New Data Sources in Road Infrastructure Management

A game-based experiment into the effects of new data sources on condition assessment and decision-making within the operations and maintenance phase of asphalt paved road infrastructures

New Data Sources in Road Infrastructure Management

A game-based experiment into the effects of new data sources on condition assessment and decision-making within the operations and maintenance phase of asphalt paved road infrastructures

in partial fulfilment of the requirements for the degree of

Master of Science
in Construction Management and Engineering

at the Delft University of Technology,
to be defended publicly
on Tuesday April 18th, 2017 at 16:00.

By Barış Can Düzgün

Thesis committee:

Prof.dr.ir. A.R.M. **Wolfert**	TU Delft
PhD.ir. P.A. **Brous**	TU Delft
dr. D.F.J. **Schraven**	TU Delft
dr.ir. H.K.M van de **Ruitenbeek**	Volker InfraDesign

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Volker InfraDesign

New Data Sources in Road Infrastructure Management

A game-based experiment into the effects of new data sources on condition assessment and decision-making within the operations and maintenance phase of asphalt paved road infrastructures

Author

Barış Can **Düzgün** | 4253108 | bc.duzgun@gmail.com

MSc Construction Management and Engineering | Section of Integral Design and Management | Faculty of Civil Engineering and Geosciences | Delft University of Technology

Graduation committee

Prof.dr.ir. A.R.M. ***Wolfert*** CHAIR

Delft University of Technology | Civil Engineering and Geosciences | Integral Design and Management

PhD. P.A. ***Brous*** FIRST SUPERVISOR

Delft University of Technology | Technology, Policy and Management | Information Technology

dr. D.F.J. ***Schraven*** SECOND SUPERVISOR

Delft University of Technology | Civil Engineering and Geosciences | Integral Design and Management

dr.ir. H.K.M van de ***Ruitenbeek*** EXTERNAL SUPERVISOR

Volker InfraDesign | a VolkerWessels Enterprise

Cover and chapter illustrations based on photography by Hossein Zare

Abstract

Professionals in the Operations and Maintenance phase of national road infrastructure projects are making decisions with large consequences based on low frequency measurements and subjectivity prone expert observations. A solution is expected in sensor innovations, IoT and User Generated Data to generate more frequent measurements and less subjective observations. The use of these methods has been researched and proven, however the main focus of these studies was often improvement of the technological capabilities or implementation in current practises with limited research into their contribution and effects on professionals in the construction sector. This research describes an experiment that tested the effects of more data and more diverse data on the decision-making of professionals in the construction industry. An attempt to answer this question is made by modelling different data sources into a Serious Game and testing the assumptions. After analysis of the gaming data, the questionnaires and the debriefing it can be concluded that in this experiment there was a correlation between better assessments and higher scores. However experts did not assess damage differently when presented with extra information, nor did they make significantly different decisions. From the qualitative section of the experiment the explanation was found that the extra information proved too much and experts were able to extrapolate with the marginal data that represents the current industry practise. This suggests that new information requires training, as our built environment gets richer in terms of data, the assessment of this built environment becomes too much for humans to cope with and solutions can be sought in the application smart algorithms, Machine Learning and Artificial Intelligence.

Keywords: User Generated Data | Crowdsourcing | Big Data | Smart Asset Management | Object Generated Data | Asset Maintenance | Road infrastructures | Internet of Things | Game-based simulation experiment

Table of contents

List of Abbreviations

AM	Asset Management
ANOVA	Analysis of Variance
BIM	Building Information Model
DAB	Closed Top-layer Asphalt | Dicht Asfalt Beton
DBFM	Design, Build, Finance and Maintain
EGD	Expert Generated Data
IRI	International Roughness Index
OE	Operating Environment
OGD	Object Generated Data
PCD	Point Cloud Data
QoS	Quality of Service
RIS	Road Inspector
RWS	Rijkswaterstaat
SE	Systems Engineering
SOI	System of Interest
SPV	Special Purpose Vehicle
UGD	User Generated Data
ZOAB	Porous Asphalt | Zeer Open Asfaltbeton

List of tables

List of figures

Summary

Introduction - A change in the way of thinking and organising large scale construction projects has taken place in the last decade. Contracting Authorities are shifting their focus within the contractual relation from product oriented towards performance oriented with a long timeframe. This shift has implications for the Contractors, who in the past did exactly what was in the requirements brief, but now have to manage all aspects, related to Design, Build, Finance and Maintenance of an asset according to predetermined performance requirements. Contractors are struggling with this new role, resulting in poor management with the consequence of in large fines, higher maintenance costs, unplanned maintenance and unsafe situations (Cobouw, 2016). One of the processes that has an influence on the delivered quality is the management of information, something relatively new to the construction sector. According to TheIAM (2008) good Asset Management processes require information on many levels and the decisions regarding an asset are only as good as the information they are based on. Meanwhile the discipline of Information Technology is growing rapidly in terms of technological capabilities and practical usability with the coming of developments such as Big Data and Internet of Things a paradigm shift in Information Management is expected. Researchers and Commercial parties are eagerly experimenting with different ways of integrating these new technologies into current practise. An example of the large scale use of IoT sensors and Big Data Analytics in city management is the Smart Santander project (Sanchez et al., 2014). One of the domains in which new technologies in information management might be interesting is the domain of road management. A domain that, at the moment, is very traditional regarding the information that is used for decision-making. Very technical and yearly measurements used for extrapolation combined with more frequently conducted visual inspections based on Expert Judgement result in an infrequently updated database with elements that are susceptible to subjectivity. An increase in information quality can be achieved by using technological advancements to get more accurate, timely and complete data. The emergence and possibilities of these information technologies and the consequences of poor information management led to the definition of a problem within the management of roads. Stating that professionals in the Operations and Maintenance phase of national road infrastructure projects are making decisions with large consequences based on low frequency measurements and subjectivity prone expert observations. A solutions is expected in IT related sensor innovations that generate more frequent measurements and less subjective observations. The potential of these technologies is acknowledged by other sectors, but they still remain unused because their contribution is unknown an there is a lack of trust compared to established methods. Leading to the following research question:

"What is the effect of new data sources on the assessment of quality deterioration and decision-making of professionals in road management?" Supported by five guiding questions, of which the first three built up towards the operationalization of the method and the last two supported the answering of the main question. Due to time and resource related limitations of real-world testing and the subjectivity of

interviews, the effects on assessment and decision-making were determined by means of an experiment. An initial scope was set focusing on surface related anomalies because of their short time frame and the applicability of new sensing methods. Anomalies related to incidental damage were excluded, focusing solely on regular deterioration. Within the sensing methods the scope was set to exclude sensors that are specifically designed for one anomaly type.

Literature review - From literature reviews of pavement related anomalies, sensing methods, contractual requirements and exploratory interviews the scope of this research was defined more clearly. Because the effect of the information source on assessment and decision-making had central place within this research, a number of additional research choices were made. The research focused on the advantages of new data from the perspective of data completeness and relevancy. The aspects timeliness and accuracy were excluded because the expected advantage when time becomes a factor is too obvious and the accuracy of measurements is an issue that lies much more in the technical domain and would much more fit a question focussing on the trust of new data instead of the benefits when fully trusted. Data consistency was included in the research design such that tested methods were presented according to current practise and standards. Five anomalies were identified as relevant and fitting the scope being, Ravelling, Roughness, Rutting, Cracking and Potholing. From the sensors a selection was made based on uniqueness of sensors, ability to detect anomalies, ease of integration in the current practise and the possible added benefits resulting in in-car sensors as counterpart of the ARAN measurements and 3D Point Clouds as counterpart of the Visual Inspections. The use of IoT sensors was omitted because the impracticality in condition measurement or observation of asphalt deterioration. However experts agreed on the usefulness of IoT sensors in non-pavement related requirements such as drainage capacity, greenery management, alignment of road furniture and external factor measurements (rainfall, sun exposure etc.).

Method - To facilitate the testing of different sensor methods, anomaly types and quality conditions a research design is created that allows for these elements to be included without requiring an unmanageable amount of participants or wasting resources. The fractional factorial experimental design was chosen with three factor combinations, base quality with two levels, the data source with sensor combinations with six levels and the deterioration rate corresponding to the chosen anomalies with three levels, this resulted in 36 design points (unique combinations of the factors sensor, anomaly and quality). A replication was built in to increase the statistical significance of the experiment, resulting in 72 test moments per participant. The participants were selected based on their job profile. The desired profile was either engineering orientated (knowledge on pavements, interpretation of data, deterioration behaviour and maintenance actions) or process orientated (contractual requirements, mechanisms and operational effects) and was even more desired if they had management experience (resource management, strategic planning and organisational goals). The optimum number of required participants according to the Resource Equation Method was between 7 and 14, this target area was reached with 11 participants. The operationalisation of the experiment was performed using the

sandbox building game, Minecraft, which allows for elaborate data logging. Combining all this information into testable statements resulted in three propositions with a total of 13 hypotheses. The first proposition suggested that the use of data sources containing new data sources additional to the current practise data sources results in more accurate assessments of quality conditions for all anomaly types. For each of the sensing methods (aimed at more frequent measurements and more objective observations) and the combination, a hypotheses is formulated. Additionally it is suggested that the use of these data sources is linked to longer assessment times. The second proposition suggested that the use of data sources containing new data sources additional to the current practise data sources results in better maintenance decisions being made. To determine the quality of a decision a construct was designed resulting in a game score per decision ranging from -2 to 2. A third proposition was suggested to make the link between the accuracy of an assessment and decision score. It was suggested that a participant that estimates a quality condition more accurately will make better decisions and will use more precise actions in their decision. Subsequently the process of serious game design was walked through according to the theories of Harteveld (2011) and Meijer (2009). The Product Design Sprint methodology (Banfield, Lombardo, & Wax, 2015; Reis, 2011) was used to perform three sprints before coming to an end product suitable for the experiment. The resulting game placed the participant in the shoes of an Asset Manager in the roads domain. They were asked to decide on the maintenance actions until the start of 2018. After 2018 large maintenance was planned to exclude long term thinking and clouding of the measurements. The objectives of the participants were to assess the quality of the road with the information they were given, to decide how large their action would be and to use as small actions as possible. At the end of the game they were given a score based on the actions they had taken, the correctness of an action and the size of the action. To fit the experiment design in the game a section of road was divided in six segments with each dedicated to a single sensing method and containing different deterioration types and base qualities. The experiment sessions were run according to a strict red line. Starting with a brief introduction, followed by a tutorial, a final verification was performed to check if all objectives and procedures were clear before the participants were left alone. After the experiment a debriefing took place, which resulted in an extensive set of qualitative information to support the quantitative gaming dat. To test the propositions and hypotheses a number of different data point were collected, being assessment, time to asses, decision, size of the decision, location of the decision, total time for a road section and the total time for a road segment.
Results - The analysis of proposition one relied on the use of a Repeated Measures ANOVA to test if within a participant there was a significant difference between the assessments and timings. Testing for the effects of new data sources proved not significant for all three hypotheses. The additional hypotheses related to timing also proved not significant for all three hypotheses. For the analysis of proposition two a Repeated Measures ANOVA was used to test if within a participant there was a significant difference between the decision scores. Testing for the effects of new data sources on assessment proved not significant for all two of the three hypotheses.

Hypothesis 6 (combination of more frequent and more objective information lead to better decision scores) did prove significant on one of the two occasions. However because it contradicted the second test within the same hypotheses the null-hypothesis was not rejected due to inconclusive results. For proposition three Kendall's Tau-b correlation test was used to test is there was a correlation between accuracy of assessment and the decision score as well as the accuracy of assessment and the size of the decision. Testing for the decision score proved significant at the p = 0.01 level (τb = .273, p = 0.000). Suggesting that participants that assess more accurately also make better decisions. Testing for the size of intervention showed that there is some correlation between accuracy and precision of the decision, however the results did not report significant. The qualitative analysis found a possible explanation of these effects and uncovered some additional effects. It appeared that a large sum of the participants did not use the information they were given apart from what they were already familiar with. They found the additional information too extensive, unnecessary for assessment and overloading such that extrapolation was unmanageable. The experts stated that they were able to extrapolate with the marginal data that represents the current industry practise. They found the process and the information to be realistic but missed the presence of contractual elements such as cost of fines, budgets and coupling of actions. Nearly all participants mentioned that the presence of these elements would have significantly changed their choices.
Discussion - Possible explanations are that new information requires some more getting used to before it can or will be used by professionals. Another explanation is that the effects of these data sources become more evident when time becomes a factor and the obvious benefits are allowed into the test setting. It may also be possible that the use of more and different sensor data does not affect decision in the management of roads if the consequences of violating requirements is not known. This suggests that decisions are based on the financial consequence and not on the technical quality. Another explanation is that as our built environment gets richer in terms of data, the assessment of this built environment becomes too much for humans to cope with and solutions can be sought in the application smart algorithms, Machine Learning and Artificial Intelligence. But until that time, more information from different data sources does not appear to be actively used by professionals.
Conclusion - Concluding from this research it can be said that in this study the use of new data sources did not show any effect on the assessment or decision-making of professional in road management. To achieve the step towards the use of new data sources within the construction industry, three recommendations are made. (1) Perform pilots studies with promising sensor types, (2) use this information to design and optimise algorithms that rely less on human capabilities, (3) verify the results by comparing them to the current methods. A fourth unofficial recommendation is based the integration of Contracting Authority, Contractors and Engineering companies to collaborate in the creation of new methods for increased industry support and incorporation in future Contracts. The future of our world will be shaped by how we are connected to our environment, this research is a small step for the civil industry towards being part of that connected world.

Acknowledgements

I would like to thank my chair Prof. dr. ir. Rogier Wolfert for his direct attitude, clearness and the connections he provided. Additionally I would like to thank him for talking with me about my future plans and giving me ideas and new inspirations to pursue. I would like to thank Paul Brous for introduction me into serious gaming, his help with building a virtual world and for providing the Minecraft servers at the start of the research. I particularly enjoyed the fish Fridays together with him and Jasper Spiegeler. I would like to express great appreciation towards Dr. Daan Schraven for his stern guidance, becoming increasingly involved as the research progressed. Your advice helped me to make significant improvements to the structure and readability of this report and my academic thinking. My grateful thanks are also extended towards dr.ir. Martinus for the inspiration you gave me and for welcoming me in the software developers family at Volker InfraDesign. Specifically our talks about that what you do in graduation shapes you a little bit as a person. Looking back on this, I can completely agree, since I have gotten increasingly interested in serious gaming and simulations as a research method. Lastly I would like to acknowledge Prof. dr. ir. Pieter van Gelder for inspiring me with his lectures on data analytics and our following conversations on Big Data, Internet of Things and crowd sensing applications. It was my main trigger for performing this research.

My gratitude goes out to everyone within Volker InfraDesign and VolkerInfra with whom I talked during the development of this research. My special thanks goes out to the interviewees and gaming participants from VolkerInfra, BAM, Boskalis and GAIM without them this research was not possible.

In special I would like to thank Jasper Spiegeler, for our endless debates in which we helped each other get completely off track. But also helped each other regain stronger than before. I hope that in the future we can do another game together.

I would also like to thank my family, friends and sports team for bearing with me through some moments in which I seemed to have vanished off the face of the earth.

Last but definitely not least, Nienke, for helping me throughout the entire process. You were my biggest supporter and a great conversational partner to philosophise about my research with, even at moments you did not want to.

1 Introduction

The construction industry is known for its lack of innovativeness and the adoption of new technologies. Changing views on the management of large infrastructure assets require Contractors to become increasingly efficient. Opportunities are seen in the employment of emerging technologies such as Internet of Things and Big Data Analytics. However their contribution to asset management processes are still unknown. The aim of this research is to contribute to the innovations and adoption of emerging technologies in the construction sector. In this chapter the elements laying at the foundation of this research will be presented giving a brief overview of the current situation leading to the problem definition, the research objectives and a presentation of the research question. The chapter will be concluded by presenting the general method and an overview of how the report is built up.

1.1 Asset Management and connectedness of the world

In this section the stage will be set along the line of topics that were considered interesting and relevant in the current world of civil engineering. These topics were a main source of inspiration in the definition of this thesis.

1.1.1 The shift in Asset Management

In the last decades the awareness to perform Asset Management (AM) has increased among asset owners. The design, construction, operation and maintenance of an asset requires large monetary, material, time and data related resources and the consequences of inadequate management are large e.g. high costs, asset unavailability, congestion and decreased safety (PIARC, 2014, pp. i-iii). To improve on the integral management their focal points have shifted from the management of uncoupled aspects towards the management of an asset's entire life cycle. This lifecycle perspective on managing infrastructure assets requires information on many organisational levels and the decisions regarding an asset are only as good as the information they are based on. Thus information management is considered to be a core activity within AM (TheIAM, 2014).

1.1.2 The connectedness of the world

At the same time the connectedness of the world is increasing at a rapid rate. A world where huge amounts of data are created by humans as well as machines that are connected with each other through many (open) platforms. A world where technological advances in Information Technology (IT) such as the Internet of Things (IoT) and crowd sourcing of data combined with Big Data Analytics are the next paradigm in Information Management. These developments are increasingly applied by researchers and commercial parties, who use connected devices, participatory platforms and Machine to Machine (M2M) communication to improve their research or businesses. Such as the use of crowdsourcing to answer questions about the world around us in a research project called NervousNet (Helbing et al., 2012). Or a community based app for real-time traffic predictions, improvement of roadmaps and detection of traffic jams to give the best traffic route as possible (Waze, 2016). The use of crowd sourced data is not only interesting in transportation engineering but also in safety engineering. Kong et al. (2015) are using smart phones for the detection of earthquakes around the world, they have written an algorithm that recognises unique earthquake vibrations with the phone's accelerometer and is able to send warnings to other areas just seconds before the shocks hit. These kind of information based innovations are rapidly increasing on many kinds of services.

1.1.3 The role of information technology in Asset Management

These emerging technologies also have the potential to improve AM processes enabling Contractors to become more efficient by increasing data quantity and quality, and improving their decisions. Currently pilot projects are being performed in the different phases of an asset's life cycle. In Design the combination of 3D models and augmented reality come together in interactive 3D-models (VolkerInfraDesign, 2016). In Construction sensors assess effects of certain parameters on final asphalt quality (Bijleveld & Doree, 2013) this information can also be used in subsequent phases. In Operation and Maintenance assessment of road conditions is done by using smart phones in cars (Seraj, van der Zwaag, Dilo, Luarasi, & Havinga, 2016) and in-car sensors (Van Geem et al., 2016). This last mentioned phase, the Operations and Maintenance phase, in the road infrastructures domain is particularly interesting for researching the applicability of IT innovations due to the many users, lack of matureness in current processes and the increasing importance of integral contracts which require adequate Management of Assets over longer time periods.

1.1.4 Need for improved Operations and Maintenance of roads

Even though current research in this field is progressive, the gap with current practice remains large. In the Operation and Maintenance phase of Dutch road infrastructures informational needs are addressed by means of two methods. Specialised measuring vehicles are deployed once per year to perform precise measurements (TNO, 2010). However their frequency of measurement is low due to time and resource intensity (Dawkins, Bishop, Powell, & Bevly, 2011, p. 4). To support the decision-making process intermediate data points are needed. These are collected by Road Inspectors (RIS) who draft reports and provide remarks and pictures in management systems. The importance of having timely and correct information becomes evident when contractual mechanisms are included. If quality conditions are not accurately estimated, incorrect decisions can be made. Underestimation of actual quality can result in premature repairs and inefficient use of resources. Overestimation of actual quality can result in unsafe road situations and road damages resulting in large contractual fines, higher maintenance costs and unplanned maintenance (Cobouw, 2016). Experts acknowledge the need for maturing and professionalization of the maintenance processes within management of roads and look at innovative technologies for the solutions (Interview-VI, 2016).

1.1.5 Opportunities for IT innovations in Road Management

New technologies have the potential to contribute in the abovementioned improvement by enabling measuring with more frequency and less subjectivity. However promising these techniques are, they still face difficulties because their contribution is unknown and there is a lack of trust compared to established methods (Brous & Janssen, 2015). The real world testing of their effects is considered time and capital intensive due to design, development and implementation aspects and interviews often lack objectivity by being too conservative (TNO, 2010) or too optimistic. The researches that have been performed yield promising results, however the main focus of these studies was often improvement of the technological capabilities or implementation in current practises (Rezgui & Zarli, 2006), with limited research into their applicability and effects on professionals in the construction sector.

1.2 Problem statement

From the different sub-problems a general problem statement is produced:

Currently decisions in the Operations and Maintenance phase of road infrastructures are based on low frequency measurements and subjective observations. Technological innovations offer possibilities but remain unused because their contribution is unknown.

1.3 Research objective and questions

The objective of this research is to understand the usefulness of data sources, related to technological advancements, which are new to the construction sector. An effect on maintenance decisions is looked for by examining the assessment of quality preceding a maintenance decision involving professionals in the operations and maintenance phase of pavements. Leading to the following research question:

What is the effect of new data sources on the assessment of quality deterioration and decision-making of professionals in road management?

A set of guiding research questions are presented to work towards the main question:

1 – How does the system of Operations and Maintenance of road management work and what are its important aspects?

2 – Which sensing methods are currently in use, in development or envisioned for the future that address the relevant distress modes, according to literature and industry standards, in roads management?

3 – What information is needed by professionals in the Operations and Maintenance Phase of roads for the purpose of decision-making?

4 – What are the differences in assessments and decisions of professionals in road management given a certain data source?

5 – How do professionals assess quality and make a decision within the context of road management?

1.3.1 Scope

Within the Dutch construction sector the most obvious domains where user and IoT based data sources can be combined with Asset Management is the transportation and buildings domain, due to the relatively large amount of unique users with private means of transportation. Both containing different subdomains of which the focus will be on road infrastructures due to practical and academic reasons. Practically it is an interesting area of research because of the extensive amount of users, the possibilities for new sensing techniques involving the wisdom of crowds and the emergence of autonomous driving, the need for different management due to long term contracts and the size in volume and cost of national roads. The perspective from contractor side is especially interesting due to the operational nature of new data sources. Academically it is an interesting area of research because a large body of work is dedicated to the Design and Construction phase as well as the technical capabilities of emerging technologies. Also there are few studies, in terms of Asset Management optimisation, that have been dedicated to the Operations and Maintenance phase and the effects on decision-making processes.

1.4 Research design and structure

The guiding research questions are answered by employing a method. Together these answers build towards the answering of the main research question. The combination of these research elements results in the research design. An overview of this research design and the report structure is presented in Figure 1-1.

1.4.1 Research design

Guiding question - 1 - How does the system of Operations and Maintenance of road management work and what are its important aspects?

The first guiding question contributes to answering the main question by giving an overview of how the Operations and Maintenance phase in Road Management works. It was answered by viewing the O&M phase from a Systems Engineering perspective. The framework of Wasson (2005) was used to define the O&M phase within road management as a system and to analyse its relevant aspects. The information needed for the analysis was obtained from a literature review and exploratory interviews. It provided a systematic and structured approach to collect and organise the relevant data. The first part of Chapter 2 is dedicated to this question resulting in a system diagram showing the most relevant aspects of the system. The comprehensive version can be found in Appendix I.

Guiding question - 2 - Which sensing methods are currently in use, in development or envisioned for the future that address the relevant distress modes, according to literature and industry standards, in road management?

The second guiding question contributes to answering the main question by giving an overview of current and envisioned sensors and the anomalies they can detect within the context of the Operations and Maintenance phase in Road Management. It was answered by means of a literature review and expert validations to identify predominant anomalies and determine the current sensing methods in road management. By means of exploratory interviews in different sectors, with researchers, and professionals and literature on emerging technologies, an overview was created of promising sensing methods. The second part of Chapter 2 presents the findings, resulting in an overview of anomalies, an overview of sensors and a sensor-anomaly cross reference with respect to industry standards. The comprehensive version for sensors in Appendix II.

Guiding question - 3 - What information is needed by professionals in the Operations and Maintenance Phase of roads for the purpose of decision-making?

The third guiding question contributes to the main question by defining what decision-making looks like within the organisation. It was answered by means of a literature review and expert validation to determine the information needed by professionals in road management. The flow of information within the organisation, the decision-making process and the measures matrix were defined. The second part of Chapter 2 ends with the answer to this guiding question and the resulting model for decision-making.

Operationalization of the research

The last part of Chapter 2 combines this information into a theoretical model and presents the constructs, propositions, variables and hypotheses. It defines the input and paves the way towards the research operationalization in Chapter 3. To answer guiding question 4 and 5 an experiment design was employed to answer the main question, given the time and resource related limitations of long term real world testing and the subjectivity of interviews. A serious gaming theory (Harteveld, 2011) was used to simplify the complexity of reality uncovered by guiding questions 1 to 3.

Guiding question - 4 - What are the differences in assessments and decisions of professionals in road management given a certain data source?

The fourth guiding question was answered by means of a quantitative data collection as part of a game-based simulation experiment. The quantitative data made the effects on assessment of quality deterioration and the following decisions measurable and allowed for the testing of the hypotheses. The results are presented in the first part of Chapter 4.

Guiding question - 5 - How do professionals assess quality and make a decision within the context of road management?

The fifth guiding question was answered by means of a qualitative data collection i.e. debriefing theory (Van den Hoogen, Lo, & Meijer, 2014) as part of a game-based simulation experiment. The qualitative data gave insight into the “black-box” of a professionals’ assessment of quality deterioration and the following decisions by means of introspection (Collins, n.d.). It allowed for the explanation of the results from guiding question 4. The results are presented in the second part of Chapter 4.

Conclusion of the research

The final chapters are dedicated to the translation of the results from guiding questions 4 and 5. In Chapter 5 the results are linked back to reality and the meaning is discussed. In Chapter 6 conclusions are drawn, the practical and academic implications are presented, the limitations are discussed and recommendations are made for the practise and for further research. The chapter is closed with some concluding remarks and a spot on the horizon.

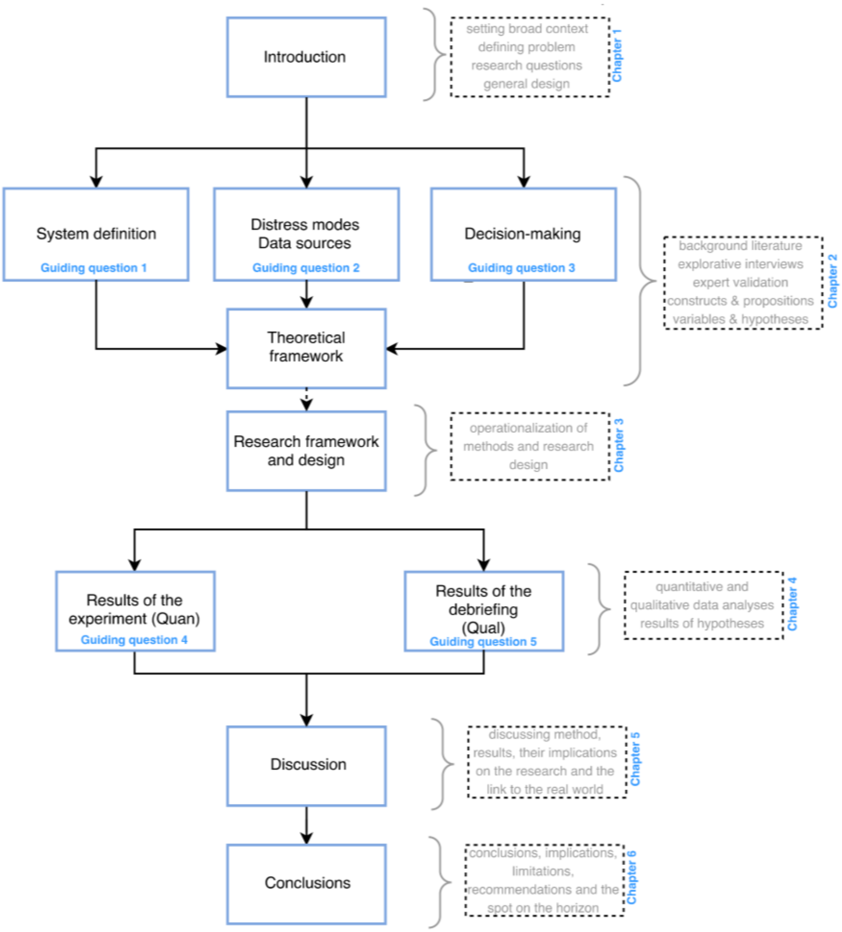


Figure 1-1: Research design and structure

1.4.2 Relevance of this research

This research is intended as a step towards using crowdsourcing in design, construction and asset management of (civil) construction projects and one of the first steps towards Smart Asset Management. With the underlying idea that thousands of daily users who experience the comfort and condition of a road can give an additional, more up to date and continuous, status of a road than expensive high precision testing vehicles who test the roads on yearly basis. In addition it contributes to the research of PhD Paul Brous, who studies the effects of IoT data sources in AM organisations and data governance on data quality. In parallel it contributes to and gains from the research of Jasper Spiegeler, who focuses on the effects of new data in management of concrete bridges. The use of gaming based simulation is a field which still offers much potential within the domain of construction, thus adding to the academic significance of this thesis. This research is geared towards the belief in a future scenario where the shift from Asset Management is made towards Smart Asset Management by combining a diversity of data sources ranging from crowd sourced to expert generated to realise a better understanding of the assets condition.

2 Literature review and preliminary analysis

The previous chapter set the stage for this research by providing knowledge on the main subjects and presenting the questions and general method via which an answer was looked for. This chapter focuses on a literature review and the findings of explorative interviews and case studies to answer the first three guiding questions. A number of theoretical elements lie at the foundation of this research, these elements form the inputs for the theoretical framework. The first element is the complex surrounding of Operations and Maintenance within the Asset Management of Road Infrastructures. It allows for the understanding of inputs and outputs and the creation of a complete picture of this phase in construction projects. The second element is related to technological advancements, the understanding of pavement quality and the link between those two. It allows for the selection of relevant sensing methods and anomalies. The third element is related to the information flows and decision-making within Road Management. It allows for the understanding of how information is processed and decisions are taken within the organization. It is concluded by presenting the theoretical framework built up from the elements defined in this chapter.

2.1 The current practice of Road Management

To unravel the current practice of road management, a Systems Engineering framework (Wasson, 2005) is deployed in order to give an answer to the first guiding question.

2.1.1 Road Management as a System

This method provides a systematic and structured way of breaking down complex information belonging to e.g. projects, objects or services by defining them as a system. The "black box" in between involves certain processes, events and relationships that together make up a characteristic way of how the system entity translates these inputs into their belonging outputs accordingly. It simplifies processes without affecting the purpose of the process. These processes, events and relationships are defined and built up by several attributes that have an impact on the system's way of working (stakeholders, roles, missions & objectives, resources, controls, opportunities, threats and physical constraints) seen in Figure 2-1. A system analysis is performed for the Operations and Maintenance phase within Road Infrastructure Management, the extensive version is found in Appendix I.

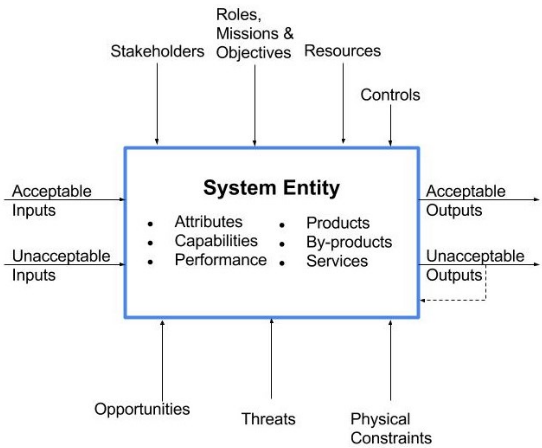


Figure 2-1: Wasson's System Entity Construct

System Roles, Missions & Objectives and Stakeholders,

The role of the system is to monitor and control the road with the intended goal of providing a qualitative and safe transportation system with high availability (Haider, Koronios, & Kumar, 2012; Rijkswaterstaat, 2017). The Assets are managed by the executive agency of the Ministry of Infrastructure and the Environment, better known as "Rijkswaterstaat" (RWS) or the Contracting Authority (CA). In practice the management of these civil infrastructures is outsourced to consortia of companies joined in a Special Purpose Vehicle (SPV) referred to as the Contractor (Engel, Fischer, & Galetovic, 2010). Based on system requirements the Contractor Designs, Builds, Finances and Maintains (DBFM) these infrastructures for long periods of time. The Contractor employs an executive organisation for the execution of maintenance and repairs in order to keep the road at maximum availability for the users of the road (Coval, Jurek, & Stafford, 2009). In order to preserve an asset's safety, comfort and total lifespan, maintenance is needed (VBW-Asfalt, 2005) together with a maintenance strategy. This maintenance strategy can be classified as one of three, being predictive, preventive or corrective maintenance (Haarman & Delahaye, 2004). Predictive Maintenance is aimed at the highest asset availability and requires an active maintenance organisation that is aware of their asset's behaviour when approaching failure. It employs (semi-)continuous monitoring to detect this near failure state. Secondly, preventive maintenance, which is either condition-based (measured with intervals and acted upon when a certain condition is reached) or planned (determined up front and executed as planned). Lastly Corrective Maintenance is the most basic strategy with the lowest availability. It uses minimal measurements and actions are undertaken after failure has occurred, as seen in Figure 2-2. An organisation that wants to move to Predictive Maintenance, needs better knowledge regarding their asset's deterioration (Morosiuk, Riley, & Odoki, 2004) and needs state of the art monitoring systems (Haarman & Delahaye, 2004). In current road management the employed strategy is often aimed at Preventive Maintenance (Interview-VI, 2016).

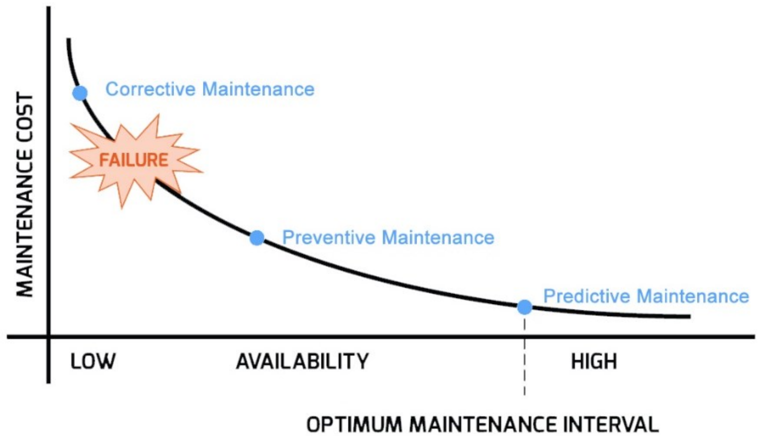


Figure 2-2: Maintenance strategies v availability, adopted from keelsolution.com

Controls

The Contractor is controlled by a diversity of mechanisms and governing principles. First and foremost the Contractor is bound by laws, regulations and municipal agreements. Secondly (and more influential in project specific terms) it is controlled by the Contract which stipulates the contractual obligations and requirements regarding a number of aspects. The most important are the Quality of Service and the Asset Availability, which are defined in the contractual requirements. The Contractor needs to traceably prove the fulfilment of these requirements and is periodically audited by the CA. These requirements are linked to the Payment Mechanism and impact the Contractor´s responsibility towards its sponsors (Kenniscentrum, 2008). To achieve these service levels and requirements the Contractor views an asset on different levels of its life cycle (Haider, 2012). Firstly, strategic, focused on the total lifespan. Secondly, tactical focused on the long term with large maintenance cycles (8-10 years). Thirdly, operational focused on short term and per year (Interview-VI, 2016). The actions on life cycle level are dedicated to a business strategy and changing needs. While the long and short term are focused on the operationalisation of the strategy and long term plan (Haider, 2012).

Resources

To achieve the fulfilment of service levels, the Contractor has Resources both internally as well as those flowing from the Contract. The main Resources are the monetary budget, the time budget for construction, the time budgets for maintenance in planned Vehicle Lost Hours, and the time budgets for rehabilitation of the asset in case of incidental unavailability. The monetary budget flowing back from the contract (in case of a DBFM) is generally expressed in Quarterly Payments and in a One-time payment (Rijkswaterstaat, 2009). Lastly the Contractor has internal budgets related to deployment of Personnel, Equipment and Construction materials.

Opportunities and Threats

Opportunities for the System are that proven Quality of Service can lead to good reputation and a higher chance of future projects, the improvement of processes which can be adopted by other projects (Dubois & Gadde, 2002), that technical components may perform better than expected and the ability to monitor and maintain quality such that discounts stay to a minimum. Threats to the system are the inability to meet requirements resulting in large discounts (Cobouw, 2016), underestimation of financial aspects related to the technical components, misjudgement of the planning in terms of time, the quality of construction and repairs, and external conditions leading to unexpected deterioration (Morosiuk et al., 2004).

Inputs - Data Resources

Next to the abovementioned resources a Contractor also employs data resources, which are the primary inputs. Asset information is fed into management systems, relying on basic tools which are easy to understand (Fujita, Kobayashi, Hoshino, & Chanseawrassamee, 2016; Interview-VI, 2016). In existing projects the CA has historic information stored in their internal information facility (abr. CIV), containing data collected by own means or external measurements. The information in Road Management often consists of used asphalt types, a road inspectors' remarks, measurement data in spreadsheets (excel), text processing (word), two dimensional maps (CAD), location related information (GIS) and occasionally information such as driving heights of constructions (Rijkswaterstaat, 2007). Though the CA provides the Contractor with this information at project commencement, in practise this information is often incorrect, outdated or missing (Interview-VI, 2016). To compensate for this, the Contractor performs a null-measurement at the start of a project to check if the provided information is correct.

Outputs

During the lifecycle of the Asset the Contractor periodically collects information on the current state of the asset and keeps track of the maintenance operations. This information is feed-backed into the short term, long term and life cycle planning and in some cases into the learning ability of the organisation as project experience e.g. deterioration curves (Interview-VI, 2016). The contractor performs actions on an asset e.g. maintenance or grey field construction, physical output, and in the process generating new data, digital output. Periodically the CA request this data to be sent back to the CIV, in digital containers called Coins, to update their asset database. This communicational output occurs internal, external and with higher order systems both hierarchical and peer level. Additional to the planned actions of the Contractor there are also unplanned actions as reactions to deviations between expected condition and measured condition (Haarman & Delahaye, 2004). The choice of these countermeasures depends on the moment in time and the severity of the situation.

2.1.2 Conclusion: System of Road Management

The first guiding question stated “How does the system of Operations and Maintenance of Road Management work and what are its important aspects?” To answer this question a system diagram is presented which contains all aspects found relevant for the core purpose of the system (Figure 2-3).

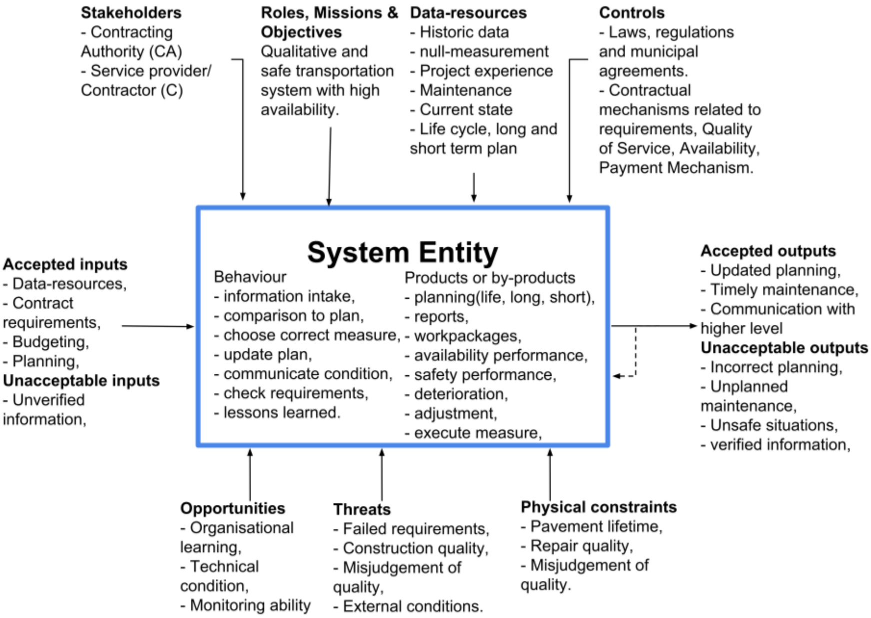


Figure 2-3: Wasson's System Entity Construct for O&M of road management

To summarise: the Operations and Maintenance of Road Management is characterised by the relation between CA and the Contractor. The Contractor is held responsible for providing a qualitative and safe transportation system with a high availability and bound to the contract and contractual mechanisms by performance requirements. The system uses data-resources to engineer maintenance plans on different levels within the life cycle and uses these plans to check if the current condition meets the expected condition. To react to its environment the system has measures and counter-measures which can be deployed in specific situations. In order to effectuate a high asset availability the system is dependent on its ability to transform data related input into realistic plans which can be followed without failing requirements and resulting in timely maintenance actions.

2.2 Current knowledge on distresses in porous asphalt

Porous asphalt paved roads are the most common road type in the Netherlands (TNO, 2010). They are similar to other heavy duty roads and, depending of their use case, consist of several layers as seen in Figure 2-4. At the base it consists of a sand bed, known as the subgrade. On top of this subgrade one or two layers of unbound granular are placed for drainage and stability purposes. The top package often contains multiple layers of asphalt and a porous top layer (Molenaar & Houben, 2002).

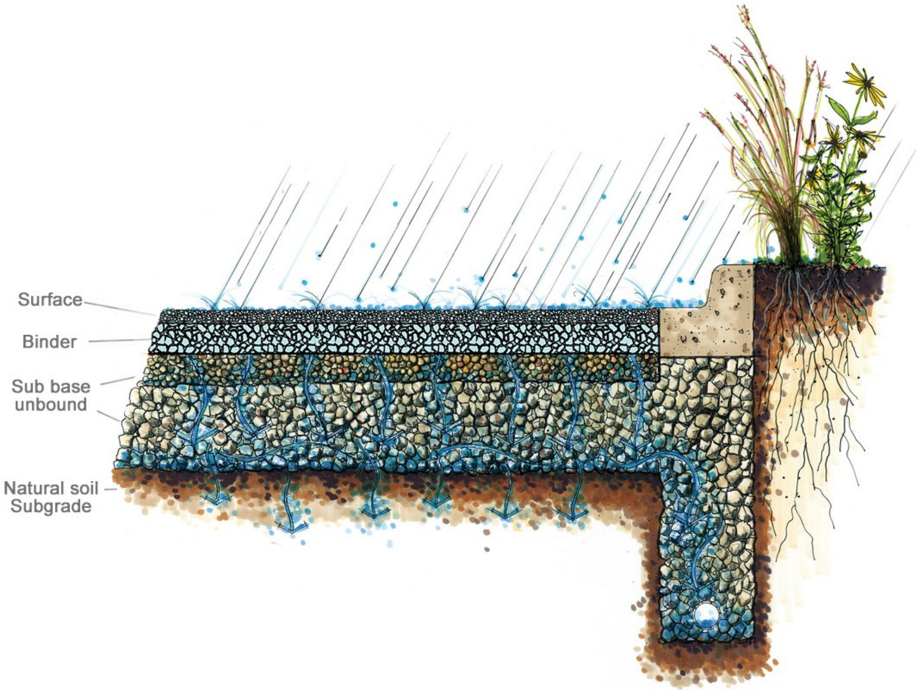


Figure 2-4: Pavement Cross Sections, altered, original by Doug Adamson

2.2.1 Main groups of distress modes

In general two main distress groups are distinguished in road management, structural and surface damages. Structural damages occur when the structural elements i.e. the foundation, cannot handle the exerted forces caused by vehicle loading. It leads to small movements in the sub layers and causes damages in the bottom layers which slowly spread to the surface. Surface related damages are those that are caused by wearing out of the binding component, the bitumen, in asphalt pavements due to contact with vehicles, natural elements (i.e. UV, moisture, oxygen) and incidents (TNO, 2010, p. 7). In the Netherlands, short term management of pavements pertains to surface distresses (KOAC-NPC, 2016) and long term to structural distresses.

2.2.2 Surface distress modes

In 1987 the World Road Association (PIARC) defined an international scale for mapping road surface irregularities in wavelengths. Ranging from Micro texture (< 0.5 mm) to Unevenness (>50m), this is displayed in Figure 2-5. Ideally the smaller wavelength irregularities of the surface (<10mm) should have large amplitudes giving them texture, providing "grip" and adding to the drainage abilities of the road. And the larger wavelength irregularities (>10 mm) should have small amplitudes giving the road smoothness and a comfortable "bumpless" driving experience (Dunford, 2013, pp. 9-84,25). Irregularities within these wavelengths result to different types of anomalies which play a part, either small or large, in determining the remaining lifetime of the pavement. The Dutch knowledge platform CROW has systematically documented these parameters in their inspection manuals. They have organised anomalies in damage groups, distinguishing several types of damage. Table 2-1 shows this systematic documentation.

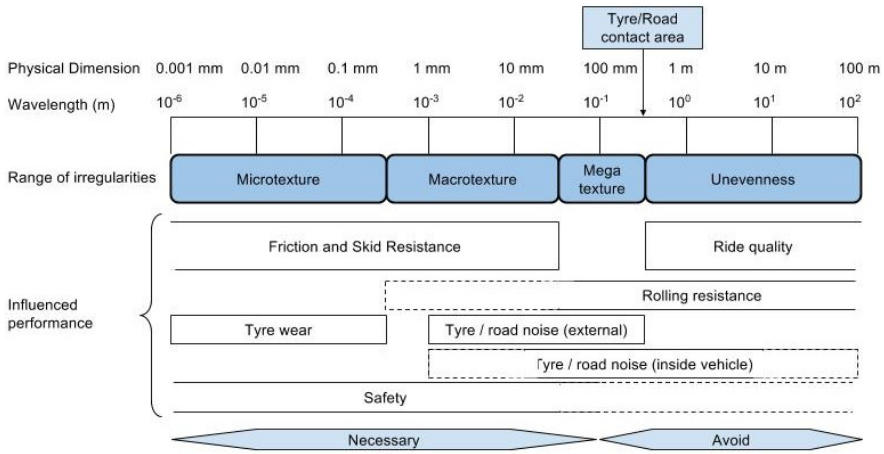


Figure 2-5 Distress wavelengths adopted from Dunford (2013)

Table 2-1: systematic classification of damage groups by CROW (2011a):

DAMAGE GROUP	ANOMALY	IRREGULARITY RANGE
Texture	Skid resistance	Micro/ macro texture
	Ravelling	Macro texture
Roughness	Longitudinal	Unevenness
	Transversal (rutting)	Mega texture/ unevenness
	Unevenness	Unevenness
Consistency	Cracking; longitudinal, transversal, crackling	Macro/ Mega texture
Shoulder	Edge damage	N.A.
Others	Ride Comfort	Unevenness
	Vibration hinder	Macro/ Mega Texture
	Traffic noise	Macro/ Mega Texture
	Potholes	Unevenness
	Welds	Mega texture/ unevenness
	Drainage	N.A.
	Berm	N.A.
	Settlement	Unevenness
	Markings	N.A.
	Furniture	N.A.

2.2.3 Anomalies acknowledged by regulating agencies

In the Netherlands most of these identified parameters are embedded in different systems and regulations regarding quality. The parameters that are acknowledged by the Dutch national road administration are longitudinal profile (Roughness), cross section profile (Rutting), Skid Resistance, Ravelling and Crackling. These parameters are used for their internal long term planning of road maintenance (Lub, 2016). They are also imposed onto commercial parties by incorporation into different contract types and quality warranties (DWW, 2002; Interview-VI). In DBFM and performance contracts the following anomalies are penalised: Ravelling, Crackling (transverse, longitudinal and craquele), Cross Sectional Profile (Rutting), Longitudinal Profile (Roughness according to IRI), Skid Resistance and Transversal Slope (DWW, 2002) where they are linked to performance related penalties.

2.2.4 Selection of predominant anomalies

Since not all parameters affect the quality and deterioration of asphalt pavements equally, a selection of parameters is made that have the most influence on a pavement's lifetime. According to literature review these parameters are the longitudinal profile, cross section profile, the International Roughness Index (IRI), macro texture, rutting, ravelling and crackling (Koch, Georgieva, Kasireddy, Akinci, &

Fieguth, 2015). Hunt and Bunker (2003) have concluded from extensive literature research that roughness is the parameter that correlates best with the condition of the road and the user's ride comfort. This is also confirmed by Dawkins et al. (2011) and validated in an interview with Prof.dr.ir. Erkens. According to them the International Roughness Index (IRI) is the best measurement factor for detecting road deterioration. Sayers, Gillespie, and Paterson (1986) have designed a range of IRI values fitting a pavement category ranging from super flat (0 , 2.0], new [1.5 , 3.5], older [2.5 , 6.0], maintained [3.5 , 10.0] and damaged pavements [4.0 , 11.0]. The bottom thresholds correspond with the Dutch regulations. Also the longitudinal profile of a road over longer distance (unevenness) is an important variable to consider. Even though it does not directly relate to the road condition, it does have many effects on the user's driving comfort and safety (Boerboom, 2013, p. 20). From literature, expert interviews, regulations and national quality requirements the following five distresses are identified as most predominant on short term in Dutch Road Infrastructure Management(Table 2-2 gives an overview and Figure 2-6 illustrates these damages):

Transversal cross section – Rutting

Rutting is an anomaly that occurs in the wheel paths and is caused gradually by a combination of heavy loading, high temperatures and/or reduced strength of the road. It occurs mostly in closed top layers and less in porous asphalt (ZOAB). Currently ~90% of Dutch national roads have an open top layer asphalt, making rutting a seldom seen anomaly type (TNO, 2010, p. 7).

Longitudal cross section – Roughness

Roughness is the lack of smoothness in longitudinal direction expressed as International Roughness Index (IRI) [m/km] which is the mathematical evaluation of long profile wavelengths. This anomaly occurs over time and is related to cracking, constructional load distribution or settling of the soil often occurring at the area of transition between the road and a civil structure (e.g. bridge or tunnel) creating an unevenness in the asphalt (KOAC-NPC, 2016). The unevenness can manifest itself gradually in different wavelengths (from smooth to rough) and can lead to vehicle damage, reduced vehicle control and shifting or loss of cargo (TNO, 2010, p. 8).

Ravelling

As a result of traffic loading, weather conditions and/or improper processing, aggregate comes loose from the asphalt package. Once ravelling has started the rate of deterioration increases rapidly, resulting in noise increase, decrease of vehicle control, skipping up of pebbles, and loss of comfort and traffic safety (TNO, 2010, p. 8; VBW-Asfalt, 2005). The presence of ravelling is one of the determinant criteria for porous asphalt lifespan and can lead to more severe anomalies such as potholes. (VBW-Asfalt, 2005, p. 13).

Cracking

Cracking is an anomaly which can manifest itself in transversal, longitudinal or combined cracks originating from the surface or the subsurface. Though surface cracks do not effect safety much they do allow for protrusion of moisture, which may lead to larger defects especially when temperature drop below freezing point. Structural cracking is an anomaly which occurs seldom in the Netherlands, however surface cracking may occur more often but is often related to closed top layer asphalt types and less frequently in ZOAB. (Pavement-Interactive, 2008; TNO, 2010, p. 7).

Skid Resistance

Skid resistance is an anomaly that is dependent on the (micro and macro) texture of an asphalt pavement. It is necessary for the intended interaction between vehicle and the road e.g. ability to steer, brake, accelerate etc. It decreases over time due to the polishing effect of contact with tyres or by greasing of the road due excess bitumen being pushed out by loading. A decreased skid resistance is, unlike most other anomalies, invisible to the driver and has a large impact on the braking ability especially during and after rainfall (KOAC-NPC, 2016). This type of distress behaves unpredictably and can deteriorate gradually or rapidly and vice versa (Interview-VI, 2016).

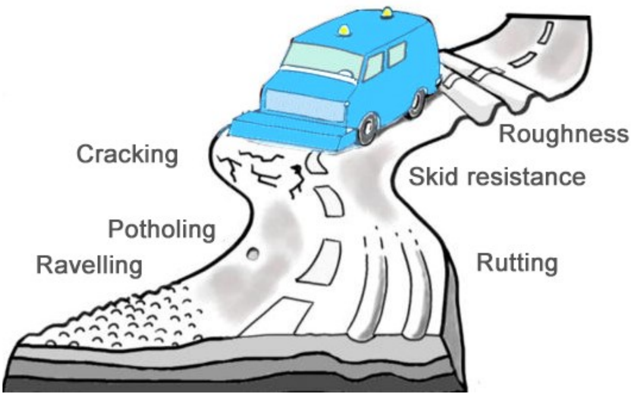


Figure 2-6: illustrated overview of distress modes

2.2.5 Treatments

The type of maintenance actions are dependent on what maintenance strategy an organisation employs, ranging from Incidental Repairs, Life Prolonging Maintenance to Large Maintenance (TNO, 2010, p. 5). When the strategy is aimed at corrective maintenance, short term and incidental repairs are often seen. When aimed at preventive maintenance, with the goal to upkeep safety standards and increase lifespan, the focus shifts towards, life prolonging and large maintenance. The different anomaly types each have a range of different treatments, however for porous asphalt the treatment is often one of three. Either a local small repair is performed by sealing the damaged area. Or locally a larger repair is performed by milling out a small section and replacing the asphalt. Or the asphalt shows damages in many areas in a larger road section, requiring for replacement of the entire road section by milling out the old road section over the full width of a lane and replacing it with new asphalt (BouwendNederland, 2012, p. 25). The type of action fitting with the specific anomaly is often something that is based on budgets and contractual obligations (Interview-VI, 2016). For each project a custom measures matrix is defined containing all types of mitigating measures and linking these to damages and severity as defined by the DWW (2002) and Appendix B2 of Performance Contracts.

Table 2-2: Acknowledged distresses, predominant distresses in bold

Damage group	Anomaly	Lit. Exp.	RWS	DBFM
Texture	**Skid resistance**	x	x	x
	Ravelling	x	x	x
Roughness	**Longitudinal**	x	x	x
	Transversal (rutting)	x	x	x
	Unevenness	x		
Consistency	**Cracking; longitudinal, transversal, crackling**	x	x	x
Others	Ride Comfort			
	Vibration hinder			
	Traffic noise			
	Potholes	x		x
	Welds			
	Drainage			
	Berm			
	Settlement	x	x	x
	Markings/Furniture			x

2.2.6 Deterioration

To improve upon maintenance within road management, knowledge on deterioration of pavements is needed. Much research has been performed on understanding asphalt behaviour resulting in numerous theoretical formulas and models. One of the most developed ones with the most extensive body of knowledge is the HDM 4 model, which has defined an equation into which all predominant anomalies can be entered. This model has been adopted by research around the globe who have found the model to be fitting with small adjustment to certain variables. Morosiuk et al. (2004) have shown in research that the model can be transformed to a general simplified model which is less dependent on the many parameters which the normal equation uses. According to Morosiuk et al. (2004) the deterioration of asphalt pavements follows the bathtub curve displaying initial densification called ‘bedding in’ then followed by gradual deterioration, which is the regular operations phase and finally showing rapid deterioration leading to failure. Even though there is much literature and research on the subject of deterioration, in practise it is not used for modelling deterioration and planning maintenance, these activities are done by Expert Judgement and experience (Interview-VI, 2016).

2.2.7 Conclusion: relevant distress modes

To work towards answering the second guiding question, a summary on distress modes is given based on literature, relevant regulations a deterioration model and expert validation. The selection of predominant distress modes resulted in five distresses considered relevant. Of these distresses Rutting and Cracking can be excluded due to their rareness in porous asphalt roads. Ravelling is considered predominant due to its unpredictableness and Skid Resistance due to its effect on road safety. The parameter for Roughness, IRI, has been mentioned by most sources as the best representation of pavement distress. Via a mathematical model all other distress modes can be linked back to this parameter making it an excellent main mode of distress.

2.3 Current practice and emerging road sensing methods

This section will give an overview of the current practise of sensing methods in road management. It will introduce the emerging technologies and present a selection of the promising methods. More information can be found in Appendix II.

2.3.1 Current practice in asphalt road sensing

The monitoring of roads in the Netherlands is done in a couple of ways, by targeted (yearly) and general (daily) visual inspections, with the locked wheel tester, Laser Crack Measurement System and with an inertial profiler (TNO, 2010). These methods combined have the ability to test for nearly all surface related issues. Visual inspections are described extensively in manuals published by a licenced authority (KOAC-NPC, 2016). The locked wheel tester is a method to measure the skid resistance of a pavement, it has been in use for many years and has been constantly in development. In 2017 the measuring standard will be improved to the Side Ways Force (SWF) method for skid resistance. Also the inertial profiler, called ARAN (automatic road analyser), that carries many measuring tools such as the High Speed Road Profiler and texture lasers. In practise professional often use different strategies based on the stringency of the contract (Interview-VI, 2016). In cases where failure has large consequences, measurements are performed on all lanes to keep information updated. In less stringent contracts, with lower consequences, often the right-most lane is leading due to the higher loading. By means of extrapolation a quality estimation is made for the other lanes. However, the amount and completeness of information is limited resulting in managers in construction and maintenance to make large capital intensive decisions based on incomplete information (deWegenscanners, 2015).

2.3.2 Emerging technology in asphalt road sensing

Next to the already existing methods there are many potential methods for anomaly detection. Some of the methods that might be applied in the years to come are the visual inspection upgraded with video analyses software. The use of high altitude imagery such as aeroplane or satellite footage used to measure depths. The 3D point cloud scanner that emits and picks up light particles to create a 3D image of the surroundings. Ground penetrating scanners which also give the ability to see the sub layers of asphalt. The use of smartphones as sensor, which is a method already used in Sweden and the United Stated. The use of in-car sensors designed to measure all kind of parameters related to handling and suspensions which can also be used to calculate the other way around and give an indication of the underground, this method could also be further upgraded with the use of ground scanners used to prepare the suspension. Also social media, which can be used to crowd sense many things. More on these methods can be found in Appendix III as well as their applicability to aid the current methods in road deterioration analysis.

2.3.3 Selection of promising sensing methods

Since not all of the methods are equally interesting for the intended goal a selection was made. This was based on an analysis to assess their ability to detect relevant anomalies (sensor vs. distress matrix, Appendix II), expected improvements or improvement potential (literature), their ease of integration in current practises (literature and interviews) and possible long term benefits (Martinsuo & Poskela, 2011) are reviewed and an overview is given in Table 2-3.

Table 2-3: Sensor overview

SENSORY METHOD	DISTRESS DETECTION	IMPROVEMENT IN NEAR FUTURE	EASE OF INTEGRATION	ADDED EFFECTS
Visual Inspections	X		X	
Video recognition	**X**	**X**	**X**	
ARAN	X		X	
Satellite Imagery				
3D point cloud data	**X**	**X**	**X**	**X**
Ground Penetrating Radar			X	X
Electromagnetic Pulse		X		
Smart phones	**X**	**X**	**X**	**X**
In-car sensors	**X**	**X**	**X**	**X**
Social media				
Embedded sensors		X		
Add on sensors	**X**	**X**	**X**	

Classification of sensors

The sensors are classified in three groups, the Expert Generated Data (EGD) contains expert measurements that are established within current AM practises as well as new methods. User Generated Data (UGD) contains data that is generated by users of an object, users might refer to people driving on a road but also own employees. Finally the Object Generated Data (OGD) which contains data generated by objects that measure autonomously, clearest example is Internet of Things IoT devices. All of the sensors can measure in one of either ways, measuring a parameter (measurement) or capturing a state (observation).

Expert Generated: Visual Inspections

Visual inspections provide an input based on what the Road Inspector (RIS) sees, hears and tests for (Klunder et al., 2010) according to the guidelines for visual inspections (CROW, 2011a, 2011b). The RIS are considered craftsman due to their extensive knowledge and experience, they are able to frequently inspect a large number distresses and are able to perform small actions independently. However the method is susceptible to subjectivity as shown by Moore, Graybeal, Rolander, Washer, and Phares (2004), who found that routine inspections by different inspectors showed significant variability (overrated). A recent development in Visual Inspections improves on the subjectivity by use of video surveying which can be re-evaluated by other experts.

Expert Generated: Video Recognition

An improvement to the use of video surveying could be the use of video recognition software and video analyses in detecting road damages (Herold, Roberts, Noronha, & Smadi, 2008). Ivetić, Mihić, and Markoski (2010) have taken it a step further by combining video and structured data into an augmented video containing extra information allowing for increased objectiveness of the retrieved information and the incremental building of a conditions database.

Expert Generated: ARAN

The Automatic Road ANalyser (ARAN) is a high speed inertial profiler that maps the underlying road with speeds up to 100 km/h (TNOKlunder et al., 2010). It can determine roughness (IRI value), characterization of pavement "macro texture" and the transverse profile used for estimation of rut depth. One of the experienced difficulties with the ARAN is that its measurements are highly susceptible to external influences resulting in differences of measurements based on e.g. weather conditions (Interview-VI, 2016).

Expert Generated: 3D point clouds

3D Point Cloud Data (PCD) consists of Cartesian coordinates measured by an extremely accurate laser scanner (Dittrich, Weinmann, & Hinz, 2017) capable of capturing many type of objects and has been piloted in different sectors (Fujita et al., 2016; Interview-VI, 2016). In road management it can be used for condition assessment (McElhinney, Kumar, Cahalane, & McCarthy, 2010) and the improve the, sometimes ambiguous, traceability of contractual requirements (Interview-VI, 2016). An experienced difficulty is its high costs of deployment, intensive post analyses and dependency on clear weather conditions.

User Generated: Smart phones as sensor

The use of smart phones as sensor for IRI calculations has been studied by Forslöf and Jones (2013); (2014), who have found it an efficient and effective method. Similar results have been achieved by Mednis, Strazdins, Zviedris, Kanonirs, and Selavo (2011) and Kulkarni, Mhalgi, Sagar Gurnani, and Giri (2014) who were able to optimise algorithms to achieve a detection rate of 90-95% accuracy compared to ARAN vehicles. When taking into account the ability to use sufficient input measurements (e.g. by Big Data and Crowdsourcing) Dawkins et al. (2011) and Lub (2016) state that the possible anomalies in smart phone data (related to in-car use of phone, differences in suspension, differences in type of smart phone etc.) can be averaged out by the large number of unique measurements, making it more reliable.

User Generated: In-car sensors

The sensor technology that is currently used to enable autonomous cars i.e. automatic parking, Adaptive Cruise Control, sign recognition, lane recognition etc. (JaguarLandRover, 2016) can also be used to measure the underlying road surface (TNO, 2010). With the advantage that clustering data from many vehicles increases the accuracy and allows for anomalies to be detected much earlier than the scheduled measurements (TNO, 2010). This potential is recognized by commercial parties developing a new service (Arcadis & Capgemini) and in parallel by governments who are testing in-car sensor data on different Belgian carrier service vehicles (Beuckeleer, 2013). Some challenges have been foreseen by TNO (2010) such as the ability to handle huge amounts of data, connectivity and privacy issues.

Object Generated: road furniture embedded sensors

The emergence of Internet of Things sensors (Machine2Machine communication) also has possibilities for application on the road side and on road furniture. A large scale test has been performed in the city of Santander where many different locations and objects were fitted with IoT sensors showing that it is possible to connect sensors in an IoT setting including the entire process of information retrieval. The researchers were able to measure rainfall, sun hours and temperature around the city. Also they were able to detect the presence of cars in parking spaces events related to public transportation (Sanchez et al., 2014). Similar sensors can be applied in Road Management to determine UV exposure, rain fall, and alignment of road signs, the performance of drainages and the growth of grass.

2.3.4 Connectivity requirements

For the Object and User Generated sensors connectivity is an important aspect which is required for transmitting collected information. Cellular network providers and countries all around the world are experimenting with the future of mobile connections, the 5G network (Eenbergen, 2016). Expectations are that the future of mobile connections will be revolutionising, due to extremely high speeds, low latency and low power consumption. (Economist, 2016). Start-ups and large companies are experimenting with street and road lighting that are able to recharge electric vehicles, transmit data, communicate with the surroundings, each other and a central control unit and also act as a Wi-Fi hot spot (LightMotion, 2016).

2.3.5 Conclusion: most promising sensing methods

The second guiding question stated: “which sensing methods are currently in use, in development or envisioned for the future that address the relevant distress modes, according to literature and industry standards in Road Management?” In Chapter 2.2 the main distress modes were identified as Skid Resistance, Ravelling and Roughness. In this section a selection was made based on a number of criteria leading to the following sensing methods that fit the distress modes: Video Recognition, 3D Point Cloud Data, Smart phones as sensor and In-Car sensors. From these sensing methods Video Recognition and 3D Point Cloud Data are similar of which 3D Point Clouds offer more potential. Smart phones and In-Car sensors are similar of which the In-Car sensors offer more potential. Additionally add on IoT sensors were selected for their ability to monitor road furniture and external influences affecting asphalt quality. Resulting in three unique types of sensing methods that fit the question.

2.4 Decision making

In order to fully understand the Operations and Maintenance phase within Road Management an understanding is needed of how information flows within the organisation and contributes to decision-making. In this section the flow of information and the decision moments are examined systematically and with the information from Chapter 2.1 a flow chart was set up and validated by experts. The place within this model where emerging sensors can contribute was defined.

2.4.1 Information flow

The management processes follow the flow diagram depicted in Figure 2-7. The process starts with determining the maintenance level of service based on the requirements and contract form in which the project is contracted. These result in a life cycle plan, long term plan and short term plans setting the base for inspections and further research (Haider et al., 2012). Once approved these plans are stored in the Asset Database i.e. the management systems preferred by the organisation. During the operations and maintenance phase periodic inspections and small maintenance actions are performed and fed back into the management system. The observed condition is checked with the expected condition for deviations. When condition is as expected the cycle is completed by feeding the information back into the database. If the conditions deviate from the expected value, actions are chosen according to the measures matrix which is defined at the start of the project. These actions are influenced by many factors, such as budgets, availability of people and materials, the amount of traffic hindrance caused and the amount and severity of damage (Interview-VI, 2016). After actions have been taken, this information is again stored in the management systems and the short term, long term and in severe cases also the life cycle plan are adjusted accordingly.

2.4.2 Data quality requirements

With the rich availability and vastness of data in our natural environment, the focus of information management within asset management processes shift from the collection of data to the assessment of data quality and the governance of data. Brous, Janssen, and Herder (2016); Hazen, Boone, Ezell, and Jones-Farmer (2014) state that good data is defined by a number of dimensions, covering both the intrinsic and contextual aspects of data. To summarise, good data should be Accurate (1) meaning, correct and not conflicting with other information. Timely (2) meaning, up to date and time between measurements fitting the purpose. Consistent (3) meaning, abiding to input and output standards. Complete (4) meaning, containing all necessary data points and Relevant (5) meaning, the data fits the intended purpose.

2.4.3 Current practise of data quality

The current practise of data collection and assessment in Road Management consists of two data collection methods, one based on observations (visual inspections) and the other based on measurements (ARAN). When this process is assessed according to data quality requirements, it can be said that the collected data are Consistent (abiding to standards), more or less Complete (containing all data points) and Relevant (containing data points that fit the intended purpose). However they lack Accuracy (correctness and non-conflicting) in two ways; technical accuracy and observational accuracy. Technical accuracy is the differentiation of measurement results due to moment of measuring, weather conditions, safety factors etc. and the Observational accuracy is the lack of objectivity of observed parameters, which is very much dependent on the expertise of the inspector and the given standards. Furthermore the measured data lacks timeliness since pavements are measured once per year, the gap between these measurements is filled by observations. Whilst the problem in accuracy also affects timeliness (if you only do one measurement per year, you would like for it to be very accurate) the solution to technical accuracy is found in the high technical domain requiring improvement of measuring equipment, deterioration models and calculation standards. An improvement of observational accuracy (subjectiveness) and measurement timeliness can be found in the sensing ability of upcoming, IoT and Big Data related techniques.

2.4.4 Affected area within the decision-making process

When looking at the decision-making process and the role of deterioration models and the use of sensor data, the expected areas of change can be suggested. Sensing methods are expected to change the data collection within Road Management, the nature of this change is dependent on the type of sensor and what its main characteristics are. Deterioration models are expected to change the analyses of data, shifting from expert judgement to calculations and modelling. Changes in these areas are not considered to be independent of each other, it can be expected that a change in data collection or deterioration modelling affects the decision-making and maintenance planning. These areas are also presented in Figure 2-7.

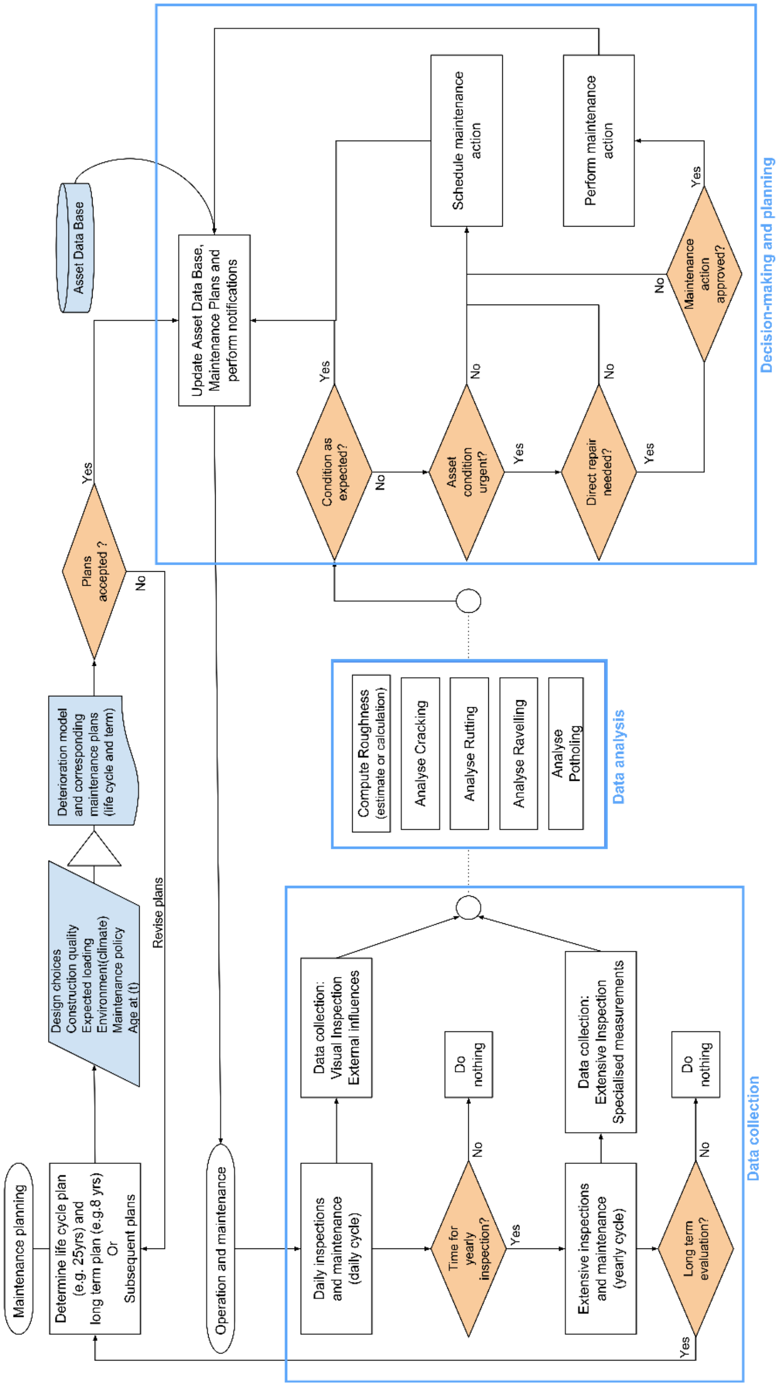


Figure 2-7: flow diagram of Road Management

2.4.5 Conclusion: decision-making

The third guiding question stated: “What information is needed by professionals in the Operations and Maintenance Phase of roads for the purpose of decision-making?” This question was answered by incrementally building on the previous two questions and determining that decision-making within the context of Road Management depends on information in three different areas. Firstly the area of data collection, where different sensing methods generate information on the current condition and build upon existing information. Secondly the area of data analysis, where condition data is analysed and quality deterioration is forecasted. Lastly the decision-making and planning area, where an expected condition is matched with the current condition with respect to the maintenance strategy and contractual requirements. The information needed by professionals in the Operations and Maintenance phase of road can be summed up as, historic information, the current quality, the deterioration of quality, contractual requirements and the maintenance plan and strategy. Within this process a role is foreseen for sensors, in the data collection area, to increase timeliness and observational accuracy of data.

2.5 Theoretical framework

With the three guiding questions being answered a theoretical base has been laid to create a framework that allows for answering the fourth and fifth question as well as the main question. This section will combine the information from the literature reviews and exploratory interviews into a theoretical framework consisting of constructs, propositions, variables and hypotheses. The section will be concluded with the presentation of the theoretical model.

2.5.1 Construct definition and propositions

The main and guiding research questions can be decomposed into more basic elements, forming the constructs of this research. Assessment of quality deterioration is defined as the asphalt quality expressed in IRI belonging to the selected anomalies as presented in Chapter 2.2. Data Sources are defined by the selection of sensing methods presented in Chapter 2.3. Decision-making is linked to the decision-making framework in which data is collected, analysed (assessed) and an action is decided upon as presented in Chapter 2.4. This section links these constructs and presents the resulting propositions.

PROPOSITION I – NEW DATA SOURCES IMPROVE THE ASSESSMENT OF DAMAGE DETERIORATION

An assessment of road quality due to deterioration is “more accurate” when information from new data sources is used in addition to the traditional data sources.

PROPOSITION II – NEW DATA SOURCES IMPROVE THE MAINTENANCE DECISION THAT IS MADE

The use of information from new data sources results in “better decisions” i.e. more fitting and results in much less unneeded actions or missed damages.

PROPOSITION III – MAINTENANCE DECISIONS ARE BETTER WHEN ASSESSMENT IS MORE ACCURATE

Decisions that are based on “more accurate” assessments are better than decision based on “less accurate” assessments. This proposition is used to corroborate the hypotheses of the first two propositions and links them together.

2.5.2 Definition of variables

In order to make the propositions measurable, the constructs need to be operationalized. This section elaborates on the previous sections by defining the variables.

Data sources

The data source variables [S] are defined by the current data sources and their areas of improvement being, the timeliness dimension and the observational accuracy dimension of data quality. The first data source variable is defined as the current data sources consisting of ARAN measurements and Visual Inspections (CUR). The second data source variable improves upon the timelines dimension by providing more frequent measurements and can be linked to Smart phones as sensor and In-Car sensors of which the latter is expect to have better performance. For control purposes this variable was split in two levels, quarterly and monthly updated (IC4 & IC12). The third and last data source variable improves upon the observational accuracy dimension by providing more objective observations and can be linked to Video Recognition and 3D Point Cloud Data Analysis, of which the latter is expect to have better performance (IR).

Asphalt quality

The asphalt quality variable [Q] is defined by the main mode of pavement distress belonging to a mathematical model being, the International Roughness Index expressed as IRI [m/km] between 0 and 16. Performance contracts stipulate the following requirements categorisation of damage for IRI: light damage (0 , 2.6), moderate damage [2.6 , 3.4], severe damage [3.5 , 4.0], critical damage > 4.0 (DWW, 2002). In which the category critical damage considered a failure, however the exact interpretation may be contract specific (Interview-VI, 2016)

Deterioration

The deterioration variable [det] main distress modes were identified as Skid Resistance, Ravelling and Roughness, from these distresses Skid Resistance has been excluded due to being an invisible distress mode which can only be approached by one sensing method. Ravelling can consist of two deterioration modes being, from gradual to rapid and from rapid to gradual, respectively fitting an exponential (EXP) and logarithmic curve (LOG). Roughness as a distress mode often proceeds gradually fitting a linear line (LIN) (Morosiuk et al., 2004; VolkerInfraDesign, 2016).

Assessment of quality

Assessment of quality has been defined as a combination of two variables being the accuracy of assessment and the assessment time. The accuracy of an assessment [A_q] is defined as the ability of the assessor to make an assessment that matches the "actual" value of the quantity being assessed. Thus an assessments that is closer to the modelled value is more accurate than one that is further. Assessment time [A_t] is defined as a time scale representing the time needed by an assessor to make an assessment, dependent on the assessment process it can be expressed as anything between seconds to days.

Decision quality

The quality of a decision is difficult to value in practice, however there are some aspects within the decision-making process which can be judged. The decision quality is defined as the correctness of a (non-)action [D_c], the magnitude (size) of an action [D_s] and the distance to modelled value [D_m] when an action is undertaken. The correctness of a (non-)action is defined with respect to the requirement, was the action needed or not? Especially focussing on the failure to meet a requirement. The magnitude of an action is defined as the resource intensity which is compared with the need, does the performed action fit the required action. The action moment is defined as the distance to failure when an action is taken, with actions closer to the point of failure more desirable than those further away.

2.5.3 Hypotheses and theoretical model

The relation between constructs is defined by the propositions, a similar relation is suggested between the variables, and these are the hypotheses that need to be tested. This section presents the suggested hypotheses and organises them according to the propositions.

PROPOSITION I – NEW DATA SOURCES IMPROVE THE ASSESSMENT OF DAMAGE DETERIORATION

HYPOTHESIS 1

A quality assessment based on information from new data sources with a higher frequency of measurement (IC4 & IC12), in addition to the traditional data sources (CUR), has a smaller deviation from the actual (modelled) quality than when based on only the traditional data sources (CUR).

$H1_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to more accurate assessments of future quality conditions*

$H1_0$ – *The use of new data sources with increased frequency of measurements do not lead to more accurate assessments of future quality conditions*

HYPOTHESIS 2

A quality assessment based on information from new data sources that are more objective (IR), in addition to the traditional data sources (CUR), has a smaller deviation from the actual (modelled) quality than when based on only the traditional data sources (CUR).

$H2_A$ – *The use of more objective data sources (IR) lead to more accurate assessments of quality condition*

$H2_0$ – *The use of more objective data sources do not lead to more accurate assessments of quality condition*

HYPOTHESIS 3

A quality assessment based on information from new data sources with a higher frequency of measurement (IC4 | IC12) and more objectiveness (IR), in addition to the traditional data sources (CUR), has a smaller deviation from the actual (modelled) quality than when based on only the traditional data sources (CUR).

$H3_A$ – *Combinations of new data sources (IC + IR) lead to more accurate assessments of quality condition given a damage deterioration*

$H3_0$ – *Combinations of new data sources do not lead to more accurate assessments of quality condition given a damage deterioration*

HYPOTHESIS 4

A quality assessment based on information from new data sources with a higher frequency of measurement (IC4 & IC12), in addition to the traditional data sources (CUR), will require more time to assess than when only the traditional data sources (CUR) are used.

$H4_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to longer assessment times*

$H4_0$ – *The use of new data sources with increased frequency of measurements do not lead to longer assessment times*

Hypothesis 5

A quality assessment based on information from new data sources that are more objective (IR), in addition to the traditional data sources (CUR), will require more time to assess than when only the traditional data sources (CUR) are used.

$H5_A$ – *The use of more objective data sources (IR) lead to longer assessment times*

$H5_0$ – *The use of more objective data sources do not lead to longer assessment times*

Hypothesis 6

A quality assessment based on information from new data sources with a higher frequency of measurement (IC4 | IC12) and more objectiveness (IR), in addition to the traditional data sources (CUR), will require more time to assess than when only the traditional data sources (CUR) are used.

$H6_A$ – *Combinations of new data sources (IC + IR) lead to longer assessment times*

$H6_0$ – *Combinations of new data sources do not lead to longer assessment times*

Proposition II – New data sources improve the maintenance decision that is made

Hypothesis 7

A decision based on information from new data sources with a higher frequency of measurement (IC4 & IC12), in addition to the traditional data sources (CUR), results in a higher decision score than when based on only the traditional data sources (CUR).

$H7_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to better decision scores*

$H7_0$ – *The use of new data sources with increased frequency of measurements do not lead to better decision scores*

Hypothesis 8

A decision based on information from new data sources that are more objective (IR), in addition to the traditional data sources (CUR), results in a higher decision score than when based on only the traditional data sources (CUR).

$H8_A$ – *The use of more objective data sources (IR) lead to better decision scores*

$H8_0$ – *The use of more objective data sources do not lead to better decision scores*

Hypothesis 9

A decision based on information from new data sources with a higher frequency of measurement (IC4 | IC12) and more objectiveness (IR), in addition to the traditional data sources (CUR), results in a higher decision score than when based on only the traditional data sources (CUR).

$H9_A$ – *Combinations of new data sources (IC + IR) lead to better decision scores*

$H9_0$ – *Combinations of new data sources do not lead to better decision scores*

Proposition III – Maintenance decision are better when assessment is more accurate

Hypothesis 10

A more accurate assessment results in fewer incorrect or unnecessary action or non-actions i.e. a higher decision score.

$H10_A$ – *More accurate assessments lead to higher decision scores*

$H10_0$ – *More accurate assessments do not lead to higher decision scores*

Hypothesis 11

A longer assessment time results in fewer incorrect or unnecessary action or non-actions i.e. a higher decision score.

$H11_A$ – *Longer assessment times lead to higher decision scores*

$H11_0$ – *Longer assessment times do not lead to higher decision scores*

Hypothesis 12

A more accurate assessment results in better fitting maintenance actions. The magnitude of the action will be smaller and the post mitigation quality will be closer to the baseline.

$H12_A$ – *More accurate assessments lead to smaller actions i.e. being more precise*

$H12_0$ – *More accurate assessments do not lead to smaller actions i.e. not more precise*

Hypothesis 13

A longer assessment time results in better fitting maintenance actions. The magnitude of the action will be smaller and the post mitigation quality will be closer to the baseline.

$H13_A$ – *Longer assessment times lead to smaller actions i.e. being more precise*

$H13_0$ – *Longer assessment times do not lead to smaller actions i.e. not more precise*

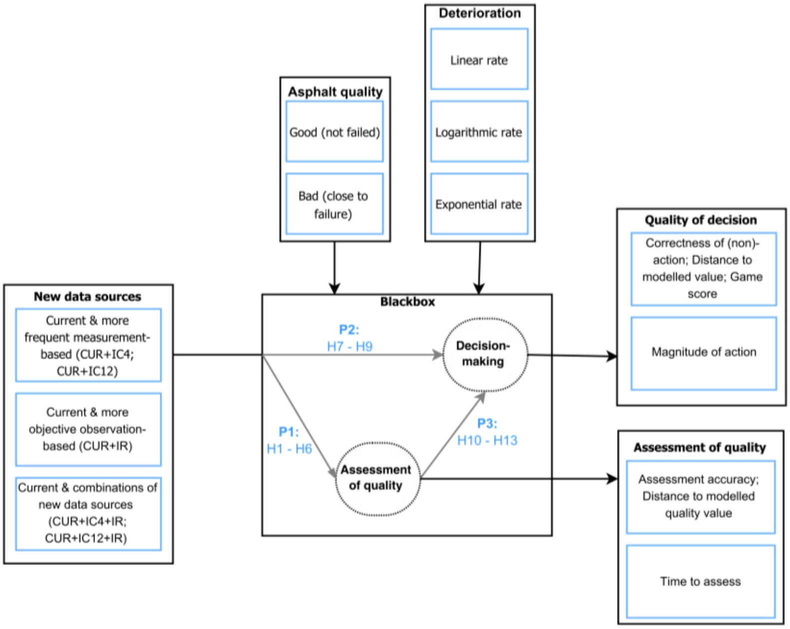


Figure 2-8: theoretical model

2.5.4 Theoretical scope

Building on the initial scope definition in Chapter 1.3, a theoretical boundary is drawn based on the findings from Chapter 2. The focus of this study is on the Operations and Maintenance phase within Asset Management of national Road Infrastructures where the main type of roads belong to the porous asphalts group. It focuses on the effect of using information from new data sources of which the effects are expected to on the operational, short term, level. More specifically in terms of data quality, it focusses on the timeliness dimension of measurement data and the accuracy dimension of observational data. Specific surface distresses with a dedicated sensing method are excluded i.e. Skid Resistance and structural distresses. The same is considered for very specific measurement methods, e.g. fall weight deflection or the Side Ways Force method. The main parameter for road quality was defined as the International Roughness Index, thus excluding all other parameters. The decision-making process is kept at a basic level without diving too deep into the domain of psychology for complexity related reasons. From the complex environment defined in Chapter 2.1 the contractual aspects relating to consequences are excluded. This refers to effects related to failing a requirement e.g. a contractual fine or budgets needed for incidental maintenance or the budgets for chosen maintenance. These elements are excluded to prevent the experiment from becoming indistinct i.e. muddy the waters.

3 Method

In the previous chapter the elements laying at the foundation of this research were presented, answering guiding questions 1 to 3. Subsequently the theoretical building blocks were defined and a set of hypotheses were formulated, concluding the chapter with the presentation of the theoretical model. This chapter builds on this model and works on its operationalization, giving an overview of the methodological framework and the experiment design. The chapter is concluded by discussing the data analysis.

3.1 Overview

To test the hypotheses, the theoretical framework needs to be operationalized. From a number of research strategies (Verschuren, Doorewaard, Poper, & Mellion, 2010) a selection was made based on the characteristics of the problem. The subjectivity involved with interviews, the lack of cases for a case study and the lack of contact with context and content that is inherent to desk research resulted in the “experiment” as most logical choice. Practise based experiments would require longitudinal studies and much resources. As an alternative research strategy the Game-based simulation experiment was found to be the logical choice because of the timeframe of this research project (Hoogen & Meijer, 2015; Lo, Hoogen, & Meijer, 2013) and the pureness of measurements. It allowed for a simplification of the reality and testing of the assumptions in a manageable and fitting manner by providing an immersive world. Furthermore it allowed for both quantitative (statistical) and qualitative (questionnaires and debriefing) data collection allowing for additional validation of the results(Hoogen, Lo, & Meijer, 2014; Van den Hoogen et al., 2014). The following sections will incrementally build the experiment design, link it to gaming and debriefing theory to design a framework fit to answer guiding questions 4 and 5.

3.2 Experiment design

From a selection of experiment designs (Fisher et al., 1960), the factorial experiment design is chosen because it is resource efficient, has good statistical power, can handle many factors and allows for the assessment of combinations of factors (Barton, 2013; Collins, Dziak, Kugler, & Trail, 2014). However a full factorial experiment with the desired factors requires a large group of participant and will test for factor combinations that are illogical or unnecessary. Improvements were made by using a fractional factorial experiment design (Collins, Dziak, & Li, 2009; Sanchez, 2005), which allows for testing of specific factor combinations and quality and deterioration types while increasing the resource efficiency. To reduce or eliminate an individual’s effects on the results (Fisher et al., 1960), the experiment is designed as a “within subjects design”, also known as “repeated measures” experiment. Which means that one participant will perform all treatment combinations, in a randomised order to eliminate possible experience or sequencing effects (Festing, Overend, Cortina Borja, & Berdoy, 2016). To further increase the experiment’s resource efficiency a repetition of was built in the experiment design, referred to as “replication” (Barton, 2013; Sanchez, 2005).

3.2.1 Experiment size

The number of required participants can generally be determined by a Power Analysis (Festing, Overend, Cortina Borja, & Berdoy, 2016) 2002. However for an experiment in which there are no prior research projects, previous statistical results and a scarcity of resources, a Power Analysis is not possible. The experiment size can then be approached by means of Mead (1988)’s Resource Equation Method seen in Equation (1). The most resource efficient amount of participants was calculated to lie between n=8 with 11 error degrees of freedom, which was deemed as the absolute minimum and n=14 with 22 error degrees of freedom, which is the maximum in terms of resource efficiency.

Mead's Resource Equation

$$E \quad = N - B - T \qquad (1)$$

Where:

$\mathrm{N} = \mathrm{n} - 1\,, \quad \mathrm{B} = \mathrm{b} - 1, \quad \mathrm{T} = \mathrm{t} - 1$

E = error degrees of freedom (10 ≤ E ≤ 20)
N = degrees of freedom for subjects including replications (n)
B = degrees of freedom for blocking (b)
T = degrees of freedom for treatments (t)

3.2.2 Profile requirements

The participants that are fit for the experiment were determined based on their knowledge and expertise in the field of asphalt pavements and road management. The participants profile is defined by three elements: (1) Engineering orientated, defined as knowledgeable on pavements, interpretation of data, deterioration behaviour and maintenance actions. (2) Process orientated, defined as knowledgeable on contractual requirements, mechanisms and operational effects. And (3) Management oriented, defined as knowledgeable on resource management, strategic planning and organisational goals. The minimum requirement for a participant was the presence of either engineering or process knowledge. More desirable was any combination of two. Most desirable was a combination of all three.

3.2.3 Participants

For this experiment 11 participants were found, meeting the resource requirements of a minimum of 8 and a maximum of 14. Selection occurred based on expert references and a short checklist which was assessed prior to participation. Due to a scarcity of professionals fitting the profile and to increase generalizability, participants were selected from three different construction companies (6 professionals) and two different engineering consultancies (five professionals). The distributions between different profiles was almost equal, with 6 professionals having a pure profile and 5 professionals having a blended profile. More specifically with 5 professionals having a (partial) engineering orientation, 6 professionals having a (partial) process orientation and 5 professionals having a (partial) management orientation. The participants considered themselves quite knowledgeable regarding Asset Management and Road Management, respectively graded with a 4.2 and 4.1 out of 5. Figure 3-1 shows a Venn diagram of the profile requirements and the areas where participants fell in.

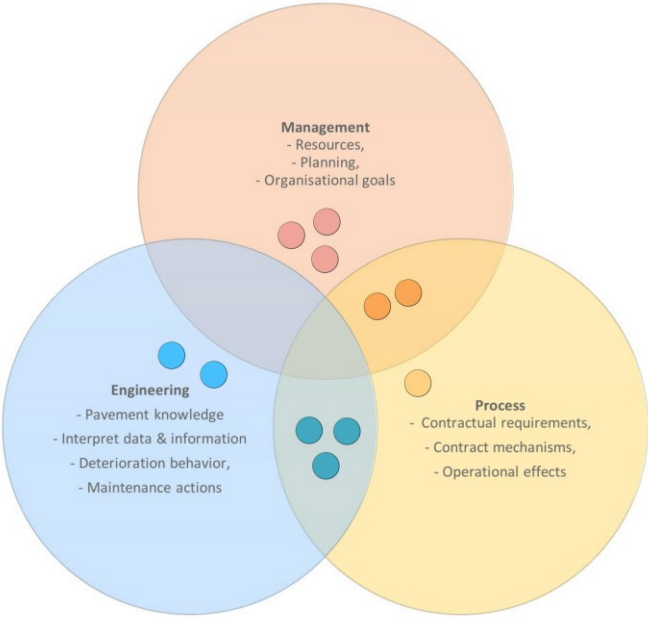


Figure 3-1: profiles of participants (the location within an area is not significant)

3.2.4 Factors, combinations and Units of Analysis

In a full factorial experiment each variable, as defined in 2.5, would be considered a different factor, with different levels per factor. In the fractional variant the sensor combinations and their level are grouped together based on logical and practice-oriented combinations, thus reducing the amount of experimental runs needed. The factors that were tested are asphalt quality at two levels (good and close to failure), deterioration rate at three levels (linear, exponential and logarithmic) and data sources at six levels (combinations of current, in-car quarterly and monthly and point cloud image recognition), resulting in $2^1 \times 3^1 \times 6^1 = 36$ experimental runs. Table 3-1 shows the experiment factors, their levels, the abbreviations and the coding. The factors levels are combined in a matrix to give a number of unique factor combinations, these combinations are the experiment's design points and are shown in Table 3-2. Each design point on its own is used to test one combination of factors while groups of design points can be used to test the main factor (L. M. Collins, Dziak, Kugler, & Trail, 2014). A participant's assessment of quality deterioration and the decision following an assessment is were chosen as Unit of Analysis (UoA). Per design point the time was measured to provide more context regarding the assessment and decision and to aid with the interpretation of the experiment outputs.

Table 3-1: experiment factors, levels, abbreviations and coding

FACTOR	LEVELS	ABR.	CODE
Data source combination	Current (control)	CUR	0
	Current + In-Car quarterly(4)	CUR + IC4	1
	Current + In-Car (12)	CUR + IC12	2
	Current + Image Recognition	CUR + IR	3
	Current + In-Car (4) + Image Recognition	CUR + IC4 + IR	4
	Current + In-Car (12) + Image Recognition	CUR + IC12 + IR	5
Asphalt quality	Good quality, IRI low	GOOD	0
	Approaching failure, IRI high	BAD	1
Deterioration rate	Linear	LIN	0
	Exponential	EXP	1
	Logarithmic	LOG	2

3.2.5 Selection of experiment operationalisation

With the experiment design determined a method for operationalisation is needed. Lo et al. (2013) have done extensive literature reviews on the designing of experiments and have identified two additional experimental settings. The traditional scale states that on the spectrum between the true experiment (laboratory setting) and the field experiment (close to reality and lacking control), there are two hybrid experiment types. They identify the artefactual field experiment as a field experiment related to the laboratory setting and the framed field experiments as a more controlled version of natural experiment. (Guala, 2002; Lo et al., 2013) and both argue that an experiment lies between a modelled system and a target system in which internal validity is generated by measuring the modelled system as purely as possible and external validity is generated by providing enough context for the experiment to represent the reality. Thus the artefactual field experiment is considered a simulation experiment (Guala, 2002), where a simulation is used to operationalize a true experiment to measure an effect as pure as possible. The framed field experiment is considered a serious game (Lo et al., 2013), where a simulation is used to operationalize a field experiment to measure an effect as realistically as possible. The objectives of this research best fit the artefactual field experiment i.e. the simulation experiment. More specifically within simulation experiments a distinction is made between computer simulation experiments and gaming simulation experiments (Lo et al., 2013). Computer simulations relate to computer algorithms performing the simulation e.g. Monte Carlo simulations or scenario analyses such as effects on greenhouse gasses due to policy changes. While gaming simulations relate to human participants geared to uncover the actions or policies of participants in given situations (Meadows, 2001). The effect of new data sources is looked for on assessment and decision-making without the external influences related to organisational or contractual mechanisms. Since the problem at hand is inherently human based, the logical choice is the gaming simulation experiment.

Table 3-2: Design points, coding and description

Coding	Design point	S	Q	Det	Description
000	**1**	0	0	0	CUR & BAD & LIN
001	**2**	0	0	1	CUR & BAD & EXP
002	**3**	0	0	2	CUR & BAD & LOG
010	**4**	0	1	0	CUR & GOOD & LIN
011	**5**	0	1	1	CUR & GOOD & EXP
012	**6**	0	1	2	CUR & GOOD & LOG
100	**7**	1	0	0	CUR + IC4 & BAD & LIN
101	**8**	1	0	1	CUR + IC4 & BAD & EXP
102	**9**	1	0	2	CUR + IC4 & BAD & LOG
110	**10**	1	1	0	CUR + IC4 & GOOD & LIN
111	**11**	1	1	1	CUR + IC4 & GOOD & EXP
112	**12**	1	1	2	CUR + IC4 & GOOD & LOG
200	**13**	2	0	0	CUR + IC12 & BAD & LIN
201	**14**	2	0	1	CUR + IC12 & BAD & EXP
202	**15**	2	0	2	CUR + IC12 & BAD & LOG
210	**16**	2	1	0	CUR + IC12 & GOOD & LIN
211	**17**	2	1	1	CUR + IC12 & GOOD & EXP
212	**18**	2	1	2	CUR + IC12 & GOOD & LOG
300	**19**	3	0	0	CUR + IR & BAD & LIN
301	**20**	3	0	1	CUR + IR & BAD & EXP
302	**21**	3	0	2	CUR + IR & BAD & LOG
310	**22**	3	1	0	CUR + IR & GOOD & LIN
311	**23**	3	1	1	CUR + IR & GOOD & EXP
312	**24**	3	1	2	CUR + IR & GOOD & LOG
400	**25**	4	0	0	CUR + IC4 + IR & BAD & LIN
401	**26**	4	0	1	CUR + IC4 + IR & BAD & EXP
402	**27**	4	0	2	CUR + IC4 + IR & BAD & LOG
410	**28**	4	1	0	CUR + IC4 + IR & GOOD & LIN
411	**29**	4	1	1	CUR + IC4 + IR & GOOD & EXP
412	**30**	4	1	2	CUR + IC4 + IR & GOOD & LOG
500	**31**	5	0	0	CUR + IC12 + IR & BAD & LIN
501	**32**	5	0	1	CUR + IC12 + IR & BAD & EXP
502	**33**	5	0	2	CUR + IC12 + IR & BAD & LOG
510	**34**	5	1	0	CUR + IC12 + IR & GOOD & LIN
511	**35**	5	1	1	CUR + IC12 + IR & GOOD & EXP
512	**36**	5	1	2	CUR + IC12 + IR & GOOD & LOG

3.2.6 Selection of gaming simulation tool

Gaming simulations can be made in many ways, they can be programmed via computational software such as Matlab or Excel (van Riel, van Bueren, Langeveld, Herder, & Clemens, 2016), geared exclusively towards data collection. Or developed in a game engine such as Unity, which is geared predominantly toward 3D game development and immersion. For this gaming simulation experiment a middle course was chosen, creating a combination of data collection and immersive aspects by means of a multiplayer sandbox-building game, Minecraft (Canossa, Martinez, & Togelius, 2013; Duncan, 2011). It focusses on constructing, survival and creativity and allowed for add-ons and scripting of own game elements for the translation of the real world to measurable outputs in the in-game virtual world. It was run on multiple servers and allowed for the storage of existing statistics and customised statistics.

3.3 Operationalization of gaming simulation experiment

The operationalization of a game-based simulation experiment used the same principles as Meijer (2009)'s serious game design theory (Lo et al., 2013). It translates the reality to a realistic game environment and facilitates the tools for quantitative data collection and links to debriefing theory to thoroughly extend and validate the information generated in the gaming simulation experiment. For this part a collaboration was set up with MSc graduate student Jasper Spiegeler and PhD student Paul Brous. This collaboration has resulted in the development of a gaming design theory review and a digital model to serve as a base for the game.

3.3.1 Serious Gaming theory on design requirements

The primary elements of a Serious Game are Reality, Meaning and Play. The game aspects should be realistic enough for the intended participants. A realistic game will induce more motivation, higher quality of gaming and thereby a less biased outcome. However, realism should not necessarily mean increased complexity in the game itself, as this muddles the meaning of the game. The Meaning of the game refers to the key skills and knowledge to successfully participate in the game. Play means that both Reality and Meaning will have to be balanced in an actual in-game environment for the game to achieve its predesigned purpose (Harteveld, 2011).

Concept

The first steps in the game design are related to defining the "concept" of the game and include abstract notions of what the rules of the game are, what message these rules convey and the link to reality. The concept is translated to specific game elements such as the storyline, gameplay elements or others (Harteveld, 2011). Meijer (2009) refers to the terms used by Gibbs (1974) to define the elements needed in a game, being, 'roles', 'rules', 'objectives', 'constraints'.

Roles

Roles make a differentiation between participants and game leaders, giving each role a different set of objectives. The game leader can 'play' a role that is thought out following a specific storyboard to influence the participants or give them specific information.

Rules

The rules of a game are the boundaries given to roles or the entire set of participants, and can be used to represent real world limitations or influence participants. The rules should be unambiguous and clear to all players. The objectives create the gaming experience by setting the goals that players have to achieve in order to win or lose. Rules tell the player what is allowed and what is not while constraints give value to variables and can be seen as the resources a player has (to earn) e.g. the amount of time, equipment and the amount and type of punishments or rewards.

Time horizon

An important and extremely influential element of the game design is the time horizon (Meadows, 2001), which must be chosen in such a way that every in-game time unit represents a real time period without neglecting important system behaviours. Too large real time periods result in skipping important moments, too short real time periods resulting in too many rounds most probably straining the attention span of participants.

Summarising

For clear communication, this can be summarised as follows: “a serious game has different or similar roles for the players, binding them to a set of individual or role related objectives that have to be achieved in order to win the game while abiding by the rules and making optimum use of the given resources”. The following sections design the gaming simulation environment along the lines of the following requirements.

Requirements for building a gaming simulation

1. *The main characteristics of the real world situation should be modelled in the game,*
2. *The game should have a clear objective (communicated to the players) and a clear research goal (not necessarily communicated to players),*
3. *The game should have an established amount of players, a timeline and a storyboard (not necessarily communicated to the players though),*
4. *Rules, roles and requirements should be clarified and unambiguous,*
5. *Have a t=0 situation, all players starting in the same role have the same objectives, the same rules, and the same constraints and resources,*
6. *Performance indicators should be established in the form of constraints e.g. money, time, requirements,*
7. *Briefing and debriefing should be incorporated.*

3.3.2 Modelling the real world

The modelling of the real world characteristics was based on the findings in the literature review and as a reference case the A4 national highway around the city of Delft was chosen. This sections focuses on the designing of an environment, the building of a deterioration model and simulating new data sources.

Modelling deterioration

The simulation of asphalt deterioration can be based on actual project details, however this results in a lack of experimental control and the project databases were not adequate for this goal. For the simulation of asphalt deterioration the model of Morosiuk et al. (2004) was used that links to the IRI parameter. The corresponding deterioration curve can be assessed by the equation formulated by Amick, Patterson, and Jorgensen (2013).

Deterioration curve equation:

$$RI_t = 1.04e^{mt} \cdot [RI_0 + 263(1 + SNC)^{-5} \cdot NE_t] \qquad (2)$$

In which:

RI_t	*= the roughness at pavement age t [m/km IRI]*
Ri_o	*= the initial roughness (m/km IRI]*
t	*= pavement age after construction or reconstruction (year)*
SNC	*= the modified structural number of asphalt composition*
NE_t	*= the cumulative ESA at age t [10^6 esa/lane]*
m	*= the environmental coefficient*

For this model the Modified Structural Number, ESAs and the environmental coefficient need to be determined. The Modified structural number can be determined by inputting the different layer thicknesses. These are project specific however industry norms state that for a national carriageway the following composition can be assumed (BouwendNederland, 2016) with additional assumptions to be made relating to layer coefficients, for the application in this case the standard values will be assumed:

- *Surf, top layer 25 mm + sub layer 45 mm* — *layer coefficient: 0.4*
- *Bind, middle layer 250 mm* — *layer coefficient: 0.14*
- *Base – bottom layer 300 mm* — *layer coefficient: 0.10*

Resulting in a theoretical structural thickness of 93 mm, the SNC is defined by inches, resulting in 3,661 inch, rounded up this results in a SNC of 4. The number of 10^6 ESAs per lane are determined by using the traffic load determined by the (CBS, 2015) resulting in a value of 0.98817725. The environmental coefficient is determined by the climate conditions of the geographical location and is 0.1 for a climate as in the Netherlands (Morosiuk et al., 2004). To determine a realistic pavement life cycle, the standards defined in Chapter 2.2 have been used and the moments of reconstruction (large maintenance) were defined based on the tolerance threshold for the national carriageway in the Netherlands, being IRI 3.5 and the absolute maximum of 4.0 before critical severity is reached. Combining this information with Equation (2) resulted in the graph shown in Figure 3-2. This model corresponds with the literature stating the average lifetime porous asphalt is 7 to 10 years (VBW-Asfalt, 2005), compared to the model which indicates a life time of roughly every 7 years.

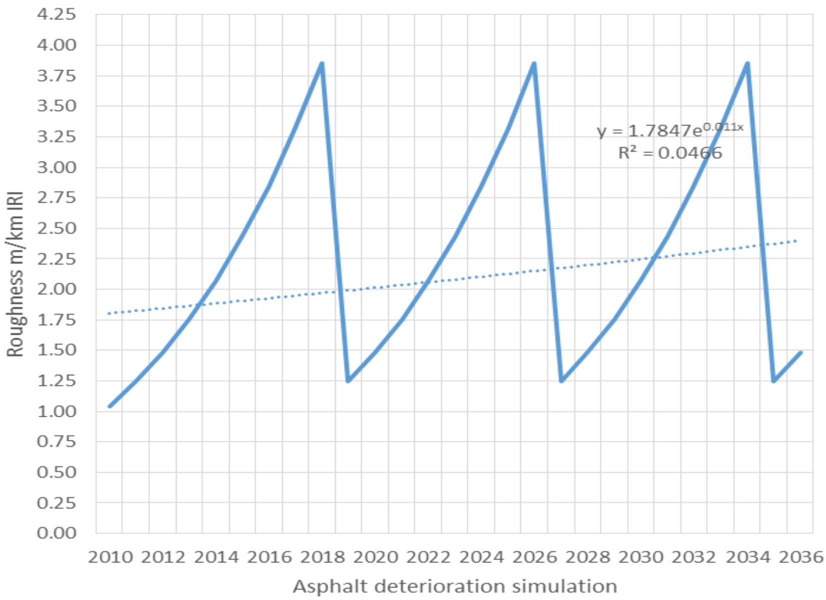


Figure 3-2: deterioration base model

This model served as a base for the deterioration modelling in which deterioration rates were linked to the model and different quality levels were simulated. Resulting in a matrix containing the monthly quality value for a period of four years for each of the 72 experimental runs (design points and replications) of the experiment design. An example of this matrix, the deterioration rate coefficients and further modelling products can be found in Appendix III.

Modelling new data sources

The modelling of new data sources was derived from the deterioration model and translated to current practice as defined in Chapter 2.1 and 2.3.13 Which is management via text processors, spreadsheets, two dimensional maps and location information in GIS. Examples of new data source visualisations were derived from literature for In-Car sensors (Forslöf & Jones, 2013; Herold et al., 2008; Koch et al., 2015; Kulkarni et al., 2014) and Point Cloud Data Image Recognition (Interview-VI, 2016; Lee, Bruce Blundell, Starek, & Harris, 2013; Puttonen et al., 2016; van Oosterom et al., 2015), additionally a link was sought with upcoming data visualisation tools (QlikSense, 2016; Tableau, 2016).

3.4 Debriefing theory

Debriefing is the final framework that is used for the experiment. It can be described as: "the process in which people who have had an experience are led through a purposive discussion of that experience" (Lederman, 1992, p. 146). For serious gaming, debriefing can sometimes even be of more value than the game itself because it can be used as an additional data collection part outside the game (both for the researcher as the participant). For the experiment simulation a debriefing structure was designed based on the six main functions identified by Van den Hoogen et al. (2014). The cooling down (1) is aimed at taking the participant out of their immersed state and talk about how they feel and how they´ve experienced it. It provided an opportunity for participants to cool down and get ready to talk about what happened. By means of discussion extra qualitative data is collected (2) aimed at going past the quantitative outputs and enables a look into the thoughts and experiences of the participant during decision-making. The reliability (3) of the game is discussed by assessing if a subsequent play would result in similar results. It tests the sensitivity of the simulation model. The internal and external validity (4) are assessed by discussing the in-game events and the relation of these to the model and the real world. It allows for discussing design choices and how more or less realism might have affected the participants. Also if, in reality, they would behave in the same manner as they did during the game simulation. The possible actions and future implications (5) are discussed to generate input for subsequent games or the practice. The session is closed by protecting the instrument (6) and evaluating the game session its link to reality in a more lightly manner. These functions make debriefing very time efficient since it generates more information which contains feedback without increasing playtime. Appendix IV contains the main topics and the debriefing structure designed for the gaming simulation experiment.

3.5 Design sprints

Before reaching the final game design, a whole process was undergone. An initial prototype was designed according to the design requirements stated in the previous section. A real world model was created in Minecraft, a deterioration model was built and data sources were simulated. The objective of the game was to place a player in a single player world where they were an asset manager responsible of managing an asset (road) for the next year while doing some additional tasks such as checking road furniture and data quality. Different types of mechanisms were designed, constraints and resources were set and an extensive briefing en debriefing were developed. Before the game simulation experiment could be used, testing was needed as well as validation of the mechanisms. This process was the actual designing process in which iteratively the design got refined. To structure the design process the Product Design Sprint cycle (Knapp, Zeratsky, & Kowitz, 2016) was applied to a certain extent. This design cycle is short cyclical and uses the power of teams to rapidly go from understanding a problem to validating an artifact i.e. a prototype in five steps. Firstly, the general idea is to understand the problem in its broadest context, as done in the first chapters. Secondly a divergence is needed to gather ideas and possible solutions. Thirdly a convergence is needed to bring together the ideas and choose the best fitting one. The fourth step requires the building of a usable product that allows for discussion and understanding the main principles. It could be called a Minimum Viable Product (Reis, 2011). The final step is to validate the prototype with actual users and to learn what works and what does not work. This process is shown in Figure 3-3.

Figure 3-3: Product Design Sprint, adopted from Thoughtbot (n.d.)

Sprint 1 – large group testing

The first design sprint was with a large group of student users, it was aimed at removing all the large flaws according to the principles of Faulkner (2003). She states that design problem discovery is related to the number of users who have tested it, suggesting that with 5 test users 55% of problems are uncovered, with 15 users 90% and with 20 users 95%. The initial test resulted in uncovering the following problems:

- The tutorial was found to be too elaborate and too much text based,
- The combination of multiple types of tasks was found to be confusing,
- No structuring in the virtual world lead to users not doing the tasks.

Sprint 2 – unit testing

The second sprint incorporated the lessons learned from the first sprint and removed the additional tasks related to data quality, shortened the tutorial and built a structured world for the participants to be more or less guided through the game. The second tests were aimed at testing decoupled elements of the game for optimisation, uncovering the following problems:

- Tutorial is still too much to fully understand,
- Structuring was good however in game orientation was lacking,
- Still very difficult for users to understand,
- Much text, information and paper, users really needed to focus.

Sprint 3 – full testing

The third sprint incrementally built upon the lessons learned from the second sprint and shortened the tutorial even further. The paper tutorial was substituted by an in-game version (Figure 3-4) and was used as a reference. Orientation elements were added and more guiding elements were provided (Figure 3-5). The text and amount of material were reduced and the element of isolation was added by letting users play in a confined space without distractions. The third test run was aimed to test the entire cycle of the experiment design from start to finish, uncovering the following:

- Users still found the tutorial and information much to handle,
- Game handling could be improved, users still struggled,
- Data collection went fluidly,

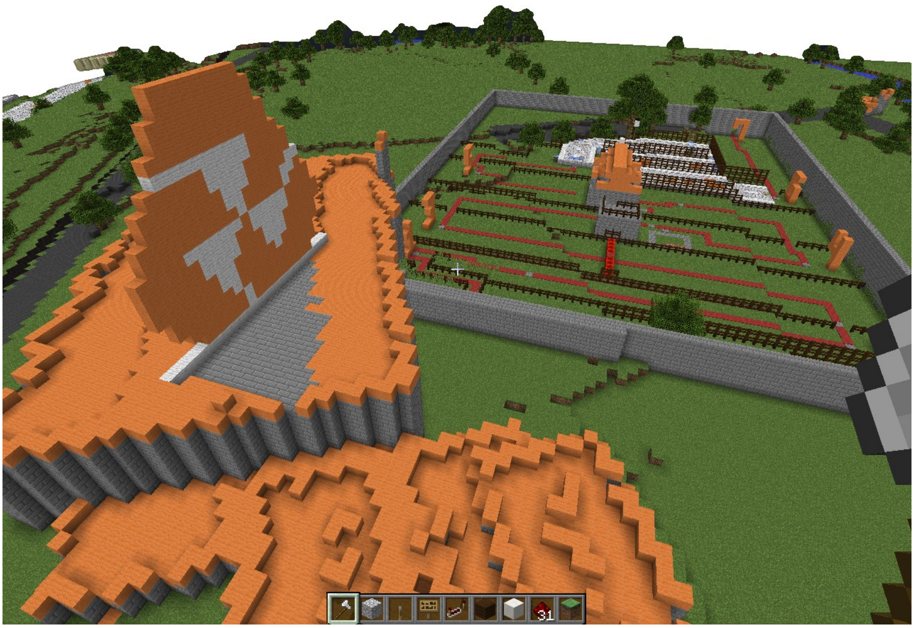


Figure 3-4: in-game tutorial playing field

Sprint 4 – expert testing

After the full test minor improvements were needed. The tutorial was improved, the reference book was shortened and an information package was made that could be sent in advance to prepare participants. The fourth test was aimed at validation by experts and split testing of certain assumptions. Split testing is the designing of two or more versions and simultaneously testing these to see the differences. For the expert validation a split test was used to check how contractual mechanisms and certain data source elements affected the user, resulting in:

- The user who received more contractual context approached the task completely different that the other users, data was unusable,
- The explanation of the scoring system triggered users to play for score instead of based on their own judgement,
- From debriefing it appeared that practice experience influenced users to a great extent, such as the choice of maintenance treatment which differed among the test users based on preference,

- The combination of multiple damage mechanism was confusing for users and made decision origins unclear in the data analysis,
- The presentation in a simulated GIS portal was too much information, it was experienced as unnecessary,
- Users who had received forecasts mentioned that this influenced their judgment, assessment and decisions conformed to the forecast instead of to own ideas.

Figure 3-5: in-game world, distress depiction and orientation landmark (right)

The final lessons learned

The last sprint uncovered many improvement points most of which were related to an abundance of information that needed processing in a short amount of time. This showed elements of information overloading as defined by Eppler and Mengis (2004). They found that limitations in the individual human information-processing capacity, insufficient contextual information, the rise of informational items, the number of alternatives, unknown information and an overabundance of irrelevant information led to this phenomenon. Thus a further simplification of the simulation design was achieved by removal of contractual elements and writing of the requirements from a technical perspective without consequence. Scoring system was hidden to participants except for the general idea of the game score. The briefing contained a section in which participants were explicitly told to ignore practical experiences related to contractual or other governing mechanisms. A practise round was included in which the researcher sat next to participants and left the room only before the real experiments started. Damage mechanisms were reduced to IRI only, other mechanisms contributing to IRI were left out and modelled into the deterioration model as to be invisible. Additionally add-on IoT sensors were selected for their ability to monitor road furniture and external influences affecting asphalt quality, however these were also experienced as clouding the results. Forecasts were removed and data sets were made more intuitively. The earlier versions were built in a simulation of GIS, however this was too distracting. The main learned lesson was, put in Einstein's words, "everything should be made as simple as possible, but not simpler". The improvements were unit and full tested to satisfaction resulting in the finalisation of prototyping and an experiment ready game.

3.6 The final game-based simulation experiment

This section presents the final version of the game-based simulation experiment, based on the storyline, the rules, the objectives, the scoring system and the experiment timeline.

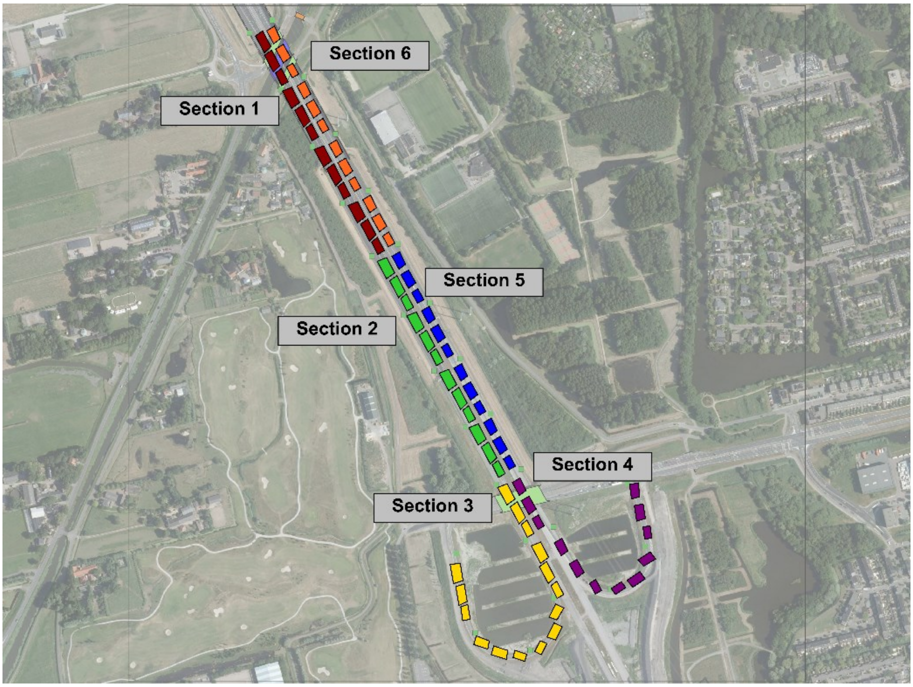


Figure 3-6: overview of the route, the segments and the sections

The experiment was performed on a virtual stretch of road, referred to as the route. It contained route segments and road sections. The route was divided into six route segments, one for each factor level of sensors, and each route segment contained twelve road sections, two for each design point (original and replication). Per segment six different combinations of quality and deterioration were tested as well as their replications to increase the accuracy of experiment output. An overview of the road section is given in Figure 3-6. The experiment timeline is given in **Error! Reference source not found.**. The entire playbook containing all gaming documents can be found in appendix IV.

3.6.1 Storyline

The participant (referred to as expert) took on the role of "Expert in Pavement Engineering" within the Asset Management department of the fictional company "MineCraftBouw". They were replacing the previous manager and had to work with the data that this manager had collected over the years. The road was currently 7 years old and some damages started appearing just recently. According to the long term planning the expert had to bridge another year with minimum repairs as to not waste resource with the upcoming large planned maintenance.

3.6.2 Rules

- The expert played in single player mode and always took on the role of Asset Manager,
- The expert could only use the information that was provided,
- They had to make an assessment for each road section, linked to the 72 design points and replications,
- They were free to decide where they planned a maintenance action,
- The assessment and maintenance action only focussed on the coming year, till the start of 2018,
- Quality was dependent on many factors, however was simplified to roughness given in IRI,
- Quality was judged in severity classes, so if a section was just below the threshold it still counted as good,
- Damaged road blocks counted for a 0.05 m/km increase of roughness,
- Repair blocks counted as 0.10 m/km decrease of roughness →quality increase,
- The game score was determined by amount of repair blocks placed and the correctness of action or non-action,

3.6.3 Objectives

The objective was to maintain the road according to the following principles:

- Perform as few action as possible,
- The road quality had to remain smaller than IRI 4.00,
- Use of information as presented to assess as accurately as possible

3.6.4 Scoring

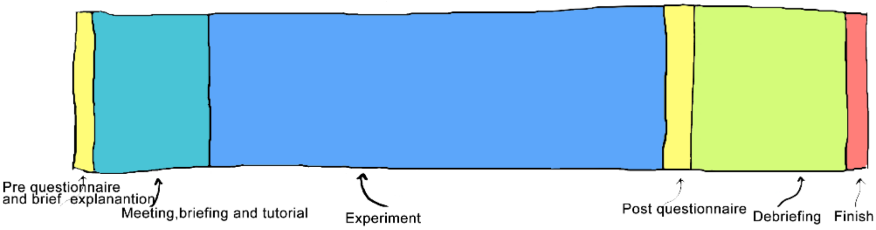
To determine the quality of a decision a construct was formulated. Roughly the score was based on the required action minus the size of the chosen actions, between 0 to infinite blocks. Where the correctness of the decision determined the direction of the score (positive or negative) and the distance to the modelled value determined the size of the score [-2,2]. Figure 3-7 displays the code used to determine the decision score. Where "val18' is the modelled value, "act" is the number of blocks placed and "res" is the result of modelled value – (number of blocks/10).

```
if val18(j) < 3.9 && act(j) == 0
    Score(j) = 1
elseif val18(j) < 3.9 && act(j) > 0
    Score(j) = 0
elseif val18(j) >= 3.9 && val18(j) < 4 && act(j) == 0
    Score(j) = 2
elseif val18(j) >= 3.9 && val18(j) < 4 && act(j) > 0
    Score(j) = 0
elseif val18(j) >= 4 && val18(j) < 4.1 && act(j) == 0
    Score(j) = - 1
elseif val18(j) >= 4.1 && act(j) == 0
    Score(j) = - 2
elseif val18(j) >= 4 && act(j) > 0 && res(j) >= 3.9 && res(j) < 4
    Score(j) = 2
elseif val18(j) >= 4 && act(j) > 0 && res(j) >= 3.8 && res(j) < 3.9
    Score(j) = 1
elseif val18(j) >= 4 && act(j) > 0 && res(j) < 3.8
    Score(j) = 0
elseif val18(j) >= 4 && act(j) > 0 && res(j) >= 4 && res(j) < 4.1
    Score(j) = - 1
elseif val18(j) >= 4 && act(j) > 0 && res(j) >= 4.1
    Score(j) = - 2
```

Figure 3-7: decision score calculation

3.6.5 Timeline

The timeline of the experiment sessions starts a number of days before the experiment the participant receives an email with some basic information and the link to the pre questionnaire. These have to be checked before the day of the experiment. On experiment day, the participant gets a brief explanation and starts with the tutorial. During this the researcher stayed with the participant and helped where needed. A practise run was performed, for the participant to understand the game and for the research to be confident about the session. When the participant was ready to start, the researcher left the room. During the experiment the research monitored the events via an online server, in case anything went wrong. After the participant had finish, they started the post questionnaire while the game score was calculated for debriefing purposes. The debriefing took place and the experiment was concluded. The average session took 2.5 hours depending on the scheduled time.



Figuur 3-8: experiment timeline, adopted from errantscience.com

3.7 Data analysis

From the data generated during the gaming simulation experiment four data sets were created to test the hypotheses. This section elaborates on the nature of the variables in a given data set, the steps taken in the analysis of the data and the corresponding statistical test per proposition.

Variables characteristics

The hypotheses belonging to the first proposition all have a single dependent variable, the distance of the assessment to the modelled value or time needed to make the assessment, which are scale variables. The hypotheses belonging to the second proposition all have a single dependent variable, the decision score, which is a scale variable (interval). The independent variables in all of these cases are the data sources, the asphalt quality and deterioration rate, which are all categorical inputs. The hypotheses belonging to the third proposition all have a single dependent variable, the size of a decision or the decision score, which are both scale variables (interval). A change in each of these dependent variables is expected due to an independent variable, being the assessment accuracy or the time to assess, which are both continuous. The experiment is set up as a Repeated Measures experiment however the effects can be measured 'between subjects'.

Assessment of data sets

Firstly the descriptive statistics were calculated to assess the data set properties for the assumptions of parametric testing. These assumptions relate to the usability of parametric tests and include measurement scale, the independence or dependence (depending on the type of test), the normality of the distribution and the homogeneity of variances. The datasets passed all tests accept the test for normality, which after closer inspection was still accepted due to the small effect. The data set containing the variables for the hypotheses of proposition III was too non-normal for the intended tests thus making parametric testing not possible for those cases.

Statistical tests

Given the characteristics and the experiment design, the One Way Repeated Measures ANOVA was deemed most suitable for the hypotheses of propositions I and II (Ellis, 2003; Field, 2009). This statistical test checks if the mean of a variable in one group is different than the means in other groups. The data was organised per design point and prior to the Repeated Measures ANOVA, Mauchly's tests of sphericity was performed. In some cases the assumption of sphericity had been violated due to the extreme sensitivity of the Mauchly's test. In these cases the values with corrected degrees of freedom were used to maintain reliability of results. The repeated measures ANOVA tested for the main effects on factor level (sensor, quality and deterioration) and for the interaction effects on each intermediate level (two-way and three-way interactions). The data set characteristics of the hypotheses of proposition III made the testing of these hypotheses fit for the use of a Kendall's tau-b correlation (Ellis, 2003; Field, 2009). This test checks if one variable changes together with another variable in a constant direction. For these tests the assumption of independence was violated because 792 data points were analysed from just 11 participants. Making the effect of one participant's results rather large compared to 792 independent data points. While this might be of interest in the case where a participants opinion was asked or there was unwillingness involved, distorting the results. In this experiment an action was measured and participants were all involved seriously, decreasing the likelihood that one person's mood might have influenced the results.

Some distinct cases

A single participant did not finish enough comparable sections however did finish the game at a later moment. All the assessment and decision data were usable, however timing did not take place per assessment but for an entire set, the time per assessment was calculated by averaging.
Two participants did finish all sections however there were errors in the data, due to the participant being interrupted resulting in duplicate values. The correct entries were retrieved from the hardcopy control sheets, timings were substituted with the mean value of that participant. In the cases where substitution takes place the statistical power is maintained however the variability decreases and covariance and correlation in data weakens. This is not the strongest method available but concerning the nature of the dataset it is the best choice for completing datasets with random gaps.

4 Results - Evaluation of data

The previous chapter elaborated on the design of the experiment and the design of the game-based simulation. It addressed the series of product design sprints to optimize and validate the simulation and concluded by giving an overview of the data analysis. In this chapter the collected data and the results of the analyses will be discussed. The first part is devoted to the quantitative analyses and is structured around the propositions and hypotheses. The second part focusses on the qualitative data analysis from the pre and post tests and the debriefing to give an insight into the black box of decision-making to support and explain the quantitative result. It resulted in the answering of the final two guiding questions.

4.1 Overview

The data generated during the gaming simulation experiment resulted in four data sets, three of which were used for the Repeated Measures ANOVAs and one that was used for the Kendall's tau-b correlation tests. The post-questionnaire resulted in a single data set containing qualitative data and marks for different statements. The debriefing resulted in a single data set containing only qualitative data.

Quantitative descriptive

The properties of the quantitative data sets were assessed by calculating the descriptive statistics Table 4-1: Descriptive statistics for main sensor groups. These showed no consistent differences between the sensor groups for assessment accuracy, time or score.

Table 4-1: Descriptive statistics for main sensor groups

Main groups	μ assess	σ assess	μ time	σ time	μ score	σ score	n
CUR	-,04909	,120365	52,689	29,2630	,780	,9190	132
CUR + IR	-,02394	,107240	49,515	32,6841	,788	,8474	132
CUR + IC4	-,04167	,117954	63,265	47,5547	,811	,9091	132
CUR + IC12	-,02356	,122182	56,955	34,3464	,780	,8677	132
CUR + IC4 +IR	-,04311	,120287	53,523	34,8916	,947	,8677	132
CUR + IC12 + IR	-,04659	,100281	58,985	33,9182	,773	,9213	132

Subsequently the descriptive statistics were calculated for all factor level combinations for assessment accuracy, time and score (Table 4-2). These revealed that generally the quality assessments were underestimated when assessing a section which had good quality, while overestimating section that had bad quality. Furthermore it appeared that in assessments were more accurate when the deterioration rate was linear or logarithmic and much less accurate when the rate was exponential. Again no consistency was found between sensor groups for the accuracy of assessment. For assessment times the values for all factor combinations appeared more or less steady. While for the linear and logarithmic deterioration rate the decision score also was steady, it appeared unstable for the exponential deterioration rate.

Lastly the quantitative section of the post-questionnaire was assessed, showing no particularly outstanding data. Participants were positive about the sessions, the instructions and task clarity. They indicated to have had sufficient guidance and information to perform the tasks and were not influenced by the game leader. Most data that they had received was understood, not all data that was received was used. Participants were quite satisfied with the work they had performed and expected good results. Motives for decision-making were rather spread, some participants focused solely on meeting requirements while others sought more balance between quality and amount of actions. They felt responsible for the asset during the game and found the simulation to be a correct representation of the real world.

Table 4-2: Descriptive statistics for all design points

Factor combinations	*μ assess*	*σ assess*	*μ time*	*σ time*	*μ score*	*σ score*	*n*
CUR & GOOD & LIN	-,019545	,1402880	59,18	39,296	,95	,213	22
CUR & GOOD & LOG	-,048636	,1044393	48,36	27,705	1,00	,000	22
CUR & GOOD & EXP	-,131818	,1387233	57,00	28,820	,95	,213	22
CUR & BAD & LIN	,010455	,0706755	57,73	29,713	1,09	,750	22
CUR & BAD & LOG	,011364	,0906745	47,86	22,935	,95	1,046	22
CUR & BAD & EXP	-,116364	,0834121	46,00	24,689	-,27	1,453	22
CUR + IR & GOOD & LIN	,005000	,1121967	43,05	41,857	1,00	,000	22
CUR + IR & GOOD & LOG	,010455	,1039449	50,41	24,160	1,00	,000	22
CUR + IR & GOOD & EXP	-,095000	,1010304	53,36	35,505	1,00	,000	22
CUR + IR & BAD & LIN	,011818	,0742990	55,09	27,881	1,00	1,024	22
CUR + IR & BAD & LOG	,034091	,0830180	47,68	30,994	,45	,858	22
CUR + IR & BAD & EXP	-,110000	,0754668	47,50	35,057	,27	1,453	22
CUR + IC4 & GOOD & LIN	-,024091	,1164007	62,09	53,879	1,00	,000	22
CUR + IC4 & GOOD & LOG	-,005455	,1317909	59,41	47,689	1,00	,000	22
CUR + IC4 & GOOD & EXP	-,115000	,1560449	58,73	44,482	,95	,213	22
CUR + IC4 & BAD & LIN	-,011818	,0424978	72,59	54,358	1,14	,710	22
CUR + IC4 & BAD & LOG	-,002273	,0764655	68,45	49,949	,91	1,019	22
CUR + IC4 & BAD & EXP	-,091364	,1067394	58,32	36,303	-,14	1,552	22
CUR + IC12 & GOOD & LIN	-,020000	,0734199	50,36	34,461	1,00	,000	22
CUR + IC12 & GOOD & LOG	-,025000	,1256507	60,95	36,704	1,00	,000	22
CUR + IC12 & GOOD & EXP	-,110455	,1102153	57,27	40,136	,95	,213	22
CUR + IC12 & BAD & LIN	,054091	,1016370	61,68	35,305	,91	1,019	22
CUR + IC12 & BAD & LOG	,044091	,0924100	59,45	36,567	,77	,973	22
CUR + IC12 & BAD & EXP	-,084091	,1343648	52,00	22,621	,05	1,397	22
CUR + IC4 + IR & GOOD & LIN	-,011364	,0925458	48,64	38,366	1,00	,000	22
CUR + IC4 + IR & GOOD & LOG	,007273	,1176833	52,00	35,141	1,00	,000	22
CUR + IC4 + IR & GOOD & EXP	-,127273	,1441620	52,82	36,659	,95	,213	22
CUR + IC4 + IR & BAD & LIN	,005455	,0780554	54,59	32,328	1,45	,912	22
CUR + IC4 + IR & BAD & LOG	,014545	,0419441	58,09	34,223	,82	1,006	22
CUR + IC4 + IR & BAD & EXP	-,147273	,1037020	55,00	35,694	,45	1,503	22
CUR + IC12 + IR & GOOD & LIN	-,048636	,0886027	65,41	57,935	1,00	,000	22
CUR + IC12 + IR & GOOD & LOG	-,022727	,1193706	51,82	28,468	,95	,213	22
CUR + IC12 + IR & GOOD & EXP	-,065455	,1357774	57,91	30,568	,91	,294	22
CUR + IC12 + IR & BAD & LIN	,000000	,0444008	58,14	25,735	1,00	1,024	22
CUR + IC12 + IR & BAD & LOG	-,013636	,0693195	56,14	22,802	,82	1,220	22
CUR + IC12 + IR & BAD & EXP	-,129091	,0624812	64,50	27,453	-,05	1,327	22

4.2 Results of quantitative data

The results of the quantitative data are presented in this section and are organized according to the proposition they belong to. The section is concluded by answering guiding question 4 – “what are the differences in assessments and decisions of professionals in road management given a certain data source?”

4.2.1 Proposition I – New data sources improve assessment of damage deterioration

Assessment accuracy

The hypotheses tested with the repeated measures ANOVA are similar in experiment size and degrees of freedom, the critical threshold of F is determined as: reject H_0 if F > 4.325. Violations were found for Mauchly’s test, indicated that the assumption of sphericity had been violated, corrections were applied to the degrees of freedom using Greenhouse-Geisser estimates of sphericity ε = .52. Testing for the effects of all data source groups on the accuracy of assessments reported no significant main effects [F (2.574, 54.055) =1.909, *p* = .147]

$H1_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to more accurate assessments of future quality conditions*
Testing for the effects of IC4 and IC12 reported no significant main effects of the type of data source on the accuracy of assessments. Contrasts revealed that there was no significant effect for CUR+IC4 [F (1, 21) = 0.532, *p*= .474]. Also contrasts revealed that the assessment was more accurate for CUR+IC12, [F (1, 21) = 4.474, *p*= .047], than for the control group. These results are contradicting, thus suggesting that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H2_A$ – *The use of more objective data sources (IR) lead to more accurate assessments of quality condition*
Testing for the effects of IR reported no significant main effects of the type of data source on the accuracy of assessments. Still contrasts revealed that the assessment was more accurate for CUR+IR, [F (1, 21) = 10.032, *p*= .005], than for the control group. This suggest that the alternative hypothesis should be accepted and the null hypothesis should be rejected.

$H3_A$ – *Combinations of new data sources (IC + IR) lead to more accurate assessments of quality condition given a damage deterioration*
Testing for the combined effects of IC4, IC12 and IR reported no significant main effects of the type of data source on the accuracy of assessments. IC4&IR, [F (1, 21) = 0.173, *p*= .681] and IC12&IR, [F (1, 21) = 0.070, *p*= .794] were similar to the control group. None of the design points testing this hypothesis were found to be significant,

this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

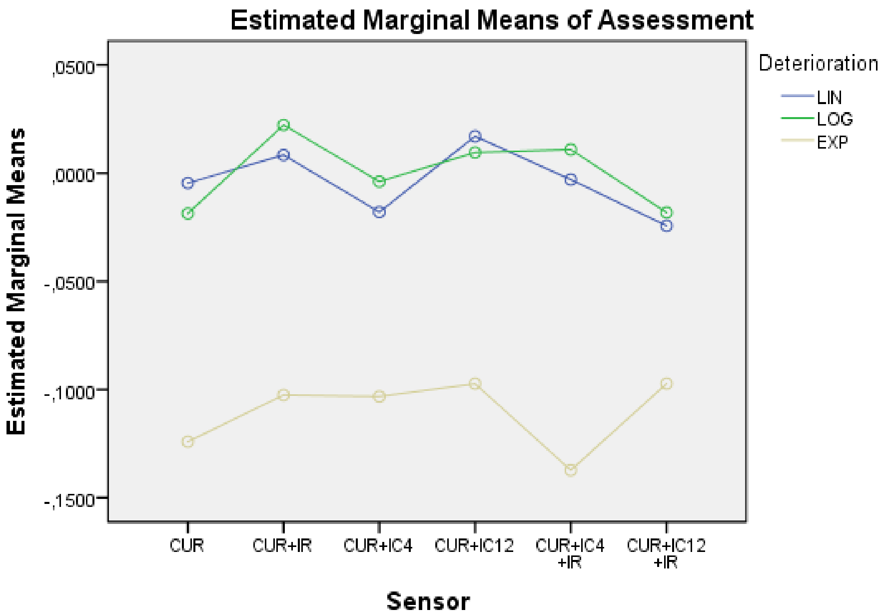


Figure 4-1: Estimated Marginal Means of Assessment

In general it seemed that the use of a new data source did not affect the accuracy of assessment much. The repeated measures ANOVA showed no visible trend for the main effect of data sources on assessment accuracy, neither did the two way interactions or the three-way interactions. Only the tests for CUR+IR and CUR+IC12 reported significant effects. Figure 4-1 shows the estimated marginal means for two-way interactions between data source and the deterioration rate, other interactions produced similar results. The cases in which significant or remarkable results were produced are mentioned per hypotheses.

Assessment time

The hypotheses tested with the repeated measures ANOVA are similar in experiment size and degrees of freedom, the critical threshold of F is determined as: reject H_0 if F > 4.325. Violations were found for Mauchly's test, indicated that the assumption of sphericity had been violated, corrections were applied to the degrees of freedom using Greenhouse-Geisser estimates of sphericity ε = .56. Testing for the effects of all data source groups on the accuracy of assessments reported no significant main effects [F (2.790, 58.585) =0.830, *p* = .475]

$H4_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to longer assessment times*

Testing for the effects of IC4 and IC12 reported no significant main effects of the type of data source on the time needed for assessment. CUR+IC4, [F (1, 21) = 2.000, *p* = .172] and CUR+IC12 [F (1, 21) =0.429, *p* = .520] compared to the control group. None of the design points testing this hypothesis have tested significant, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H5_A$ – *The use of more objective data sources (IR) lead to longer assessment times*

Testing for the effects of IR reported no significant main effects for the type of data source on the time needed for assessment. CUR+IR, [F (1, 21) = 0.231, *p* = .636] compared to the control group. The test found no significant effects, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H6_A$ – *Combinations of new data sources (IC + IR) lead to longer assessment times*

Testing for the effects of the combined factors CUR+IC4+IR and CUR+IC12+IR reported no significant main effects of the type of data source on the time needed for assessment. CUR+IC4+IR, [F (1, 21) = 0.016, *p* = .899] and CUR+IC12+IR, [F (1, 21) = 2.216, *p*= .151] compared against to the control group. All tests found no significant effects, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

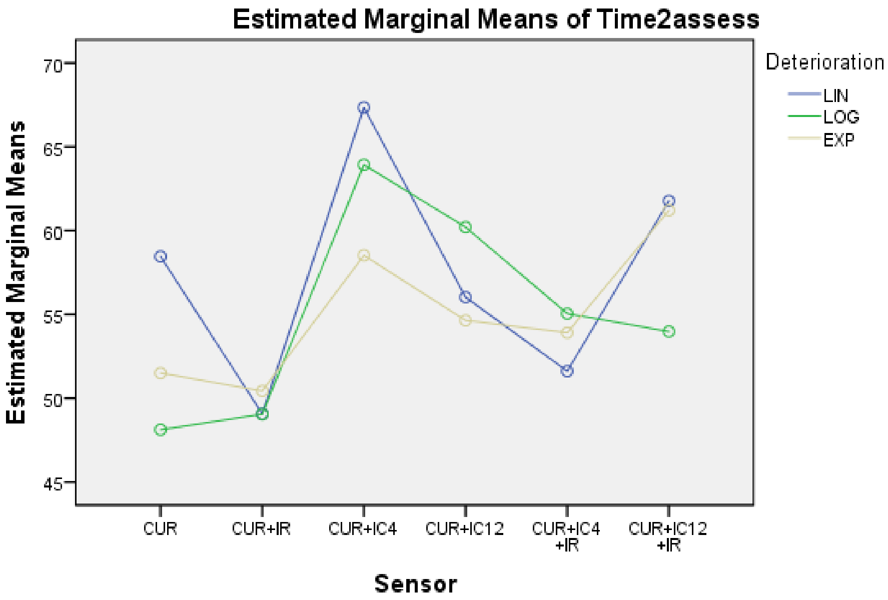


Figure 4-2: Estimated Marginal Means of Time2Assess

The repeated measures ANOVA showed no visible trend for the main effect of data sources on assessment accuracy, neither did the two way interactions or the three-way interactions, which all had comparable result. The use of CUR+IC4 appeared to have an effect on assessment time, however this same effect was not seen in the other data sources which build on CUR+IC4, suggesting that it might be due to an in-game element. Figure 4-1 shows the estimated marginal means for two-way interactions between data source and the deterioration rate, other interactions produced similar results. The cases in which significant or remarkable results were produced are mentioned per hypotheses.

4.2.2 Proposition II – New data sources improve the maintenance decision that is made

The hypotheses tested with the repeated measures ANOVA are similar in experiment size and degrees of freedom, the critical threshold of F is determined as: reject H_0 if $F > 4.325$. Testing for the effects of all data source groups on the accuracy of assessments reported no significant main effects [F (5, 105) =1.370, *p* = .242]

$H7_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to better decision scores*

Testing for the effects of IC4 and IC12 reported no significant main effects of the type of data source on the decision score. The tests reported IC4&IR, [F (1, 21) = 0.140, *p*= .712] and IC12&IR, [F (1, 21) = 0.000, *p*=1.000], which are comparable to the control group. All test have failed to reject the null-hypotheses, indicating that the use of IC4 or IC12 did not show an effect on the quality of the decision. None of the design points testing this hypothesis have tested significant, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H8_A$ – *The use of more objective data sources (IR) lead to better decision scores*

Testing for the effects of IR reported no significant main effects of the type of data source on the decision score, IR, [F (1, 21) =0.005, *p*= .943]. The mean scores for this group were comparable to the control group. All tests indicated that the use of IR did not show an effect on the quality of the decision. None of the design points tested significant, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H9_A$ – *Combinations of new data sources (IC + IR) lead to better decision scores*

Testing for the combined effects of IC4, IC12 and IR reported no significant main effects of the type of data source on the accuracy of assessments. Still contrasts revealed that the score for CUR+IC4+IR, [F (1, 21) = 5.022, *p*= .036], was significantly different than for the control group. Further examination of the means show that the direction of this difference is positive, which can be seen in Figure 4-3.

CUR+IC12+IR, [F (1, 21) = 0.009, *p*=.923], appeared to be very similar to the control group. Further examination showed that the means were exactly the same suggesting that there is no difference. The test for CUR+IC4+IR proved significantly different, while the test for CUR+IC12+IR reported similar values to the control group. These data sources are quite similar and which makes the results contradicting, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

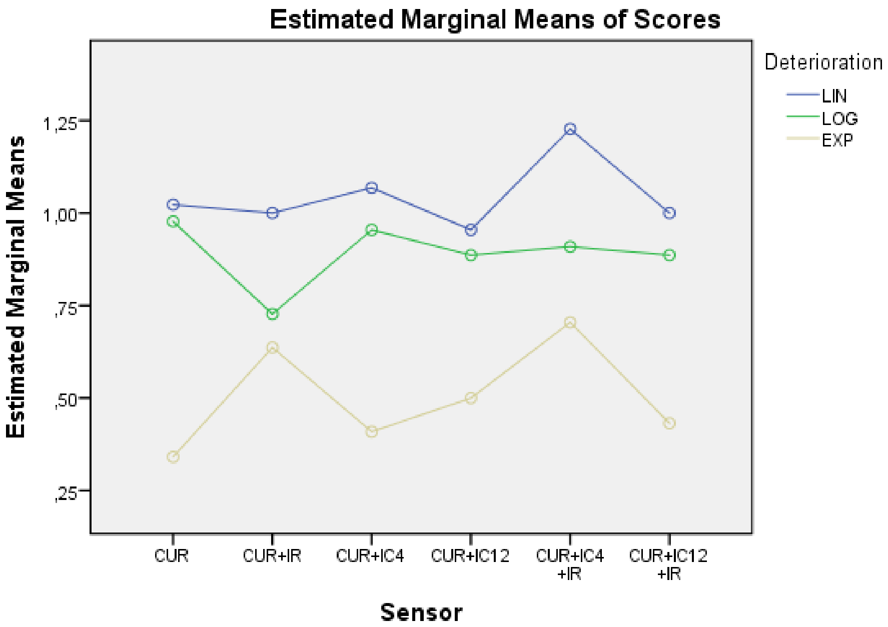


Figure 4-3: Estimated marginal means of decision score

In general it seemed that the use of a new data source did not affect the decision score much, however the deterioration rate did have effect. The repeated measures ANOVA showed no visible trend for the main effect of data sources on decision score, neither did the two way interactions or the three-way interactions. Figure 4-3 shows the estimated marginal means for two-way interactions between data source and the deterioration rate, other interactions produced similar results. The cases in which significant or remarkable results were produced are mentioned per hypotheses.

4.2.3 Proposition III – Maintenance decision are better when assessment is more accurate

When a decision is based on an accurate assessment it is better than when a decision is based on a less accurate assessment. This proposition is used to corroborate the hypotheses of the first two propositions. Linking the two together. In the testing of these hypotheses the cases in which the base quality was good were omitted due to lack of variance within groups (the score was often constant). The accuracy of an assessment was converted to negative absolute values as to correct the scale (from 0 distance being accurate to -x distance being inaccurate). A Kendall's tau-b correlation was used in four distinct cases to determine the relations between (1) the accuracy of assessment and the decision score and between (2) the assessment time and the decision score. Similarly a relation is looked for between (3) the accuracy of assessment and the precision of a decision and between (4) the assessment time and the precision of a decision for 396 decisions. The tests showed a light correlation for the relations relating to accuracy of assessment and no correlation for the relations relating to assessment times. The cases in which significant or remarkable results were produced are mentioned per hypotheses.

$H10_A$ – *More accurate assessments lead to higher decision scores*
Testing for a correlation between accuracy of assessment and decision scores showed that there was a moderate, positive correlation between accuracy of assessment and the received score. This effect was statistically significant at the p = .01 level (τb = .273, $p < .05$). These results suggest that there is a positive correlation between the accuracy of assessment and the decision score, seen in Figure 4-4. Participants who assessed a condition more accurately also received higher decision scores, suggesting that the alternative hypothesis needs to be accepted, rejecting the null hypothesis.

$H11_A$ – *Longer assessment times lead to higher decision scores*
Testing for a correlation between assessment time and decision scores showed that there was no correlation between accuracy of assessment and the size of the decision taken, the effect was statistically not significant at the p = .05 level (τb =.007, p = .856). Longer assessment times appeared not to be correlated with decision scores, his suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H12_A$ – *More accurate assessments lead to smaller actions i.e. being more precise*
Testing for a correlation between accuracy of assessment and the size i.e. precision of an action showed that there was a light, negative correlation between accuracy of assessment and the size of the decision taken. However the effect was statistically not significant at the p = .05 level (τb = -.065, p = .104). Based on these results it appears that participants who have assessed less accurately tend to take bigger decisions, however this effect is neither strong nor significant. This suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

$H13_A$ – *Longer assessment times lead to smaller actions i.e. being more precise*
Testing for a correlation between assessment time and the size of an action showed that there was no correlation and the effects were not statistically significant at the p = .05 level (τb = -.008, p = .843). Longer assessment times appeared not to result in

more precise actions, this suggest that the alternative hypothesis should not be accepted that the null hypothesis should not be rejected.

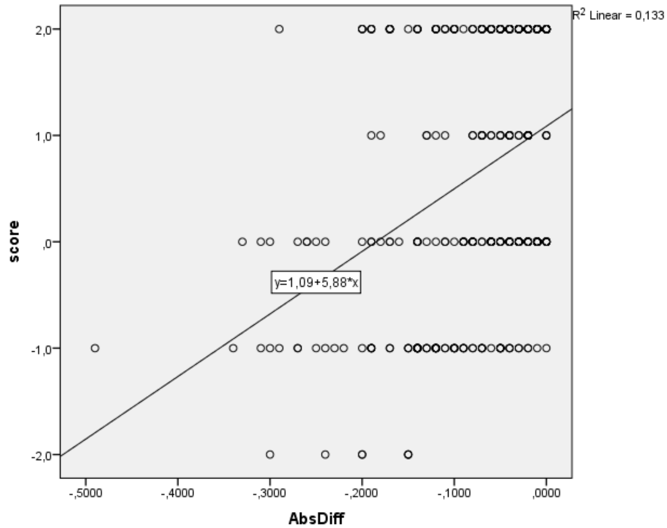


Figure 4-4: correlation between assessment accuracy and decision score

4.2.4 Other significant effects

The results of the Repeated Measures ANOVA did find other non-hypotheses bound significances. These are noted in this paragraph for additional contextual understanding. The main effects of deterioration on accuracy of assessment reported significant on the Exp. deterioration type. Contrasts revealed that the assessment was less accurate for EXP, [F (1, 21) = 79.532, *p* = .001] compared to LIN. While being similar for LOG, [F (1, 21) = 0.484, *p* = .494] compared to LIN.

4.2.5 Summary: differences in assessment and decisions

The fourth guiding question stated: “What are the differences in assessments and decisions of professionals in road management given a certain data source?” This question was answered by analysis of quantitative data from the gaming simulation experiment organised according to the propositions. The hypotheses corresponding to proposition I proposed a relation between new data sources and both, the accuracy of assessment, as the assessment time. This resulted in six hypotheses being tested by means of a repeated measures ANOVA. The statistical tests did not report any significant differences between groups of data sources, suggesting that the alternative hypotheses be rejected for both accuracy of assessment as well as assessment times. The hypotheses corresponding to proposition II proposed a relation between new data sources and the quality of a decision. This resulted in three hypotheses being tested by means of a repeated measures ANOVA. The statistical tests did not report any significant differences between groups of data sources, suggesting that the alternative hypotheses be rejected. Lastly, the hypotheses corresponding to proposition III proposed a relation between both, the accuracy of assessment or the assessment time and the decision score or precision of an action. This resulted in four hypotheses being tested by means of a Kendall's tau-b correlation test. The statistical tests reported a significant moderate correlation between the accuracy of an assessment and the quality of a decision. The remaining tests did not report any significant correlations, suggesting that the alternative hypotheses be rejected. An explanation for these results is looked for in the next section.

4.3 Qualitative data from the surveys and debriefing

The experiment sessions contained three points of qualitative data extraction, the pre session questionnaire, which is to get a better understanding of the participant and to explain for any unusual inputs. The post session questionnaire, which is to assess how the session was and to find out if any unwanted factors played a role. Also the debriefing which is to understand what participants did and why they did it. The latter two were used in the answering of guiding question 5 which stated, "How do professionals assess quality and make a decision within the context of road management?" With the intended goal of supporting and giving meaning to the quantitative results by giving insight into the "black-box" of a professional's assessment of quality deterioration and the following decisions. It relies on the use of introspection, which is defined as the "examination of one's own mental processes" (Collins, n.d.).

4.3.1 Type of information used

Participants were asked to reflect on what information was found most useful during the exercise. They were free to enter anything they wanted and an answer was not needed. Figure 4-5 gives an overview of what was said and by how many people. It is notable that a large part of the participants relied mostly on the use of the table with numeric data, the remarks from the road inspector and the graphical depiction of the historic data. The intermediate data points, as the point cloud analysis, statistical data and trend lines were used much less or not at all.

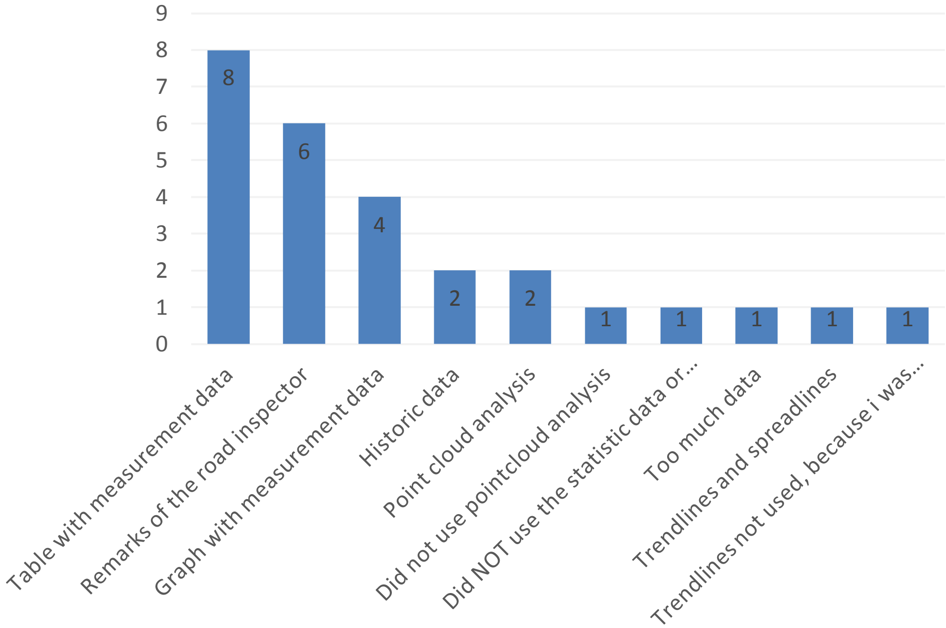


Figure 4-5: most useful information according to participants

From the debriefing it became clear that participants generally just looked at the yearly measurements data and ignored other measurements. When the decision became more difficult some participants resorted to checking with the road inspector to reflect on the correctness of their own analysis. While others completely ignored the road inspector and used the extra data to reflect on their choices. The graphs and tables with less data were easier to comprehend and use for extrapolation, while the sections containing more data made it difficult to extrapolate. It can be concluded that for making assessments and decisions the yearly data and graphs were the main mode of analysis and the extra data or road inspector remarks served for reflection.

4.3.2 Difference with the practice and the effects

Participants were asked what their focus was during the experiment, what information they missed and how they would have acted if they would have that information. Subsequently in the debriefing a similar question was asked about what elements the game missed and how that would have affected their play. From Figure 4-6 it can be seen that there are quite a number of different aspects and specific information which the experts would have used in practice. Two terms stand out more than others, being the cost of fines and discounts and the budgets for vehicle lost hours. While during the experiment most participants focused on quality, some on balancing quality and the use of resources and only one participant had focused on limiting use of resources.

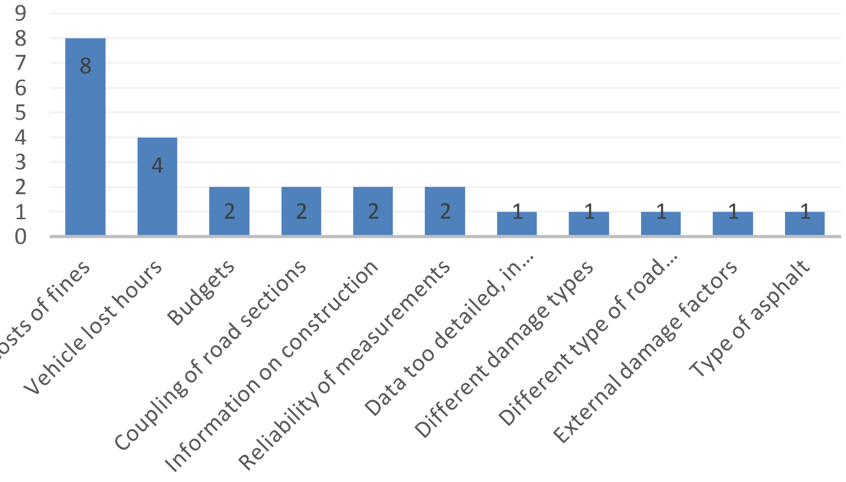


Figure 4-6: what would have affected the choices in practise

In the debriefing these same topics were discussed including their effect on the game play. Many of the participants started with the lack of fines and other budgets in the game. The cost of fines was one of the most mentioned aspects and has a significant effect on the choices made according to the experts.

"In some cases it is cheaper to receive a fine from the contracting authority then to repair a road immediately".

"If fines were incorporated, as I know them, then I would have played the game much more defensively and would have taken much less risk because I'm constantly thinking about the fines."

Also the effect of Vehicle Lost Hours was mentioned quite often, this term refers to the availability of the road which is predetermined in the contract.

"In practise most of the cost is in closing down the road, not the repairs, so it is often cheaper to plan maintenance as a network by coupling as many pieces together instead of just repairing what is in need of repair".

The participants found it particularly difficult to deal with the idea of not knowing what would happen if a quality threshold would be violated. Which shows the high dependency of the assessments and the choices on what the consequences are. Many of the participants explained that they would firstly start with determining what the contractual requirements are and what the consequences are of not meeting those requirements. This would be used to form a strategy and to determine the risk profile to be used for the regarding project. Resulting in a decision matrix with what type of measure to be taken when a specific type of damage occurs. The cycle mentioned by the experts corresponds with the flowchart explained in Figure 2-7.

The main points that became clear are that the technical requirements are only a fraction of the decision to be made. Firstly the contractual requirements and consequences of those requirements are considered, secondly if it fits the long term maintenance planning including the planned Vehicle Lost Hours and thirdly the coupling of adjacent road sections. Other aspects which are mentioned also play a role however appear not to be predominantly in the minds of the experts.

4.3.3 How do experts see the future

When asked how the participants saw the future of road management and asset management in general with regards to new data sources not many participants had an answer ready. However the answers that were given all corresponded with one other. Experts acknowledged that even though it is not a very big part of the industry at the moment, the same the potential of more data. Not as a primary business but as a tool for decision-making support. However many of the experts remarked that data is a difficult thing, it can be very useful but the techniques to analyse it and interpret it must be solid or else it will not have added value. Some experts remarked about the information the game but also data science in general that large quantities of information should not be judged by people. Without reliable algorithms and software data cannot be used to determine concrete actions and sensors don't have much added value compared to what craftsman can deliver. So it is crucial to first work on what can be done with data instead of the mass collection of data. Lastly, the understanding of why people do what they do, if the choices people make can be understood and modelled, than that information can be combined with measurement data to not only give a representative depiction of the world around us but also of what choices could be made with respect to many other input factors.

It was stated by a majority of the experts that progress in the industry cannot be accomplished by single parties and in secrecy. It requires collaboration of the entire chain from contracting authority to sub-contractor.

4.3.4 Summary: how do professionals assess quality

The fifth guiding question stated: "How do professionals assess quality and make a decision within the context of road management?" This question was answered on the basis of some topics discussed in the debriefing and post questionnaire. The experts indicated that (1) the information that was most used was the table with IRI values in which usually the yearly values where assessed. Secondly the remarks of the road inspector were used as a verification of their own assessment. Lastly the graphical depictions of measurement data were used. The remaining data sources were mentioned in either one or two cases. When asked for (2) the difference compared to the practise most experts firstly mentioned the cost of fines and discounts, secondly they mentioned the costs of vehicle lost hours. The remaining differences were only mention once or twice and related mostly to technical information, strategic planning or reliability of measurements. The final noteworthy result was found when asked for (3) the role of new data sources in the future and the link to current practise. Firstly the experts that had a clear idea about this question mentioned that the current practise is far from the ideal practise. Secondly, they remarked that data is a difficult concept which needs to be understood with respect to the context of road management. Emphasising the need for strong algorithms in translating the data to useable information in which the potential role of artificial intelligence and machine learning was acknowledged. Lastly, it was mentioned that there needs to be more attention for the human aspects of decision-making within road management. If these processes are understood they can be combined to increase the strength and reliability of computer models in decision-making. Many parallels are found between the qualitative data collected from experts and the quantitative data collected from the simulation experiment discussed in Chapter 4.2. These will be discussed in the next chapter and an answer will be provided for the main research question.

5 Discussion

In the previous chapter the results from the data analysis were presented for both the quantitative data and the qualitative data the answering of the fourth and fifth guiding questions. In this chapter an overview is given of the guiding questions and their link to literature, the limitations of the research are presented and the answers to the guiding questions are discussed with respect to the research limitations. The chapter is concluded by presenting suggestions for further research.

5.1 Overview of the results

The goal of this study was to find out if and how professionals in road management would assess quality differently and if they would change their decisions if they were given information from new data sources. To achieve this goal five guiding research questions were formulated supporting a main question. The first three were used to define the research design and the last two to answer the main research question. A game-based simulation experiment was built in which eleven professionals, fitting a specific profile, participated. These sessions were used to generate quantitative data to enable answering of the fourth question which relates to the measurable difference in output and qualitative data to enable answering the fifth question which relates to finding an explanation for the quantitative results. The following sections are designed around these last two questions and will build towards answering the main research question. The results of answering guiding questions 4 and 5 will be discussed, the limitations of this research will be given and the chapter will be concluded with recommendations for further research.

5.1.1 Assessments and decision-making of professionals

This section will discuss to the fourth guiding question along the line of three propositions and will accept or reject each of the hypotheses. The fourth guiding question stated: "What are the differences in assessments and decision-making of professionals in road management given a certain data source?"

Proposition I – accuracy of assessment

The first proposition suggested a positive relation between new data sources and the accuracy of assessment and resulted in the formulation of three hypotheses. An effect was looked for in the quantitative data, however the statistic tests found no consistent effects, thus failing to reject the null-hypotheses H1 to H3. This contradicts the expectations and the general belief in practice that the use of IoT sensors and Big Data aid an organisation in become more predictive in the management of their assets. The results imply that the use of information from new data sources does not allow for more accurate assessment of assets behaviour.

Proposition I – assessment time

The first proposition also suggested a positive relation between new data sources and the time to assess, also resulting in the formulation of three hypotheses. Similarly for these hypotheses no effects of information from new data sources on the assessment time were found, thus failing to reject the null-hypotheses H4 to H6. This contradicts the intuitive expectation that more information leads to a higher time requirement, it additionally contradicts with the experiences from the design sprints where the amount of information was found to increase session times. The results imply that the use of information from new data sources do not increase assessment times. Which might be positive for the use of big data in decision-making.

Proposition II – decision quality

The second proposition suggested a positive relation between new data sources and the decision quality and resulted in the formulation of three hypotheses. An effect was looked for in the quantitative data, however the statistic tests found no effects, thus failing to reject the null-hypotheses H7 to H9. Similarly as in proposition I this contradicts the expectations and the general belief that the use of IoT sensors and Big Data aid an organisation in become more predictive in the management of their assets. Additionally it contradicts the literature in the importance of information which is timely and accurate as a key part within asset management (Brous & Janssen; Haarman & Delahaye, 2004; TheIAM, 2008). The results imply that the use of information from new data sources does not allow for better maintenance decisions.

Proposition III – assessment accuracy and decision quality

The third proposition suggested a positive relation between assessment accuracy and the decision quality, and the action size, resulting in the formulation of a hypothesis for each. An effect was looked for in the quantitative data and the statistic tests found that more accurate assessments resulted in better decisions, thus rejecting H10 and accepting the alternative hypothesis. This is in line with the expectations and the literature on asset information. The tests also found a light positive relation between assessment accuracy and size of the action, however this effect was not statistically significant thus failing to reject the null-hypothesis H11. This result contradicts with the effect found between assessment accuracy and improved decision quality. But on the other hand the light positive relation between assessment accuracy and size of the action, even though not significant, is in line with the effect found between assessment accuracy and improved decision quality.

Proposition III – assessment time and decision quality

The third proposition also suggested a positive relation between assessment time and the decision quality, and the action size, resulting in the formulation of a hypothesis for each. An effect was looked for in the quantitative data and the statistic tests found that longer assessment times did not result in better decisions, nor did it result in smaller actions, thus failing to reject the null-hypotheses H12 and H13. This result contradicts the expectation that if more time is needed (possibly due to use of data sources) it results in better decisions and more exact actions.

Summarising

It appeared that the use of new data sources in addition to current data sources did not result in much difference in either assessment or decision-making. It was found that if an assessment is more accurate it improves decision quality (tested significant) and reduces actions size (tested not significant). No effect was found to support the claim that longer assessment times lead to better decision or exacter actions. Table 5-1 gives an overview of the found results.

Table 5-1: summary of hypothesis testing

Alternative hypothesis	*Result*	*Result*
$H1_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to more accurate assessments of quality*	H_0 not rejected	H_A not accepted
$H2_A$ – *The use of more objective data sources (IR) lead to more accurate assessments of quality*	H_0 not rejected	H_A not accepted
$H3_A$ – *Combinations of new data sources (IC + IR) lead to more accurate assessments of quality*	H_0 not rejected	H_A not accepted
$H4_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to longer assessment times*	H_0 not rejected	H_A not accepted
$H5_A$ – *The use of more objective data sources (IR) lead to longer assessment times*	H_0 not rejected	H_A not accepted
$H6_A$ – *Combinations of new data sources (IC + IR) lead to longer assessment times*	H_0 not rejected	H_A not accepted
$H7_A$ – *The use of new data sources with increased frequency of measurements (IC4 & IC12) lead to better decision scores*	H_0 not rejected	H_A not accepted
$H8_A$ – *The use of more objective data sources (IR) lead to better decision scores*	H_0 not rejected	H_A not accepted
$H9_A$ – *Combinations of new data sources (IC + IR) lead to better decision scores*	H_0 not rejected	H_A not accepted
$H10_A$ – *More accurate assessments lead to higher decision scores*	H_0 rejected	H_A accepted
$H11_A$ – *Longer assessment times lead to higher decision scores*	H_0 not rejected	H_A not accepted
$H12_A$ – *More accurate assessments lead to smaller actions i.e. being more precise*	H_0 not rejected	H_A not accepted
$H13_A$ – *Longer assessment times lead to smaller actions i.e. being more precise*	H_0 not rejected	H_A not accepted

5.1.2 Explaining the black box

This section will discuss the fifth guiding question, which stated: "How do professionals assess quality and make a decision within the context of road management?"

Expert strategy and use of information

The main finding from guiding question five is that participants mostly used the information that was known to them and is most relatable to the current practise. The experts generally employed a similar strategy which was based on a three step approach. The general strategy was (1) the use of measurement values in which only the yearly data points were used for assessment (by most experts). This information was used to (2) linearly extrapolate the last measured value to find the value which was asked for. Some experts noted that they applied a correction factor when they noticed heavier increase in the value (i.e. exponential deterioration rate). This effect was noticeable in the additional graphs but their use was not mentioned by the participants. Some even explicitly mentioned ignoring them because their strategy was to approach the questions mathematically. Lastly, the found value was (3) validated with the remarks made by the road inspector. In some cases an additional graph was added containing objective measurements corresponding with the remarks of the road inspector. However also the use of these graphs was minimally mentioned and in some cases even ignored. The new data sources were experienced to be too much and irrelevant for the assessment and decision-making. It was mentioned that this information would have more value if calculations were used, suggesting the applicability in computational models.

Expert focus

The experts were asked what their focus was during the experiment, most expert responded with focusing on quality or balancing quality and resources. While when asked for the difference compared to the practise, the most mentioned aspects were costs of fines and costs of vehicle lost hours. Which are both mechanism to respectively guarantee a certain quality and availability. Multiple experts mentioned that if a fine due to lack of quality is cheaper than repairs, the fine will be chosen. Suggesting that costs are in fact the main drivers in practise.

5.2 Limitations

Following from the application of the method, the analysis of the data, the analysis of the results and the discussion of the guiding questions three limitations are formulated which might have an effect on the discussion on the general results and the formulation of a conclusion.

Research design

The experiment and the field experiment are both very strong methods of testing hypotheses. Where the former has significant power and the capability of measuring at a variable level with a large certainty that what is intended to be measured is actually measured (internal validity). The latter has significant power in the sense that it is very good at mimicking the real world, especially if a quasi-experiment is employed. However what it makes up for in realness (external validity) it lacks in the amount of control over what is measured. This research employ a hybrid method by using a simulation as being an artefact. Defined by Lo et al. (2013) as the artefactual field experiment, which is a field experiment related to the laboratory setting, resulting in the outcomes of this research being valid but less generalizable to the real world. Because the artificial setting inherent to experiments, in which variables are systematically excluded and locked, may alter the behaviours or responses of participants. Additionally the exclusion of factors increases the clearness of causality generates a less realistic effect than when more factors could be considered.

Time advantage

A second limitation is the due to the equality of grounds principle which is a strong element to have in a research design, since it does not offer unfair advantages to any of the groups under research. However it does have the effect of taking away advantages that might be present in a real world situation. In this experiment the element of Time was not taken into account. E.g. a participant that has monthly data can extend the exact moment of decision-making because they can review a section in month 11 while a participant who assess yearly will have to make the decision up front and will have to survey visually what the deterioration does. The participants in this research were provided with information which made assessment and decision-making possible on equal grounds. By doing so removed an advantage of sensor data, which is its on-demand availability.

Testing of hypotheses

In the testing of the hypotheses a limitation was found which has two origins. The testing of one hypothesis (H12) did not produce results which were significant. On the one hand this might be due to the size of the experiment. But more likely this is due to the operationalisation of the construct used to measure it i.e. the size of action. It is believed that with a larger scale of actions, a clearer result would have been seen.

5.3 Discussion of results

In the results section it was found that the use of sensor data with a higher frequency of measurement, additional objective measurements and the combination of both do not affect the assessment of quality deterioration made by professionals within a game based simulation experiment. Nor does it affect their in-game decisions. Some possible explanations were found in the qualitative debriefing. Also the limitations of this research were addressed. With this information the results of the research can be discussed in a broader context.

Information load

From de debriefing it was found that the participants stated that the additional data sources were too much. It made it (1) difficult for them to extrapolate with the quantity of data and they could not (2) apply the trend lines and other graphs in their assessment process. They found that it was easier and quicker to use the basic information for assessment. These findings lightly correspond to the research of Eppler and Mengis (2004) who have defined the effects and exhibitions of information overloading in people. Some of these characteristics were also mentioned in the debriefing, being the tendency to (1) lose control of information and (2) ignore information deemed irrelevant, becoming selective. This also corresponds with the experiences from the design sprints in which data quantity and complexity was subsequently decreased. This explanation builds on the theory of Brous (2017) who states that the shift to data driven asset management requires for trust to be built and the integration of IoT data in the decision-making processes, by emphasising the importance of good algorithms that assist human assessors without overloading them with information and sufficient training in the handling of data.

Organisational level

An alternative explanation is that the use of added new sensor data do not affect the decision-making at this level within the organisation of road management because the advantages of new data sources are more heavily related to the short term road management cycle as defined by (Haider, 2012) and discussed in the limitations(time advantage). In reality the current data sources, i.e. the inertial profiler data, require some time to be analysed and reported, becoming available months after the measurements have taken place. Whereas the new data sources, aim at a semi-continuous monitoring system which can be viewed at any given moment. Introducing this added advantage of these data sources could show significant changes in assessment accuracy and decision-making. This corresponds with the remarks of some participants during the qualitative debriefing in which they state that intermediate measurements would be very helpful in expectation management. They would result in quality conditions being accessible at any given moment. It would prevent chaos and hasty decision-making as a result of unexpected (bad) quality conditions.

Effects of secondary requirements

A different alternative explanation is that the use of added new sensor data did not affect the decision-making of participants because technical quality might not be of importance in maintenance decisions. From the debriefing the suggestion was raised that the most important driver in maintenance decisions is linked to the monetary factors both direct and indirect (resp. cost of action and discounts due to violated requirements or purchase of additional VLHs). It is expected that, with the addition of those factors, the decisions would significantly change. This corresponds strongly with the findings from the design sprints in which the participant who had contractual requirements was much more risk averse and cautious in decision-making compared to the participant who had a target value. It also corresponds with the qualitative debriefing in which nearly all participants made a statement about the effect of money or contractual requirements on the choices. It seems that the more stringent a contract the more decisions are driven by monetary factors to a point where violating a requirement is preferred above fast rehabilitation because the cost of the fine is cheaper than the cost of direct rehabilitation.

Man versus machine

A final alternative explanation is that the use of added new sensor data does not affect the assessment of quality deterioration made by professionals because it is inherent to human nature to employ heuristics when performing assessments and especially when making decisions (Kahneman, 2011). This corresponds with the findings from initial interviews and the design sprints in which participants mentioned the preference dependency of maintenance actions. In projects this is countered by the use of decision matrices made at project commencement to take away Subjectiveness towards a measure. An addition may be made to this explanation by relating it to the expectation that machines might be better at assessing quality deterioration and forming decision, given that they are trained correctly. This corresponds with statements from the qualitative debriefing in which participants suggested that it should be computers who process and assess the new information.

5.4 Further research

Building on the limitations and results, suggestion for further research are given.

From game-based simulation experiment to experimental simulation game

Building on the limitation that this research was performed in a laboratory-like setting and the aspect of time was excluded. Further research is suggested to determine if new data sources have any impact on assessment and decision-making when the operations and maintenance phase is viewed in its totality (including contractual obligations, monetary and time related factors as well as multiple stakeholders). It could be used to include advantage of time as mention in the limitations

Man and machine or man versus machine

Building on the experience that participants did not use the data sources, the mentioning of a need for better algorithms and machine learning. Further research is suggested in two directions based on the use of smart algorithms and advanced computing technologies such as big data analytics, machine learning and artificial intelligence to analyse datasets. The first suggestion is towards the use of a hybrid process, where machines, based on current and added sensor data, make the assessments and suggest decisions based on the organisation’s goals and budgets. This suggestion is reviewed by a human and used for further decision-making. A comparison between participants aided by Machine Learning algorithms and participants who do use only the human capacity can show the expected added benefits of Machine Learning algorithms. The second is a more extreme suggestion based on a fully automated process by making the comparison between human made assessments and decision and those made by a computer employing Machine Learning algorithms.

6 Conclusions

The previous chapter presented an overview of the results and linked these back to the testing of the hypotheses and to literature. It elaborated on the qualitative data and lead to the limitations of this research. Subsequently the results their implications were linked to the limitations to discussed in a more integral context. It concluded by pointing out areas of interest which emerged from this research or were not incorporated. This chapter looks back on the research and combines the arguments from the discussion into a final conclusion with academic and practical implications, and business recommendations. The chapter is concluded by philosophising about the place of this research in the bigger picture and presents the author's spot on the horizon extending beyond road management.

6.1 Overview of the research

In the introduction a problem was defined which stated that: "Currently decisions in the Operations and Maintenance phase of road infrastructures are based on low frequency measurements and subjective observations. Technological innovations offer possibilities but remain unused because their contribution is unknown." Which lead to the main research question being formulated and five guiding questions being stated to give direction to the research with the goal of uncovering what the effects are of new data sources on the assessment of quality and the decision-making of professionals in road management.

The second chapter was structured around the first three guiding questions, providing the context to understand the processes and elements of road management, and diving deeper into road maintenance, sensor technology and decision-making within the context of road management. The second chapter concluded by combining the information and linking it to a theoretical framework and a set of hypotheses.

Chapter three focussed on the building of a research method, for which an experiment was chosen as the main manner of testing the hypotheses. The characteristics of this problem lead to the choice of a simulated experiment and serious gaming as a framework was used to operationalize the experiment design. A simulation model was built and a game was designed to enable a selection of professionals i.e. experts in road management, to use this instrument for the purpose of data collection. To aid in understanding the quantitative data, a qualitative data extraction tool was used. The briefing theory generated the context in which the data could be analysed and understood. The game-based simulation experiment sessions resulted in two outputs, being: a (1) quantitative data set containing information on assessment and decision-making and (2) a qualitative data set enabling a look into the "black box" of decision-making. The data was analysed by means of a within subject analyses of variances and certain relations were looked for by means of a correlation analyses.

The results of this data analysis were presented in chapter four and an overview was given of the main findings.

Chapter five reviewed the fourth and fifth guiding questions. It answered what the differences in assessment and decision-making are when using new data and it presented the limitations of this research. The chapter concluded by binding all the results and answers to guiding questions including their limitations into a discussion which lead to suggestions for further research. The main research questions was prepared for answering.

6.2 Answering the main question

The question stated in the first chapter of this thesis stated:

"What is the effect of new data sources on the assessment of quality deterioration and decision-making of professionals in road management?"

Based on the results and discussion of this research, it can be said that this research did not find any effects of new data sources on the assessment of quality and the decision-making of professionals when presented with this information. The professionals in road management who participated in this research did not use the new data sources sufficiently or did not experience the new data sources as beneficial. The discussions held during debriefings support this conclusion by laying the emphasis on finding a balance between collecting data and building the tools to use this data.

6.3 Implications

From the experience and knowledge gained in this research a number of practical implications and business recommendations can be given to the hosting company. These are structured on the method of research and the research content.

6.3.1 Academic implications

Even though no formal conclusion could be presented, the findings do have implications for the academic world. This research contributes to the academic world by testing the effects of simulated innovations which are eagerly developed by many researcher, but are not implemented by practice. It suggests there is a gap between what is possible, what is desired and what is practical. Secondly, it contributes to the serious gaming body of knowledge by adding to the game based simulation experience in this field. Within this domain, the use of serious games as a quasi-experiment or field experiment is growing increasingly, however the use of a simulation based gaming experiment (the artefactual field experiment) is remains relatively low.

6.3.2 Business implications

In parallel it also contributes to practice by emphasizing on the need for more research and pilots with big data, machine learning etc. In practice it was found that innovative techniques are not easily used because the methods are not sufficiently developed. At the same time the methods are not sufficiently developed because the innovations are unusable. This leads to a number of business recommendations.

Serious gaming as a professional tool

In the numerous conversations and discussion serious gaming has always been an interesting topic. In this research gaming was used as the operationalisation of an experiment. In this role the game became a very fragile tool that served solely the purpose of information collection. However in a different context, that of learning tool or simulation tool, it can have much business advantages (Oprins, Bakhuys-Roozeboom, Visschedijk, & Kistemaker, 2013). As a learning tool it can be used to train junior asset managers the line of thought in long term contracts and the effects of choices. While as a simulation tool it can be used to run through different scenario's and experience first-hand what effects design and operational choices have in the long term.

Choices in sensing methods

Experts are well known with the defects of current methods, being: the time and capital intensiveness of specialised measurements, their susceptibility to external factors and the subjectivity around visual inspections. Even though this research has concluded that new sensor data has no effect when used without proper instrumentation, the use of Big Data and IoT based sensors can aid in decision-making by providing more continuity regarding quality conditions and more precisely the moment of failure. The testing of these methods should happen in parallel to the current methods so experts can gain trust in the methods and fine tuning can be done based on current standards. Furthermore the choice in sensory tools should be focussed on making everything around the road smart while keeping the road "unintelligent" because of the ease of implementation in other areas. This development is connected to the use of machine learning tools.

6.4 Concluding remarks

From this research a bigger picture can be drawn. The use of more data appeared not create any effect in the assessment and decision-making of professionals. However it was shown that better assessments lead to better decisions. This is in line with what much literature suggests and implies the building of a process in which current information sources have a role but there is also room for new information sources. The current practise shows a large gap in this field, both from the client side in terms of regulations as from the contractor side in terms of focus on the financial component. The step towards a data driven management process is not only about measuring an assets condition, predicting its failure and thus gaining financial advantage. It is about learning from our environment and using these lessons to improve upon existing processes e.g. to improve upon design principles by learning from user behaviour and gaining a better understanding of road comfort and perceived road safety. The emergence of autonomous driving and connected devices offers the world a vast amount of sensors which do not need to be plugged in or incorporated in the design of assets. Thus allowing for future roads and structures to be kept only as smart as they need to be but utilising the smart sensors that are already around to make the shift towards smart asset management. The future of our world will be shaped by how we are connected to our environment, this research is a small step for the civil industry towards being part of that connected world.

7 References

Amick, R. Z., Patterson, J. A., & Jorgensen, M. J. (2013). Sensitivity of tri-axial accelerometers within mobile consumer electronic devices: a pilot study. *International Journal of Applied, 3*(2).
Beuckeleer, K. d. (2013). Thesis vergemakkelijkt onderhoud wegen. *Het Nieuwsblad*. Retrieved from http://www.nieuwsblad.be/cnt/dmf20131204_00871553
Bijleveld, F., & Doree, A. (2013). Aspari is paving forward, professionalisering uitvoeringsproces 2.0. *Asfalt, 40*, 10-12.
Boerboom, S. (2013, july 2013). CROW gaat certificering langsvlakheids metingen verzorgen. *CROWetcetera*.
BouwendNederland. (2012). Richtlijn tweelaags ZOAB. *Vakgroep Bitumineuze Werken*. Retrieved from
BouwendNederland. (2016). Rijchtlijn 2laagsZOAB. *Vakgroep Bitumineuze Werken*.
Brous, P. (2017).
Brous, P., & Janssen, M. (2015). *A Systematic Review of Impediments Blocking Internet of Things Adoption by Governments.* Paper presented at the Conference on e-Business, e-Services and e-Society.
Brous, P., Janssen, M., & Herder, P. (2016). Coordinating Data-Driven Decision-Making in Public Asset Management Organizations: A Quasi-Experiment for Assessing the Impact of Data Governance on Asset Management Decision Making. In K. Y. Dwivedi, M. Mäntymäki, M. N. Ravishankar, M. Janssen, M. Clement, L. E. Slade, P. N. Rana, S. Al-Sharhan, & C. A. Simintiras (Eds.), *Social Media: The Good, the Bad, and the Ugly: 15th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2016, Swansea, UK, September 13–15, 2016, Proceedings* (pp. 573-583). Cham: Springer International Publishing.
Canossa, A., Martinez, J. B., & Togelius, J. (2013). *Give me a reason to dig Minecraft and psychology of motivation.* Paper presented at the Computational Intelligence in Games (CIG), 2013 IEEE Conference on.
CBS. (2015). Verkeersintensiteiten op rijkswegen.
Cobouw. (2016). VolkerWessels heeft primeur HoloLens. *Cobouw*.
Collins. (n.d.). introspection. *Collins English Dictionary, Complete and Unabridged, 12th Edition 2014.* Retrieved from http://www.thefreedictionary.com/introspection
Collins, L. M., Dziak, J. J., Kugler, K. C., & Trail, J. B. (2014). Factorial experiments: efficient tools for evaluation of intervention components. *American journal of preventive medicine, 47*(4), 498-504.
Coval, J., Jurek, J., & Stafford, E. (2009). The economics of structured finance. *The Journal of Economic Perspectives, 23*(1), 3-25.
CROW. (2011a). Publicatie 146a: Handboek visuele inspectie 2011.

CROW. (2011b). Publicatie 146b: Handleiding globale visuele inspectie 2011.
Dawkins, J., Bishop, R., Powell, B., & Bevly, D. (2011). Investigation of Pavement Maintenance Applications of Intellidrive SM (Final Report): Implementation and Deployment Factors for Vehicle Probe-based Pavement Maintenance (PBPM). *Auburn University*.
deWegenscanners. (2015). *wegonderzoek volgens de wegenscanners* Retrieved from
Dittrich, A., Weinmann, M., & Hinz, S. (2017). Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data. *ISPRS Journal of Photogrammetry and Remote Sensing, 126*, 195-208. doi:http://dx.doi.org/10.1016/j.isprsjprs.2017.02.012
Dubois, A., & Gadde, L.-E. (2002). The construction industry as a loosely coupled system: implications for productivity and innovation. *Construction Management and Economics, 20*(7), 621-631. doi:10.1080/01446190210163543
Duncan, S. C. (2011). Minecraft, beyond construction and survival. *Well Played: a journal on video games, value and meaning, 1*(1), 1-22.
Dunford, A. (2013). *Friction and the texture of aggregate particles used in the road surface course.* University of Nottingham.
DWW. (2002). Schadebeoordeling en interventieniveaus voor het verhardingsonderhoud. *Dienst Weg- en Waterbouwkunde*.
Economist, T. (2016). Wireless: the next generation. *The Economist*. Retrieved from http://www.economist.com/node/21693197/print
Eenbergen, C. v. (2016). Huawei focust vooral op 4.5G, 5G en Internet of Things. *Techzine*. Retrieved from http://www.techzine.nl/blogs/redactioneel/918/huawei-focust-vooral-op-4-5g-5g-en-internet-of-things.html
Ellis, J. L. (2003). *Statistiek voor de psychologie*: Boom.
Engel, E. M., Fischer, R. D., & Galetovic, A. (2010). The economics of infrastructure finance: Public-private partnerships versus public provision. *EIB papers, 15*(1), 40-69.
Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society, 20*(5), 325-344. doi:10.1080/01972240490507974
Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, 35*(3), 379-383.
Festing, M., Overend, P., Cortina Borja, M., & Berdoy, M. (2016). The Design of Animal Experiments: Reducing the use of animals in research through better experimental design.(Revised and updated edition): Sage Publications Ltd.
Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Forslöf, L., & Jones, H. (2013). *Roadroid: continuous road condition monitoring with smart phones.* Paper presented at the IRF 17th World Meeting and Exhibition, Riyadh, Saudi Arabia. Available from http://www. roadroid. com/common/References/IRF.

Fujita, Y., Kobayashi, I., Hoshino, Y., & Chanseawrassamee, W. (2016). Development of Attribute-Assign-Editor for Road Surface Point Cloud Data. *International Journal of Engineering and Technology, 8*(3), 170.

Gibbs, G. I. (1974). Handbook of Games and Simulation Exercises.

Guala, F. (2002). Models, simulations, and experiments *Model-based reasoning* (pp. 59-74): Springer.

Haarman, M., & Delahaye, G. (2004). Value Driven Maintenance. *Dordrecht: Mainnovation*.

Haider, A. (2012). Information and Operational Technologies Governance Framework for Engineering Asset Management *Engineering Asset Management and Infrastructure Sustainability* (pp. 299-313): Springer.

Haider, A., Koronios, A., & Kumar, A. (2012). An object oriented road asset information model *Engineering Asset Management and Infrastructure Sustainability* (pp. 283-297): Springer.

Harteveld, C. (2011). *Triadic game design: Balancing reality, meaning and play*: Springer Science & Business Media.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics, 154*, 72-80.

Helbing, D., Giannotti, F., Pedreschi, D., Pentland, A., Lukowicz, P., Kossmann, D., & Crowley, J. (2012). A planetary nervous system for social mining and collective awareness. *The European Physical Journal Special Topics, 214*(1), 49-75.

Herold, M., Roberts, D., Noronha, V., & Smadi, O. (2008). Imaging spectrometry and asphalt road surveys. *Transportation Research Part C: Emerging Technologies, 16*(2), 153-166. doi:http://dx.doi.org/10.1016/j.trc.2007.07.001

Hoogen, J. v. d., Lo, J., & Meijer, S. (2014). *Debriefing in gaming simulation for research: opening the black box of the non-trivial machine to assess validity and reliability*. Paper presented at the Proceedings of the 2014 Winter Simulation Conference, Savannah, Georgia.

Hoogen, J. v. d., & Meijer, S. (2015). Gaming and Simulation for Railway Innovation. *Simulation & Gaming, 46*(5), 489-511. doi:doi:10.1177/1046878114549001

Hunt, P. D., & Bunker, J. M. (2003). Time series analysis of pavement roughness condition data for use in asset management.

Interview-VI. (2016).

Ivetić, D., Mihić, S., & Markoski, B. (2010). Augmented AVI video file for road surveying. *Computers & Electrical Engineering, 36*(1), 169-179. doi:http://dx.doi.org/10.1016/j.compeleceng.2009.08.003

JaguarLandRover. (2016). Pothole detection technology. *Jaguar Land Rover Advanced Research Center*.

Jones, H., & Forslof, L. (2014). *Roadroid continuous road condition monitoring with smartphones.* Paper presented at the Proceedings of the 5th SARF/IRF Regional Conference, Pretoria, South Africa.

Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.

Kenniscentrum, P. (2008). DBFM Handboek: Ministerie van Financiën, Den Haag.

Klunder, G., Aalst, W. v., Willemsen, D., Vliet, D. v., Leegwater, G., & Ruijs, P. (2010). *Auto als Sensor – Quick-scan naar detectiemogelijkheden van wegdekschade* Retrieved from

Knapp, J., Zeratsky, J., & Kowitz, B. (2016). *Sprint: how to solve big problems and test new ideas in just five days*: Simon and Schuster.

KOAC-NPC. (2016). Wegoppervlakeigenschappen. Retrieved from http://www.koac-npc.com/Flex/Site/Page.aspx?PageID=26492&Lang=NL

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics, 29*(2), 196-210.

Kong, Q., Kwon, Y.-W., Schreierz, L., Allen, S., Allen, R., & Strauss, J. (2015). *Smartphone-based networks for earthquake detection.* Paper presented at the Innovations for Community Services (I4CS), 2015 15th International Conference on.

Kulkarni, A., Mhalgi, N., Sagar Gurnani, D., & Giri, N. (2014). Pothole Detection System using Machine Learning on Android.

Lee, H., Bruce Blundell, S., Starek, M. J., & Harris, J. G. (2013). Airborne lidar point cloud-based below-canopy line-of-sight visibility estimator. *Journal of Applied Remote Sensing, 7*(1), 073510-073510. doi:10.1117/1.JRS.7.073510

LightMotion. (2016). The all-in-one street light. Retrieved from http://lightmotion.nl/

Lo, J., Hoogen, J. v. d., & Meijer, S. (2013). *Using gaming simulation experiments to test railway innovations: implications for validity*. Paper presented at the Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World, Washington, D.C.

Lub, R.-J. (2016). Smartphones als meetinstrument voor langsonvlakheid.

Martinsuo, M., & Poskela, J. (2011). Use of Evaluation Criteria and Innovation Performance in the Front End of Innovation*. *Journal of Product Innovation Management, 28*(6), 896-914. doi:10.1111/j.1540-5885.2011.00844.x

McElhinney, C., Kumar, P., Cahalane, C., & McCarthy, T. (2010). *Initial results from European Road Safety Inspection (EURSI) mobile mapping project.* Paper presented at the ISPRS Commission V Technical Symposium.
Mead, R. (1988). The design of experiments: Cambridge University Press, Cambridge.
Meadows, D. L. (2001). Tools for understanding the limits to growth: Comparing a simulation and a game. *Simulation & Gaming, 32*(4), 522-536.
Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., & Selavo, L. (2011). *Real Time Pothole Detection Using Android Smartphones with Accelerometers.* Paper presented at the 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS).
Meijer, S. (2009). The organisation of transactions. *Studying supply networks using gaming simulation. Wageningen University*.
Molenaar, A., & Houben, L. (2002). *Geometrisch en constructief ontwerp van wegen en spoorwegen. Deel C: constructief ontwerp van wegen*: Technische Universiteit Delft.
Moore, M., Graybeal, B. A., Rolander, D. D., Washer, G. A., & Phares, B. M. (2004). Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability. *Journal of Bridge Engineering, 9*(4), 403-413. doi:doi:10.1061/(ASCE)1084-0702(2004)9:4(403)
Morosiuk, G., Riley, M., & Odoki, J. (2004). Modelling road deterioration and works effects-version 2-Highway Development and Management-HDM-4. *Highway development and management series*.
Oprins, E., Bakhuys-Roozeboom, M., Visschedijk, G., & Kistemaker, L. (2013). *Effectiviteit van serious gaming in het onderwijs*: Soesterberg: TNO.
Pavement-Interactive. (2008). Flexible pavement distress. Retrieved from http://www.pavementinteractive.org/category/pavement-management/pavement-distresses/flexible-pavement-distress/
PIARC, W. R. A. (2014). *The importance of road maintenance*
Puttonen, E., Briese, C., Mandlburger, G., Wieser, M., Pfennigbauer, M., Zlinszky, A., & Pfeifer, N. (2016). Quantification of Overnight Movement of Birch (Betula pendula) Branches and Foliage with Short Interval Terrestrial Laser Scanning. *Frontiers in Plant Science, 7*(222). doi:10.3389/fpls.2016.00222
QlikSense. (2016). QlikView - Powerful data discovery and interactive guided analysis. Retrieved from http://www.qlik.com/us/resource-library/qlikview-overview
Reis, E. (2011). The lean startup. *New York: Crown Business*.
Rezgui, Y., & Zarli, A. (2006). Paving the Way to the Vision of Digital Construction: A Strategic Roadmap. *Journal of Construction*

Engineering and Management, 132(7), 767-776. doi:doi:10.1061/(ASCE)0733-9364(2006)132:7(767)
Rijkswaterstaat. (2007). Evaluatie en optimalisatie DTM-Asfalt.
DBFM overeenkomst, (2009).
Rijkswaterstaat. (2017).
Sanchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., Gutierrez, V., . . . Pfisterer, D. (2014). SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks, 61*, 217-238. doi:http://dx.doi.org/10.1016/j.bjp.2013.12.020
Sayers, M. W., Gillespie, T. D., & Paterson, W. D. (1986). *Guidelines for conducting and calibrating road roughness measurements*.
Seraj, F., van der Zwaag, B. J., Dilo, A., Luarasi, T., & Havinga, P. (2016). *RoADS: A road pavement monitoring system for anomaly detection using smart phones.* Paper presented at the Big Data Analytics in the Social and Ubiquitous Context: 5th International Workshop on Modeling Social Media, MSM 2014, 5th International Workshop on Mining Ubiquitous and Social Environments, MUSE 2014, and First International Workshop on Machine Learning for Urban Sensor Data, SenseML 2014, Revised Selected Papers.
Tableau. (2016). Answer questions at the speed of thought with Tableau Desktop. Retrieved from https://www.tableau.com/products/desktop
TheIAM. (2008). Key Principles. Retrieved from https://theiam.org/knowledge/diagrams/key-principles
TheIAM. (2014). The "Big Picture".
Thoughtbot. (n.d.). Introduction to Product Design Sprints. Retrieved from https://thoughtbot.com/playbook/product-design-sprint/introduction
TNO. (2010). *Auto als Sensor - quick-scan naar detectiemogelijkheden van wegdekschade*. Retrieved from
Van den Hoogen, J., Lo, J., & Meijer, S. (2014). The debriefing of research games: A structured approach for the validation of gaming simulation outcomes. *The shift from teaching to learning*, 88-99.
Van Geem, C., Bellen, M., Bogaerts, B., Beusen, B., Berlémont, B., Denys, T., . . . Hellinckx, P. (2016). Sensors on Vehicles (SENSOVO)–Proof-of-concept for Road Surface Distress Detection with Wheel Accelerations and ToF Camera Data Collected by a Fleet of Ordinary Vehicles. *Transportation Research Procedia, 14*, 2966-2975.
van Oosterom, P., Martinez-Rubi, O., Ivanova, M., Horhammer, M., Geringer, D., Ravada, S., . . . Gonçalves, R. (2015). Massive point cloud data management: Design, implementation and execution of a point cloud benchmark. *Computers & Graphics, 49*, 92-125. doi:http://dx.doi.org/10.1016/j.cag.2015.01.007
van Riel, W., van Bueren, E., Langeveld, J., Herder, P., & Clemens, F. (2016). Decision-making for sewer asset management: Theory and practice. *Urban Water Journal, 13*(1), 57-68.

VBW-Asfalt. (2005). Asfalt onderhoudstechnieken : voorkomen is beter dan genezen. Retrieved from

Verschuren, P., Doorewaard, H., Poper, R., & Mellion, M. (2010). *Designing a research project* (Vol. 2): Eleven International Publishing The Hague.

VolkerInfraDesign. (2016). Primeur met Hololens op nieuwe N18. Retrieved from https://www.volkerwessels.com/nl/nieuws/detail/primeur-met-hololens-op-nieuwe-n18

Wasson, C. (2005). *System analysis, design, and development concepts, principles, and practices.* . Hoboken, N.J.: : Wiley-Interscience.

Waze. (2016). Outsmarting traffic, together. Retrieved from www.waze.com

8 Appendices

Appendix I	System of Road Management
Appendix II	Current and promising sensors in asphalt pavements
Appendix III	Experiment base mode, deterioration, coding and scoring
Appendix IV	Game set
	Reference