions due to actions and by observability of the actual state.

	Control over transitions	States fully observable
Markov chain	×	✓
Markov decision process	✓	✓
Hidden Markov model	×	×
Partially observable Markov decision process	✓	×

Hidden Markov models account for the fact that the information about a system's state is often imperfect. This is the case, for example, if the measurements could be subject to errors (Baum and Petrie, 1966). The uncertainty can be modeled by assuming an underlying Markov chain, which is unobservable—its states are hidden. However, observations are made which relate to the real states according to a predefined probability distribution. Using Bayesian updating, a probability distribution of the possible sequences of state transitions for a given sequence of actions and observations can be computed to define the most probable hidden state and reduce the uncertainty. This probability distribution is called the belief state and encapsulates all information about the past.

A MDP involves the possibility to perform certain actions, which are associated with different state transition probability matrices (Puterman, 1994). Assuming perfect information, policies for performing optimal actions—maximizing the expected rewards—can be determined by value iteration or policy iteration algorithms.

POMDPs combine both approaches (Kaelbling et al., 1998; Cassandra et al., 1994). Thus, it can be modeled that states are not completely observable and

actions can be taken to control the state transitions. A description of a POMDP is provided in the following section.

9.3 Research Design

This section addresses the components of a POMDP and approaches for solving it. This is followed by a description of the scenarios to be examined in this chapter.

9.3.1 Partially Observable Markov Decision Processes

A formal description of a POMDP is provided below to indicate the components and necessary parameters for its application in this chapter. A POMDP can be formally described as a tuple $(ST, AS, TR, RE, \Omega, OB, df)$, where

- ST is a finite set of states that are not completely observable,
- AS is a finite set of actions that can be taken to control the state transitions,
- TR is a probabilistic state-transition function $ST \times AS \mapsto \Pi(ST)$,
- RE is an immediate reward function $ST \times AS \mapsto \mathbb{R}$,
- Ω is a finite set of observations,
- OB is a probabilistic observation function $ST \times AS \mapsto \Pi(\Omega)$,
- $df \in [0, 1]$ is a discount factor.

The question to be answered by a POMDP is which action is optimal at a certain stage of the process given the belief state and the current observation. Therefore, the expected cumulated reward associated to each action is iteratively determined. Information on the history of observations and actions is inherently contained in the belief state, which is updated in each iteration of the process. The computational effort for solving a POMDP increases exponentially with the number of states, actions and observations.

9.3.2 Value Iteration

For a discrete state-space POMDP, there exist several algorithms that provide exact or approximate solutions, such as Enumeration (Sondik, 1971), Two-Pass (Sondik, 1971), Witness (Cassandra et al., 1994), Linear Support (Cheng, 1988) and Incremental Pruning (Cassandra et al., 1997; Zhang and Liu, 1996). All of them are based on an iterative forward-backward approach. The belief

state is calculated for a given time horizon and subsequently the optimal value function—that maximizes the expected reward associated to actions and states—is computed recursively.

9.3.3 Scenario Setup

To assess the relevance of crowdsensing-based road condition inspections to road authorities, three different scenarios are defined in this chapter for comparison. The scenarios differ in terms of the type and frequency of inspection—laser-based and crowd-based—performed. Laser-based inspections are performed at four year intervals as it is the case for the German federal road network. Crowd-based road condition measurements are gathered within one year and are aggregated at the end of each year for having a single robust crowd-based inspection. Combining these inspection types, following three scenarios result: *scenario 1*—laser-based inspection every fourth year; *scenario 2*—crowd-based inspection every year and *scenario 3*—laser-based inspection every fourth year and crowd-based inspections every remaining year. For each scenario a POMDP is applied to determine the optimal maintenance policy. Given these policies, a road authority knows in which year and at which road condition a certain maintenance decision should be performed for minimizing the expected total costs—maintenance costs and RUC—for a time horizon of 100 years. Its parametrization is described in the following section.

9.3.4 Parametrization

Using the POMDP definition from Section 9.3.1, the states ST reflect road conditions, the actions AS reflect different maintenance tasks. Transition probabilities TR describe the effects of maintenance tasks and road deterioration. The observation probabilities OB indicate the accuracy of the inspection methods. The immediate rewards RE consist of both cost components, maintenance and RUC. The states, actions, transition probabilities and rewards are defined in correspondence to the work of Gao and Zhang (2013). The observation probabilities for the crowd-based inspections are defined based on the findings in Chapter 3. All these parameters are defined and described below.

Road Condition States

For defining the road condition states $ST = \{s_1, s_2, s_3, s_4, s_5\}$ the ride quality index (RQI) is derived from the international roughness index (IRI) by using Equation 9.1 and splitting the resulting range into whole numbers, which facilitates classifying IRI values into discrete quality states

$$RQI = 6.122 - 1.963\sqrt{IRI}. \quad (9.1)$$

The states are shown in Table 9.2. A low IRI and a high RQI indicate a good road condition. Thus, s_1 stands for good condition, whereas s_5 stands for complete infrastructure failure.

Table 9.2: Road condition states ST are defined by whole numbers of the RQI (Gao and Zhang, 2013). Next to the RQI and the corresponding IRI ranges, each state is provided with representative RQI and IRI values, which are considered within the analysis.

	IRI range [m/km]	RQI range	Representative IRI [m/km]	Representative RQI
s_1	0.327 – 1.169	$4 \leq RQI \leq 5$	0.683	4.5
s_2	1.170 – 2.529	$3 \leq RQI < 4$	1.784	3.5
s_3	2.530 – 4.409	$2 \leq RQI < 3$	3.405	2.5
s_4	4.410 – 6.808	$1 \leq RQI < 2$	5.544	1.5
s_5	6.809 – 9.726	$0 \leq RQI < 1$	8.202	0.5

Maintenance Actions

There are three maintenance actions $AS = \{a_1, a_2, a_3\}$ that can be scheduled by road authorities. These are reconstruction a_1 , resurfacing a_2 and just doing minor tasks or nothing a_3 . Reconstructing a road segment is considered as resetting it in the best possible state s_1 . According to Ouyang and Madanat (2004), a resurfacing action on an asphalt road section improves the condition, as described in Equation 9.2

$$IRI_t - IRI_{t+1} = \frac{0.66 \cdot IRI_t}{7.15 \cdot IRI_t + 18.3} \delta. \quad (9.2)$$

Thus, the IRI after resurfacing IRI_{t+1} depends on the IRI before the maintenance task IRI_t and on the resurfacing thickness δ in mm. In this analysis, a thickness of 40 mm is chosen to keep comparability with the study from Gao and Zhang (2013). It is obvious that performing no maintenance action has no direct effect on the road condition.

Transition Probabilities

Transition probabilities TR are determined by the maintenance actions performed and the road deterioration. According to Ouyang and Madanat (2004) and Gao and Zhang (2013), the annual deterioration of an asphalt road segment can be described by Equation 9.3

$$IRI_{t+1} = (IRI_t + \alpha)e^\beta. \quad (9.3)$$

In this equation, $\alpha = 0.2$ is defined as $\epsilon(1 - e^{-\beta})$, where ϵ depends on road type and traffic and $\beta = 0.0153$. This parametrization is analogous to Gao and Zhang

(2013). The actions, which are described above, are performed at the beginning of each year. If a maintenance action is performed— a_1 or a_2 —, first, the condition is improved, as described above, and second, the annual deterioration is considered based on the maintained road. This combination of maintenance and deterioration probabilities results in the combined transition probabilities, which are shown in the columns headed with a_1 and a_2 in Table 9.3. If no maintenance

Table 9.3: Transition probabilities TR composed of rehabilitation due to maintenance action AS performed at the beginning of the year and annual deterioration (Gao and Zhang, 2013).

	a_1 (reconstruction)					a_2 (resurfacing)					a_3 (nothing)				
	s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
s_1	0.741	0.259	0	0	0	0.741	0.259	0	0	0	0.741	0.259	0	0	0
s_2	0.741	0.259	0	0	0	0.741	0.259	0	0	0	0	0.825	0.175	0	0
s_3	0.741	0.259	0	0	0	0.276	0.614	0.110	0	0	0	0	0.858	0.142	0
s_4	0.741	0.259	0	0	0	0	0.191	0.700	0.109	0	0	0	0	0.873	0.127
s_5	0.741	0.259	0	0	0	0	0	0.093	0.794	0.113	0	0	0	0	1

action is performed at the beginning of a year, the road is expected to deteriorate according to the probabilities provided in the right columns a_3 of Table 9.3. These deterioration probabilities are directly derived from Equation 9.3.

Observations

A finite set of possible observations $\Omega = \{o_1, o_2, o_3, o_4, o_5\}$ is defined, which corresponds to the set of states. Observation probabilities OB depend on the inspection method—laser- and crowd-based. For the laser inspections a 100 % accuracy is assumed. This means that given an RQI value observation, it is assumed that it reflects the actual road condition. The observation probabilities for this certain inspection method can be modeled as an identity matrix, which is depicted in the first laser-based columns of Table 9.4. For the crowd-sensed inspections, an accuracy in observing the actual state of 85 % is assumed. This accuracy results from Chapter 3. If in Chapter 3 classifications are applied instead of regressions, for example, 15 % of the s_1 states are falsely classified as s_2 state. Since the road link considered for the analyses in the former part is a district road in Germany, which was in an overall good condition, an extrapolation of the hereby determined accuracy from the states s_1 and s_2 to the states s_3, s_4 and s_5 is performed. These empirically determined observation probabilities are provided in the crowd-based columns of Table 9.4.

Since it is also required to model scenarios that have a recurring pattern every four years, the states set is extended by the cross product with the four years treated differently $AN = \{a_1, a_2, a_3, a_4\}$. Likewise, the transition and observation probabilities are extended in a way that a state $ST_{i,j}$ will always result in a

Table 9.4: Observation probabilities OB for laser-based inspections are defined as an identity matrix since certainty is assumed. For crowd-based inspections, the probabilities are derived from the findings from Chapter 3.

	Laser-based					Crowd-based				
	o_1	o_2	o_3	o_4	o_5	o_1	o_2	o_3	o_4	o_5
s_1	1	0	0	0	0	0.850	0.150	0	0	0
s_2	0	1	0	0	0	0.075	0.850	0.075	0	0
s_3	0	0	1	0	0	0	0.075	0.850	0.075	0
s_4	0	0	0	1	0	0	0	0.075	0.850	0.075
s_5	0	0	0	0	1	0	0	0	0.150	0.850

state $s_{i,k}$ with one year offset, thus, $k = (j + 1) \text{ modulo } 4$. For convenience, in the description states are addressed without mentioning a certain year.

Rewards

Below both cost components considered in this study, road maintenance and RUC, are described. The maintenance costs depend on the maintenance action AS performed at the beginning of a year. A reconstruction a_1 is expected to cost €45.32/m² and a resurfacing a_2 €17.43/m² (Gao and Zhang, 2013). Assuming a 200 m asphalt road segment having a width of 4 m, the aggregated maintenance costs per year are given in Table 9.5. The annual RUC are also provided in this

Table 9.5: Annual maintenance costs and RUC in Euro for one road segment depending on segment state and maintenance action performed at the beginning of the year. Rewards RE are defined as the sum of maintenance costs and RUC. Costs are from Gao and Zhang (2013) and are translated from Hong Kong Dollar into Euro at the closing exchange rate 0.1162 on 02.11.2016.

	a_1 (reconstruction)		a_2 (resurfacing)		a_3 (nothing)	
	MC [€/year]	RUC [€/year]	MC [€/year]	RUC [€/year]	MC [€/year]	RUC [€/year]
s_1	36 249.91	158 037.99	13 942.30	87 739.89	0	82 891.46
s_2	36 249.91	158 037.99	13 942.30	87 739.89	0	193 477.04
s_3	36 249.91	158 037.99	13 942.30	157 234.91	0	357 251.23
s_4	36 249.91	158 037.99	13 942.30	324 099.64	0	573 788.56
s_5	36 249.91	158 037.99	13 942.30	555 065.01	0	823 950.81

table. They are composed of vehicle operating costs (VOC) and costs caused due to additional travel time. As for the road user side assessment, the VOC are composed of expenses for fuel, tire wear and vehicle repair. The costs for travel

delay depend on maintenance actions, since work zones can cause traffic jams and thus force road users to stop or drive slower. The determination of the RUC aggregates is provided by Gao and Zhang (2013), while considering the annual average daily traffic, the road's capacity of three lanes, the duration and effects on speed of work zones for the middle lane and the wear of two vehicle types, passenger cars and trucks. Figure 9.2 provides all road user effects (RUE), while the ones considered for the road authority side assessment are highlighted. While

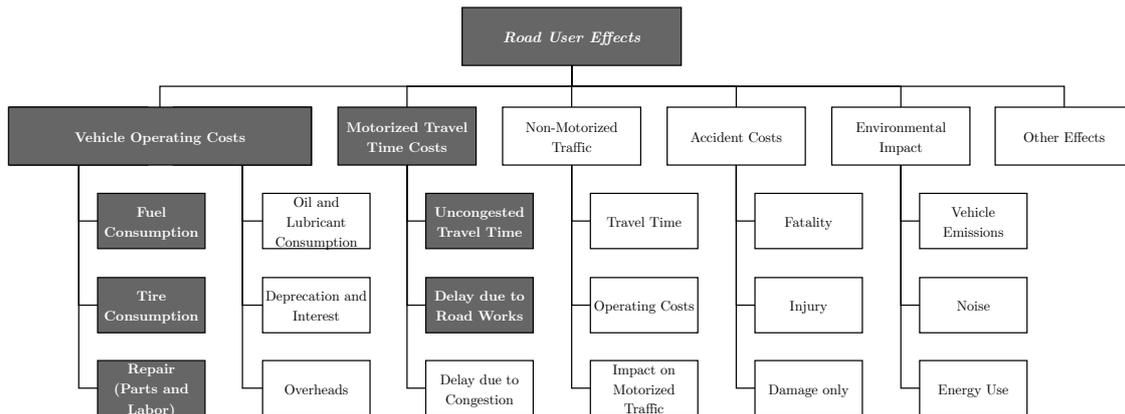


Figure 9.2: Components of RUEs (Bennett and Greenwood, 2003), which are included in the road authority side assessment. The highlighted components are considered by Gao and Zhang (2013) and likewise in the road user side assessment conducted in this chapter.

considering multiple years in the analysis a discount factor $df = 0.95$ is applied to future rewards.

9.4 Results

This section presents the results of the POMDP analyses for the three scenarios—laser-based inspections, crowd-based inspections and combination of both inspection types. Hereby, Research Question 7 is answered.

9.4.1 Savings Potential

The scenario specific policies lead to the following expected total costs—maintenance costs and RUC, discounted for a 100 years horizon and for a 200 m asphalt segment: *scenario 1*—€1 938 836; *scenario 2*—€1 824 365 and *scenario 3*—€1 803 470. Figure 9.3 provides the corresponding average annual total costs for each scenario and per segment. It can be seen that performing laser-based inspection every fourth year—as it is currently done in Germany—causes the

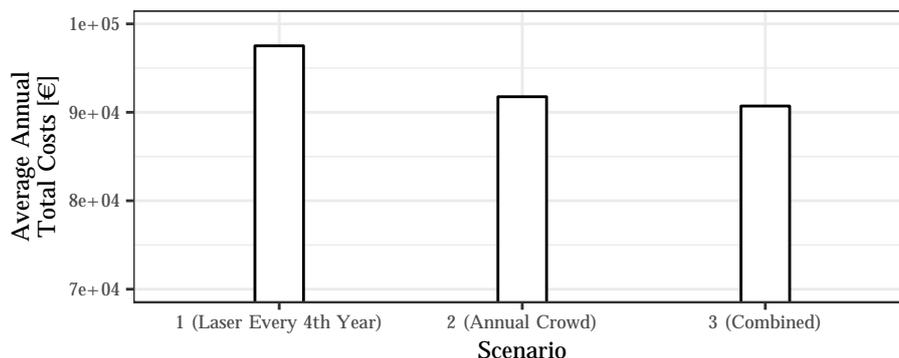


Figure 9.3: Average annual total costs—maintenance costs and RUC—per road segment for optimal maintenance policies of investigated inspection scenarios.

highest annual total costs (€97 519.17). Relying on crowd-based inspections instead reduces the costs to €91 761.53. A further cost reduction to €90 710.55 can be achieved by complementing four year laser-based inspections with annual crowd-based inspections.

Given these results, Research Question 7 can be answered. The results show that having annually, but less accurate crowd-based inspections reduces maintenance costs and RUC by 5.9 % compared to performing laser-inspection at four year intervals. This finding shows that from a managerial perspective the inaccuracy of a crowdsensing-based monitoring is overcompensated by its potential of providing measurements more frequently. Furthermore, it is shown that the combination of both inspection types reduces the total costs by 6.98 %.

9.5 Conclusion

To assess the implications of crowdsensing-based road condition monitoring to road authorities, POMDPs were applied to determine optimal maintenance policies for different scenarios. The scenarios differ in the inspections' accuracy and frequency. For considering inaccuracy in the scenarios that model crowdsensing inspections, it was made use of observation probabilities, which were derived from Part II. The results show that with respect to a decision support system for road maintenance, the application of a crowdsensing-based road condition monitoring can be recommended as a substitute and as a supplement to the today's laser-based monitoring. The assessment in this chapter also has limitations. The observation probabilities are empirically determined based on road segments, which are in condition states s_1 and s_2 . Thus, an extrapolation was required to determine the uncertainty for the other states. Higher external validity can be achieved if an empirical determination of the observation probabilities for the residual states is conducted. Furthermore, except to the differentiation in inspec-

tion methods, a fixed set of parameter values is considered. Even though, they are carefully chosen, an analysis of the results' sensitivities to parameter variations, such as different road types, lane types and the traffic volume, can extend the current findings.

Part IV

Finale

Chapter 10

Conclusion

To conclude the thesis at hand, this chapter first summarizes the answers to the research questions and the consequential contributions. Thereafter, propositions for future research are provided before the main conclusions drawn from the thesis are provided.

10.1 Contributions

The road infrastructure is of great importance to road users and constitutes a valuable asset for economies. For safety, comfort and economic reasons, the availability of accurate and current information about the road condition and its deterioration is essential. However, due to scarce resources the road network is often monitored at very low frequency if road condition inspections are performed at all. The thesis at hand focuses on technical issues and economic questions in order to facilitate a crowdsensing-based monitoring system for a frequent and nationwide provisioning of road condition information. With regard to the research questions of the thesis formulated in Section 1.2, the contributions of the thesis are discussed.

Answer to Research Question 1 – Self-Calibration: A major technical issue in the implementation of a crowd-based road monitoring by using smart devices that are carried within vehicles, is the variety of such devices and vehicles. For example, each vehicle—and at least each vehicle type—has specific characteristics determined by its weight, suspension system, tires, chassis, etc. In order to ensure that the effects derived from the sensor readings that provide information about the road surface are not masked by vehicle-specific effects, an individual calibration is required for each new participant of the crowd-based system. In order to avoid approaches with manual effort, such as from Nitsche et al. (2014), because they would prevent an easy participation and thus avoid a large number of participants, a self-calibration procedure is described in Chapter 3. The self-calibration allows to learn the vehicle characteristics while driving on public roads, for which the condition is known from previous measurements. The procedure comprises map-matching and multiple interpolation steps to map the

collected data to a road network. A subsequent feature extraction also reflecting the frequency content allows the supervised training of estimation models. As an evaluation, seven test drives were performed with a smartphone attached to the dashboard measuring the acceleration and rotation rates. Random forest regressions (RFRs) and support vector regressions (SVRs) were applied with the former achieving an out-of-sample R^2 of 0.68 on average. This result shows that even single vehicles with consumer devices can contribute to a crowdsensing-based road condition monitoring without integration effort by means of the self-calibration approach.

Answer to Research Question 2 – Data Reduction: The presented self-calibration procedure, which allows an easy integration of new participants into a crowdsensing-based system, involves the processing of the gathered sensor data in a backend system. Because of limited memory and transmission resources of mobile devices, data reduction potentials have to be investigated. Accordingly, Chapter 4 is concerned with the reduction of the amount of data to be sent to the backend. The permutation importance (PI) as an embedded feature importance metric—which can be applied during the training phase of the RFR models—was used to reduce the number of features and likewise the amount of data by following a backward elimination. In addition, a sensitivity analysis was carried out to determine the effect of a reduction in the sampling frequency on the estimation performance. In both approaches, it was found that a data reduction up to approximately 50% can be achieved without a loss in the estimation performance. Furthermore, the results reveal that, given the setting of the study, consumer devices achieve nearly maximum estimation performance even at a sampling frequency of 25 Hz. Accordingly, virtually any device equipped with an inertial measurement unit (IMU) can contribute to the crowd-based road condition monitoring system.

Answer to Research Question 3 – Combination of Estimations: Since the self-calibration approach enables an efficient participation of new cars and sensors, the question arose how to best make use of the existence of multiple condition estimates for a single road segment. In Chapter 5, various linear combination alternatives were examined, which differ in the weights assigned to the estimates from the single drives. In addition to weightings based on quality metrics, the simple average (SA) and optimal weights (OW) were examined. The latter two approaches represent extremes. While the SA is purely a combination, applying OW aims at reducing the in-sample error by an additional training phase. In order to address overfitting issues resulting from the additional estimation, the OW were shrunk towards the SA. This regularization showed that the application of the SA results in the best out-of-sample performance. It was shown that an increase in the R^2 from 0.68 to 0.75 can be achieved by taking the SA of all seven test drives instead of relying on the estimation of a single drive. Accordingly, it was

shown how the crowdsensing-based road condition monitoring can be improved with an increase in participants.

Answer to Research Question 4 – IT Infrastructure: To avoid the need to determine the initial dimensioning of the backend IT infrastructure—which processes features from the sensors as well as estimates derived from these features, as described in the Chapters 3, 4 and 5—in advance and to be able to dynamically address uncertainties in the adoption of the system, infrastructure as a service (IaaS) cloud computing is a suitable solution. The application of IaaS allows the utilization of scaling technologies to dynamically adapt the resources to the load. However, it was unclear which scaling technology is the most efficient in terms of serving a load with a daily seasonality—as it can be expected for the crowdsensing-based system. Chapter 6 describes the implementation of a generic resource scaling simulation model. In order to allow a comparison between simulation scenarios, within the model criteria can be defined—such as the fulfillment of a certain quality of service (QoS) per time slot—which have to be met by all scenarios. Using this simulation model, a real load pattern was given to model to simulate and examine the resource consumption of horizontal and diagonal scaling technologies. The results of the simulations show that by using diagonal scaling, a resource reduction of 4.16 % to 8.05 % compared to horizontal scaling can be achieved—depending on the minimum contract duration. This is mainly caused by the fact that with diagonal scaling, fast scaling ups within a virtual machine (VM) can be performed and thus a faster adaption to rapid increases in load can be addressed without violating the QoS.

Answer to Research Question 5 – Integrated Service: The crowdsensing-based road condition monitoring presented in this thesis is aimed primarily at road users and road authorities as customers which have a demand for reliable and current road condition information. The crowd, on the other hand, is the primary data provider. A challenge that needs to be overcome is that an individual participant from the crowd cannot market its data directly to the customers, since the raw sensor data of a single participant does not provide value to them. The analysis steps described in Chapters 3, 4 and 5 must be carried out beforehand in order to derive meaningful road condition information by calibrated models on the one hand and to increase the reliability and spatial coverage by combining measurements from multiple participants on the other hand. Initially and in recurring intervals, also the acquisition of accurate laser-based measurements from engineering companies is necessary for calibration. To address the resulting need of orchestrating the data suppliers—basically, the crowd and engineering companies—on the one hand, and deriving a service that is useful for road users and road authorities, on the other hand Chapter 7 describes an integrated service that introduces a new intermediary between the data providers and the service customers. The integrated service described as a service map was on

the one hand analytically evaluated in terms of its ability to provide timely data for decision support using spatio-temporal hotspot analyses and on the other hand evaluated descriptively to determine its utility for the participants. The hotspot analyses showed that the crowdsensing approach can detect significant road anomalies 8.37 months earlier than a conventional approach with annual inspection intervals. The evaluation of the participants' utility revealed concrete use cases and incentives for them. Use cases for road users and road authorities were assessed in the subsequent Part III of the thesis.

Answer to Research Question 6 – Road User Side Assessment: Even though, the affect of road roughness on vehicle operating costs (VOCs) was investigated (Chatti and Zaabar, 2012; Tan et al., 2012) and some VOCs, such as fuel expenses, are often considered in vehicle navigation, it is new to consider VOCs caused by rough roads for navigation. In Chapter 8, the cost savings potential an individual road user can achieve while taking a detour on a smoother road were determined. The underlying idea of the assessment is to analyze the trade-off between additional VOCs due to taking a detour for being able to drive on a smoother road and the hereby achieved costs savings in tire consumption and repair through less vehicle wear. Multiple baseline paths with a predefined roughness were successively compared to multiple longer and smoother alternative paths. The VOCs were determined for medium cars and light trucks. The trade-off analyzes provided in Chapter 8 show that there are saving potentials for different vehicle types—medium cars and light trucks—when considering road roughness for vehicle navigation. Basically, the worse the overall road condition, the higher is the annual savings potential. This means for a baseline road roughness at an international roughness index (IRI) of 3 m/km, there is just a modest saving potential. The results also show that for light trucks the absolute savings are higher compared to the savings for medium cars in most cases. Thus, the approach can be expected to be of interest for truck manufacturers in developing countries or logistic companies. The savings potentials also vary depending on the fuel price. For most cases, considering a lower fuel price leads to higher savings. This is an interesting finding, especially in view of the today's increasing adoption of electric vehicles. If it turns out that the energy costs are permanently cheaper through the use of electric vehicles, compared to current fuel costs, a road condition-dependent navigation would result in even more value.

Answer to Research Question 7 – Road Authority Side Assessment: In addition to the utility for road users, it is also essential that the crowdsensing-based approach is beneficial for road authorities. Although it seems reasonable that the consideration of crowdsensing-based road condition information leads to a larger database and therefore to better decisions regarding road maintenance, this has not been assessed in the past. However, an assessment was particularly relevant since the use of the crowd allows fast acquisition of data about the road

conditions, but it is less accurate and reliable than the inspections carried out today with special vehicles using high-precision laser technology. To determine the benefit from a road authority's perspective, partially observable Markov decision process (POMDP) models have been applied in Chapter 9 to determine maintenance costs and road user costs (RUC), given different inspection methods—accurate but rare laser-based inspections, frequent but less accurate crowd-based inspections and a combination of both inspection types. The results show that using the crowd for inspections causes a reduction in total costs of 5.9% compared to the current inspection procedures on the German federal road network. If both inspection types are combined, a cost reduction of 6.98% is achieved.

10.2 Outlook

As the previous section described the contributions of the thesis at hand with regard to a crowdsensing-based road condition monitoring, now an outlook is provided. Further future work can go into various directions to enhance this thesis. A list of selected research activities follows.

Data sources: The supervised machine learning method RFR allows the consideration of multiple dimensions. Thus, the procedure presented in this thesis allows further characteristics to be taken into account when estimating the road condition in a crowdsensing-based manner. Accordingly, other data sources can be involved in the determination of the road condition. Thus, an enhancement of the monitoring can be achieved through the usage of data from the increasing number of sensors that are readily built in current vehicles. To mention a few: (stereoscopic) front cameras, radar sensors, ultrasonic sensors, rain sensors, sensors that determine the slip of the tires or sensors that measure the driver's steering behavior, such as the angle of the steering wheel or the position of the pedals. In addition, it seems interesting to leverage data sources from outside the vehicle to obtain additional environmental information, which may influence the sensing. For instance, snow-covered roads are expected to cause different vehicle movements and likewise different sensor readings than uncovered roads or light conditions depending on the position of the sun, the rain and the coverage by clouds influence the camera-based sensing. The use of additional data sources would therefore allow these effects to be taken into account in order to increase the accuracy and robustness of the crowdsensing-based road condition monitoring. Furthermore, the external validity of the presented calibration procedure can be confirmed by examining its feasibility to handle the heterogeneity of vehicles and sensors in the crowd. As a consequence, the complete fulfillment of the criterion calibration—according to Section 2.2.5—could be verified.

Investigated objects: The crowdsensing-based estimation of the road condition in nearly real-time is covered in this thesis. Future work could address temporal aspects and thus, focus on forecasting the road condition. Even though there has been work done in the field of determining the road condition deterioration, the frequent road condition inspections made possible by crowdsensing represent a new and promising data basis to this field (Morosiuk et al., 2004; Svenson, 2014). On the one hand, more detailed information about the long-term degradation of the road condition and its dependence on other factors, such as traffic load, weather conditions, etc., can be used to infer insights into the durability of road construction materials and construction methods. On the other hand, the accurate data base can be used to forecast sudden and unexpected changes in road conditions. For example, erosion of the road base layers or aforementioned "blow-ups"—which are until now unforeseen bucklings of the concrete elements of a road—could be detected beforehand, allowing interventions, such as speed limits or road closures. In addition to increasing the estimation performance and extending the validation by the consideration of further data sources, as mentioned in the previous item, new data sources—especially camera sensors—allow to focus on other objects not directly related to road conditions. Thus, in addition to an extension to detection of objects that can be recognized by vehicle vibrations, such as ruts and cracks, also guardrails, road signs, streetlights, lane markings, etc.—which are of high interest to road authorities for inventory purposes—can be addressed by applying additional sensors, such as camera sensors.

Utility to road users and road authorities: As stated in Section 2.2.5, the assessment of the system's utility is not regarded as completely fulfilled in this thesis. This is partly due to the fact that, especially from the point of view of road authorities, further investigations should be carried out. First, the costs related to each inspection—laser-based and crowd-based—can be considered additionally. This would allow a more comprehensive view of the related costs on the one hand and would allow the analysis to be extended to determine optimal inspection intervals on the other hand. Secondly, the cost rates taken into account in the analysis could be adapted to the German case in order to allow for a more coherent assessment. In particular, this applies to the RUC and the costs of the maintenance actions and inspection tasks. Moreover, it should also be noted that, even though, the RUC should also be taken into account when determining optimal maintenance policies (Ouyang and Madanat, 2004; Bennett et al., 2006), it may be assumed that road authorities focus more on reducing their direct costs. To take this aspect into account, the utility of the crowdsensing-based condition monitoring for road users and road authorities could be revised, as-

suming that road authorities exclusively minimize their maintenance and inspection costs.

10.3 Conclusions

With regard to the findings of the thesis, three main conclusions are drawn. First, a self-calibration of new vehicles in a crowdsensing-based monitoring system is facilitated by applying machine learning methods and automatically aligning the features generated from the vehicles with training data. As a result, the limitations of current approaches can be overcome, in which either the individual vehicle characteristics are not taken into account or additional high-precision sensors are required. Thus, this self-calibration enables the integration of new participants in an easy and effortless manner, which is crucial for a high number of participants and likewise a high spatio-temporal coverage. In addition to the fact that considering individual vehicle characteristics allows a direct increase in estimation accuracy, the now possible spatio-temporal coverage can be leveraged to further increase the estimation accuracy. This can be achieved by combining several estimations from single drives to an overall more robust and accurate estimation, which was also investigated in this thesis.

Second, road users should consider precise and up-to-date information on the road condition—as it becomes available through the presented crowdsensing-based system—for their route planning. Driving on smoother roads reduces vehicle wear. The trade-off analyses indicate that, particularly for road networks that are in an overall bad condition, relatively high savings can be achieved by taking a detour on a smoother route. In addition, it was found that trucks have a higher savings potential than cars. Therefore, it is especially recommended for developing countries or logistics companies to consider road conditions for route guidance. In comparison, for road networks that are in an overall good condition—as it is the case in Germany—the currentness of the condition information considered for vehicle navigation is of high importance. Since the savings potential for good road networks is modest, it should be avoided to expect savings by taking a detour based on outdated information. A crowdsensing-based condition monitoring can prevent that a meanwhile deterioration of the preferred route may result in a situation where the expected savings are not realized.

Third, road authorities and engineering companies, which nowadays conduct road condition inspections with specially equipped vehicles, should consider the operation of a crowdsensing-based monitoring system in view of the fact that significant cost savings can be achieved. Even considering the lower accuracy of crowd-based inspections, a significant reduction in the overall maintenance and RUCs can be achieved by replacing the today's inspections performed every four years with frequent crowd-based inspections. This is because on the basis of more frequent inspections, maintenance tasks can be adapted more efficiently to the

actual road condition. In addition, the combination of conventional and crowd-based inspections was investigated, which resulted in a further reduction of overall costs. A crowdsensing-based solution can be realized, as examined, by means of relatively few IT resources. With this in mind, it is also entirely plausible that new intermediaries may establish, providing a crowd-based decision support for road authorities and road users. In regards of these findings, road users as well as road authorities should consider the deployment of a crowdsensing-based road condition monitoring for a real-time and widespread information basis for decision support.

Part V
Appendix

Appendix A

Feature Importance

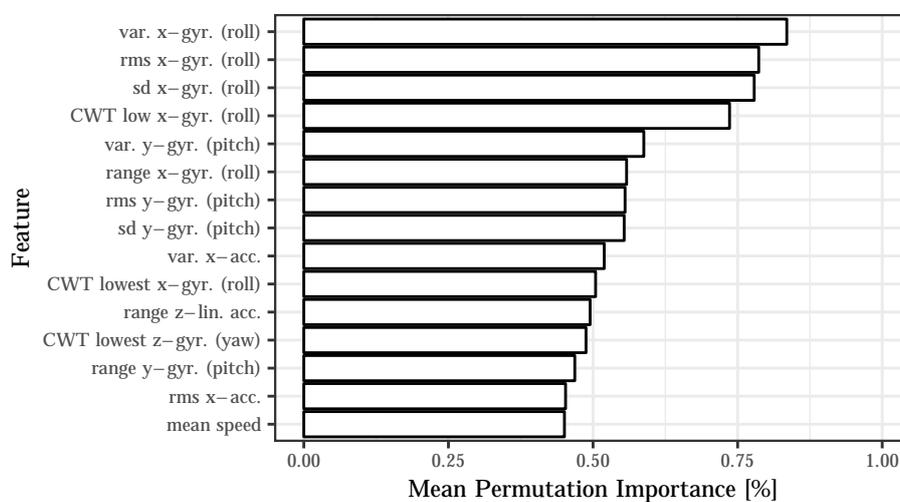


Figure A.1: Mean PI of the 1 to 15 most important features.

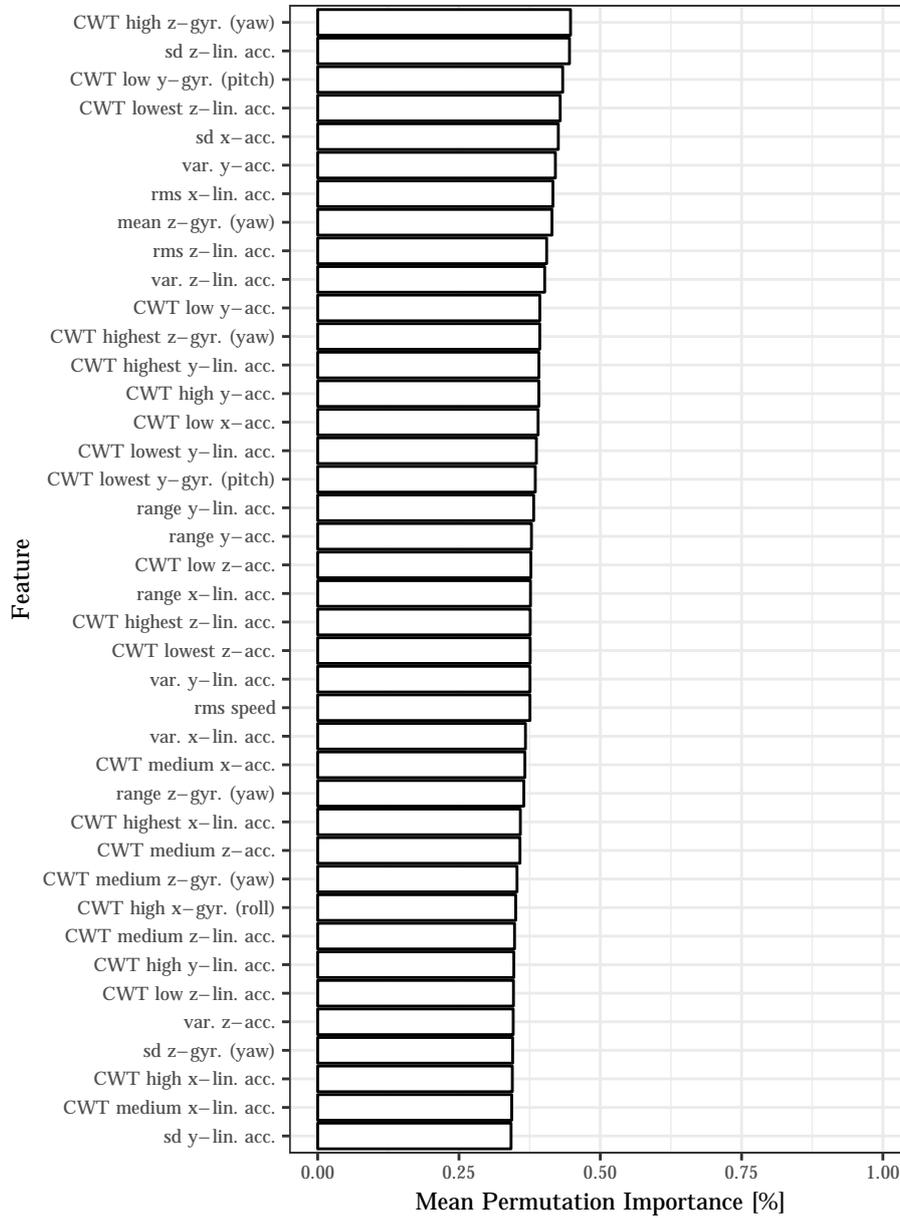


Figure A.2: Mean PI of the 16 to 55 most important features.

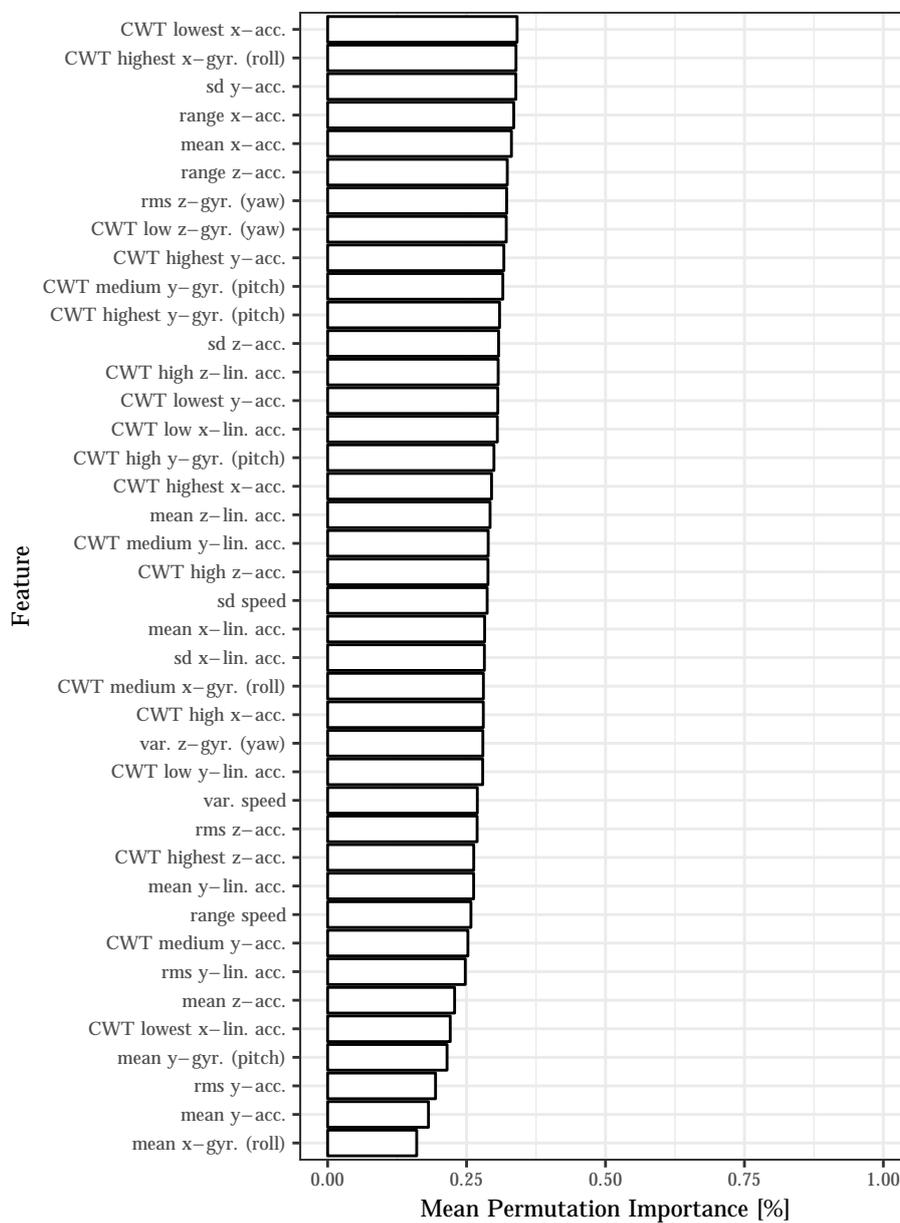


Figure A.3: Mean PI of the 56 to 95 most important features.

Appendix B

Prediction Errors and Combination Weights

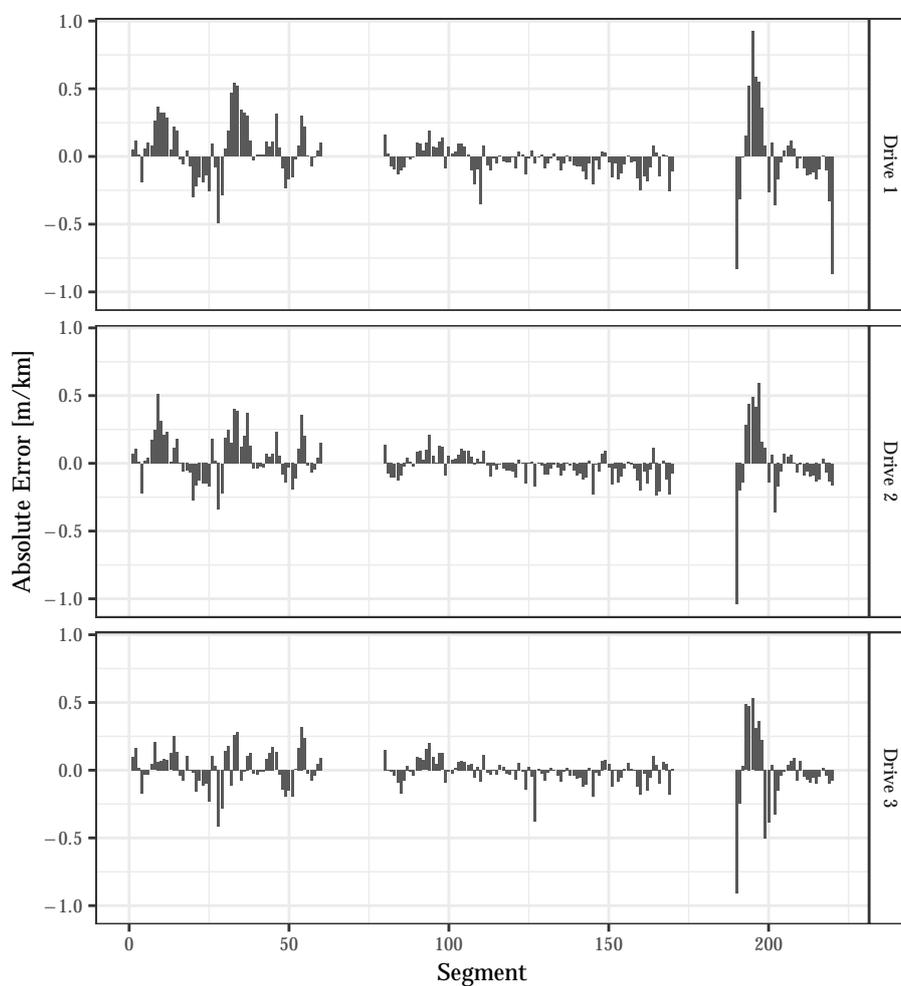


Figure B.1: Absolute errors of in-sample predictions for drives 1 to 3.

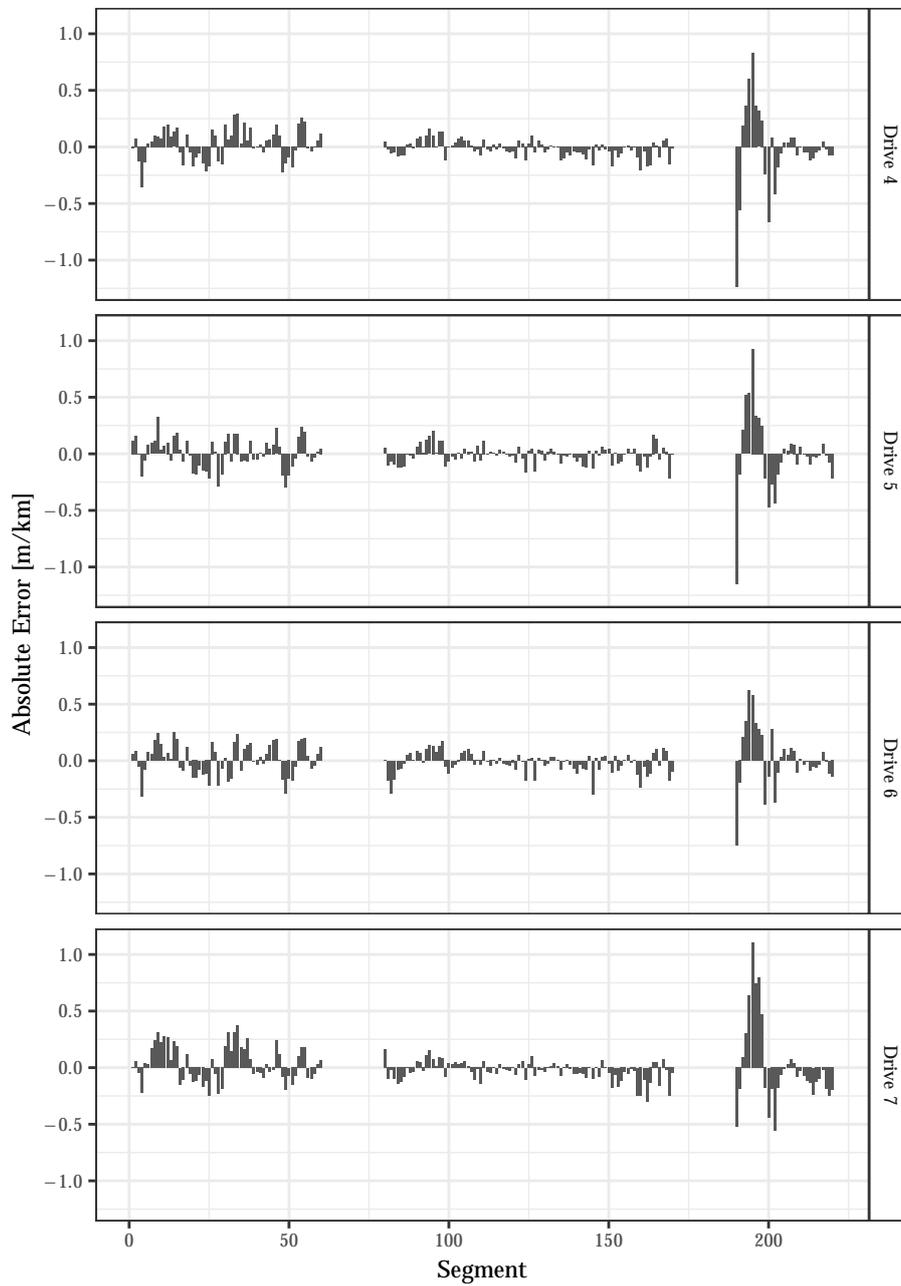


Figure B.2: Absolute errors of in-sample predictions for drives 4 to 7.

Table B.1: Weights of different combination methods for each drive.

Drive	\bar{P}	\bar{P}_{R^2}	\bar{P}_{RMSE}	\bar{P}_{Best}	\bar{P}_{OW}	\bar{P}_λ
1	0.1429	0.1503	0.1475	0	-0.2362	0.1429
2	0.1429	0.1338	0.1486	0	0.3794	0.1429
3	0.1429	0.1397	0.1444	0	0.4068	0.1429
4	0.1429	0.1500	0.1400	0	-0.3167	0.1429
5	0.1429	0.1270	0.1251	1	-0.0538	0.1429
6	0.1429	0.1646	0.1353	0	0.7735	0.1429
7	0.1429	0.1346	0.1590	0	0.0470	0.1429

Appendix C

Server Load

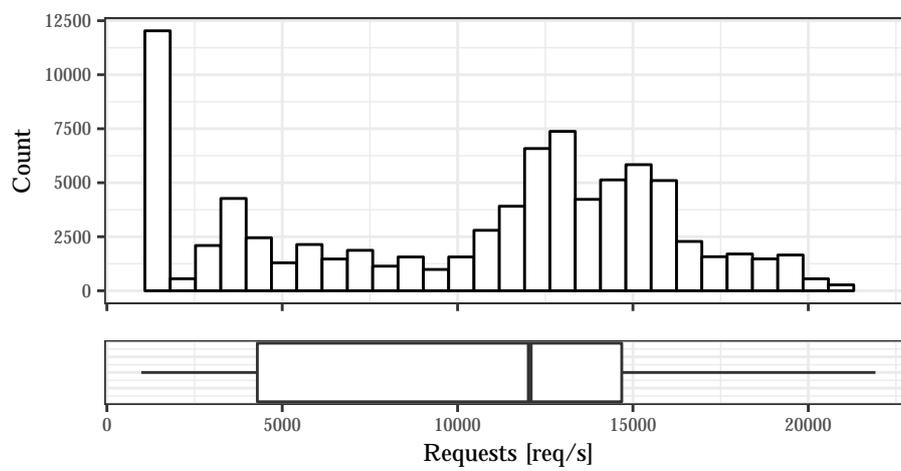


Figure C.1: Histogram and boxplot of server load (15 minutes aggregates).

References

- Allmendinger, G. and R. Lombreglia (2005). Four Strategies for the Age of Smart Services. *Harvard Business Review* 83(10), 1–11.
- Amaldi, E. and V. Kann (1998). On the Approximability of Minimizing Nonzero Variables or Unsatisfied Relations in Linear Systems. *Theoretical Computer Science* 209(1), 237–260.
- Andrade, P. R. M., R. G. Araujo, J. C. Filho, T. R. Pereira, A. B. Albuquerque, and N. C. Mendonca (2015). Improving Business by Migrating Applications to the Cloud Using Cloudstep. In *Proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications Workshops*, pp. 77–82.
- Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis* 27(2), 93–115.
- Anttiroiko, A.-V., P. Valkama, and S. J. Bailey (2014). Smart Cities in the New Service Economy: Building Platforms for Smart Services. *AI & Society* 29(3), 323–334.
- Auffarth, B., M. López, and J. Cerquides (2010). Comparison of Redundancy and Relevance Measures for Feature Selection in Tissue Classification of CT Images. In *Advances in Data Mining: Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, pp. 248–262. Heidelberg: Springer Berlin / Heidelberg.
- AWS (2010). AWS CloudFormation: User Guide API (Version 2016-11-15). <http://docs.aws.amazon.com/AWSCloudFormation/latest/APIReference>.
- AWS (2017). Amazon Elastic Compute Cloud User Guide for Linux Instances (Version 2010-05-15). <https://aws.amazon.com/documentation/ec2>.
- Bapna, R., A. Barua, D. Mani, and A. Mehra (2010). Research Commentary – Cooperation, Coordination, and Governance in Multisourcing: An Agenda for Analytical and Empirical Research. *Information Systems Research* 21(4), 785–795.
- Barile, S. and F. Polese (2010). Smart Service Systems and Viable Service Systems: Applying Systems Theory to Service Science. *Service Science* 2(1-2), 21–40.
- Bates, J. M. and C. W. J. Granger (1969). The Combination of Forecasts. *Journal of the Operational Research Society* 20(4), 451–468.

- Baum, L. E. and T. Petrie (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics* 37(6), 1554–1563.
- Bein, P. (1993). VOC Model Needs of a Highways Department. *Road and Transport Research* 2(2), 40–54.
- Ben-Akiva, M., A. De Palma, and K. Isam (1991). Dynamic Network Models and Driver Information Systems. *Transportation Research Part A: General* 25(5), 251–266.
- Bennett, C. R. and I. D. Greenwood (2003). *Modelling Road User and Environmental Effects*, Volume 7 of *Highway Development and Management Series*. Puteaux: The World Road Association (PIARC).
- Bennett, C. R. and W. D. O. Paterson (2000). *A Guide to Calibration and Adaptation*, Volume 5 of *Highway Development and Management Series*. Puteaux: The World Road Association (PIARC).
- Bennett, C. R., H. D. Solminihac, and A. Chamorro (2006). Data Collection Technologies for Road Management. Technical Report 30, The World Bank, Washington, D. C.
- Berndt, P. and A. Maier (2013). Towards Sustainable IaaS Pricing. In *Economics of Grids, Clouds, Systems, and Services*, pp. 173–184. Cham: Springer International Publishing.
- Bhoraskar, R., N. Vankadhara, B. Raman, and P. Kulkarni (2012). Wolverine: Traffic and Road Condition Estimation Smartphone Sensors. In *Proceedings of the 4th International Conference on Communication Systems and Networks*, pp. 1–6.
- Blanc, S. M. and T. Setzer (2016). When to Choose the Simple Average in Forecast Combination. *Journal of Business Research* 69(10), 3951–3962.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45(1), 5–32.
- Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and Regression Trees*. Monterey: CRC Press.
- Bruns, J. and V. Simko (2017). Stable Hotspot Analysis for Intra-Urban Heat Islands. *GI-Forum, Journal for Geographic Information Science* 1(1), 79–92.
- Byun, J. and S. Park (2011). Development of a Self-Adapting Intelligent System for Building Energy Saving and Context-Aware Smart Services. *IEEE Transactions on Consumer Electronics* 57(1), 90–98.

- Cassandra, A., L. P. Kaelbling, and M. L. Littman (1994). Acting Optimally in Partially Observable Stochastic Domains. In *Proceedings of the 12th AAAI National Conference on Artificial Intelligence*, pp. 1023–1028.
- Cassandra, A., M. L. Littman, and N. Zhang (1997). Incremental Pruning: A Simple, Fast, Exact Method for Partially Observable Markov Decision Processes. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pp. 54–61.
- Cha, S.-H. (2007). Comprehensive Survey on Distance / Similarity Measures Between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences* 1(4), 300–307.
- Chatti, K. and I. Zaabar (2012). *Estimating the Effects of Pavement Condition on Vehicle Operating Costs*. Washington, D. C.: Transportation Research Board (TRB).
- Chebrolu, S. B. (2011). Assessing the Relationships among Cloud Adoption, Strategic Alignment and Information Technology Effectiveness. *Journal of Information Technology Management* 22(2), 13–29.
- Cheng, H.-T. (1988). *Algorithms for Partially Observable Markov Decision Processes*. Ph. D. thesis, University of British Columbia.
- Chieu, T. C., A. Mohindra, A. A. Karve, and A. Segal (2009). Dynamic Scaling of Web Applications in a Virtualized Cloud Computing Environment. In *Proceedings of the IEEE International Conference on e-Business Engineering*, pp. 281–286.
- Chugh, G., D. Bansal, and S. Sofat (2014). Road Condition Detection Using Smartphone Sensors: A Survey. *International Journal of Electronic and Electrical Engineering* 7(6), 595–601.
- Claudiel, C. G., A. Hofleitner, N. Mignerey, and A. M. Bayen (2008). Guaranteed Bounds on Highway Travel Times Using Probe and Fixed Data. In *Proceedings of the 88th Transportation Research Board Annual Meeting*, pp. 1–18.
- Cristianini, N. and J. Shawe-Taylor (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.
- De Weille, J. (1966). *Quantification of Road User Savings*. Washington, D. C.: Johns Hopkins Press.
- Deflorio, F. P. (2003). Evaluation of a Reactive Dynamic Route Guidance Strategy. *Transportation Research Part C: Emerging Technologies* 11(5), 375–388.

- Demirkan, H. and D. Delen (2013). Leveraging the Capabilities of Service-oriented Decision Support Systems: Putting Analytics and Big Data in Cloud. *Decision Support Systems* 55(1), 412–421.
- Dibbern, J., T. Goles, R. Hirschheim, and B. Jayatilaka (2004). Information Systems Outsourcing: A Survey and Analysis of the Literature. *ACM SIGMIS Database* 35(4), 6–102.
- Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* 1(1), 269–271.
- Douangphachanh, V. and H. Oneyama (2013). A Study on the Use of Smartphones for Road Roughness Condition Estimation. *Journal of the Eastern Asia Society for Transportation Studies* 10(1), 1551–1564.
- Dutta, S., S. Gera, A. Verma, and B. Viswanathan (2012). SmartScale: Automatic Application Scaling in Enterprise Clouds. In *Proceedings of the 5th IEEE International Conference on Cloud Computing*, pp. 221–228.
- El Kihal, S., C. Schlereth, and B. Skiera (2012). Price Comparison for Infrastructure-as-a-Service. In *Proceedings of the 20th European Conference on Information Systems*, pp. 1–12.
- Ericsson, E., H. Larsson, and K. Brundell-Freij (2006). Optimizing Route Choice for Lowest Fuel Consumption – Potential Effects of a New Driver Support Tool. *Transportation Research Part C: Emerging Technologies* 14(6), 369–383.
- Eriksson, J., L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan (2008). The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. In *Proceedings of the 6th ACM International Conference on Mobile Systems, Applications, and Services*, pp. 29–39.
- Estellés-Arolas, E. and F. González-Ladrón-de Guevara (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science* 38(2), 189–200.
- Farhangi, H. (2010). The Path of the Smart Grid. *IEEE Power and Energy Magazine* 8(1), 18–28.
- FGSV (2001). *Richtlinien für die Planung von Erhaltungsmaßnahmen an Straßenbefestigungen*. Köln: FGSV-Verlag.
- Fleming, W. J. (2008). New Automotive Sensors – A review. *IEEE Sensors Journal* 8(11), 1900–1921.
- Forslöf, L. and H. Jones (2015). Roadroid: Continuous Road Condition Monitoring with Smart Phones. *Journal of Civil Engineering and Architecture* 9(4), 485–496.

- Fujimura, K. (1996). Path Planning With Multiple Objectives. *IEEE Robotics & Automation Magazine* 3(1), 33–38.
- Gallivan, M. J. and W. Oh (1999). Analyzing IT Outsourcing Relationships as Alliances Among Multiple Clients and Vendors. In *Proceedings of the 32th Hawaii International Conference on Systems Sciences*, pp. 1–15.
- Ganti, R., F. Ye, and H. Lei (2011). Mobile Crowdsensing: Current State and Future Challenges. *IEEE Communications Magazine* 49(11), 32–39.
- Gao, H. and X. Zhang (2013). A Markov-Based Road Maintenance Optimization Model Considering User Costs. *Computer-Aided Civil and Infrastructure Engineering* 28(6), 451–464.
- Goldberg, M., A. Kieninger, and H. Fromm (2014). Organizational Models for the Multi-sourcing Service Integration and Management Function. In *Proceedings of the 16th IEEE Conference on Business Informatics*, pp. 101–107.
- Goldberg, M., A. Kieninger, G. Satzger, and H. Fromm (2014). Transition and Delivery Challenges of Retained Organizations in IT Outsourcing. In *Proceedings of the Springer International Conference on Exploring Services Science*, pp. 56–71.
- Goldberg, M., G. Satzger, and A. Kieninger (2015). A Capability Framework for IT Service Integration and Management in Multi-Sourcing. In *Proceedings of the 23th European Conference on Information Systems*, pp. 1–16.
- Gonzalez, L. C., R. Moreno, H. J. Escalante, F. Martinez, and M. R. Carlos (2017). Learning Roadway Surface Disruption Patterns Using the Bag of Words Representation. *IEEE Transactions on Intelligent Transportation Systems* 18(11), 1–13.
- González-Gurrola, L. C., F. Martínez-Reyes, and M. R. Carlos-Loya (2013). The Citizen Road Watcher – Identifying Roadway Surface Disruptions Based on Accelerometer Patterns. In *Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*, pp. 374–377.
- Goovaerts, P. and G. M. Jacquez (2005). Detection of Temporal Changes in the Spatial Distribution of Cancer Rates Using Local Moran's I and Geostatistically Simulated Spatial Neutral Models. *Journal of Geographical Systems* 7(1), 137–159.
- Gouhier, T., A. Grinsted, and V. Simko (2016). R-Package "biwavelet": Conduct Univariate and Bivariate Wavelet Analyses. <https://CRAN.R-project.org/package=biwavelet>.
- Granger, C. W. J. and R. Ramanathan (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting* 3(2), 197–204.

- Guyon, I. and A. Elisseeff (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 703(3), 1157–1182.
- Hainen, A. M., S. M. Remias, T. M. Brennan, C. M. Day, and D. M. Bullock (2012). Probe Vehicle Data for Characterizing Road Conditions Associated with Inclement Weather to Improve Road Maintenance Decisions. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 730–735.
- Hall, J. W., K. L. Smith, J. C. Wambild, T. J. Yager, and Z. Rado (2009). *Guide for Pavement Friction*. Washington, D. C.: Transportation Research Board (TRB).
- Han, D., K. Kaito, and K. Kobayashi (2014). Application of Bayesian estimation method with Markov hazard model to improve deterioration forecasts for infrastructure asset management. *KSCE Journal of Civil Engineering* 18(7), 2107–2119.
- Han, R., L. Guo, M. M. Ghanem, and Y. Guo (2012). Lightweight Resource Scaling for Cloud Applications. In *Proceedings of the 12th IEEE / ACM International Symposium on Cluster, Cloud and Grid Computing*, pp. 644–651.
- Hand, J. R. M. and B. Lev (2003). *Intangible Assets: Values, Measures, and Risks*. Oxford: OUP Oxford.
- Hanson, T., C. Cameron, and E. Hildebrand (2014). Evaluation of Low-Cost Consumer-Level Mobile Phone Technology for Measuring International Roughness Index (IRI) Values. *Canadian Journal of Civil Engineering* 41(9), 819–827.
- Hara, T., T. Springer, K. Muthmann, and A. Schill (2014). Towards a Reusable Infrastructure for Crowdsourcing. In *Proceedings of the 7th IEEE / ACM International Conference on Utility and Cloud Computing*, pp. 618–623.
- Hart, P. E., N. J. Nilsson, and B. Raphael (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics* 4(2), 100–107.
- Hasenfratz, D., O. Saukh, S. Sturzenegger, and L. Thiele (2012). Participatory Air Pollution Monitoring Using Smartphones. In *Proceedings of the 2nd International Workshop on Mobile Sensing*, pp. 1–5.
- Heinze, T., V. Pappalardo, Z. Jerzak, and C. Fetzer (2014). Auto-Scaling Techniques for Elastic Data Stream Processing. In *Proceedings of the International Conference on Data Engineering*, pp. 296–302.
- Henry, J. J. (2000). *Evaluation of Pavement Friction Characteristics*. Washington, D. C.: Transportation Research Board (TRB).

- Herz, T. P., F. Hamel, F. Uebernickel, and W. Brenner (2010). Deriving a Research Agenda for the Management of Multisourcing Relationships Based on a Literature Review. In *Proceedings of the 16th Americas Conference on Information Systems*, pp. 1–11.
- Hoppengarten, A., B. Griesheim, Buslaps, Klinhart, Krause, Oertelt, Ohmen, Schniering, Socina, and Ueckermann (2006). *ZTV Asphalt-StB 01: Zusätzliche technische Vertragsbedingungen und Richtlinien für den Bau von Fahrbahndecken aus Asphalt*. Köln: FGSV-Verlag.
- Howe, J. (2006). The Rise of Crowdsourcing. *Wired Magazine* 14(6), 1–5.
- Hu, S.-C., Y.-C. Wang, C.-Y. Huang, and Y.-C. Tseng (2009). A Vehicular Wireless Sensor Network for CO2 Monitoring. In *Proceedings of the IEEE Sensors Conference*, pp. 1498–1501.
- Iqbal, W., M. N. Dailey, D. Carrera, and P. Janecek (2011). Adaptive Resource Provisioning for Read Intensive Multi-Tier Applications in the Cloud. *Future Generation Computer Systems* 27(6), 871–879.
- Jamshidi, P., A. Ahmad, and C. Pahl (2014). Autonomic Resource Provisioning for Cloud-Based Software. In *Proceedings of the 9th ACM International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pp. 95–104.
- Jang, J., Y. Yang, A. W. Smyth, D. Cavalcanti, and R. Kumar (2017). Framework of Data Acquisition and Integration for the Detection of Pavement Distress via Multiple Vehicles. *Journal of Computing in Civil Engineering* 31(2), 04016052-1–04016052-15.
- Janisch, D. (2015). An Overview of Mn/DOT's Pavement Condition Rating Procedures and Indices. Technical report, Minnesota Department of Transportation, Maplewood.
- Jazar, R. N. (2014). *Vehicle Dynamics: Theory and Application*. New York: Springer New York.
- Jokela, M., M. Kuttila, and L. L. L. Le (2009). Road Condition Monitoring System Based on a Stereo Camera. In *Proceedings of the 5th IEEE International Conference on Intelligent Computer Communication and Processing*, pp. 423–428.
- Jun, S., L. Jian-yuan, C. Han, and W. Xi-li (2008). Study on Near-Optimal Path Finding Strategies in a Road Network. *Algorithms & Computational Technology* 2(3), 319–333.
- Kaelbling, L. P., M. L. Littman, and A. R. Cassandra (1998). Planning and Acting in Partially Observable Stochastic Domains. *Artificial Intelligence* 101(1–2), 99–134.

- Kalyvianaki, E., T. Charalambous, and S. Hand (2009). Self-Adaptive and Self-Configured CPU Resource Provisioning for Virtualized Servers Using Kalman Filters. In *Proceedings of the 6th ACM International Conference on Autonomic Computing*, pp. 117–126.
- Kerali, H. G. R., J. B. Odoki, and E. E. Stannard (2006). *Overview of HDM-4*, Volume 1 of *Highway Development and Management Series*. Puteaux: The World Road Association (PIARC).
- Koch, C. and I. Brilakis (2011). Pothole Detection in Asphalt Pavement Images. *Advanced Engineering Informatics* 25(3), 507–515.
- Kohavi, R. and G. H. John (1997). Wrappers for Feature Subset Selection. *Artificial Intelligence* 97(1–2), 273–324.
- Kohlmann, F., R. Börner, and R. Alt (2010). A Framework for the Design of Service Maps. In *Proceedings of the 16th Americas Conference on Information Systems*, pp. 1–9.
- Lane, N., E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell (2010). A Survey of Mobile Phone Sensing. *IEEE Communications Magazine* 48(9), 140–150.
- Laubis, K., M. Konstantinov, V. Simko, A. Gröschel, and C. Weinhardt (2018). Enabling Crowdsensing-Based Road Condition Monitoring Service by Intermediary. *Electronic Markets*, 1–16. (forthcoming).
- Laubis, K., V. Simko, and A. Schuller (2016a). Cloud Adoption by Fine-Grained Resource Adaptation: Price Determination of Diagonally Scalable IaaS. In *Advances in Service-Oriented and Cloud Computing*, pp. 249–257. Cham: Springer International Publishing.
- Laubis, K., V. Simko, and A. Schuller (2016b). Crowd Sensing of Road Conditions and its Monetary Implications on Vehicle Navigation. In *Proceedings of the IEEE International Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress*, pp. 833–840.
- Laubis, K., V. Simko, and A. Schuller (2016c). Road Condition Measurement and Assessment: A Crowd Based Sensing Approach. In *Proceedings of the 37th AIS International Conference on Information Systems*, pp. 1–10.
- Laubis, K., V. Simko, A. Schuller, and C. Weinhardt (2017). Road Condition Estimation Based on Heterogeneous Extended Floating Car Data. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1582–1591.

- Laubis, K., V. Simko, and C. Weinhardt (2016). Weighted Aggregation in the Domain of Crowd-Based Road Condition Monitoring. In *INFORMATIK 2016*, pp. 385–393. Bonn: Gesellschaft für Informatik (GI).
- Le, M. D. and C. M. Tan (2013). Optimal Maintenance Strategy of Deteriorating System Under Imperfect Maintenance and Inspection Using Mixed Inspection-scheduling. *Reliability Engineering & System Safety* 113(1), 21–29.
- Leavitt, N. (2009). Is Cloud Computing Really Ready for Prime Time? *Computer* 42(1), 15–20.
- Lin, G., D. Fu, J. Zhu, and G. Dasmalchi (2009). Cloud Computing: IT as a Service. *IT Professional Magazine* 11(2), 10–13.
- Luxen, D. and C. Vetter (2011). Real-Time Routing with OpenStreetMap Data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 513–516.
- Ma, H., D. Zhao, and P. Yuan (2014). Opportunities in Mobile Crowd Sensing. *IEEE Communications Magazine* 52(8), 29–35.
- Maerschalk, G., A. Ueckermann, and S. Heller (2011). *Berichte der Bundesanstalt für Straßenwesen – Längsebenheitsauswerteverfahren "Bewertetes Längsprofil"*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Maglio, P. P., S. L. Vargo, N. Caswell, and J. Spohrer (2009). The Service System is the Basic Abstraction of Service Science. *Information Systems and e-Business Management* 7(4), 395–406.
- Malatras, A. and L. Beslay (2015). A Generic Framework to Support Participatory Surveillance Through Crowdsensing. In *Proceedings of the Federated Conference on Computer Science and Information Systems*, pp. 1135–1146.
- Mao, M. and M. Humphrey (2012). A Performance Study on the VM Startup Time in the Cloud. In *Proceedings of the 5th IEEE International Conference on Cloud Computing*, pp. 423–430.
- Mednis, A., G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo (2011). Real Time Pothole Detection Using Android Smartphones with Accelerometers. In *Proceedings of the International Conference on Distributed Computing in Sensor Systems and Workshops*, pp. 1–6.
- Menasce, D. A. (2002). QoS Issues in Web Services. *IEEE Internet Computing* 6(6), 72–75.

- Merlino, G., S. Arkoulis, S. Distefano, C. Papagianni, A. Puliafito, and S. Papavasiliou (2016). Mobile Crowdsensing as a Service: A Platform for Applications on Top of Sensing Clouds. *Future Generation Computer Systems* 56(1), 623–639.
- Miller, J. S. and W. Y. Bellinger (2014). Distress Identification Manual for the Long-Term Pavement Performance Program. Technical Report FHWA-HRT-13-092, Federal Highway Administration, Georgetown Pike.
- Mohamed, A., M. M. M. Fouad, E. Elhariri, N. El-Bendary, H. M. Zawbaa, M. Tahoun, and A. E. Hassanien (2015). RoadMonitor: An Intelligent Road Surface Condition Monitoring System. In *Intelligent Systems'2014*, pp. 377–387. Cham: Springer International Publishing.
- Mohan, P., V. N. Padmanabhan, and R. Ramjee (2008a). Nericell: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, pp. 323–336.
- Mohan, P., V. N. Padmanabhan, and R. Ramjee (2008b). TrafficSense: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphones. In *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems*, pp. 1–29.
- Morosiuk, G., M. J. Riley, and J. B. Odoki (2004). *Modelling Road Deterioration and Works Effects*, Volume 6 of *Highway Development and Management Series*. Puteaux: The World Road Association (PIARC).
- Můčka, P. (2016). Current Approaches to Quantify the Longitudinal Road Roughness. *International Journal of Pavement Engineering* 17(8), 659–679.
- Newson, P. and J. Krumm (2009). Hidden Markov Map Matching Through Noise and Sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 336–343.
- Nitsche, P., C. Van Geem, R. Stütz, I. Mocanu, and L. Sjögren (2014). Monitoring Ride Quality on Roads with Existing Sensors in Passenger Cars. In *Proceedings of the 26th ARRB Conference – Research Driving Efficiency*, pp. 1–13.
- Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers* 47(2), 617–644.
- Ord, J. K. and A. Getis (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27(4), 286–306.
- Ortenzi, F. and M. A. Costagliola (2010). A New Method to Calculate Instantaneous Vehicle Emissions Using OBD Data. Technical Report 2010-01-1289, SAE Technical Paper, Warrendale.

- O'Sullivan, D. and D. Unwin (2010). *Geographic Information Analysis*. New Jersey: John Wiley & Sons, Inc.
- Ouyang, Y. and S. Madanat (2004). Optimal Scheduling of Rehabilitation Activities for Multiple Pavement Facilities: Exact and Approximate Solutions. *Transportation Research Part A: Policy and Practice* 38(5), 347–365.
- Paterson, W. D. O. and T. Scullion (1990). Information Systems for Road Management: Draft Guidelines on System Design and Data Issues. Technical Report INU 77, The World Bank, Washington, D. C.
- Pham, H. and H. Wang (1996). Imperfect Maintenance. *European Journal of Operational Research* 94(3), 425–438.
- Prpić, J. (2016). Next Generation Crowdsourcing for Collective Intelligence. In *Proceedings of the Collective Intelligence Conference*.
- Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New Jersey: John Wiley & Sons, Inc.
- Rajamaki, J. and M. Vuorinen (2013). Multi-Supplier Integration Management for Public Protection and Disaster Relief (PPDR) Organizations. In *Proceedings of the IEEE International Conference on Information Networking*, pp. 499–504.
- Ratcliffe, J. H., T. Taniguchi, E. R. Groff, and J. D. Wood (2011). The Philadelphia Food Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots. *Criminology* 49(3), 795–831.
- Raza, M. H., A. F. Adenola, A. Nafarieh, and W. Robertson (2015). The Slow Adoption of Cloud Computing and IT Workforce. *Procedia Computer Science* 52(1), 1114–1119.
- Revuz, D. (1984). *Markov Chains*. Amsterdam: North-Holland.
- Rode, S. S., S. Vijay, P. Goyal, P. Kulkarni, and K. Arya (2009). Pothole Detection and Warning System: Infrastructure Support and System Design. In *Proceedings of the IEEE International Conference on Electronic Computer Technology*, pp. 286–290.
- Sahlholm, P. and K. Henrik Johansson (2010). Road Grade Estimation for Look-Ahead Vehicle Control Using Multiple Measurement Runs. *Control Engineering Practice* 18(11), 1328–1341.
- Savera, A., A. Zia, M. S. Edhi, M. Tauseen, and J. A. Shamsi (2016). BUMPSTER: A Mobile Cloud Computing System for Speed Breakers and Ditches. In *Proceedings of the 41st IEEE Conference on Local Computer Networks Workshops*, pp. 65–71.

- Sayers, M. W., T. D. Gillespie, and C. A. V. Queiroz (1986). The International Road Roughness Experiment – Establishing Correlation and a Calibration Standard for Measurements. Technical Report WTP 45, The World Bank, Washington, D. C.
- Sayers, M. W. and S. M. Karamihas (1998). *The Little Book of Profiling*. Michigan: University of Michigan.
- Schöbi, R. and E. N. Chatzi (2016). Maintenance Planning Using Continuous-State Partially Observable Markov Decision Processes and Nonlinear Action Models. *Structure and Infrastructure Engineering* 12(8), 977–994.
- Schölkopf, B., A. J. Smola, F. Bach, et al. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge: MIT Press.
- Sedaghat, M., F. Hernandez-Rodriguez, and E. Elmroth (2013). A Virtual Machine Re-Packing Approach to the Horizontal vs. Vertical Elasticity Trade-Off for Cloud Autoscaling. In *Proceedings of the ACM Cloud and Autonomic Computing Conference*, pp. 1–10.
- Seraj, F., B. J. van der Zwaag, A. Dilo, T. Luarasi, and P. Havinga (2014). RoADS: A Road Pavement Monitoring System for Anomaly Detection Using Smart Phones. In *Big Data Analytics in the Social and Ubiquitous Context*, pp. 128–146. Cham: Springer International Publishing.
- Simko, V. and K. Laubis (2017). R-Package "rroad": Road Condition Analysis. <https://CRAN.R-project.org/package=rroad>.
- Simpson, A. L. (1999). Characterization of Transverse Profile. *Transportation Research Record: Journal of the Transportation Research Board* 1655(1), 185–191.
- Smilowitz, K. and S. Madanat (2000). Optimal Inspection and Maintenance Policies for Infrastructure Networks. *Computer-Aided Civil and Infrastructure Engineering* 15(1), 5–13.
- Smith, J. and K. F. Wallis (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics* 71(3), 331–355.
- Sondik, E. J. (1971). *The Optimal Control of Partially Observable Markov Decision Processes*. Ph. D. thesis, Stanford University.
- Steenberghen, T., T. Dufays, I. Thomas, and B. Flahaut (2004). Intra-Urban Location and Clustering of Road Accidents Using GIS: A Belgian Example. *International Journal of Geographical Information Science* 18(2), 169–181.
- Stock, J. H. and M. W. Watson (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* 23(6), 405–430.

- Strazdins, G. and A. Mednis (2011). Virtual Ground Truth in Vehicular Sensing Experiments: How to Mark it Accurately. In *Proceedings of the 5th International Conference on Sensor Technologies and Applications*, pp. 295–300.
- Sugumaran, R., S. R. Larson, and J. P. DeGroot (2009). Spatio-Temporal Cluster Analysis of County-Based Human West Nile Virus Incidence in the Continental United States. *International Journal of Health Geographics* 8(43), 1–19.
- Sun, X., S. Hu, L. Su, T. F. Abdelzaher, P. Hui, W. Zheng, H. Liu, and J. A. Stankovic (2016). Participatory Sensing Meets Opportunistic Sharing: Automatic Phone-to-Phone Communication in Vehicles. *IEEE Transactions on Mobile Computing* 15(10), 2550–2563.
- Svenson, K. (2014). Estimated Lifetimes of Road Pavements in Sweden Using Time-to-Event Analysis. *Journal of Transportation Engineering* 140(11), 04014056-1–04014056-8.
- Tai, Y.-c., C.-w. Chan, and J. Y.-j. Hsu (2010). Automatic Road Anomaly Detection Using Smart Mobile Device. In *Proceedings of the 15th Conference on Artificial Intelligence and Applications*, pp. 1–8.
- Tan, F., T. Thoresen, and C. Evans (2012). Review of Vehicle Operating Costs and Road Roughness: Past, Current and Future. In *Proceedings of the 25th ARRB Conference – Shaping the Future: Linking Policy, Research and Outcomes*, pp. 1–24.
- Tedeschi, A. and F. Benedetto (2017). A Real-Time Automatic Pavement Crack and Pothole Recognition System for Mobile Android-Based Services. *Advanced Engineering Informatics* 32(1), 11–25.
- Thiagarajan, A., L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson (2009). VTrack: Accurate, Energy-aware Road Traffic Delay Estimation Using Mobile Phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pp. 85–98.
- Timmermann, A. (2006). Forecast Combinations. In *Handbook of Economic Forecasting*, pp. 135–196. Amsterdam: Elsevier.
- Titterton, D. H. and J. L. Weston (2004). *Strapdown Inertial Navigation Technology*. Cambridge: Academic Press.
- Torrence, C. and G. P. Compo (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79(1), 61–78.
- Ueckermann, A. and B. Steinauer (2008). The Weighted Longitudinal Profile. *Road Materials and Pavement Design* 9(2), 135–157.

- Unterharnscheidt, P. and A. Kieninger (2010). Service Level Management – Challenges and their Relevance from the Customers’ Point of View. In *Proceedings of the 16th Americas Conference on Information Systems*, pp. 1–8.
- Urgaonkar, B., G. Pacifici, P. Shenoy, M. Spreitzer, and A. Tantawi (2005). An Analytical Model for Multi-tier Internet Services and Its Applications. *ACM SIGMETRICS Performance Evaluation Review* 33(1), 291–302.
- Vapnik, V., S. E. Golowich, and A. Smola (1997). Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, pp. 281–287.
- Vasić, N., D. Novaković, S. Miučin, D. Kostić, and R. Bianchini (2012). DeJaVu: Accelerating Resource Allocation in Virtualized Environments. *ACM SIGARCH Computer Architecture News* 40(1), 423–436.
- Watanatada, T., C. Herral, W. Paterson, A. Dhareshwar, A. Bhandari, and K. Tsunokawa (1987). *The Highway Design and Maintenance Model: Description of the HDM-III Model*, Volume 1–2 of *The Highway Design and Maintenance Standards Series*. Washington, D. C.: U. S. Department of Transportation.
- Yagi, K. (2014). Collecting Pavement Big Data by Using Smartphone. In *Proceedings of the 1st IRF Asia Regional Congress*, pp. 1–8.
- Yazdanov, L. and C. Fetzer (2012). Vertical Scaling for Prioritized VMs Provisioning. In *Proceedings of the 2nd IEEE International Conference on Cloud and Green Computing*, pp. 118–125.
- Yi, W., K. Lo, T. Mak, K. Leung, Y. Leung, and M. Meng (2015). A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems. *Sensors* 15(12), 31392–31427.
- Zenonos, A., S. Stein, and N. R. Jennings (2016). An Algorithm to Coordinate Measurements Using Stochastic Human Mobility Patterns in Large-Scale Participatory Sensing Settings. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pp. 3936–3942.
- Zhang, N. and W. Liu (1996). Planning in Stochastic Domains: Problem Characteristics and Approximation. Technical Report HKUST-CS96-31, The Hong Kong University of Science and Technology, Hong Kong.
- Zhang, X., Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao (2016). Incentives for Mobile Crowd Sensing: A Survey. *IEEE Communications Surveys & Tutorials* 18(1), 54–67.

Zhao, D., H. Ma, L. Liu, and X.-Y. Li (2015). Opportunistic Coverage for Urban Vehicular Sensing. *Computer Communications* 60(1), 71–85.

Zhao, Y. (1997). *Vehicle Location and Navigation Systems*. Norwood: Artech House.