Effect of environmental conditions on the longevity of thin noise-reducing asphalt top layers; A case study in the province of Gelderland

[Effect van omgevingsfactoren op de levensduur van dunne geluidsreducerende asfaltdeklagen; Een case study in Provincie Gelderland]

Master Thesis end report, 07-11-2018,
Laboratory of Geo-information Science and Remote Sensing,
Wageningen University & Research

Personal information

Name	ing. TTN (Thijs) van Loon
Student number	940512528020
Email address	thijs.vanloon@wur.nl
Master	Geo-information Science
Study advisor	ing. WT (Willy) ten Haaf

Graduation committee

First supervisor	ir. HJ (John) Stuiver	Wageningen University & Research
Second supervisor	dr.ir. A (Arend) Ligtenberg	Wageningen University & Research
First supervisor	ir. H (Hans) Kooreman	Provincie Gelderland
Second supervisor	REJ (Remco) Hermsen	Provincie Gelderland

Figure 1.1; Ramboll Laser-RST measurement vehicle (Source; Ramboll.se, edited)

Colophon

Document information

Title	Effect of environmental conditions on the longevity of thin noise-reducing asphalt top layers
Subtitle	A case study in the province of Gelderland
Shortened title	Effect of environment on noise-reducing asphalt top layers
Pagecount	63
License	Open Access, CC BY-SA 4.0
Publication date	07-11-2018
Reference number	GIRS-2018-54

Author information

Name	ing. TTN (Thijs) van Loon
Student number	940512528020
Email address	thijs.vanloon@wur.nl
Master	Geo-information Science
Study advisor	ing. WT (Willy) ten Haaf

Graduation committee

First supervisor	ir. HJ (John) Stuiver	Wageningen University & Research
Second supervisor	dr.ir. A (Arend) Ligtenberg	Wageningen University & Research
First supervisor	ir. H (Hans) Kooreman	Provincie Gelderland
Second supervisor	REJ (Remco) Hermsen	Provincie Gelderland

Keywords

pavement ageing, remaining lifetime, asphalt longevity, noise-reducing asphalt, DGD, ZOAB, environmental factors, ravelling, texture measurements, ISO13473, ISO13565-2, MPD, core roughness depth, decisegment approach, road inspections, visual inspections, CROW

Preface

Almost a year ago, I was in the stage of finding a subject for my master's thesis. I was starting to be active in a community of voluntary map makers where I introduced my background in civil engineering, road inspections and GIS. One of the members got me into contact with the Province of Gelderland. Several meetings later, the proposal of this research was a fact. I never could have asked for a subject more fitting with my background.

The research presented in this report was conducted at the departments 'information and automation' *[I&A]* 'Management and maintenance of public roads' *[BOW]* at the Province of Gelderland. This study is targeted at helping future decisions made by *BOW*, and to show which further research opportunities are present with the vast amount of public and private data available at the province.

Acknowledgements

I want to acknowledge various people for their contribution to this project. I very much appreciated the help of my supervisors at the university of Wageningen, John Stuiver and Arend Ligtenberg. Especially for the help in any geo-related challenge and the guidance in writing this report.

I especially want to thank my supervisors, Hans Kooreman and Remco Hermsen, and my colleagues at the province of Gelderland for their engagement and pro-active thinking. I gracefully acknowledge the funding received from the province of Gelderland, which made it possible to execute this research more thoroughly. Because of the space and time that the Province of Gelderland was able to provide, this research subject has become a passion for me.

I should not forget to thank for the help and open-mindedness of the interviewees. Their expertise in different specific parts of this research is both impressive and inspiring. This research wouldn't be able to reach the current level of success without their knowledge. Only by sharing knowledge and working together, we can elevate the science of road maintenance.

Finally, I would like to express my gratitude to my close relatives for their support and patience during the entire thesis period.

Summary

The department of ‘Management and maintenance of public roads’ *[BOW]* of the province of Gelderland is entrusted with the task of ensuring road safety. To spend their budget well and to prevent nuisance, it is important to plan maintenance thoughtfully. This study focusses on predicting the end-of-life of *DGD* pavement on provincial roads. This porous type of asphalt covers a quarter of the provincial acreage and is very susceptible to ‘ravelling’; the loss of stones in the surface of asphalt pavements.

Ravelling is currently assessed by visual inspections. These estimations are less suitable for statistical analysis due to their subjectivity and bad repeatability. The province of Gelderland puts a lot of effort into collecting and archiving data. Since several years, special laser measurement vehicles are deployed. Which record 2D-profiles of the asphalt texture with millimeter precision at traffic speeds. In this research a correlation is made between visual inspection results and texture measurements.

This research ultimately tries to exploit these valuable datasets with the goal to predict the end of the civil lifetime due to ravelling. This is done by associating the growth of ravelling with environmental conditions where the asphalt is subjected to.

Deriving ravelling from texture measurements

The goal of the first research question is to come up with a model to measure ravelling by texture measurements. This will be done via ‘the decisegment approach’; Visual inspections state the severity and extent of ravelling, and the most severe patch of ravelling is normative. The inspected area, a road segment, is divided into ten deci-segments. The roughness of the most severe decisegment can then be correlated with the inspected severity class. The extent class is given by the number of damaged decisegments.

Roughness can be derived from 2D texture measurements in multiple ways. 23 options have been assessed. The ‘Core roughness depth (Rk)’ shows the best ability to distinguish the severity of ravelling. The model was able to find 26% of the segments that reached end-of-life.

It is unknown whether this low accuracy is caused by the visual inspections or texture measurements. Since the patches with the highest roughness are correlated with the severity class the roughness thresholds are relatively high. Depending on the accuracy needed for the final use of this model it is advised to alter the probability thresholds. It is recommended to circumvent the decisegment approach by using more accurate data from detailed visual inspections.

Correlating environmental conditions

The second research question focusses on 1) gathering possible environmental factors, 2) expressing these factors in geospatial data, and 3) correlate this data to ravelling.

First, a list of environmental factors which are expected to have influence is set up. The international literature on the ageing of porous pavements is limited and mostly aimed at highways. Interviews with professionals have been conducted. They noted location-dependent sources of damage such as agricultural traffic, tannic acids from leaf litter.

Secondly, geospatial information was gathered to represent these environmental factors. Ultimately, the following predictors are used; Age, tree cover, days of frost, hours of rain, heavy vehicles per day, light vehicles per day and the presence of levering forces.

Hereafter these environmental factors were correlated to both the visual inspections and the texture measurements. End-of-life according to texture measurements could not directly be predicted. The roughness per decisegment is predicted. Which is converted to an end-of-life diagnostic by means of ‘the decisegment approach’. The texture-based model performed slightly worse than the inspection-based model, but both showed significant results.

Environmental scenarios giving rules of thumb

The influences of environmental factors are quantifiable by comparing different scenarios. The base scenario has no overhanging trees, no nearby crossings, an average amount of traffic, and a moderate climate. Nine other scenarios were set up where these factors are altered. The differences in the growth of ravelling are noticeable and followed expectations. The test error of the model is high relative to these differences. Rules of thumb have been set up, such as “When a segment of 5 years old is below a tree, it shows the same amount of ravelling as a segment which is 6 years old.”.

Use of the outcome

Recommendations towards the province of Gelderland are made. Currently, warranties demanded for road longevity are expressed in CROW inspection classes. This research shows that the use of texture measurements to quantify ravelling has the potential to be more accurate than the subjectivity of inspections. Several recommendations on improving this quantification have been noted.

It is advised to use the environmental prediction model of this research to generate shortlists. This shortlist could then be inspected more regularly, in turn preventing emergency repairs and providing time to apply rejuvenation cures or search for integrative approaches.

Contents

Abbreviations and definitions

Abbreviations

AoI	Area of Interest, defined as through-going roads with *DGD* pavement inside the maintenance area of the province of Gelderland
AUC	Area under the ROC curve, measure for model performance
BGT	Topographical key register *[Basisregistratie grootschalige topografie]*
CROW	Dutch institute for collaboration and knowledge sharing between professionals in the sector of maintenance and management of the public space. They prescribe guidelines and recommendations for local authorities. Local authorities regularly cite these guidelines as requirements in contracts.
DGD	Thin noise-reducing asphalt top layer *[Dunne geluidsreducerende deklaag]*
EoL	End of lifetime *[einde levensduur]*
ETD	Estimated profile depth
ISO	International Organization for Standardization
ML	Machine Learning
MPD	Mean profile depth
MTD	Mean texture depth
NEN	Dutch standardization institute *[Nederlands Normalisatie-instituut]*
PA	Porous asphalt
RF	RandomForest, an example of a machine learning model
RMSE	Root-mean-square error
ROC	Receiver Operating Characteristic, graph of sensitivity vs. specificity
SMA-8g+	Stone mastic asphalt, acoustically optimized, a.k.a. 'Gelders mixture' *[Steenmastiekasfalt, akoestisch geoptimaliseerd, 'Gelders mengsel']*
ZOAB	Very open asphalt concrete *[Zeer Open Asfalt Beton]*

Road element definitions

Road	A road defined by its road prefix and number, e.g. N781 *[weg]*
Road trajectory	A part of road with a corresponding property *[wegtraject]*
Road segment	Segment of road between two hectometer posts, typically 100m long *[wegvak]*
Road decisegment	A tenth of a road segment, typically 10m long *[deci-wegvak]*
Road subsegment	Any lengthwise division of a road segment, for example a road decisegment
Baseline	A 100mm±10mm stretch of road, as defined and used in ISO 13473-1

1 Introduction

The department of ‘Management and maintenance of public roads’ *[BOW]* of the province of Gelderland is entrusted with the task of ensuring road safety while providing a smooth flow of traffic. Every year this department spends around 20 million euros in the management and maintenance of its provincial roads (Provincie Gelderland, 2017a, 2018). This is 1/8th of the total infrastructure-related expenses of the province (CBS, 2017b). To spend this money well it is important to plan maintenance. Planned maintenance is less expensive and more durable than emergency repairs and decreases the amount of disturbance for road users. To decrease maintenance costs in the long run as well, it is crucial to make the right choices for road construction.

In the past years the province of Gelderland has put a lot of effort into collecting and archiving information about road quality. Over time, this opens possibilities to analyze such datasets. This, in turn, can help the department of *BOW* in making choices in road construction and maintenance. This study will focus on predicting the end-of-life (EoL) moment of *DGD* pavement as caused by the phenomenon ‘ravelling’. This type of pavement covers about 25% of the provincial acreage and is very susceptible to ravelling.

The amount of ravelling is now visually assessed, which is subjective and therefore less suitable for statistical analysis. Experience on coarse *ZOAB* pavements has shown that high-resolution texture measurements are a good substitute. These methods may be applicable to the finer *DGD* pavements as well.

1.1 Research setup

To reach the goal of predicting the EoL moment of *DGD* pavement, the following research setup is followed. The research setup is more thoroughly described in the research proposal, which can be found in the digital supplements of this document.

1.1.1 Problem definition

To decrease maintenance costs, it is crucial to improving planning. For this, a better understanding of the lifetime of asphalt will be needed. Using the present knowledge, it is hard to predict the remaining lifetime *[restlevensduur]* due to ravelling in case of the fine-graded noise-reducing asphalt mixtures. However, it is known that ravelling is caused by the degradation of the bitumen adhesive. And that this degradation is highly depending on both the construction method and the local environmental conditions (Hagos, 2008). Which of these environmental factors has influence in the specific case of noise-reducing asphalt in Gelderland is yet to be explored.

Having more insight into the source of ravelling has multiple benefits. First of all, if a prediction on the remaining lifetime could be done with sufficient certainty it will reduce overall costs. This will lengthen the usable period of a pavement by preventing premature maintenance. If ravelling can be detected beforehand, control measures can be adopted. Which also opens up possibilities to integrative approaches *[integrale aanpak]* with other projects. Secondly, designers can assess the applicability of noise-reducing pavement as a measure against noise nuisance in a specific local environment. And thirdly further research to asphalt mixture optimizations can be targeted better if the importance of different environmental factors is known.

1.1.2 Research objectives

This research will try to gain insight in the longevity of *DGD* pavement as they are applied on provincial roads in the Netherlands, such that the occurrence of one or several environmental factors can be expressed in a decrease in lifetime. Which should make the End-of-Lifetime moment predictable. Such prediction should be specific enough to be incorporated into maintenance policies. To accomplish this objective, this research will try to correlate the decrease in lifetime with local environmental factors.

Main question

"Is a certain texture property of thin noise-reducing asphalt top layers indicative for the end of the civil lifetime due to ravelling, and is the influence of local environmental factors on its longevity measurable?"

Question 1; Which of the available methods of deriving ravelling from texture measurements is most suitable to classify the severity and extent of ravelling in the case of DGD asphalt?

a) What is the definition of ravelling and how does it arise in the case of *DGD* asphalt?
b) Which of the known derivative methods is most suitable to correlate a 2D texture measurement with visual inspections regarding ravelling at the CROW end-of-lifetime threshold?
c) What is the performance of the method to predict visual inspection outcomes from the most suitable texture derivative?

Question 2; How well is the influence of local environmental factors on the longevity of DGD asphalt measurable?

a) Which local environmental factors may be of influence on the longevity of *DGD* asphalt, and how?
b) What accuracy can be achieved by means of a predictive model using local environmental factors?
c) How much influence do local environmental factors have on the longevity of *DGD* asphalt?
d) Does the use of texture measurements above visual inspections improve the accuracy of the predicted moment of end-of-lifetime?

1.1.3 Research scope

- Most types of noise-reducing asphalt are porous and are therefore more susceptive to ravelling. This research will focus on the thin noise-reducing asphalt top layer *[Dunne geluidsreducerende deklaag, DGD]*. Which covers about 25% of the provincial acreage.

- As the main cause of a decrease in the lifetime of *DGD* is caused by ravelling, the causes of ravelling are explored. The end-of-lifetime point will be based on the current maintenance policy on ravelling of 'E1 and/or M3 or higher' (CROW, 2011c), as this is seen as a normative threshold for ravelling.

- As this research focusses on the local environmental factors during the maintenance period of the pavement, other factors are out of scope. As will be described in the literature study, the circumstances during the application of the asphalt may be of influence on longevity as well.

- The texture data is made available for three transects of the rightest lane of a road. Due to time restrictions, an optimization in choosing the best transect or transect combination will not be executed. The most right transect will be used. It is expected that the most right transect will show ravelling earlier, due to superelevation *[verkanting]* whereby the effective weight of trucks is higher and more rainwater passes by, more tree cover, and a less even temperature during the application of the asphalt.

1.2 Reading notice

- *[Dutch jargon]*; For clarity for Dutch readers, non-translated Dutch jargon is formatted in italic and/or placed between square brackets.
- (Model numbers); The conclusions in this document are primarily based on the outcomes of calculation models. The full scripts and data used by these models are to be found in the digital supplements, where they are numbered according to the step in the methodology. Where possible these models are referenced in-text by their model number.

2 Literature study

The Netherlands has one of the densest road networks in the world. Together with a high traffic intensity, this can lead to noise nuisance to nearby living residents. The province of Gelderland has set up an action plan against noise nuisance *[actieplan geluid]* as part of the Noise Abatement Act *[Wet Geluidhinder]* and the third environmental plan of Gelderland *[derde Gelders Milieuplan]* (2008). They have declared to tackle noise hotspots near provincial roads in an accelerated manner. Since 1990, the province has successfully applied noise-reducing asphalt to reduce nuisance for nearby living residents (Provincie Gelderland, 2008). And in the above-stated plan, they determined to construct 300km of noise-reducing asphalt between 2008 and 2013.

2.1 Types of noise-reducing asphalt

Of the twelve provinces in the Netherlands, Gelderland has the biggest provincial road network with 1350km in length (CBS, 2017a). At this point40% of the provincial roads in Gelderland is constructed using noise-reducing asphalt. Whereof 60% or 290km is constructed with a so-called *DGD*[1]. *DGD* or *DGAD* stands for '*Dunne Geluidsreducerende Asfalt Deklaag*', a thin noise-reducing top layer. This type of top layer reduces noise by a specific texture whereby tires generate less noise. And with a porosity of 12 to 25%, these pores help to absorb the sound generated (Werkgroep Stille wegdekken, 2010). In more high-strain situations *ZOAB [Zeer Open Asfalt Beton]* is applied. Which has a higher noise reduction and a higher porosity (>20%) and in some cases an extra underlayer. Increasingly the asphalt mixture 'SMA-NL 8G+' is used, which is developed by Provincie Gelderland. It has a smaller noise-reducing effect while it is more durable relative to *DGD* (Kersten, Bobbink, & Reinink, 2014).

Porous asphalt is popular in Europe, Japan, New Zealand and North-America and. However, there are both large and subtle differences in the mixtures. The international literature is therefore mostly inapplicable to the Dutch asphalt mixtures. Because noise-reducing asphalt is a relatively new technology, most research so far is done by commercial parties.

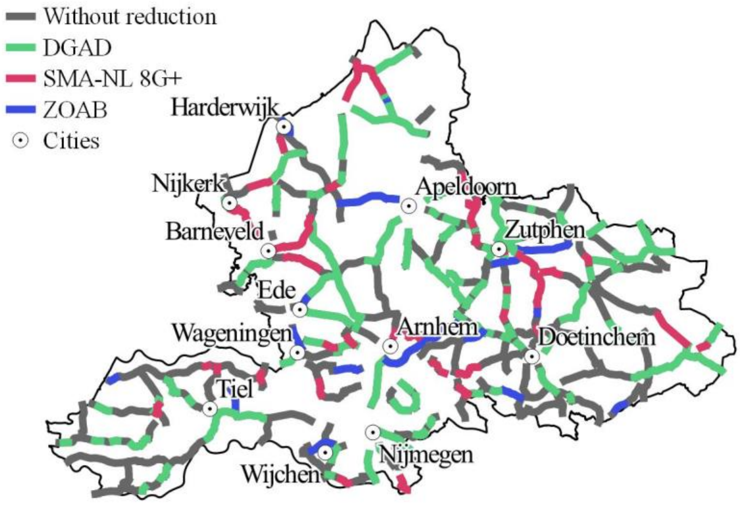


Figure 2.1; Noise-reducing asphalt in the province of Gelderland
(Data sources; Pavement type; Public GeoServer province of Gelderland, Cities; Open Street Maps)

---

[1] Interpretation of the Geoserver of Provincie Gelderland (2017b), retrieved on 9-10-2017
Total length of roads; Layer 'Intensiteit', sum of length_m; 1151km
Noise-reducing asphalt; Layer 'Geluidreductie verharding', sum of length_shape; 457km
DGD; Layer 'Geluidreductie verharding', where srt_toplaag='DGAD', sum of length_shape; 290km

2.2 Lifetime and Ravelling

A disadvantage of asphalt mixtures with high porosity is the shorter lifetime. The CROW institute (Werkgroep Stille wegdekken, 2010) describes the difference between the acoustic lifetime and the civil lifetime of an asphalt pavement. Where reference is made to lifetime, the civil lifetime is meant. The average civil lifetime of a noise-reducing pavement is 8 to 10 years. However, experience has shown that the actual lifetime has a lot of variation[2]. This makes the prediction of a lifetime more difficult. Experience has also shown that the phenomenon 'ravelling' is the main cause of the end of the lifetime for porous asphalt mixtures (Kersten et al., 2014). 90% of the maintenance to the Dutch national highways is executed because of an excess of ravelling (van Reisen, Erkens, v.d. Ven, Voskuilen, & Hofman, 2008). Porous pavements on provincial roads show similar results[3]. Other phenomena like cracking *[scheurvorming]*, rutting *[spoorvorming]* or polishing *[polijsting]* are less prevalent or may form in a later state.

Asphalt mixtures consist of stone aggregate which is bonded by bitumen. Ravelling is defined as the loss of aggregate in the surface of the pavement (Zhang & Leng, 2017). This occurs due to the ageing of the bitumen binder (Mo, Huurman, Wu, & Molenaar, 2009). Aged bitumen is less flexible and has a lower retention force whereby small stone aggregates can disappear out of the road surface. Hagos (2008) has done extensive research to how this ageing of bitumen develops chemically. He concluded that the degradation of bitumen when in conjunction with traffic load and freeze-thaw cycle forces greatly raises the chance of damage. Bochove *(2014)* states especially lower-order roads such as provincial roads undergo more wrenching forces *[wringing]* due to the high amount of exits, crossings, roundabouts, adjacent driveways and subsequent breaking forces.

The method of production and application of the asphalt mixture is also of great importance. An inaccurate production can cause bitumen to age faster, for example due to overheating. A bad application can cause the asphalt matrix *[korrelskelet]* to be incomplete which causes higher internal forces during a loading. For this reason, Provincie Gelderland states additional requirements for contractors. For example the use of a *shuttle-buggy*, which prevents de-mixing of aggregate and bitumen and equalizes the temperature of the mixture[2]. The process of installing asphalt in the case of provincial roads is less continuous than it is for highways and often needs manual modifications in bends and corners. Therefore application quality is harder to guarantee (Bochove, 2014).

A small increase in the amount of ravelling causes an increase in noise production and rolling resistance while decreasing the resistance against skidding (Nagelhout, Wennink, & Gerritsen, 2004). This poses a risk to road users. As a result of ravelling without prompt maintenance, potholes can occur fast. Especially in combination with high traffic loading and low temperatures (Opara, Skakuj, & Stöckner, 2016). If left untreated, the top layer can disintegrate where after the underlayer is exposed to wear. As replacing an underlayer is more expensive due to its thickness, maintenance of the top layer is a form of capital preservation[4].

[2] Correspondence with REJ Hermsen, Provincie Gelderland, 13-10-2017
[3] Correspondence with REJ Hermsen, Provincie Gelderland, 13-9-2018
[4] Correspondence with REJ Hermsen, Provincie Gelderland, 06-02-2018

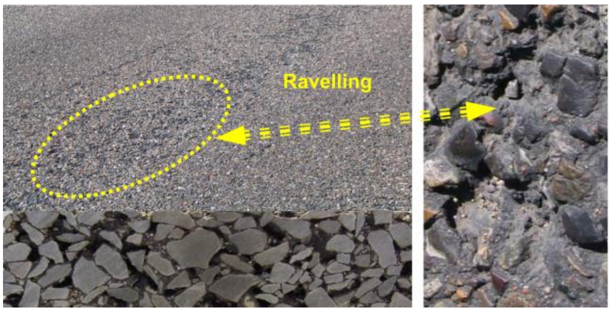


Figure 2.2 (left); Loss of stones (ravelling) of PA surfacing layer (Source; Hagos (2008))
Figure 2.3 (right); Sample of DGD asphalt, no ravelling occurred. Note the fine gradation and high porosity whereby air can flow through.

2.3 Current assessment of ravelling

At this moment the risk of ravelling is contained by the regular execution of visual inspections. By assessing pictures of the road, the amount of ravelling is classified. The exact method for such assessment and the subsequent maintenance actions are described by the CROW in their publication 'manual for visual inspections' *[Handboek Visuele Inspectie]* (CROW, 2011a).

In this publication, classes are described to be used with the visual inspections to the amount of ravelling. The amount of ravelling is expressed in 1) the severity of ravelling as a percentage of a normative square meter of road (L/M/E), and 2) the extent of the ravelling as a percentage of the area of the 100m stretch of road (1/2/3). Depending on the policy and budget of the authority a certain class can be deemed as intervention limit. Provincie Gelderland for example states demands to ravelling in *DGD* asphalt as a warranty from contractors. Ravelling of class 'E1 and/or M3 or higher' may not occur within 5 years after completion (Kersten et al., 2014)[3]. Ravelling is considered as of category E1 if a normative square meter shows more than 20% loss of stone aggregate and this normative square meter occurs on 5-30% on the 100m stretch of road.

Although visual inspections are exclusively executed according to the standard and only by trained and experienced personnel, the results are an estimate done by a human and therefore subjective (Tsai & Wang, 2015). The Dutch research firm KOAC-WMD (Nagelhout et al., 2004) has investigated the spread of the results of the severity and extent of ravelling of a visual assessment on non-porous asphalt. Three surveyors have been inspecting multiple samples of 1m². If the amount of ravelling for a normative square meter was around 15 to 60%, the spread in the results was significant.

This subjectivity can lead to disputes between the client and contractor. Especially in the case of 'Design, Build, Finance & Maintain' contracts[5]. In such casec the contractor is accounted as financially responsible for both the construction and the maintenance for multiple years. There can be disagreements about warranty and road surface quality. Besides this, these estimations are less suitable for statistical analysis due to the spread in the results and the repeatability.

[5] Graduation Thesis J. Boersma in collaboration with Heijmans, 'Predicting the degradation of surface pavement due to ravelling', 2017

2.4 Texture measurements

The photos used for the visual inspection which are described above are taken from a specially designed vehicle. Since several years such camera vehicles are also equipped with laser-based texture measurement devices to measure the texture of the asphalt surface. These vehicles are deployed on the provincial roads of Gelderland on a yearly basis since 2007[6].

One or several lasers measure the texture of the road surface below the vehicle. Most often this is done in both wheel tracks and in the middle of the lane. In the most modern measuring vehicles this is measured at a maximum sampling frequency of 32kHz[7] up to 64kHz[8]. The actual sampling frequency is depending on the travelling speed of the vehicle. This is adapted to get a height measurement every 1mm in travelled distance[7,8]. With 32kHz, the measuring vehicle would still be able to drive 80km/h. The vertical resolution is 0.1mm or higher[9]. This gives a very detailed 2D height transect.

Out of this 2D road profile several road properties can be gathered or estimated. For example noise production, ravelling, rolling resistance, driving comfort, grip and water drainage capacity (Ueckermann, Wang, Oeser, & Steinauer, 2015; Werkgroep Stille wegdekken, 2010; Woodward, Millar, & McQuaid, 2014). The reliability of these derived properties is depending on the type of pavement, the gradation of stone aggregate and the porosity. Generally speaking the derivatives of asphalt mixes with a fine or porous gradation are harder to determine (Nagelhout et al., 2004).

Recent developments

The laser-altimetry on a road-level to be used for this research is a relatively new technology. Although this technology is used since the eighties, for a long time there was no possibility to save the data generated because of a lack of storage capacity (Arnberg et al., 1991). Next to being able to save the measured data and the availability of open data the possibilities to process these datasets have also been increasing. An increase in computing power made it possible to execute advanced statistical methods on big datasets.

In the past several types of research have been executed to correlate texture measurements with ravelling in *ZOAB* (Nagelhout et al., 2004; van den Bol-de Jong, Bouman, van Ooijen, & Verra, 2003). In the case of the thin noise-reducing asphalt pavements of provincial roads, the grains of asphalt is much finer. Whereby it is needed to have a finer texture measurement to distinguish ravelling. During the preliminary study, no results have been found in measuring ravelling using similar laser texture measurement technology with such fine-grained asphalt types. Although the texture data suggests stones and stone loss should be distinguishable.

[6] Correspondence REJ Hermsen, Provincie Gelderland, 2-10-2017
[7] Ramboll measuring vehicle, out of presentation ‘Estimation of stone-loss on network condition surveys by use of multiple texture lasers’, T. Wahlman, P. Ekdahl
[8] ARAN measuring vehicle, http://www.roadware.com/related/Smart-Texture_2014.pdf
[9] NEN-ISO 13473-3, table 6 (NEN, 2002)

3 Methodology

The stated main question is two-fold. The research questions follow this division. Although it cannot be said that these questions are unassociated. The second research question is highly depending on the results of the first question; It uses the dataset generated including age, ravelling condition and locations of *DGD* pavement. And most importantly, the model to convert texture measurements to a ravelling condition. The section numbers and names will follow the section number and names in the result chapters.

3.1 Method for research question 1; Texture to ravelling class model

The goal of the first research question is to come up with a model to convert texture measurements to a ravelling condition. This condition will be expressed in ravelling classes as defined by the CROW standard. First, data is gathered and prepared. Then several methods of deriving a value for roughness from texture measurements will be applied. These values are then correlated with ravelling inspection results following 'the decisegment approach'.

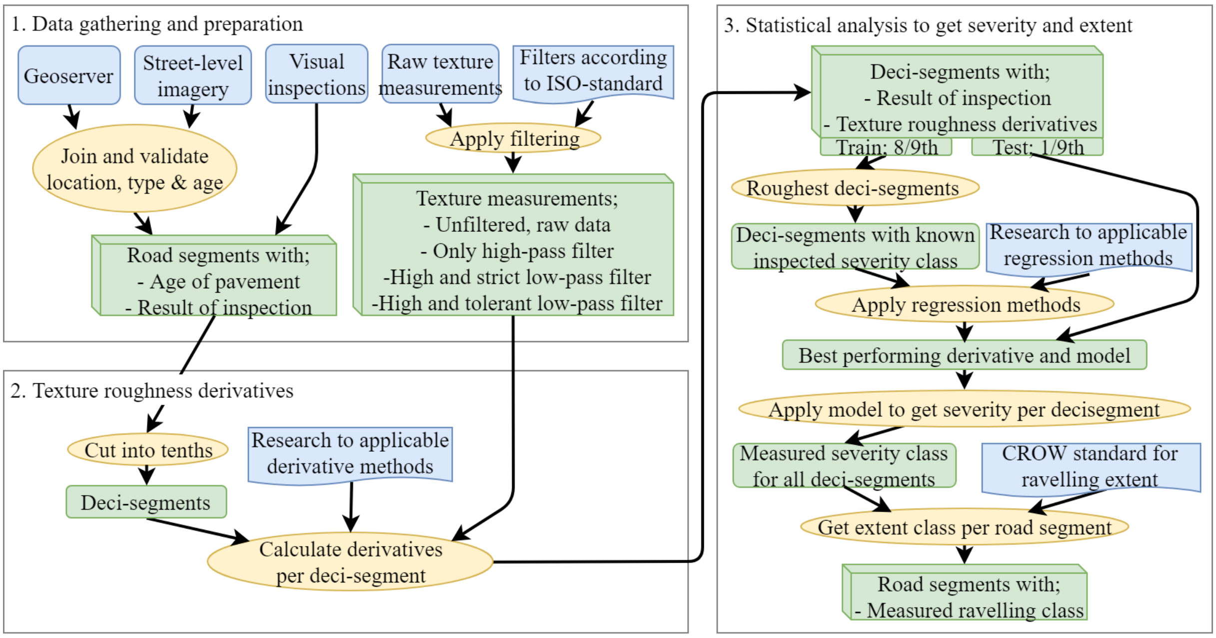


Figure 3.1; Flow chart of research question 1

3.1.1 Data gathering and preparation

The main interest of this part in research is to check for reliability and validity of the source data. As this can hugely affect the outcome of this research. Therefore, the value of every source will be discussed. For this first research question, three major data sources are used;

- ***DGD* pavement location and age;** Built up from polygon shapes with spatially joined attributes of pavement type and age. These locations of *DGD* pavement are bound to change over time due to maintenance or replacement and will therefore be checked thoroughly with street-level imagery.
- **The results of visual inspections;** At this stage, this dataset is temporarily seen as ground-truth and will therefore not be altered. They will be spatially joined to the dataset described above. It should be noted that each side of the road is inspected every other year.
- **The texture measurements;** Validation of this dataset will be focusing on the measurement standard used and the georeferencing done as a preparation step. Another important data preparation step is filtering the raw signal. The ISO 13473-1 standard is set up for this specific purpose and will be used. Multiple filtering options are presented in this standard. As their effect on further results is unknown by current literature, all four options are considered.

3.1.2 Texture roughness derivatives

Literature study will give several derivative methods to get a value for roughness from a texture measurement. A simple measurement for roughness would be the standard deviation of the height. 2D-measurements regarding roughness are used in multiple professions. Such as engine optimization, plastic finishing, microscopy, archaeology, etcetera. The methods found will shortly be described and their effect hypothesized.

The derivative methods that are suitable will be translated into code. Some combinations between derivative methods and signal filters are illegitimate according to the ISO standards.

3.1.3 Statistical analysis to get severity and extent

An important concept of the method used in this research is 'the decisegment approach'. By the current CROW standard, a 100m[1] long road segment is deemed to be end-of-life if the damage is of a certain severity and a certain extent. The most severe patch of ravelling is normative for the severity class. Hereafter the extent of the chosen severity class is defined. Where exactly this severe patch is located inside the road segment is not noted down. To replicate this using texture measurements, a method should be set up to get both values given a single ravelling derivative. In this research, the following approach is used;

The decisegment approach

1. First, every 100m road segment will be divided into 10m decisegments. See figure 3.2.
2. For every decisegment, the roughness will be calculated based on the texture measurements. The inspected severity class per decisegment is yet unknown. What is known though, is that the worst roughness value per inspected segment shows the value for the severity because the severity class was normative.
3. The roughness values and known severity classes can now be correlated by regression. Although the correlated ravelling severity is based on a single line measurement, it is assumed the ravelling occurring at such location is normative for the entire decisegment.
4. The extent class is yet to be found. The CROW states clear thresholds on the extent classes, see table 4.4. As the area per decisegment is known, these thresholds can be applied.
5. The severity class and the extent class are now known. Combining these gives the ravelling class. By applying a threshold, the road segments which are end-of-life can be found.

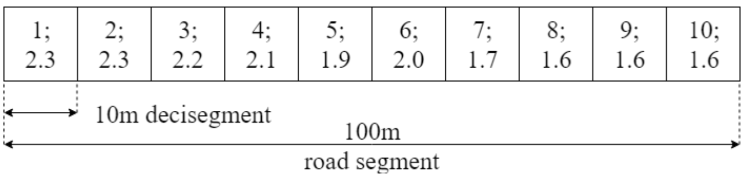


Figure 3.2; Numerical example of deriving severity and extent class from decisegments. Values show the maximum Rk values of N310 near Uddel

Regression technique

The regression technique applied in step 3 of the decisegment approach will be done with a basic machine learning algorithm. The algorithm should be able to receive a single predictor in the form of continuous values and predict a ravelling class. As this research focusses on the end-of-lifetime moment of a road segment, mainly the errors to class 'severe ravelling' should be as low as possible. As class 'moderate ravelling' can also result in EoL, it is impossible to simplify this step to a binary problem. Furthermore, this model should be as easy to interpret as possible to facilitate future implementations. Which model fits these criteria while giving optimal performances is part of the analysis.

Validation

The model should be validated to conclude which signal filter technique, derivative method, and regression technique suits best. As there is no shortage in observations the dataset will be divided in a train and test subset. As the goal is to predict the EoL moment, the testing dataset should cover enough data points that crossed this point. I.e. the diversity should be adequate. Of the ±7000 input data points, ±400 were above this EoL threshold. To have adequate diversity a minimum of 50 above-EoL datapoints is deemed to be satisfactory. Therefore a 400:50 or 8:1 ratio is considered at minimum. Rounding up to have a nice number to work with, 1000 data points are considered as test data and the rest as training data. The results will be presented in a confusion matrix displaying the classes L1-E3, which in turn will be aggregated to EoL in a binary confusion matrix.

3.2 Method for research question 2; Environmental factors

The second research question focusses on gathering possible environmental factors, expressing these factors in geospatial data, and correlate this data to ravelling. The step of correlating this data will be done for two data sources; the visual inspections and the texture measurements. This is done to be able to answer sub-question 2b.

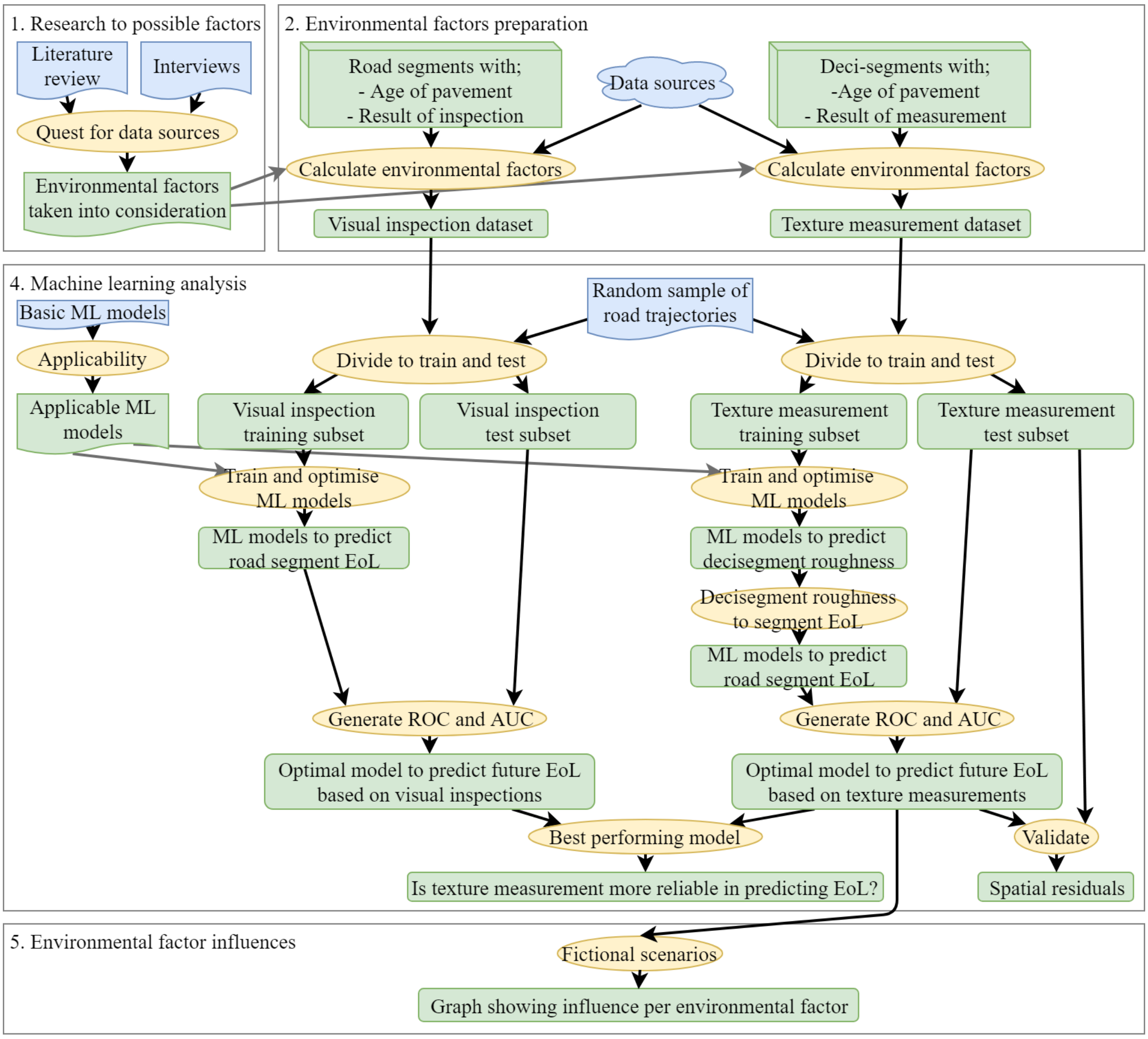


Figure 3.3; Summarized flow chart of question 2

3.2.1 Research into possible environmental factors

The goal of this step in research is to come up with a number of environmental factors that are expected to have an influence on road longevity. And for every factor, hypothesize the mechanism and the extent of the influence. Two resources will be used, literature review and interviews.

First, an in-depth literature review will be executed. This research will not be limited to Dutch pavements but will be limited to porous pavements. The research background and methods will briefly be described, where after the conclusions that are applicable to this research are noted.

Secondly, interviews with several field experts will be executed. The interviews will be executed in a semi-open matter to accommodate for the specific knowledge of the interviewees. Just before the interview, the research goals and research questions will be described. Possible factors found previously will not be communicated before the interview to prevent bias. A summary of the interviews will be added to the report, and full transcriptions will be added as supplements.

3.2.2 Preparation of possible environmental factors

As this will be a geospatial analysis, the availability of suitable geospatial data is key. For the environmental factors found in the previous step, it will be researched whether it is possible to find geospatial information. The source and preprocessing steps will be described. The final data quality and resolution will be evaluated. These environmental properties will be calculated both per road segment and per decisegment.

3.2.3 Possible environmental factors not taken into consideration

There will be possible factors where the availability of explaining data will be insufficient or unsuitable. In some cases the lack of physical examples may be problematic.

3.2.4 Machine learning analysis

The goal of this step in research is to find the environmental factors that cause a road to reach EoL. The age and environmental properties are taken as predictors. In the case of the visual inspections at road segment level, the model will directly predict EoL. In the case of texture measurements, the ML method will predict the roughness at decisegment level. The model set up by the first research question will then be used to convert the decisegment roughness values to a road segment ravelling class. Which in turn can be converted to the EoL diagnostic.

For understandability of the problem and interpretability of the outcome, generic machine learning methods will be applied. The predictors involved in this machine learning problem are both classified and continuous. Not all generic ML methods can cope with this. More advanced models such as survival analysis are more involved to implement are less easy to interpret, which is therefore out of the scope of this research.

The train- and test division should be made at a road trajectory level, as the initial study has shown that road segments show high local relations. The exact division is depending on the number of predictors and their distribution of values, aka ‘binning’. Applicable models will be applied and optimized.

Comparing model performances

As is known beforehand, only a few segments have reached EoL. This means the province of Gelderland is doing a good job at maintaining their roads. But this also means the input data for this research is highly unbalanced. 5.7% of the road segments have reached EoL according to the visual inspections, while 1.3% were deemed to have reached EoL according to the texture measurements.

During the preliminary research, the data was subjected to a logistic regression analysis. The model did not have enough confidence to classify any EoL segments. This was due to the unbalanced data, giving a very low prior probability to the minority class of EoL=True. It is therefore useless to compare the model performances based on classification matrices, as a confidence threshold of p=0.5 will never predict EoL=True.

An alternative measure for model performance which is not depending on a choice of confidence threshold is the ROC characteristic (Lobo, Jiménez-Valverde, & Real, 2008). Which can be defined as a plot of test sensitivity versus its 1-specificity or false positive rate (Park, Goo, & Jo, 2004). The sensitivity and specificity of a range of confidence thresholds will be calculated. This characteristic can only be used for binary outcomes, i.e. diagnostic tests.

The curve that appears is called the ROC-curve. The area below this curve (AUC) is a measure of how well the model performs. An AUC of 0.5 is nothing better than a random guess. An AUC of 1 would be a perfect model.

Spatial residuals

For the test subset, both the measured and the independently predicted roughness is known. The difference between these values is called 'error' or 'residual'. These residuals per decisegment can be viewed spatially. If there are any spatial trends, there could be a spatial predictor that has been unnoticed in this research. These residuals will be evaluated together with pavement professionals.

3.2.5 Environmental factor influences

As stated in literature, age is a very important predictor for ravelling. Using this as a base, the influence of other environmental factors could be assessed by means of several different scenarios. At the decisegment level, the roughness is predicted instead of the ravelling class. One of the advantages of this model above the model predicting visual inspections is the resolution of the output. The numerical output of roughness has a higher resolution than the inspection class. Ten scenarios will be set up. Scenario 0 would be a base scenario, where all environmental factors are average. In other scenarios, factors can be altered, and their influence can be made visible. Rules of thumb can be determined which are useful for maintenance engineers.

4 Results question 1; Texture to ravelling class model

In this chapter the analysis and the results of the first research question will be described. This chapter will follow and execute the methodology as described in section 3.1.

4.1 Data gathering and Preparation

At this stage, three data sources are used and described; *DGD* pavement location and age, the results of visual inspections and the properties of texture measurements.

4.1.1 DGD pavement location and age

Validation of age and location

At the end of 2017, employees of the province aggregated multiple project information sources to set up a general map with the pavement type and age of the provincial roads. This data source consisted of line segments with their surface pavement type and the year of execution.

This dataset is manually validated against street-level imagery in both time and location. This has been a tedious job. But it was a vital step as the line segments were often rounded to the nearest mile post, cluttering the source data. *DGD* pavement is visually very hard to differentiate from other fine-graded pavement types such as SMA8g+. But in most cases stretches of dense pavement are replaced by *DGD* pavement. Most transitions are therefore between *DGD* and non-porous asphalt, often very noticeable as can be seen below.

Figure 4.1; DGD to SMA 0/11 transition at the N844 near Malden (Source; CycloMedia Technology B.V.)

After this manual selection, a couple of segments were still doubtful. In those cases, an expert was consulted. These segments were judged by colleagues who knew the history of the road. If there still was any doubt, these segments were discarded.

Total pavement, length of axis	1150km
↳Total pavement length defined as being *DGD*	235km
↳ Manual validation; *DGD*	**220km**
↳ Manual validation; Non-*DGD*	8km
↳ Manual validation; Uncertain	7km
↳ Expert view; Discarded	1.5km
↳ Expert view; Uncertain	5.5km
↳ Expert view; Included	**0km**

Table 4.1; Road length in Area of Interest

Road axes to polygons

For the spatial analysis of question two, information about road segments will be needed as polygons instead of axis line segments. For example to calculate the number of overlapping trees. The province of Gelderland takes part in the 'BGT' program, which is a digital map with nationwide coverage. This map has a spatial accuracy of at least 20cm and is updated daily. An extension of this standard is IMGeo *[InformatieModel Geografie]*. This standard improves the exchangeability of data between public authorities. Part of this extension is the *'Objectenhandboek BGT | IMGeo'*. This manual defines which attributes a segment should get.

The province of Gelderland maintains a PostgreSQL database with all segments which are to be maintained by the province. These segments are updated daily. A selection has been made based on IMGeo-attributes with the goal to select through roads. It was not possible to select all segments needed without missing segments based on attributes. For example the situation in figure 4.2. The selection has been manually adapted. The selected segments are clipped to the area of interest, which is hereby defined by the line segments of *DGD* pavement inside the maintenance area of the province of Gelderland. This causes multiple problems. Segments which are shorter than 100m are formed for example. Segments smaller than 5m² are therefore dropped, and the length markings are recalculated (model f1e, see appendices).

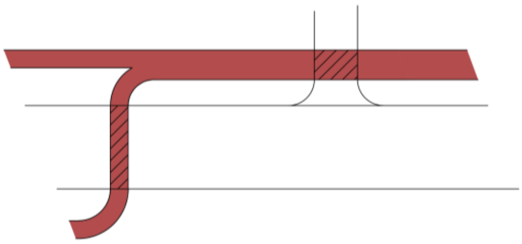


Figure 4.2 (left); Both hatched segments are a bicycle path on a road, but only the bottom one should be included
Figure 4.3 (right); Small road segments due to clip to Area of Interest

4.1.2 Visual inspection results

Each road segment maintained by the province is inspected every other year. To do this a street-level camera car takes several pictures of the pavement. These high-resolution images are later inspected on a computer[10]. These inspection results are made available for this research as line segments adjacent to the road axes. The inspection results saved in these line segments are spatially joined to the road segments (model f1d, see appendices).

The roads of the province of Gelderland are inspected by one roadside per year. This means that in 2016 the road has been inspected in the up-counting direction of mile posts called the 'right' side. In the uneven years, 2015 and 2017, the 'left' side is inspected.

4.1.3 Texture measurements

The visual inspection results will be compared to the asphalt texture measurements. These texture measurements were available to the Province. But already had undergone several calculations. The MPD, a derivative for roughness, was given per road segment. For the decisegment approach, a higher spatial resolution will be needed. And secondly, it was unknown whether the MPD is the best option to compare the texture measurement to ravelling. The unprepared, raw texture measurements were not available at the province of Gelderland. Instead, this data was requested by the company which executed these measurements, Ramboll. These texture measurements are planned such that they are executed in the same season and location as the visual inspections.

[10] After interview T. Jansen, summary can be found in paragraph 5.1.2. Transcript available in supplement 3

The LaserRST measuring device

High-speed road profilers such as Ramboll's LaserRST, Fugro's ARAN and KOAC-NPC's HSRP are measuring according to the same standard[11]. According to the CROW (2016) positioning of the laser on the vehicle can differ and may cause differences. However, it can be expected that this positioning has less influence than lateral differences[12] due to the steering of the driver. It is therefore expected that the following methodology is applicable to any of the high-speed road profilers. For this research, only the data of Ramboll's LaserRST is used.

Figure 4.4; Ramboll LaserRST high-speed road profiler. Lasers are seated in the bumper. (Source; Ramboll.com)

Data format and georeferencing

The data obtained by Ramboll was several gigabytes in size. About 1.5 billion height measurements in the area of interest were sent as .csv-files. Every file contains measurements of a left, middle and right transect over about 10km. The elapsed distance is measured by a pulse generator mounted on the hub of the vehicle's right front wheel and is therefore influenced by tire temperature and inflation pressure (Arnberg et al., 1991). These files can be georeferenced by the linear referencing tool in ArcGIS.

At this stage the longitudinal accuracy was hard to assess. In the measurements of 2017, the driven distance and the distance as stated at the mile posts differed with 1% per 10km on average. Half of the measurement trajectories differed less than .5% on 10km (model f3b.4). These differences can be explained by bends and roundabouts. But the linear reference system of mile posts is not exact either. The system inaccuracy of the linear referencing is farther assessed in paragraph 4.4.2; Accuracy of texture measurements.

Due to the unknown linear referencing accuracy described above and to prevent interpolation artefacts, the texture data is not stretched or squeezed to the needed length but used as-is. When a road segment was not measured or partially measured the measurement of that year at that segment is discarded.

Signal filtering

Before using the data, the raw texture measurements will go through a preprocessing step in the form of a signal filter, filtering out low and high frequencies. Like audio crossovers, a high-pass filter passes through high-frequency changes in amplitude. While dampening low-frequency changes in amplitude. In the case of these texture measurements, the effects of the suspension of the measurement vehicle have a low frequency and high amplitude relative to the passing stone granules of the asphalt mixture. The most noticeable effect of a high-pass filter is that the measurements become mean-centered. This can be seen in the first graph of figure 4.5.

---

[11] Mostly according to the requirements of a class D-laser defined by ISO 13473-3 (NEN, 2002)

[12] Lateral accuracy; distance relative to the center of the road, i.e. left to right

There is an existing standard for such signal preprocessing, the ISO 13473-1 chap. 7.3. This standard is followed in calculation f2.2. A 2nd order Butterworth filter with a cutoff frequency of 10 cycles/m[1] will be applied. The higher the order, the less gentle a cutoff is. A higher order would cause artefacts at frequencies close to the cutoff boundary and is computationally more intense. As the difference in frequency of the suspension and granules is great, the boundary is not very critical and therefore a low order is preferred.

The requirements of a low-pass filter are depending on its use. The goal of such filter is to suppress outliers and high-frequency noise. Which in turn changes the roughness derivative values, which may give a better correlation to ravelling. No researches were found addressing the effects of low-pass filters on texture measurements of fine-graded asphalt such as *DGD*. However, in the case of ravelling in thin noise-reducing pavements, the details of interest are the asphalt granules of 0-5mm[1], while the sampling interval is 1mm[1]. Therefore low-pass filters should be applied with caution. Hence two low-pass filters were tested; a filter with a more abrupt cutoff by using an 8th order function with a relatively low cutoff frequency (220 cycles/m[1]) and a more tolerant filter with a higher cutoff frequency (400 cycles/m[1]) and more roll-off due to a 1st order function. Both were within the specifications given in ISO 13473-1 chap. 7.4. Filters with higher cutoff frequencies are meaningless as such frequencies approach the sampling frequency of 1000 cycles/m[1].

For each derivative method, one or more of the following filtering combinations will be considered;

1. The unfiltered raw laser data
2. Only a high-pass filter
3. A high-pass filter with the strict low-pass filter
4. A high-pass filter with the more tolerant low-pass filter

Similar to moving windows, when applying a Butterworth filter the first and last few height measurements are sacrificial due to transient effects. To overcome this the selection is expanded with 10m[1], well above the largest wavelength unaffected by a high-pass filter[13]. If this data was not available, for example at the very beginning of trajectories, the available data is mirrored. After filtering this sacrificial buffer is erased.

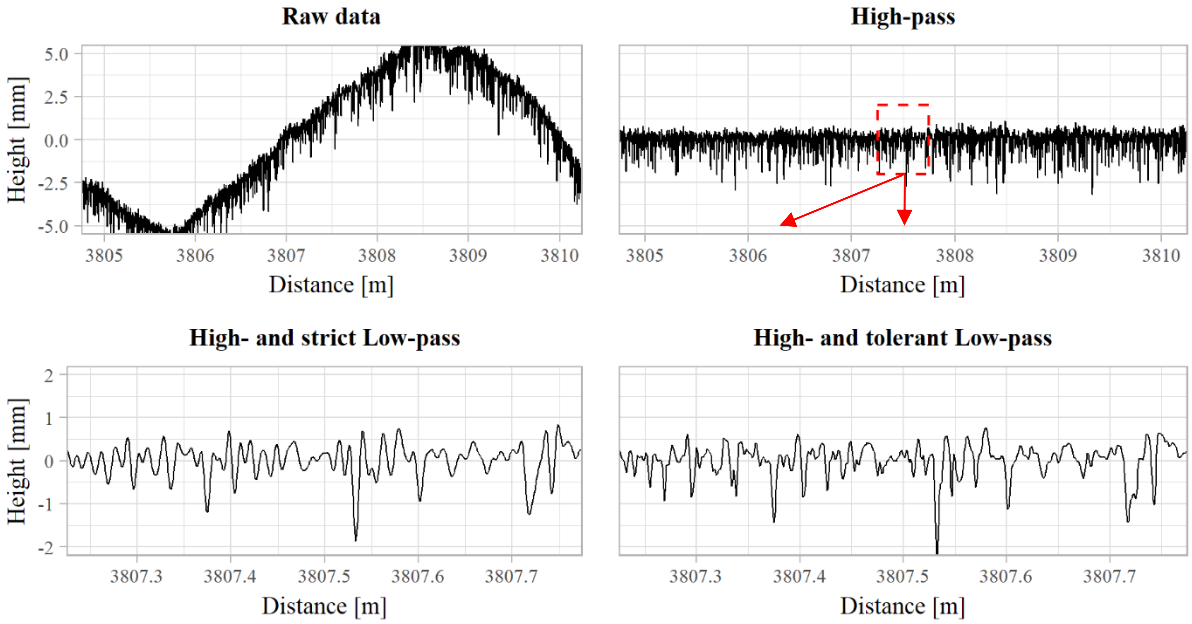


Figure 4.5; Frequency-based filtering (model g5e)

[13]At 3.16 cycles/m (wavelength of 0.32m), a 2nd order Butterworth high-pass filter with cutoff at 10 cycles/m would decrease the amplification of such frequency with 20dB, equal to a factor of 1/100th. Frequencies lower than this are attenuated even more and therefore have negligible effect.

4.2 Texture roughness derivative methods

The existing literature about texture measurements defines a multitude of roughness derivative methods. The theories behind such derivatives are often insensitive to scale and are therefore applicable to pavement roughness.

4.2.1 Baseline-based derivatives

The international standard ISO 13473-1 is specifically targeted to asphalt pavements and defines the MPD and the RMS derivative. These values are calculated per baseline, defined by a 100mm±10mm stretch of road. These values per 10cm-baseline are averaged per decisegment. This averaging also suppresses the effects of measurement outliers.

Because these baselines are so short, the effects of the car suspension and therefore the benefits of the high-pass filter may be negligible. The effects of the low-pass filter are unknown. For this reason, the baseline-based derivatives are executed with all four filtering combinations.

Mean Profile Depth

At this point, the Mean Profile Depth is the ravelling derivative most often used. Preliminary studies show decent relations between MPD and ravelling severity. The MPD standard is set up to replace the Mean Texture Depth (MTD), which is based on the 'sand patch method'. Whereby a known volume of sand is physically spread out over the asphalt and the area of such patch is measured. The MPD standard tries to mimic this field experiment by applying the following procedure.

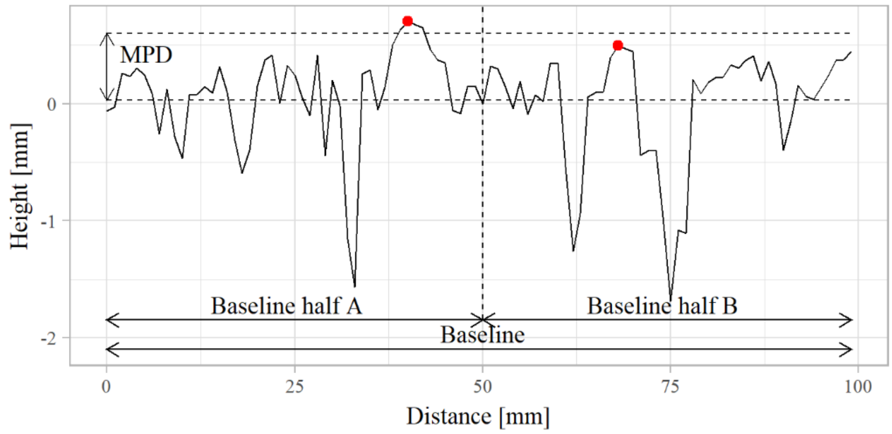


Figure 4.6; Explanation of calculation of MPD (model g5c)

To calculate the MPD the highest peaks (red dots) of the first and second baseline halves are selected. The height of these peaks is averaged, as indicated with the upper dashed line. The difference between the averaged peaks and the mean height is called the MPD (NEN, 2004).

The standard also suggests a conversion formula between MPD and MTD, existing of a scale and an offset. The statistical models used are insensitive to such transformations, and therefore the MTD derivative (also known as ETD, Estimated Texture Depth) is not taken into consideration.

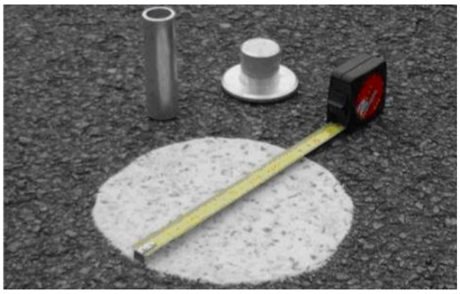


Figure 4.7; Sand patch method (Source; Prevost (2013) CC-BY 4.0)

Root Mean Squared

Several sources show the calculation of the RMS value per baseline (Izevbekhai, Watson, Clyne, & Wong, 2014; Rasmussen, Sohaney, Wiegand, & Harrington, 2011). This derivative effectively describes the deviation in height measurements. The RMS is calculated by taking the root of the mean of the squared height measurements.

4.2.2 Cumulative distribution derivatives

In the automotive industry, the study of piston cylinder roughness is done by standards defined in the ISO-13565-2 (NEN, 1998). Applying this standard to pavements was proposed by Rasmussen et al. (2011). The workflow described in this standard can be summarized as described below. This workflow is replicated in a formula defined in model f2.2, which may be useful for further researches.

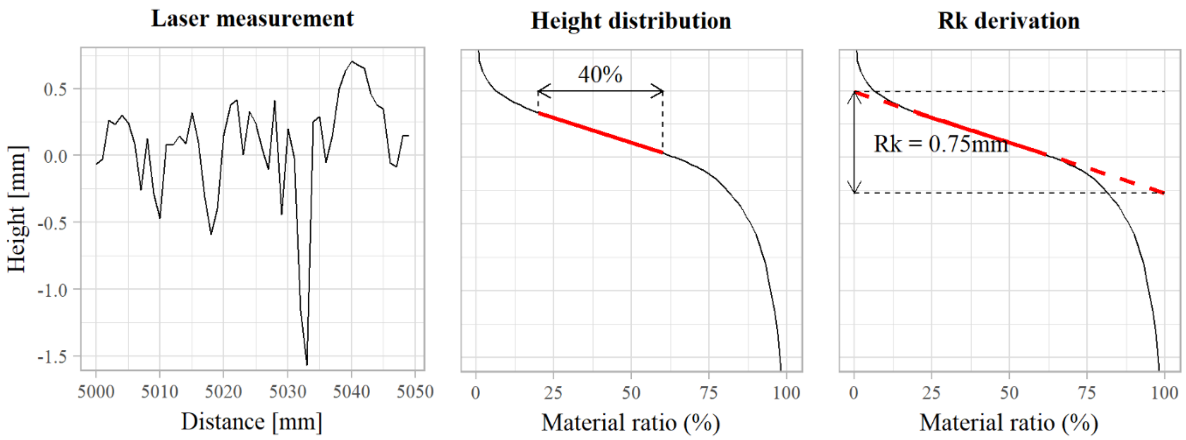


Figure 4.8; Explanation of calculation of Rk (model g5b)

In the first graph, the texture measurement after filtering is shown. A high-pass filter is applied as the measurement lengths considered in this calculation, the decisegment length of 10m, is long enough to see the effects of the car's suspension travel. This causes unwanted skewness to the cumulative height distribution function. Low-pass filters are not applied as described in ISO 13473-1, note 16 and 17.

In the second graph the cumulative height distribution is shown. Simply put, the amplitude measurements are ordered from high to low and scaled from 0 to 100%. For example, a material ratio of 25% corresponds with 0.3mm. Meaning that 25% of the measurements is higher than 0.3mm. The red line is called the secant and is defined as a line containing 40% of the height distribution while having the lowest gradient. Extrapolating the secant to 0 and 100% of the material ratio gives the 'equivalent straight line', which is the dashed red line in graph 3. The following values are derived from this line.

Rk; Core roughness depth

The height covered by the 'equivalent straight line' can be seen as the core of the pavement. The value Rk is hereafter defined as the height difference of this core. It thereby represents the roughness of the pavement, adaptively neglecting peaks and valleys.

Mr1 and Mr2; Material ratio at top and bottom of core

The material ratios at the intersection of the two black dashed lines and the cumulative distribution are the material ratios at the top (Mr1) and the bottom (Mr2) of the core. Mr1 therefore represents the percentage of peaks, and Mr2 defines the percentage of non-valleys.

Rpk and Rvk; Reduced peak height and valley depth

The areas above and below the core are defined as the peaks and valleys. Intuitively, an increase in ravelling would show as a decrease of material in the core and an increase of valleys. Therefore, it is expected that especially Rvk can be a useful derivative. These values are defined as the cross-sectional area of the profile peaks and valleys that protrude out of the core. These values are thereby very sensitive for outliers. To partly restrict this effect, the peak height is reduced by means of calculating the height of a right-angle triangle with an equal area and base width.

4.2.3 Normal distribution derivatives

Several standard height distribution values are evaluated as well. Per decisegment, the standard deviation, kurtosis and skewness are calculated. It is expected that an increase in ravelling would cause the standard deviation of the height to increase. And at light ravelling, the height distribution would tilt left, giving a negative skewness value. This effect may be overruled by low-pass filtering, therefore these derivatives are calculated with and without low-pass filtering. The effect of ravelling on kurtosis is unknown but is included anyway.

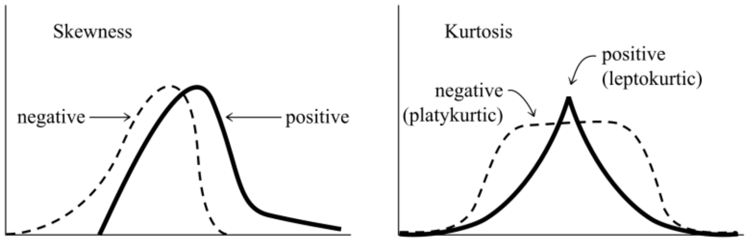


Figure 4.9; Definition of skewness and kurtosis (Source; Press, Teukolsky, Vetterling, and Flannery (1992))

4.2.4 Conventional derivative methods not taken into consideration

In the research to derivative methods, several methods have been found to be inapplicable for this research.

- The Stone(a)Way model (Ooijen & Bol, 2004; van den Bol-de Jong et al., 2003) was described in the research proposal as being an applicable model. However, this model was too hard to implement given the duration of this research. This model counts the locations where a circle with a given radius physically fits between the measured points. In other words, the detection of free space. This is a hard to code and computationally heavy task, and the source code was not available. The Coin algorithm set up by TNO is similar to Stone(a)way as it also detects free space in the model but is meant for 3d data. This model tries to find locations where a coin with a given radius and depth could be fitted[14].
- The German standard 'Integral der Differenzen' is an answer to the subjective visual inspections on ravelling. The change in texture before and after 700.000 standardized wheel loads are compared. The research of Huurman (2008) stated the outcomes of this method show a good correlation with ravelling. This standard is to be executed in a controlled environment to have a texture measurement before and after the wheel loadings and is therefore inapplicable for this research.
- The International Roughness Index (IRI) is a much-used standard to quantify ride quality. Wavelengths between 2.4 and 15m have the most influence on the IRI (CROW, 2017), such wavelength greatly overlooks the wavelength at which ravelling occurs (0-10mm). Therefore, the IRI is not taken into consideration.

[14] After interview M. Nagelhout, summary can be found in paragraph 5.1.2. Transcript available in supplement 3

4.2.5 Derivative similarities

Combining multiple derivative alternatives is out of the scope of this research. Several combinations of two derivative alternatives have been researched briefly by means of scatterplots but did not show possibilities for better classification.

The MPD values of all four filtering techniques are very similar. MPD is therefore insensitive to filtering. Rk and MPD are linearly correlated by MPD=0.07+0.63*Rk, having an RMSE of 0.07mm. For this analysis the same train and test set as defined in paragraph 3.1.3; 'Statistical analysis to get severity and extent' is used.

4.3 Statistical analysis to get severity and extent

All derivatives described above are calculated for every decisegment. This is done with multiple filtering options, as described in paragraph 4.1.3. For most derivatives, it is not known whether more ravelling shows as a higher or a lower value. The highest and the lowest derivative value per road segment are hereafter correlated to its severity class.

At this stage, a dataset is created with +/- 6500 road segments having attributes containing the inspected severity class (0/L/M/E) and 92 roughness derivatives. These are set up as follows;

Derivative method		**Raw data**	**High-pass**	**High- and strict low-pass**	**High- and tolerant low-pass**
Mean Profile Depth	MPD	x	x	x	x
Root Mean Squared	RMS	x	x	x	x
Core roughness depth	Rk		x		
Top core material ratio	Mr1		x		
Bottom core material ratio	Mr2		x		
Reduced peak height	Rpk		x		
Reduced valley depth	Rvk		x		
Standard deviation	Rsd	x	x	x	x
Skewness	Rsk		x	x	x
Kurtosis	Rku		x	x	x

Table 4.2; Combinations of derivatives and filtering techniques

For every combination of filter technique and derivative method, the maximum value, the minimum value, the mean of the maximum two values and the mean of the minimum two values per road segment are taken. This counts down to 23 combinations of derivatives and filtering techniques * 4 extreme values = 92 derivative values per road segment.

4.3.1 Choice of algorithm for severity

The dataset created at this point will be used to correlate each of the 92 derivative alternatives to the inspected severity class. Three basic classification algorithms as shown below are taken into consideration. Models such as KNN or RandomForest are deemed to be too flexible and too uninterpretable to be applied to this rather simple classification problem.

Gaussian Naive Bayes classifier

The Gaussian Bayes model is a relatively simple model based on Bayes' theorem. This model assumes that the input data is Gaussian distributed. All credible[15] derivative alternatives were more or less gaussian distributed at class 'severe'. Outliers were often present. The 'naive' extension means it does not take effects between predictors into account. But in the case of this research, there is only one predictor so therefore this is not relevant.

[15] The predicted outcome is the severity class (0/L/M/E), which are ordinal levels. If the model with a specific derivative alternative is not even able to order these classes correctly, it is not seen as a credible alternative.

For every class the prior probability, mean and variance are calculated. With these values, the fitted Gaussian distribution can be defined. For a new observation, the conditional probability for each class is calculated based on the Gaussian distributions. The class with the highest probability is the class predicted for this new observation.

As a Gaussian distribution is fitted to the input data, the input data is generalized. As the prior probability of class ‘severe’ was much smaller than the other classes due to unbalanced data, the point at which a predictor value would be classified as ‘severe’ was very high. This caused this model to omit a lot of segments that were inspected to have severe ravelling.

Nearest mean classifier

A simpler model that has been tried out was the nearest mean classifier. Similar to the Gaussian Bayes classifier, it calculates the mean value of every class. It does not take the variance or prior probability per class into consideration. For any new value, the class of the closest mean value is assigned.

This classifier was able to identify 30 of the 52 severely classified segments. However, it classified 91 other segments as having ‘severe’ ravelling as well. The rough generalization caused more features than expected to be classified as being ‘severe’, as the decision boundary of a value being ‘severe’ was relatively low.

Multinomial logistic regression

The multinomial logistic regression model is an extension of the logistic regression model. The difference is that the outcome can be categorical instead of binary. This model is in essence very similar to the Naive Bayes classifier. A multinomial regression is less ‘naive’ when dealing with multiple (collinear) predictors, but this comparison step only uses a single predictor at a time. The core of this model relies on the logistic function, which is very similar but not the same as a normal distribution. Experiments by Ng and Jordan (2002) have shown that multinomial logistic regression often has a lower asymptotic error, but Naive Bayes converges faster. As it is unexpected that there is a lack of training features, the multinomial logistic regression is still applicable.

As the multinomial logistic regression has a continuous value as output and gives back a probability per class, this data can be represented intuitively in a graph. These graphs are found in the appendix 9.1. As expected, this model shows very similar results to the Naive Bayes classifier. It also was subject to the unbalanced data problem. On the other hand, very few segments were wrongly classified as having ‘severe ravelling’. In the case of the Rk_Max alternative, 13 of the 52 segments were correctly classified as being severe. Where 5 segments were wrongly classified as being severe.

Multinomial logistic regression with balanced data

To overcome the unbalanced data problem, the source data can be under-sampled in such a way that every class is represented equally. This selection is done randomly, the smallest class is normative for the total training size. The ‘severe’ class is the smallest class, with 252 features. This makes the training dataset having 1008 features. As the Rk_Max derivative has shown good results in the case of ordinary multinomial logistic regression, only this derivative is considered. It is arguable whether it is justifiable to balance the incoming data. As in such case, the spread in results in classes similar to the smallest class is neglected. In reality, the goal of road inspectors is not to select the worst fixed number of roads, but the roads that need to be replaced.

As expected, this method gives lower threshold values between classes than the unbalanced model gives. As this model effectively doesn't take the prior probability into account, the results are very similar to the nearest mean classifier. The difference is that it does take the variance per class into account.

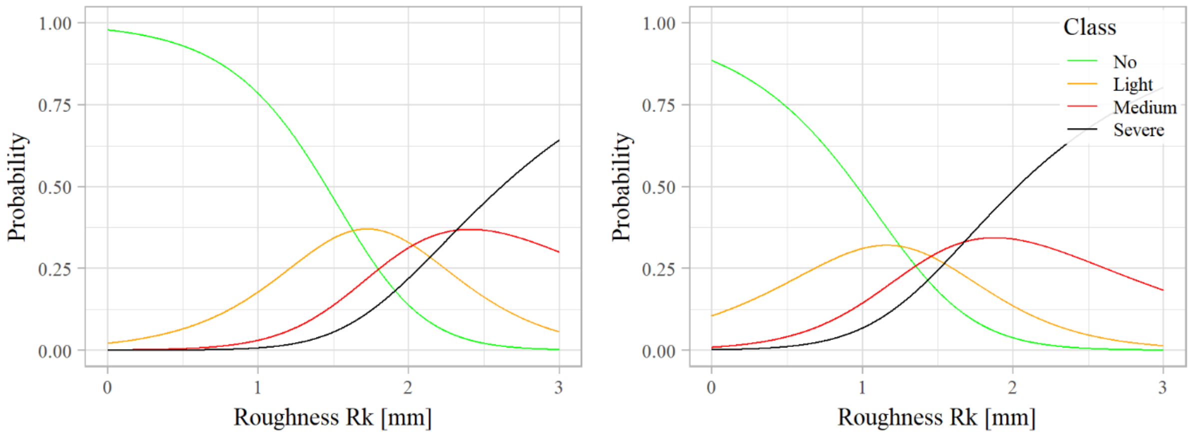


Figure 4.10; Severity class probabilities for given Rk, at unbalanced (left) and balanced (right) dataset (model g5a)

Optimal severity algorithm and derivative

In appendix 9.3, a subset of misclassification matrices is shown. By means of these classification matrices it is chosen to make use of the multinomial logistic regression, as it shows the best overall performance while being well interpretable. It is also concluded that the Rk derivative is the best option to quantify ravelling, especially in class 'severe'. The MPD derivative is a close second. While looking at the distribution of Rk and MPD values, this small difference can be explained. The distribution of Rk_Max in appendix 9.1 shows a slight bulge of values with high Rk. Together with a lower class probability to class 'moderate', this shows this derivative is better in distinguishing severe ravelling.

	Roughness severity thresholds for Rk [mm]		
Model	**Class 0-L**	**Class L-M**	**Class M-E**
Bayesian	1.591	1.901	2.202
Nearest mean	1.252	1.466	1.686
Multinom	1.630	2.027	2.316
Balanced multinom	1.253	1.455	1.677

Table 4.3; Roughness severity thresholds of Rk_Max, generated by model f2.5

To make the model less sensitive to outliers, the average value of the two maximum decisegment values is also correlated with the severity per segment. However, this did not show better results. To select segments with a known severity, the extremes of every road segment are taken. The average value of these selected segments is therefore always more extreme than the general average value. The threshold values are therefore extremes as well. The unbalanced data problem also contributes to this effect. Given the available data, there is no other method to have decisegments with a known ravelling severity class.

It should, therefore, be advised to take a certain probability threshold and calculate the according roughness thresholds, but such probability threshold would be based on management reasons rather than data analysis. Another option would be to execute the same method on well-defined stretches of asphalt with a known severity. Such data is generated with detailed visual road inspections.

4.3.2 Choice of algorithm for extent

Next to the severity class, an extent should also be defined to assign end-of-lifetime. The CROW standards are applied for this at a decisegment scale.

Area [%]	Extent class
0-5	Class 0; Very small extent *[Zeer geringe omvang]*
5-30	Class 1; Small extent *[Geringe omvang]*
30-50	Class 2; Some extent *[Enige omvang]*
50-100	Class 3; Large extent *[Grote omvang]*

Table 4.4; Ravelling extent classes (Source; Handleiding globale visuele inspectie (CROW, 2011b))

4.3.3 Results

By predicting the severity per road segment and calculating the extent, the left confusion matrix based on the test dataset can be generated. As stated before, the EoL moment is reached when it shows moderate ravelling over more than 50% of the road segment (M3) or as a segment shows severe ravelling (E1, E2, E3). With that, the left confusion matrix can be aggregated to the right binary confusion matrix.

Predicted \ Inspected	0	L1	L2	L3	M1	M2	M3	E1	E2	E3	Sum
0	562	170	46	1	56	4	1	17	3	1	861
L1	22	15	4	1	16	1	0	5	1	0	65
L2	3	2	2	0	5	1	0	1	0	0	14
L3	7	9	2	0	5	1	0	0	0	0	24
M1	2	1	0	0	3	1	0	8	0	1	16
M2	0	0	0	0	0	0	0	1	0	0	1
M3	0	0	0	0	0	0	0	0	0	1	1
E1	0	2	2	0	1	0	0	5	3	3	16
E2	0	0	0	0	0	0	0	0	1	0	1
E3	0	0	0	0	0	0	0	0	0	1	1
Sum	596	199	56	2	86	8	1	37	8	7	1000

Predicted \ Inspected	Good	EoL	Sum
Good	942	39	981
EoL	5	14	19
Sum	947	53	1000

Table 4.5; Test confusion matrix between visual inspections and multinomial regression on Rk_Max

After aggregation to the EoL diagnostic, (942+14)/1000*100 = 95.6% of the segments were correctly labelled. However, due to the unbalanced data this is not a good measure in itself. 53 segments were labelled as EoL and the model was only able to find 14 of them, which is 26%.

4.4 Sensitivity analysis

In this segment, the physical and statistical concerns are discussed. These concerns can originate from literature, discussions with professionals, analysis and experiences.

4.4.1 Accuracy of ground truth

The ground truth data, in this case, is the location of segments, the location of inspections and the inspected ravelling class. The following actions are undertaken to prevent inaccurate source data;

- The location of *DGD* segments has all been checked as described in par. 4.1.1. Uncertain segments have been dropped. Errors could have been made, for example in the case of continuously paved asphalt type transits. This is often done near crossings.
- Some road segments were shortened. Road segments smaller 10m² were discarded, as an associated inspection result wouldn't be relevant any more.
- The spatial resolution of the visual inspections is high. After a visit at the executive company, the spatial accuracy giving the extent class is roughly estimated to be below $10m^2$.
- The visual inspections are very subjective, as previous researches have shown (Tsai & Wang, 2015). Improving this is out of the scope. It should be noted that the output accuracy can never be higher than the subjectivity at the input.

4.4.2 Accuracy of texture measurements

The texture measurements can be inaccurate in its lateral, longitudinal and relative vertical direction.

- Lateral or sideways accuracy; For this research, only the right transect of the road is taken into consideration. Therefore ravelling spots in the middle or at the left side of a lane will be neglected. This will cause underestimations relative to the ground truth, causing overestimations in roughness thresholds after modelling.
- Longitudinal or lengthwise accuracy; By means of several physical properties such as speed bumps, cattle grids *[wildroosters]*, and pavement transitions the longitudinal accuracy of the texture measurements can be estimated[16]. The average inaccuracy between the texture measurement and the physical situation over 25 noticeable points was 4.9m, which is deemed to be acceptable. For this research higher precision is favorable but not necessary as the 10m-based measurements are aggregated to 100m.
- Relative vertical accuracy; The absolute vertical height, for example a height relative to sea level, is not needed for the calculations done. The relative vertical accuracy of the texture measurements is defined and validated according to ISO-13473-2. The vertical resolution relative to the car is 0.1mm[7].

4.4.3 Physical situations affecting roughness

Dirt, debris, markings, cattle grids, concrete speedbumps, small repairs, cracks and rejuvenating cures *[verjongingskuur]* can all affect roughness. A simple branch on the road would cause an extreme roughness value but would be neglected by road inspectors. Such artefacts are hard to filter out if possible at all and should therefore be accepted in the model.

[16] Table in digital supplement f3b.3

5 Results question 2; Environmental factors

At this point, ravelling is mainly predicted by the pavement age and typical longevity. We can now quantify the amount of ravelling by both its inspected severity class and by the core roughness depth Rk. Plotting this data against the age of pavement gives us the following graphs;

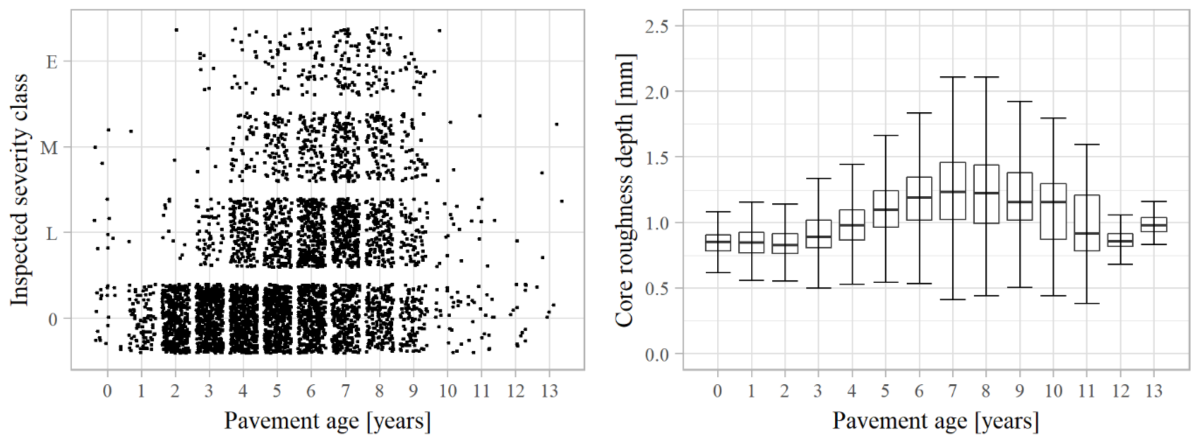


Figure 5.1; Left; Jittered scatterplot of pavement age vs. inspected severity class.
Right; Tukey boxplot of pavement age vs. measured Rk. Box contains 2nd and 3rd quartile of data. Line in box shows median, whiskers show 1.5×interquartile range. (model g5f)

As can be seen, the spread in roughness increases over time. Inferring that factors other than age are at play. This research looks at the influence of environmental factors at such spread. The goal of this research question is to correlate environmental factors with pavement longevity. And secondly, conclude whether it is advantageous to do this based on the texture measurements above visual inspections. This chapter will follow and execute the methodology as described in section 3.1.

5.1 Research into possible environmental factors

For this correlation it is needed to know which environmental factors should be considered. To get a full overview, both literature and professionals are consulted.

5.1.1 Literature review

As the knowledge of environmental factors on porous asphalt is rather limited, this review is sorted chronologically.

Research by Rijkswaterstaat (2007) gave advice in applying *DGD* pavements in non-highway situations. Just like *ZOAB*, *DGD* is discouraged at locations with levering forces. At the point of this document, no long-term observations regarding frost sensitivity were known. It is stated that porous pavements can be more rigid against ravelling as water will flow away and therefore it doesn't stay in the pores and freeze. But this is only the case when water drainage is sufficient.

Extensive research on ravelling in *ZOAB* has been done by Huurman (2008). He accounted shear stresses at curves, bad workmanship or fuel leakage for sections with bad performance. In the case of highways, the emergency lane is not trafficked. In such cases he accounted environmental factors such as temperature differences, UV light, water and air to cause stresses and ravelling.

Hagos (2008) had done in-depth research on the effect of binding ageing on *ZOAB* pavement. He proposed a new artificial ageing protocol to more realistically simulate ageing in lab environments. It was concluded that traffic induced stresses in the mortar are not affected by ageing. But the combination of age, traffic and low temperatures causes a lower resistance against fractures and reduces the self-healing effect.

The Dutch CROW (Werkgroep Stille wegdekken, 2010) has published a booklet with experiences with silent pavements of fifteen Dutch professionals. They advise to check for good weather conditions and prevent manual modifications during application and to avoid locations with levering forces before planning the application of porous pavements.

Research by the New Zealand Transport Agency lead by Henning and Roux (2012) on open-graded porous asphalt has set up a logistic regression model. This model was based on 2500 segments with their age, daily traffic loadings, pavement strength expressed by falling weight deflection and cracking inspections. Environmental loadings were not included, and to predict in the future only the daily traffic loading would be known.

Opara et al. (2016) have done research on the effects of a special additive to de-icing brine to ravelling of porous pavements. This additive keeps the ice slushy, even at very low temperatures, therefore reducing ravelling by 10-17%. He concluded ravelling increases faster at a higher age, and heavy traffic accounted for 25-35% more ravelling on the right lanes.

Research done by Zhang and Leng (2017) has looked at the ageing of bituminous mortar in the porous asphalt of both Dutch and Hong Kong mixtures. They correlated ravelling with rheological properties of the bitumen. The modified bitumen used in Hong Kong was more resistant to ravelling.

5.1.2 Interviews to possible environmental factors

For the interviewsc several professionals in the field of road inspections and measurements have been consulted. Six interview questions have been set up, leaving room for in-depth questions during the interview. The full transcription is to be found in the digital report supplements. These transcriptions have been checked for acceptance with the interviewees.

Summary of interview with A. Blanken

A. Blanken is known to the ARAN road analyzer of KOAC-NPC. He is currently involved in the development of a model to detect ravelling using the LCMS (Laser Crack Measuring System) at Rijkswaterstaat, and the ‘platform of road measurements’ *[platform wegmetingen]* of the CROW.

He stated that the longevity is very depending on the quality at the time of application. In the case of environmental factors, he expects that agricultural traffic causes the pores to clog up, whereby the road stays wet longer which causes the bonds to wear. The presence of a parallel road *[ventweg]* can be a predictor. Secondly, he expects that crossings combined with high traffic intensity cause levering forces, which would break bonding bridges. Thirdly he presumed the effect of trees and its tannic acids from rotting leaf litter. Similar to roadkill, these acids dissolve the bonding bridges, which is often noticed but never proven. According to past experiences of A. Blanken, traffic is the dominant factor. He advises splitting the traffic predictor into multiple classes.

Summary of interview with T. Jansen

T. Jansen has worked analyzing the roads from the footage made by the ARAN measuring device for 15 years. She has set up her own company which is specialized in road inspections, inspections of civil structures and inspections of accompanying assets such as signs and roadsides. Most of which is done remotely based on high-resolution images.

As a cause of ravelling she outlined the effect of UV-radiation and age. Bitumen hardens due to UV radiation, therefore young pavements in shadows can be vulnerable. In the case of environmental factors, she experienced a multitude of possible sources of ravelling. Of which levering traffic at intersections and water drainage have been noted before by A. Blanken. She expects that trees can have influences in three ways. It blocks UV, releases tannic acids, and keeps the road wet for a longer amount of time. She noted that the effect of trees and rain would only be prevalent in winter. As in summer, water will almost immediately evaporate or be sucked out by passing traffic. T. Jansen has not noticed differences in effects between different species of trees.

Summary of interview with M. Nagelhout

M. Nagelhout has been involved in the production of the DRAFT *[detectie van rafeling door middel van textuurlasermetingen]* ravelling detection model. This model uses the texture measurement data of vehicles such as the ARAN and Ramboll's LaserRST.

He describes that ravelling is caused by the weakening of the bitumen bonding bridges. When levering forces are acting regularly the bitumen weakens like metal fatigue. Similar to a comb, you can break it by folding it in half or you can break it by bending it slightly back and forth long enough. Next to the levering forces, Nagelhout noted that roadkill acids and mechanical damages can also cause premature ravelling. Mechanical damages during construction could be investigated. Such as roller compactors covering existing surfaces and abrasive blasting.

5.2 Preparation of possible environmental factors

From the environmental factors named above, it has been researched whether it was possible to find geospatial information. In this section, the factors which are taken into consideration are described. Possible environmental factors that are not taken into consideration are described in section 5.3. The environmental factors are calculated for both road segments and decisegments and are in some cases depending on pavement age.

Factor	Unit	5th percentile	Mean	95th percentile
Age	Years	2	5	9
Tree cover	Percentage	0	5	39
Mean days of frost	Days/year	41	55	63
Mean hours of rain	Hours/year	125	140	160
Heavy traffic	MVT/etm[17]	124.100	405.659	977.835
Light traffic	MVT/etm[17]	1.320.570	3.944.939	6.878.790
Leverage area	Percentage	0	11	59
		Class true	**Class false**	
Has parallel road	Boolean	16%	84%	

Table 5.1; Properties of environmental factors for road segments. Decisegments to be found in appendix 9.4.

5.2.1 Age of pavement

The age of the pavement is inferred from the dataset described in paragraph 4.1.1 and is verified by street-level images. As segments are inspected every other year, and the data is gathered in 3 years, half of the segments have double records of different ages. The ages of road segments are normally distributed between 2 and 9 years, with extremes of 0 and 13 years. In essence, road damage would be a product of age and a certain environmental condition. Therefore other environmental conditions should not have pavement age included as this is a form of data occlusion.

[17] MVT/etm stands for *motorvoertuigen per etmaal*, vehicles per 24 hours

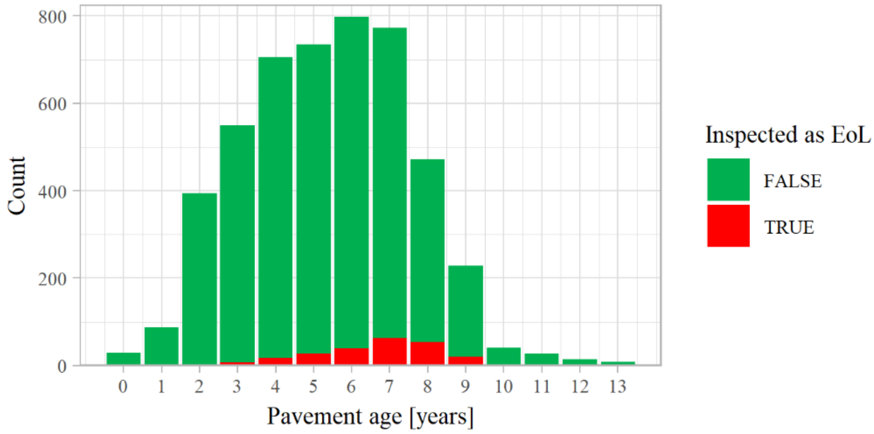


Figure 5.2; Histogram of age and EoL, generated on training dataset by model g5f

5.2.2 Tree cover

The available literature does not describe effects between ravelling and trees. This could be because most literature is targeted at highways, where trees are less prevalent. However, experiences of professionals state otherwise. Both T. Jansen and A. Blanken state that trees could have an influence on the amount of ravelling. Either due to dropping shadows, leaf litter or prolonged presence of moisture. Because of these different possible effects, it is chosen to take the direct percentage of overlap by trees as an environmental predictor. And not a buffer or offset due to drips or shadow.

The website boomregister.nl shows a raster image of 7.5x7.5m per pixel, with the height of trees. It is generated using the AHN2 and aerial images and therefore has a suitable spatial accuracy. The full dataset was made available by the province. As this raster file of 18.6Gb was too large for use in ArcGIS, the raster file was ‘flattened’ using FME. It converted the tree height to a single bit for presence/absence, to only 2.3Gb. The percentage of overlap by area was calculated. In case of the 10m^{1}-sized decisegments, it is chosen to convert the tree cover to binary. This was done as 89% of the decisegments weren’t covered or were barely covered (<10% by area).

5.2.3 Frost

The research of Opara et al. (2016) has shown that ravelling increases more in winter than it does in summer. According to Hagos (2008), this is due to both the freeze-thaw effects and the duration of frosts pausing the self-healing effect. Spatially explicit data of freezing asphalt can be derived from two sources.

- Local measurements from Slipperiness Warning Systems *[gladheidsmeldsysteem]*. Tens of measuring stations are spread out over the province, measuring the asphalt temperature.
- The KNMI data portal provides a service[18] where the interpolated daily minimum temperatures can be found. The KNMI measures these temperatures at 1.5m above a grassy plain.

After an enquiry with J. van der Beek, project leader of winter road maintenance at the province of Gelderland, the asphalt temperature data was not readily available over multiple years. This source was therefore inapplicable. The dataset of the KNMI is used. These daily maps can be ‘stacked’ to calculate the number of days with frost per year per pixel. Some interpolation glitches are noticed after summing multiple measurements per year. The temperature threshold used is 0°C, although a road surface could be colder as temperatures fall close to the ground or it could be higher due to sunshine[19].

[18] data.knmi.nl/datasets/Tn1/2
[19] Verbal communication J. van der Beek, 6th of June 2018

As roads are paved, measured and inspected in summer, the days of frost are calculated per winter season. For example, a road constructed in 2015 and inspected in 2016 will have undergone the days with frost from July 2015 until June 2016. As the total days of frost are very depending on pavement age, data occlusion could come at play. It is chosen to divide the total days of frost in the history of a segment by its age.

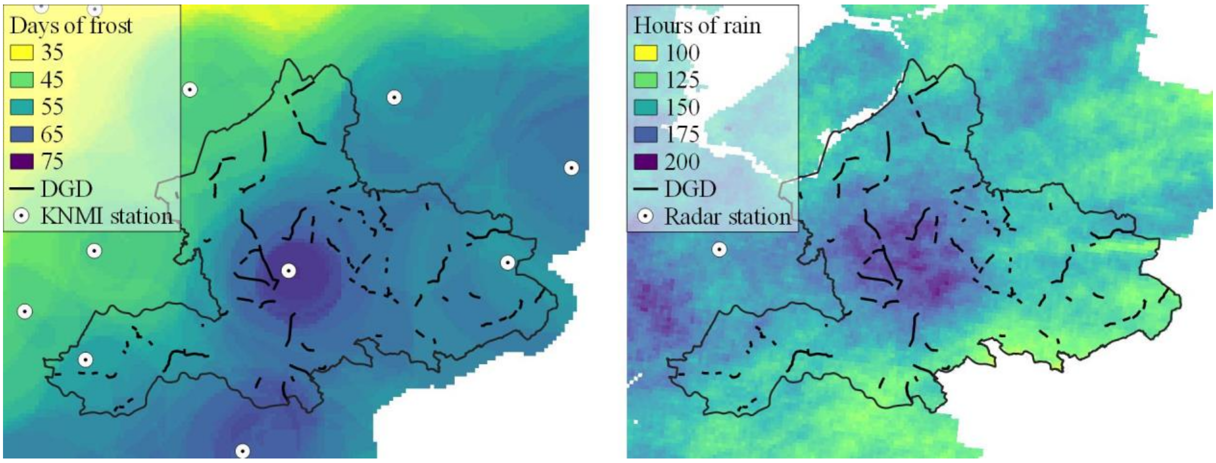


Figure 5.3; Days of frost and hours of rain in October 2014 - March 2015

5.2.4 Rainfall

Similar to the days of frost, the KNMI serves the historical images of the precipitation radar, better known as the *Buienradar*. These images provide an estimation of the precipitation per 1x1km, based on radar measurements taken in De Bilt (<‘17), Herwijnen (>’16) and Den Helder. These estimations are corrected with ground measurements and accumulated per hour[20].

Based on the advice of interviewee T. Jansen, the precipitation during the winter season is has the most impact, as in summer it is expected that water will rapidly evaporate or be sucked out by passing cars. Therefore, the amount of precipitation from October until March is summed. The *Buienradar* often measures small amounts of precipitation, which could be dust or drizzle *[miezer]*. A practical experiment has shown that 1mm of precipitation during an hour hardly wettings a dry pavement. A lower threshold of 1mm/hour is therefore chosen.

5.2.5 Traffic intensity

Different studies have noted or shown that traffic plays a key role in road deterioration. It is known that not only the traffic intensity but also the type of traffic is of influence. As proposed by A. Blanken, the traffic intensity predictor will be split up to the intensity of heavy traffic (trucks) and of light traffic (cars). This data is provided by the province of Gelderland and is measured by permanent and temporary counting points. Heavy traffic is distinguished by the length of the passing vehicle. On average, one in ten vehicles is a truck.

The change in traffic intensity per year is marginal, a general rule of thumb says traffic increases by 1.5% per year[21]. The differences in traffic are therefore way stronger by location than by time. Next to that, the way of measuring local intensities has changed over time. Therefore this increase in time is not considered, and the intensity of 2016 is considered for all ages.

[20] data.knmi.nl/datasets/rad_nl25_rac_mfbs_01h/2.0
[21] https://www.gelderland.nl/GeldersVerkeer

5.2.6 Agricultural traffic

A. Blanken stated that agricultural traffic can cause ravelling, and the presence of a parallel road *[ventweg]* could be a predictor for such effect. The province of Gelderland fortunately provides an accurate dataset of line segments where a parallel road is present[22]. The start and end points were digitized with sub-10m precision with respect to aerial and street-level images. The presence (yes/no) is joined to the road segments by a spatial join.

5.2.7 Levering forces

Several sources and all interviewees accounted levering traffic *[wringend verkeer]* for a decrease in service life. Evidently, Rijkswaterstaat (2007) and CROW (2010) advise applying dense-graded instead of porous asphalt near crossings, roundabouts and deceleration lanes. This advice is followed by the province, as SMA is applied instead of *DGD* in the same pass. This raises the question whether there is enough statistical evidence of ravelling in *DGD* near crossings.

The provincial Geoserver is very helpful again, containing a point dataset of crossings and their type. IMGeo was used to find bus stops *[haltekommen]*. The buffer distances were based on expert-view. Analogous to the percentage of tree cover, 88% of the decisegments did not have any leverage interference. In this case, having more than 10% of the area is considered as being subjected to levering forces.

Type	Source	Type	Join type
Crossing	GeoServer	Point	Area within buffer of 25m
Roundabout	GeoServer	Point	Area within buffer of 50m
Crossing with traffic lights	GeoServer	Point	Area within buffer of 50m
Bus stops	IMGeo	Polygon	Adjacent road axis expanded with 10m, within flat-end buffer at same side of road

Table 5.2; Sources of areas with levering forces, as used in models g1g and g2g

5.3 Possible environmental factors not taken into consideration

Either due to lack of examples, lack of geodata or lack of supporting theories, the following environmental factors are not taken into consideration.

5.3.1 Freeze-thaw cycles

The 'days of frost' predictor is substituted for the amount of freeze-thaw cycles. As described before, the province and Rijkswaterstaat have Slipperiness Warning Systems, which measures the temperature of the asphalt at multiple locations with sufficient temporal resolution. However, the data is not readily available for multiple years. This may be available at *Meteogroup*, the software provider. Besides that, local differences were deemed to be too big due to windy locations, shadows and gritting. This variance by location may not be covered better than the variance in the 'days of frost' predictor.

The KMNI-dataset about the daily minimum temperature had an insufficient temporal resolution. The minimum temperature was given daily but freeze-thaw cycles can occur more often.

5.3.2 Road alignment

It is understandable that uphill facing roads deteriorate faster than downhill facing roads. But this is not underlined by current literature or noted by the interviewees. Secondly, not too many roads in the province of Gelderland show decent slopes.

[22] https://opendata.gelderland.nl/dataset/ngr-wegen-parallelwegen-langs-provinciale-wegen-provincie-gelderland

Like road slopes, road curves are expected to wear faster as well. This was noted by Hagos (2008) but was not underpinned by other sources. There are quite a lot of curves in the alignments of provincial roads, but the spread in radii is big. To prevent too much 'binning', it wouldn't be desirable to include these predictors.

5.3.3 Mechanical damages

M. Nagelhout noted some experiences with mechanical damages during road construction.

- Roller compactors damaging existing adjacent pavements. Situations could be found where an adjacent trajectory is newer, but this effect is too localized to be detectable. The same goes for manual modifications during application (Werkgroep Stille wegdekken, 2010).
- Abrasive blasting *[gritstralen]* of temporal markings, but after further inquiry with R. Hermsen this is never done as it damages the road below.
- Track blasting *[kogelstralen]* if the initial grip *[aanvangsstroefheid]* is too low. This is rarely done. Instead, the driver is warned, and/or the speed limit is lowered temporarily.

Next to the reasons stated above, it is also doubtful whether enough examples can be found to have enough statistical evidence. Other sources of mechanical damages can be accidents, roadkill *[faunaslachtoffers]*, snow ploughs, scraping parts of (agricultural) vehicles, etcetera. Such sources are unpredictable, therefore future predictions cannot be made.

5.4 Machine learning analysis

The goal of this research is to find the environmental factors that cause a road to reach EoL. The age and environmental factors are taken as predictors. In the case of predicting the visual inspection outcomes, the outcome would be binary; End of life or not. In the case of texture measurements, the outcome is the roughness value per decisegment and the model of Q1 can be used to convert this to the EoL diagnostic. In this preliminary study, a generic machine learning method will be used. For better understandability of the problem, and interpretability of the outcome.

5.4.1 Applicable machine learning models

The following basic models are considered in this research;

- **GAM's;** At decisegment level a continuous output (a roughness value) will be predicted. Fitting a linear regression formula in the form of y=ax+b would therefore be applicable. Generalized Additive Models are based on such linear regression and allow for non-linear functions for each predictor (James, Witten, Hastie, & Tibshirani, 2013). GAM's are not able to combine predictors by itself. Variables with expected interactions, for example age*days of frost, or age*tree*hours of rainfall, will therefore have to be defined manually. This does make the model high-parametric and therefore hard to implement. An applicable model with fewer parameters is desired.
- **Multiple logistic regression;** This model is expanded from regular logistic regression to be able to predict a class based on multiple inputs. As this is a relatively simple model, it is less data hungry. I.e. it converges faster but may not have the best accuracy. Logistic regression is also not able to find interactions between predictors by itself.
- **Support vector classifier;** At road segment level a support vector machine is applicable. This model tries to define a boundary between EoL and non-EoL observations. The flexibility of such boundary can be defined by the SVC kernel. As the input data is not well-separable, having an inflexible kernel is not expected to show good results. At the other side of the spectrum, a very flexible kernel with a high dimensional model can lead to overfitting.

- **K-nearest neighbors;** This model looks for the K-number of most similar observations. The most prevalent class of these neighbors will be the prediction. This gives the model the possibility to act very flexibly. However, it does depend on the Euclidean distances and therefore is very sensitive to scaling. As different units are used, and their mutual importance is unknown, this model is less applicable. A high K will quickly result in a model never predicting EoL as this is a minority class.
- **RandomForest;** Random forest models are widely used, because of their flexibility and a small amount of hyperparameters. It generates a number of decision trees, but at every split in a tree, it only considers a random sample of all predictors. This prevents correlations between the generated trees. Which in turn makes the average result of these trees more reliable while preventing overfitting (James et al., 2013). Secondly, by this random sampling interactions between predictors are found automatically.

The following applicability scores are given based on applicability, interpretability and practicality. For those reasons, Logistic Regression and RandomForest will be considered in further analysis.

	Applicability	
Model	**Visual inspections**	**Texture measurements**
GAM; Generalized additive model	0	+
Logistic regression	+++	0
SVC; Support vector classifier	++	+
KNN; K-nearest neighbors	+	0
RF; RandomForest	+++	+++

Table 5.3; Applicability of machine learning models. 0 for inapplicable.

5.4.2 Train and test division

To assess the performance of a model it will have to be validated. There are several techniques to validate a model. As there is no lack of observations, the data will be divided between a train and a test dataset. The available dataset contains 6990 observations at 4690 different road segments.

To do a decent assessment of the performance, the test set will have to be large enough to inherit most unique combinations of predictors. However, it is desirable to keep the training dataset as large as possible. Finding the optimal ratio is guesswork, but the effort is taken to make an educated guess;

Division ratio

To have a decent training dataset, it is assumed that most combinations of predictors should have at least 5 features. And to make a decent test, most combinations should contain at least 1 feature.

Of the 8 predictors, 3 are binary or binary-like; TreeCover, HasParRoad and LeverageArea. Other predictors are continuous and therefore do not add 'bins' as categorized values do.

- 750 segments (11%) of the observations have more than 10% tree cover
- 1058 segments (15%) of the observations have a parallel road
- 1650 segments (23%) of the observations have more than 10% levering area

The smallest bin is therefore the bin where all three above cases are true. Of the 6990 observations, 13 observations are covered by a tree while having a parallel road while having a levering area. This smallest bin does not contain any EoL segments and contains observations at ages between 2 and 7 years. This is not beneficial as it is preferred that there are EoL segments in all bins.

Dropping predictors

Dropping insignificant predictors can be beneficial. As will be shown below in paragraph 5.4.3, the predictor “HasParRoad” is the least significant predictor. Dropping this predictor greatly increases the size of the smallest bin. The smallest bin is now defined by the tree cover > 10% and levering area > 10%. This bin contains 7 observations being EoL according to the texture measurements.

As stated before, 5 observations are deemed to be enough as training data, and about 1 observation should be left for testing. Therefore a test/train ratio between 5:2 and 6:1 is needed. A ratio of 3:1 (75:25) is aimed for (model g3b.2, see appendices).

Sampling method

Observations at road segments are highly spatially and temporally correlated. A total random sampling of observations is therefore not applicable. Sampling is done on road trajectory level instead (model g3c.1). Segments are randomly sampled based on road name. The observations at the segments of N310, N315, N335, N788, N796 and N839 are selected as test subset. A train to test ratio of 2.8:1 is found.

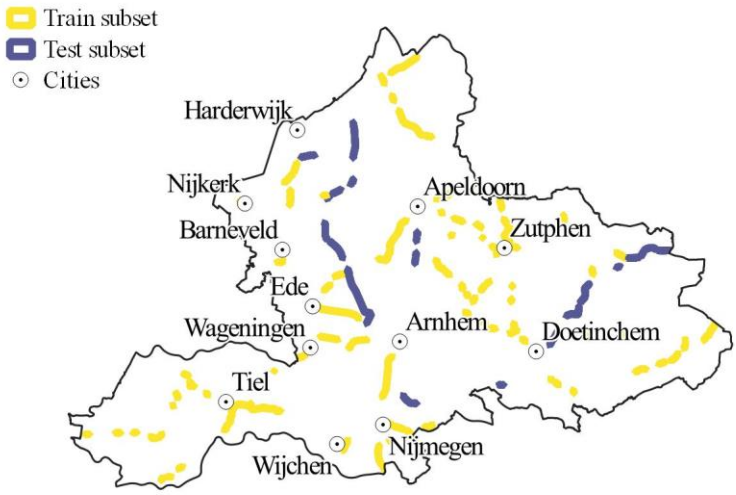


Figure 5.4; Train and test subsets sampled by road name

5.4.3 Modelling visual inspection results

Inspections are done at a road segment level. The goal of these models is to predict the inspected EoL diagnostic.

Multiple logistic regression predicting inspected End-of-Lifetime

As explained, the logistic regression model is not able to find or use interactions between predictors. And it can be beneficial to exclude insignificant values. The following steps have therefore been taken to come to an optimal model (model g3c.3, see appendices);

A. **Only age**; As a 0-hypothesis, this model is trained based on only the age of pavement. It had an AUC of 0.781, which is already relatively high.
B. **Included environmental factors;** The calculated environmental factors were included as-is. The predictor “ENV_HasParRoad” has been excluded, as described in paragraph 5.4.2. AUC=0.719. This value is slightly worse, which can be explained by overfitting to the training data.
C. **Dropping insignificant predictor;** To overcome the slight overfitting, the predictor with the highest p-value is dropped. In this case, it is the HeavyTraffic predictor. As this predictor is highly correlated with LightTraffic, this high p-value could be explained by data occlusion. Hereafter LightTraffic had the highest p-value, which was unexpected. AUC=0.729

D. **Traffic as a function of age;** The traffic intensity itself does not have much influence. Instead, traffic intensity * age does. This didn't improve the AUC; 0.707
E. **Weather as a function of age;** The same interaction is expected for DaysOfFrost and HoursOfRain, which were divided by its age previously. This did improve the model considerably; AUC=0.842
F. **Traffic and Weather as a function of age;** A small improvement, AUC=0.849

In linear models such as logistic regression, balancing the input offsets the output probabilities. The AUC is insensitive to this. Dropping other predictors and including several other interactions has been tried, but did not show any major improvements above model F.

RandomForest predicting inspected End-of-Lifetime

The advantage of a RandomForest model is that it can find interactions by itself. It, therefore, doesn't have as many hyperparameters to alter or optimize. The two required hyperparameters are;

- **Ntree;** Number of trees grown. Enough trees should be grown to get to a stable out-of-bag error. Too many trees cannot lead to an overfit, as each tree is uncorrelated from another due to different predictors at Mtry and, if applied, different bootstrap subsets of the training dataset.
- **Mtry;** Number of randomly selected predictors per split in a tree. As it is expected that the interactions between predictors are at most with 3 (e.g. tree*rain*age), mtry=3 is chosen.

The following models were generated;

A. **Only age;** Again, this is taken as a 0-hypothesis. This model only puts out EoL=F, no matter the age. This is because all trees grown are very similar as only a single predictor is used, decomposing the RandomForest model to a bagged tree (James et al., 2013). The only difference between the grown trees is the bootstrapped training subset. But as expected, for every subset, EoL=F is the majority class. To overcome this and to be able to make a 0-hypothesis, we should apply stratified sampling.
B. **Stratified sampling with only age;** By means of stratified sampling, each tree is grown with a fixed amount of EoL and non-EoL observations. In the training dataset, 261 observations were given EoL. For each tree, a random selection of 261 non-EoL observations was added. Essentially balancing each tree by undersampling. AUC=0.844
C. **Stratified sampling on environmental predictors;** At this stage, it can be shown that the parallel road predictor has a negligible effect on the outcome and accuracy of the model. A measure of importance per predictor is the decrease in Gini before and after a split in a tree. Where Gini is a measure of purity. As can be seen in table 5.4, HasParRoad gives the lowest average decrease and is excluded to decrease the amount of binning described in paragraph 5.4.2. AUC with all environmental predictors; 0.771, AUC excluding HasParRoad; 0.762
D. **Stratified sampling, exclude unimportant predictors;** What has been done with HasParRoad can be done another time. Tree cover has the second lowest decrease in Gini. Excluding this predictor did not give better results. Hereafter decrease in Gini values were very similar. AUC when excluding TreeCover; 0.753.
E. **No stratified sampling;** The problem described at model A was a lack of differences between trees. As multiple predictors are used, this problem is not apparent any more. Stratified sampling may cause lower predictive performance. It is concluded that this is not the case; AUC with all predictors excluding HasParRoad; 0.694
F. **Altering Mtry;** For the best performing model so far, the Mtry has been altered. A Mtry=1 gave the best result with AUC=0.778, gradually decreasing to 0.744 with Mtry=6.

The best performing model based on RandomForest was the one with age as the only predictor. Including environmental factors did not increase the predictability of the inspection outcomes in the case of RandomForest models. Relative to logistic regression, RandomForest can behave very flexible, leading to overfitting. This may be the reason why RandomForest behaved worse than multiple logistic regression.

Predictor	Visual inspections Mean decrease Gini	Texture measurement Increased Node Purity
ENV_MeanDaysOfFrost	50.79	648.85
ENV_MeanHoursOfRain	41.27	495.03
ENV_HeavyTraffic	37.01	440.97
ENV_LightTraffic	36.48	696.36
Year_Age	35.62	694.84
ENV_LeverageAreaPercentage	27.10	67.75[23]
ENV_TreeCoverPercentage	14.49	35.58[23]
ENV_HasParRoad	**4.85**	70.12

Table 5.4; Increase in purity per predictor for RandomForst models (model g3c.1)

5.4.4 Modelling texture measurement results

The ravelling measurements have been generated at road decisegment level. At this level, only the severity of ravelling is known. A road is however rejected based on the severity and extent. The same methodology as used for the inspection results cannot be followed. Instead, the roughness per decisegment will be predicted based on the environmental factors. This will be done by a RandomForest model. The severity class of every decisegment will be calculated. Which in turn can be aggregated to a road segment EoL prediction by the decisegment approach.

In the search for the optimal RandomForest model, the MSE statistic between prediction and measurement roughness will be used. This is done for the train and test dataset. For the same reasons as stated before, Mtry=3 will be taken as a default. With the high amount of observations, calculation times rose fast with higher Mtry. Ntree was therefore set at 250 while repeatedly checking for convergence.

A. **Only age;** Taking only the age as a predictor, a train MSE of 0.070mm is found. Which is quite high, but as this is an inflexible model the test MSE is similar, 0.089mm.
B. **Environmental factors including HasParRoad;** This model should not be used as the test dataset doesn't cover enough segments with parallel roads. This model is added for completeness, such that the influence per predictor is compared in table 5.4. Train MSE=0.022mm, test MSE=0.084.
C. **Environmental factors without HasParRoad;** This would be the base model. It is the most flexible model as all predictors are included. This would give a big difference between the train and test error. Train MSE=0.022, test MSE=0.086mm
D. **Exclude worst predictor;** Decreases flexibility, as it can be expected that this model is behaving too flexible. Like the road segment model, TreeCover is excluded. Train MSE=0.022, test MSE=0.087mm, so no significant difference.
E. **Vary Mtry;** The best performing applicable model so far is model C. With decreasing Mtry, the train MSE increased due to the model being less flexible. The lowest test MSE was reached with Mtry=2 at MSE=0.085mm. The importance values per predictor were not very different from Mtry=3. These results are marginally better. The test set is too small to distinguish noise from improvement. As it is still expected that at most 3 factors are interacting with each other, mtry=3 is used for the next model.

[23] Percentages <10% are converted to boolean for decisegments, see also par. 5.2.2 and 5.2.7

F. **Increase Ntree;** With increasing Ntree, the average MSE of the grown trees should decrease asymptotically. With Ntree=250 the MSE was still slightly decreasing. Growing more trees could therefore be advantageous. However, no major improvements were shown. With Ntree=500, the train MSE=0.023mm and the test MSE=0.084mm.
G. **With Ntree=1000**; the train MSE went down to 0.022 and the test MSE went up to 0.086.

After roughness is predicted, the accompanying severity class at varying confidence thresholds is calculated. This is done to get the ROC-curve and calculate the AUC. The AUC of the model only using age was 0.789. As the input of this model is limited to the existing ages, levels 0 to 13, the output also only has 14 different levels. This expresses itself into a 'blocky' ROC curve. This can explain the relative high AUC. The optimized model F returned an AUC of 0.756. This output was not showing this effect of lack of input variance. And is therefore seen as a plausible model.

5.4.5 Model performances

Without making a choice in the sensitivity-specificity tradeoff, a ROC-curve is used to assess the model performance. The resulting ROC curves are shown in appendices 9.5 – 9.7. The following Area Under Curves are found;

EoL according to	Model	Area under Curve [-]	Test MSE [mm]
Visual inspections	logistic regression, only age	0.781	
Visual inspections	logistic regression, optimized	**0.849**	
Visual inspections	RandomForest, only age	0.844	
Visual inspections	RandomForest, optimized	0.778	
Texture measurements	RandomForest, only age	0.789	0.089mm
Texture measurements	RandomForest, optimized	**0.756**	0.084mm

Table 5.5; Model performances

Spatial residuals evaluation

At the independent test set, the predicted and the actual roughness per decisegment is known. The difference between these values is the error, or model residual. These residuals can be viewed spatially as well. If any spatial relations can be found, this may infer that a spatial predictor is overlooked. Several pavement professionals have looked at these residuals. As humans are very good at finding patterns, patterns were found.

- One of which was that at straight roads, there was a bad segment every 50m or so. This pattern did not hold up over longer lengths or other places. Neither was it explainable by for example the application methods used. A single truckload of asphalt only paves several meters, and lunch breaks are less frequent.
- The residuals were often of the same scale for a complete road trajectory. These differences could not be correlated with any environmental property. Further research could be done to assess these differences with, for example, conditions during application.

The optimized RandomForest model that was used to predict the decisegment roughness had a test MSE of 0.084, and therefore an RMSE of sqrt(0.084) ≈ 0.3mm. For the images below, any residual smaller than 0.3mm is within the expected errors of the model. Values higher than this value are deemed to be over- or underestimated and are colored yellow to red. The full residual map including the measured roughnesses and the expected roughnesses for 2020 are made available online or alternatively via a WMS server.

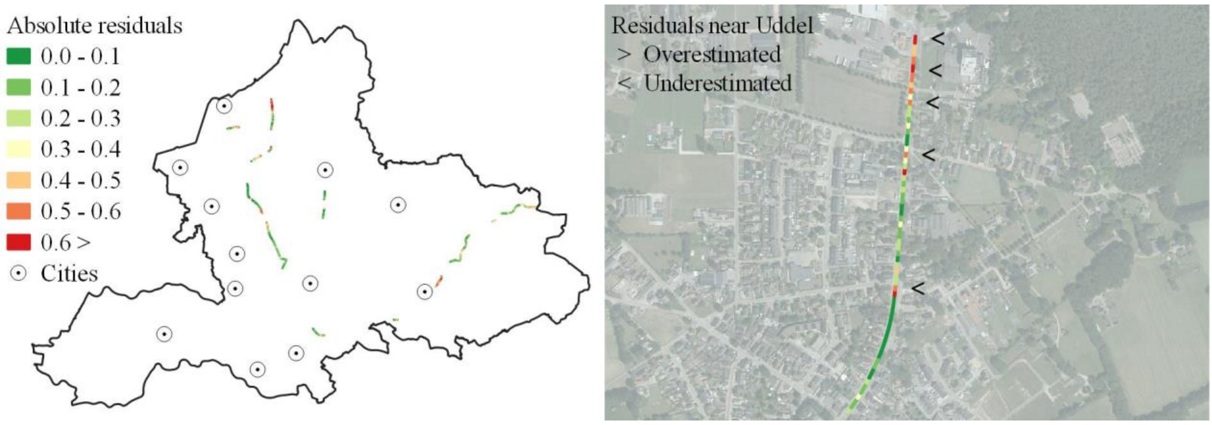


Figure 5.5; Residuals of test subset

5.5 Scenarios for influences of environmental factors

At the decisegment level, the roughness per decisegment is predicted based on the environmental factors. As can be seen in table 5.4, the pavement age is a very important predictor. Using this as a base, the influence of other environmental factors could be assessed by means of several different scenarios. At ages below 2 and above 9 years, the number of observations were limited and must be considered with caution.

0. **Base scenario;** The base scenario would be a 'normal' road. No overhanging trees, no nearby crossings, an average amount of traffic, and a moderate climate. As can be seen, the roughness makes a sort of S-curve. This same curve is found in the research by Hagos (2008) as well.
1. **High rain;** The hours with rain differs slightly over the province. A 10% increase is already at the higher end of the spectrum. Rain on itself did not have much influence.
2. **High frost;** The same goes for the days with frost. Frost on itself did not have much influence.
3. **High rain and frost;** A combination of high rain and high frost strengthen each other. The certainty of this conclusion is high, as these predictors also had a high importance as displayed in table 5.4. With a pavement older than 6 years, locations of high frost and rain show 0.1mm more roughness relative to the base model. With an average increase in Rk of (1.3-0.8mm / 9y) = 0.055mm/y at the base model, this is a difference of little less than 2 years.
4. **Leverage area;** An increase of ravelling near crossings was expected but is not prevalent in these models or the importance displayed in table 5.4. The reason behind this is part of the discussion. It should be noted that some crossings in the data have less deflecting traffic than some driveways have. Especially in the case of agricultural or industrial plots. Secondly, the residual map shows that the texture measurements are sometimes too misaligned with respect to the actual location. Thirdly, lower texture roughness values have been noticed near crossings. Which can be explained by compaction due to these levering forces, or a misclassified pavement type.
5. **Tree;** Having overhanging trees has shown to have a low influence according to table 5.4, although the figure below states otherwise. The presence of having overhanging trees does have quite some influence. When a segment of 5 years old is below a tree, it shows the same amount of ravelling as a segment without a tree which is 6 years old.
6. **Tree and high rain;** When comparing scenarios 2, 5 and 6, tree and rain do not seem to amplify each other.

7. **2/3rd traffic intensity;** According to experts, traffic has a big impact on road longevity. This can be noticed; all traffic scenarios show large deviations from scenario 0. This scenario is predicted by a road segment having 2/3rd of normal traffic, for both light and heavy traffic. Half the normal traffic has been tried but gave doubtful results that can be explained by a lack of observations at such intensity. The data shows that a pavement with light traffic intensities of 6 years old is just as damaged as a normal road of 4,5 years old.
8. **Traffic 10% more than normal;** Looking at scenarios 7 and 9, the traffic nicely follows the expected behavior of 'more traffic gives more damage'. The 10% increase from normal traffic shows otherwise. This shows the high spread in the results of the model.
9. **Double traffic;** This scenario is the opposite, double traffic, of which enough segments are observed to make a decent prediction. What is interesting is that the real effect of traffic only shows after about 5 years. A hypothesis for further research could be that up to 5 years, the bitumen is flexible enough to have a self-healing effect. Hereafter it is hardened and shows ravelling which is accelerated by the forces of passing traffic. This theory is supported by the conclusions of Hagos (2008) in the case of *ZOAB* pavement.

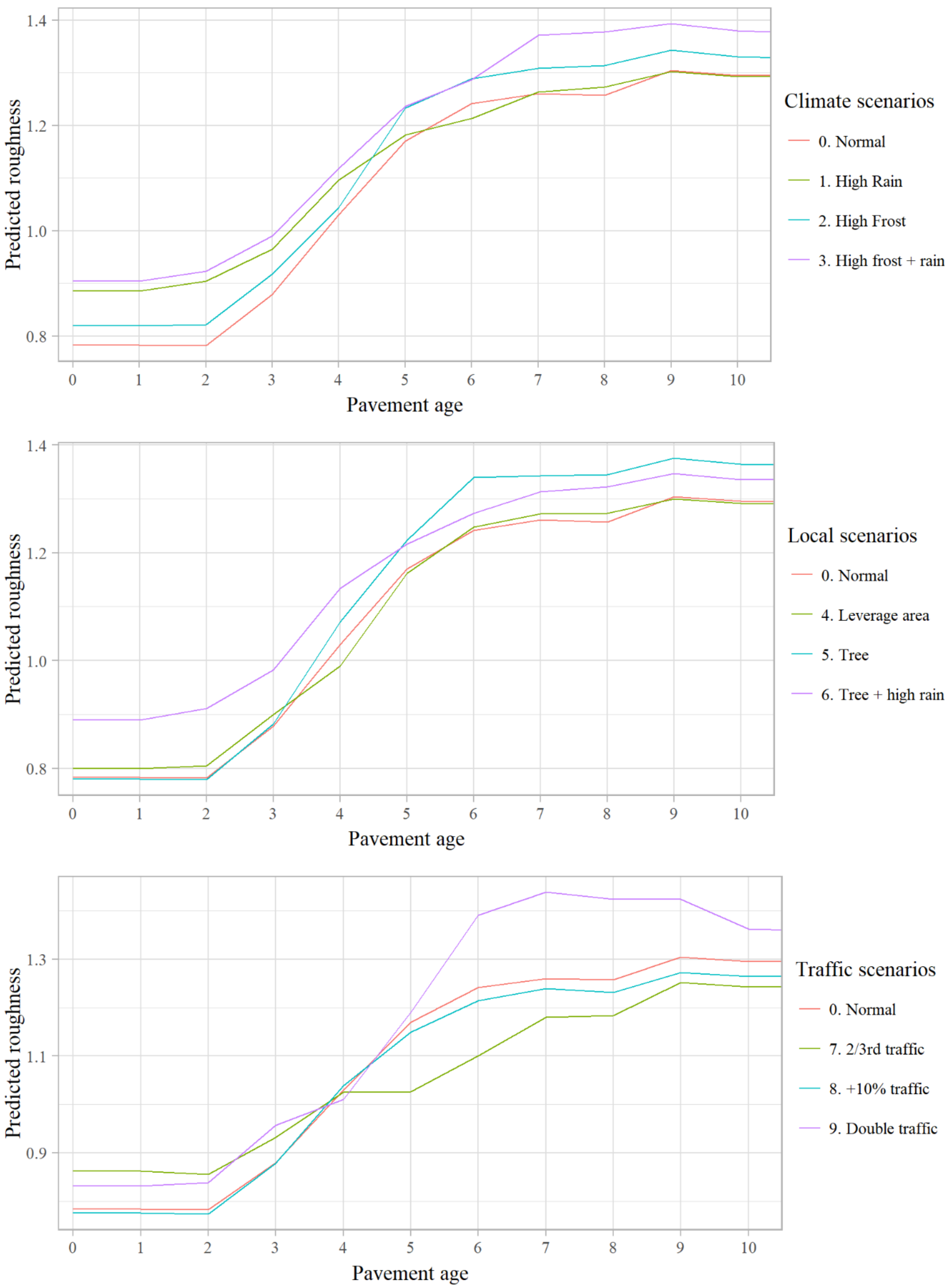


Figure 5.6; Growth of roughness indicating ravelling under different environmental scenarios (model g5a)

6 Conclusions

As described in the research setup, this research uses two research questions to answer the main question. Each question has been divided into sub-questions. These sub-questions have been answered by following the described methodology. Therefore, this chapter is divided into two sections as well, following the two research questions. The last paragraph of each section will answer the research question.

6.1 Question 1; Texture to ravelling class model

The goal of the first research question was to find a way to correlate texture measurements with ravelling. Which thereby can be a source of objective ravelling information with a high spatial resolution.

Q1a; What is the definition of ravelling and how does it arise in the case of DGD asphalt?
First, the term 'ravelling' in the scope of this project has to be defined.

Ravelling is defined as the loss of aggregate in the surface of pavement (Zhang & Leng, 2017). This occurs due to the ageing of the bitumen binder (Mo et al., 2009). Aged bitumen is less flexible and has a lower retention force whereby small stone aggregates can disappear out of the road surface.

Q1b; Which of the known derivative methods is most suitable to correlate a 2D texture measurement with visual inspections regarding ravelling at the CROW end-of-lifetime threshold?
The second research question was aiming at finding a method to correlate texture measurements to ravelling severity classes. It focusses at severe ravelling, as that is the main reason why a road is deemed to have reached his end-of-life.

Of the 23 derivative methods applied, the 'Core roughness depth (Rk)' with use of a high-pass filter as defined by the ISO-13565 shows the best ability to distinguish the severity of ravelling on a decisegment level. This derivative is known in the field of combustion engines and is chosen to be used in further analysis. The more accepted derivative 'Mean Profile Depth (MPD)' shows very similar results but is slightly worse in distinguishing severe ravelling. For reference, the Rk to MPD formula proposed in paragraph 4.2.5; derivative similarities is very usable.

Q1c; What is the performance of the method to predict visual inspection outcomes from the most suitable texture derivative?
Alongside the most suitable method, it is important to know how well this conversion method works overall. This is key for the comparisons of the second research question, as it is depending on this conversion method.

Of the 53 segments in the test set that are found to have reached end-of-lifetime according to visual inspectors, 14 were found correctly by means of texture measurements using 'the decisegment approach'. 5 road segments were wrongly classified as being EoL. The Rk thresholds are quite high in comparison with the actual measured values.

As the accuracy of the visual inspections is unknown, the added value of the predictions is hard to assess. With the true negative rate of (14/53)*100=26% and a very good negative predictive value of (14/(14+5))*100=74%, there is statistical evidence that this method has potential. But at this stage, the model is too inaccurate to be practically useful. It can, however, be an indicative measure of the state of the road which is more detailed than a severity class.

Question 1; Which of the available methods of deriving ravelling from texture measurements is most suitable to classify the severity and extent of ravelling in the case of DGD asphalt?
At this stage the sub-questions of this first research question have been answered, whereby the answer to the first research question can be concluded.

The Rk derivative as defined in ISO-13565-2 in combination with unbalanced multinomial regression is the most suitable method to predict severe ravelling and end-of-life. The thresholds on the extent as stated in the CROW publication 146c are directly applicable and show suitable results. The overall accuracy is too low to make the model practically useful to conclude need of repair, but it does show potential for further research.

6.2 Question 2; Influence of environmental factors

The first research question has provided a dataset of *DGD* pavement with inspection- and measurement results. And has provided a method to convert one into another. This is a key input for this second research question.

Q2a; Which local environmental factors may be of influence on the longevity of DGD asphalt, and how?
The data provided by the first research question shows that the spread in ravelling increases over time. Inferring that factors other than age are at play. Several types of research and interviews have been used to come up with a list of possibly influential environmental properties.

According to the available literature, it is expected that traffic has the most influence on the longevity of *DGD* asphalt. The bonding bridges that keep the stone aggregate together are weakened by the regular stresses of traffic. Similar to a comb, you can break it by folding it in half or you can break it by bending it slightly back and forth long enough. Levering traffic near bus stops and crossings may amplify this effect.

Damages during winter seasons can be related to frost. Low temperatures pause the self-healing effect of asphalt and expanding ice can induce extra stresses. The effects of trees are not described in existing literature as most literature is aimed towards highways where overhanging trees are less prevalent. Experts have often seen correlations between trees and ravelling. Trees block UV radiation, which causes the asphalt to be flexible for a longer amount of time. This could have positive and negative effects. Moisture and leaf litter chemically damages the bonding bridges, possibly shortening pavement longevity below trees.

Q2b; What accuracy can be achieved using a predictive model using local environmental factors?
The possible environmental factors have been converted to geodata. This data has been correlated with the inspected and measured amount of ravelling using machine learning.

The goal of this research was to be able to predict when segments have reached their end-of-lifetime. About 5% of segments have reached this point according to the visual inspections. This means the province of Gelderland is doing a good job at maintaining their roads. But this also means the input data for this research is highly unbalanced. Models stating that all segments would be OK would be 95% accurate, but still useless. To overcome the unbalanced data problem, the performance per model has been assessed by the area below the ROC-curve.

The outcome of visual inspections could be predicted with an AUC of 0.849, which is significant. This was done with a logistic regression model using all available environmental predictors. The prediction accuracy is not believed to be good enough to depict future maintenance, but it can generate a subset of road segments which deserve extra attention. This is a recommended use of this model.

The outcome of texture measurements was predicted in two steps, which is referred to as ‘the decisegment approach’. Per 1/10th of a road segment the ‘core roughness depth’ Rk [mm] was predicted with an RMSE of 0.3mm. Secondly, these roughness values were converted to a segment EoL using the model generated in the first research question. An AUC of 0.756 was found, which also can be seen as a significant outcome.

Q2c; How much influence do local environmental factors have on the longevity of DGD asphalt?
The influences of environmental factors can be assessed by importance factors. This does however not show the effects of predictor interactions. Scenarios have been set up to show such effects, and confirm or disprove assumptions from literature and experts.

Conclusions were made relative to a base model with average traffic, no trees or nearby crossings and a moderate climate. These conclusions can be used as a rule-of-thumb;

- Rain or frost in itself did not have powerful effects. The combination of these two did. At pavement ages above 6 years, a 10% increase in hours of rain and days of frost caused an increase in damage worth a little less than 2 years.
- The influences of levering traffic are not found by the model. Multiple explanations have come up. It is expected that the texture measurements do not have enough spatial accuracy.
- Overhanging trees had a low importance factor according to the models. In the scenarios, it did have a significant influence. When a segment of 5 years old is below a tree, it shows the same amount of ravelling as a segment which is 6 years old.
- The influences of traffic were very noticeable. The influence becomes significant after about 5 years of age, as supported by Hagos (2008).

The overall influences of environmental factors were quite prevalent in the data, and follow expectations set in interviews. However, the test RMSE of 0.3mm is quite high relative to the differences between scenarios. Meaning that a lot of ravelling could not be explained by the model. These residuals were plotted on a map and viewed by several experts. The residuals differed on the trajectory level and could not be explained by any environmental property.

Q2d; Does the use of texture measurements above visual inspections improve the accuracy of the predicted moment of end-of-lifetime?
This question can be answered by assessing the differences in the ROC curve between the inspection model and the measurement model.

As concluded by subquestion b, the visual inspection model had an AUC of 0.849 and the texture measurement model had a lower AUC of 0.756. Both models show significant results. It is unknown whether this difference in AUC is significant with respect to the differences in the characteristics of these models. One of these characteristics is the prior probability. 5.7% of the road segments have reached EoL according to the visual inspections, while only 1.3% were deemed to have reached EoL according to the texture measurements. Further research could investigate the AUC if these prior probabilities are similar.

Question 2; How well is the influence of local environmental factors on the longevity of DGD asphalt measurable?
Models based on the environmental factors predicting the amount of ravelling gave results with significant accuracy. However, this accuracy is very depending on the required specificity. No conclusion on prediction accuracy could be made without making a subjective choice in required specificity.

In the case of texture measurements, the influences of environmental factors are quantifiable by comparing different scenarios. Using differences between these scenarios, rules of thumb have been set up. The influence is noticeable and mostly follows expectations but has a high spread, as the test RMSE was relatively large. Rules of thumb have been set up. But these rules will not be specific enough to do a decent prediction on the end-of-life point of a road segment.

7 Discussion

Along the execution of this research, several assumptions and concessions have been made. Looking at the final conclusions, some assumptions were incorrect and, in some cases, concessions had a seemingly large impact on the results.

Visual inspections as ground truth

Subquestion 1c concluded that the method of comparing roughness of the right measurement transect with the inspected severity class shows potential, but that the added value is unknown. The results of the first research question could not conclude on the accuracy of the visual inspections, which is seen as a ground truth.

The second research question was able to conclude that the texture measurements show better predictability based on environmental factors. Which could be a reason to believe that the subjectivity of visual inspections is greater than the measuring inaccuracy of texture measurements. The use of subjective inspections as ground truth is disputable, but at this moment it is the only applicable dataset available.

Spatial accuracy of texture measurements

The spatial accuracy of the texture measurements was sufficient for comparison with geodata with a lower spatial resolution. The levering traffic predictor was very local, and therefore very depending on the accuracy of texture measurements. The spatial accuracy with respect to locations with levering traffic was shown to be insufficient.

The accuracy could be increased by less crude ways of georeferencing. For example, it is expected that the (RTK-)GPS tracks of these trucks is available. Whereby the measurements could be referenced with higher accuracy by interpolation of time and location. Or the measurement tracks could be scaled to fit the mile post distances.

Use of multiple laser measurement transects

Due to computational limitations and time constraints, only the right transect of the three laser measurement transects has been used. This concession has implications in two ways.

1. If spots of ravelling occurred at the middle or left side of the road, the measured roughness derivative of the right transect is incorrect relative to the inspected severity class. This causes noise in the results, increasing the deviation of values in a class. Which in turn causes more overlap between classes. Whereby the roughness thresholds of minority classes are higher than expected. This can be avoided by balancing the input dataset which effectively eliminates minority classes, as is done in paragraph 4.3.1. But this had other disadvantages.
2. The extent class is based on the CROW thresholds, as shown in table 4.4. Now, every road segment is divided into 10 decisegments, therefore representing 10% of a road segment. If multiple transects are considered, a road segment can be divided into 30 smaller segments; 10 in lengthwise and 3 in transverse direction. This causes the extent class to be more precise as well.

The decisegment approach

The previous point of discussion stated that the roughness thresholds are relatively high. This effect is amplified by concessions made by the decisegment approach. For every road segment, the highest roughness value of the decisegments is compared with the severity class. The highest value per decisegment will always be higher than the average value of decisegments. This is probably the main reason why there are way fewer EoL segments according to texture measurements than there are according to the visual inspections. As the goal was to find an applicable method, optimizing this model was out of the scope.

If a patch of ravelling occurs at the boundary of two decisegments, its effect is averaged and therefore may be misclassified. This can be averted by smaller sub-segments than the decisegments of 10m[1]. Or, as proposed by M. Nagelhout, using a moving window. This is especially effective if higher accuracy inspection data is available. This could lead to bypassing the decisegment approach in its entirety.

Use of AUC as measure for model performance

Lobo et al. (2008) criticize the use of the AUC statistic as a mean to assess model performance. By using the AUC statistic, models are assessed at confidence levels where such model would rarely operate. It also weights omission and commission errors equally. At this point, it is hard to foresee what the needed confidence level for a future use case would be, or whether omission or commission is deemed to be worse.

Precision of climate scenarios

One of the major conclusions of this research are the rules-of-thumb for *DGD* longevity in specific environmental conditions. These were based on the results of different climatic scenarios. What is concluded based on the test- and train division is that the overall error of the model is relatively high. One major shortcoming of the current methodology based on a RandomForest model, is the lack of prediction intervals. Which limits the ability of assessing the accuracy of these rules-of-thumb. An improvement would be to find the prediction intervals by a quantile regression forest[24]. This will effectively save all responses of every tree in the RandomForest, and is able to calculate the 95% intervals per prediction accordingly.

[24] https://CRAN.R-project.org/package=quantregForest

8 Recommendations

Recommendations after this research can be directed to the principal, in this case the province of Gelderland. Or more broadly to anyone who bases their research on the conclusions made in this research. Therefore, the recommendations are split up accordingly into two sections. Which does not mean that the province of Gelderland should not be the one to initiate further research.

8.1 Recommendations towards the province of Gelderland

For further research on this subject, the following advice are directed towards the province of Gelderland.

State roughness thresholds for longevity warranty

Currently, warranties demanded for road longevity are expressed in CROW inspection classes. These inspections are deemed to be subjective. This research shows that the use of texture measurements to quantify ravelling has the potential to be more accurate than the subjectivity of inspections. It is therefore recommended to do further research to roughness thresholds regarding ravelling.

The disadvantage of stating roughness thresholds as a warranty would be that every pavement type has a different texture, and therefore shows different roughness values. Every type of pavement would, therefore, need its own roughness intervention thresholds. As three transects are measured and the middle transect is often undamaged relative to the side transects, a maximum difference in roughness could be stated as well.

It is advised to either use the Rk or the MPD parameter to express such roughnesses. Rk has shown to have better-distinguishing capabilities, whereas the MPD parameter is easier to integrate while having similar results. The MPD parameter doesn't strictly need filtering and is more accepted in the field of asphalt texture measurements.

Generate shortlists of road trajectories expected to reach EoL to intensify inspections

At this stage roads are inspected every other year. It is advised to use the EoL prediction model of this research to generate shortlists of road trajectories to be inspected more frequently. For example, one could generate a shortlist of segments which have a 10% chance of reaching EoL in the next year. This shortlist could then be inspected more regularly, in turn preventing emergency repairs and providing time to apply rejuvenation cures. Segments having reached EoL will still be missed as the model isn't watertight, for example in the case of incidental damages, but it will decrease costs in the long run.

Make use of data sciences with external parties

The province of Gelderland is very rich in its available (geo)data. It is up-to-date, well documented and a majority is publicly available. During the execution of this research, there were a lot of use cases found for such datasets. The personnel of the Province is open to sharing their in-depth knowledge, and external parties such as universities can provide the needed analytical skills.

8.2 Recommendations for further research

The following recommendations are aimed towards anyone who will base part of their research upon the conclusions made in this research.

Focus on ravelling data quality

One of the disadvantages of the method used in this research is the way measured roughnesses were correlated to segments with a known severity class, referenced as 'the decisegment approach'. As discussed in the previous chapter, a higher spatial resolution of visual inspection severity results could circumvent this approach. The common CROW standard of severity and extent classes is not spatially accurate enough.

The CROW (2011a) also states standards for detailed road inspections. After the interview with T. Jansen, executing detailed inspections instead of global inspections is not very involved in the current inspection workflow. In the case of detailed road inspections, the location of patches of ravelling is sketched and the severity class is given. Which would directly circumvent 'the decisegment approach'.

Making lack of spatial accuracy of texture measurements less prevalent

An alternative to using results of detailed road inspections, the overall lack of spatial accuracy can also be solved by only correlating bigger patches of ravelling. In this research the decisegment with 1) the single worst roughness measurement and 2) the average of the two worst measurements are correlated with the known ravelling severity class. Alternatively, the number of worst segments to correlate with the severity class can be related to the known extent class. This especially makes sense when the road subsegments are smaller than the decisegments used in this research.

In addition, road segments with a low extent class (e.g. L1, M1 or E1) could be neglected as the chance to miss the patch of ravelling is greater if it is less prevalent. But this is only advisable if there is plenty of source data available.

Redefine predictors

Two predictors are deemed to be imperfect given the available data sources. Due to time constraints, the following possible improvements could not have been implemented;

The DaysOfFrost predictor has shown to have quite some influence on the road longevity when combined with the HoursOfRain predictor. Calculation of the DaysOfFrost predictor was rather crude, as it was based on interpolated temperature measurements at 1m above a grassy plain. The asphalt temperature could be very different. Research done by Qiao, Flintsch, Dawson, and Parry (2013) incorporated a heat balance model to approximate the actual temperature of the asphalt. Which included air temperature, solar radiation, radiation of the asphalt surface and wind speed.

The second predictor with room for improvement is the levering traffic predictor. In this research, this predictor was set up by given buffer distances around points of crossings and lines of bus stops (see also table 5.2). No selection has been applied based on the amount of traffic using these crossings. Minor crossings, even of dead-ending streets, have been included as well. While driveways of agricultural or transportation businesses could have much more impact. It is advised to exclude minor crossings and to include driveways of major businesses.

Lifetime optimization; Environmentally optimized mixtures

As described in the problem definition, the conclusions of this research could be used for lifetime optimization. Especially the rules of thumb for *DGD* longevity in specific environmental conditions. It can be beneficial to optimize the overall lifetime of a road trajectory by developing and applying *DGD* mixtures that are optimized for the local environmental conditions. Chemical additives may be effective in making the pavement more resistive to tannic acids *[looizuren]* from trees. Or adding more bitumen may be effective in making the pavement more resistive to the effects of rain and frost. Further research should show whether it is financially profitable to apply such environmentally optimized mixtures.

Lifetime optimization; Optimize homogeneity of roughness

Another approach in lifetime optimization could be to reject a road based on local ravelling homogeneity instead of a fixed roughness threshold. This has been proposed by T. Wahlman, P. Ekdahl[25] and M. Nagelhout[26]. A stretch of road is often rejected based on the worst patch of ravelling, which is often local. This homogeneity can be defined in the transverse direction, i.e. the difference in measurements by the middle laser and side lasers. Or the homogeneity can be defined in the longitudinal direction, as the spread in measuring results per road trajectory. Or a combination of these two.

Taking homogeneity as a factor also greatly reduces the problems of having different types of *DGD* pavement, states of rejuvenating cures, and application circumstances (local differences in binder content) which can alter the texture properties. And decreases the need of stating different thresholds for every type of pavement, as foreseen in the recommendation of 'State roughness thresholds for longevity warranty'. In case this method is used to replace inspections, it will be very hard to define a threshold on homogeneity which is both universal to every pavement type and clear to understand.

[25] Presentation 'Estimation of stone-loss on network condition surveys by use of multiple texture lasers', T. Wahlman, P. Ekdahl. See also note 7.

[26] After interview M. Nagelhout, summary can be found in paragraph 5.1.2. Transcript available in supplement 3

9 Appendices

9.1 Overview of digital supplements

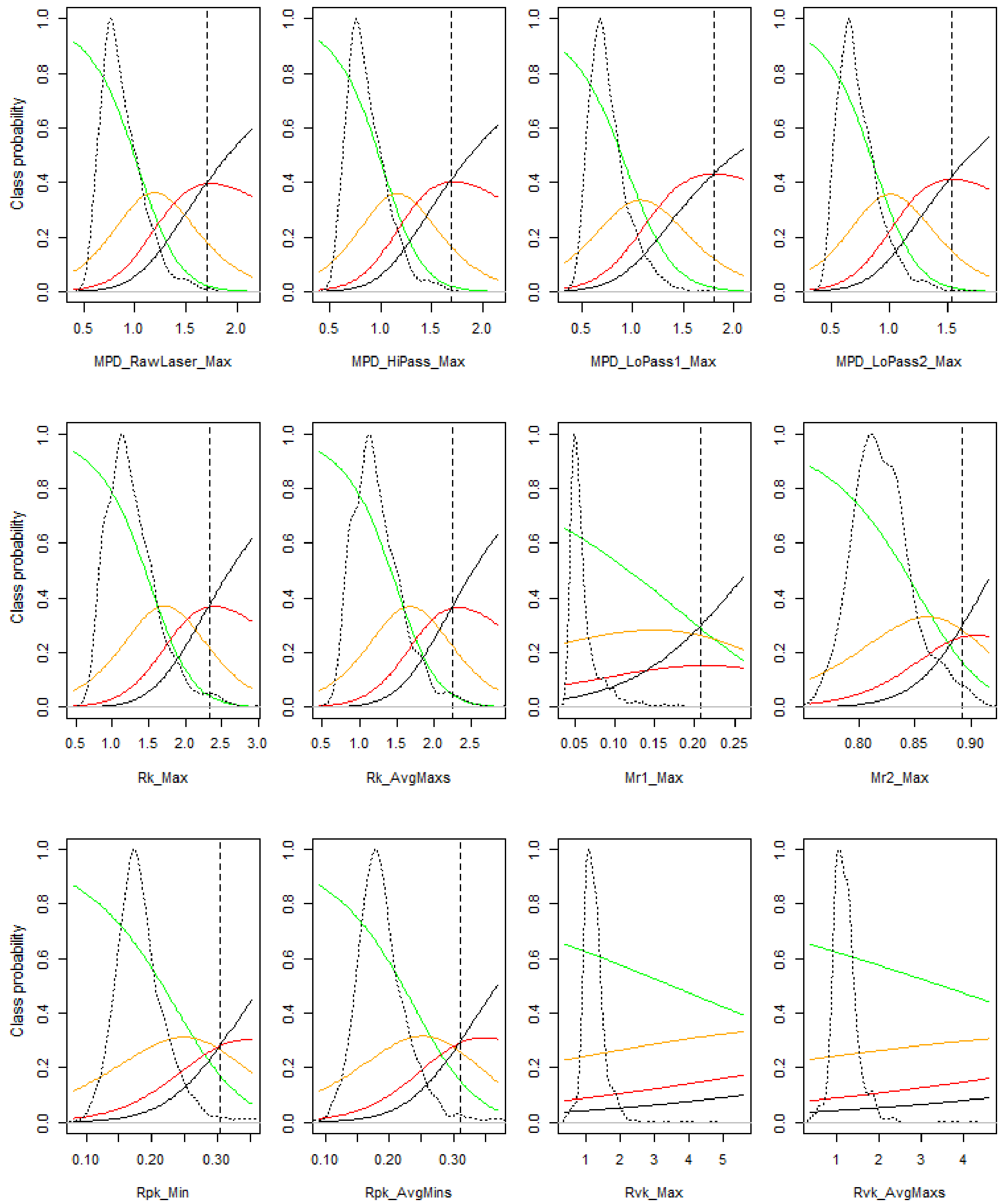
Due to their size or format, these supplements are not printed. Instead, they are provided digitally.

1. Main report in digital format
2. Report supplements in digital format
    1. Approved research proposal
    2. Midterm presentation
    3. Approved minutes of interviews with professionals
    4. Dutch executive summary
    5. Dutch final presentation
    6. Dutch news article; as internally shared at the Province
    7. Final presentation
3. Source geodata
4. Source texture measurements
5. Models of research part 1; Texture to ravelling class model
6. Models of research part 2; Environmental factors
7. Project data; as used for conclusions

9.2 Multinomial logistic regression models

Generated from model f2.5; SeverityModel

	No ravelling
	Ravelling severity class L; Light ravelling
	Ravelling severity class M; Moderate ravelling, EoL if extent >50% (M3)
	Ravelling severity class E; Severe ravelling, EoL no matter the extent
	Roughness derivative density
	Class M-E roughness derivative threshold value



9.3 Misclassification matrices of severity class

Three machine learning algorithms were tried out on each of the 92 ravelling derivative alternatives. A subset of misclassification matrices is shown.

MPD_HiPass_Max

Mean Profile Depth, after high-pass filtering, maximum value of road segment per decisegment

Naive Bayes

Predicted \ Inspected	0	L	M	E	Sum
0	562	226	67	23	878
L	29	26	26	9	90
M	2	3	2	14	21
E	3	2	0	6	11
Sum	596	257	95	52	1000

Nearest mean

Predicted \ Inspected	0	L	M	E	Sum
0	417	117	19	8	561
L	92	57	24	3	176
M	57	53	27	12	149
E	30	30	25	29	114
Sum	596	257	95	52	1000

Multinom

Predicted \ Inspected	0	L	M	E	Sum
0	565	226	69	23	883
L	27	26	25	10	88
M	3	4	1	18	26
E	1	1	0	1	3
Sum	596	257	95	52	1000

Balanced Multinom

Predicted \ Inspected	0	L	M	E	Sum
0	419	118	19	8	564
L	87	49	20	3	159
M	51	55	24	11	141
E	39	35	32	30	136
Sum	596	257	95	52	1000

Rk_Max

Core roughness depth, after HiPass filtering, maximum value road segment per decisegment

Naive Bayes

Predicted \ Inspected	0	L	M	E	Sum
0	557	212	56	19	844
L	33	29	28	8	98
M	5	12	10	8	35
E	1	4	1	17	23
Sum	596	257	95	52	1000

Nearest mean

Predicted \ Inspected	0	L	M	E	Sum
0	413	109	16	8	546
L	109	64	25	3	201
M	46	50	25	11	132
E	28	34	29	30	121
Sum	596	257	95	52	1000

Multinom

Predicted \ Inspected	0	L	M	E	Sum
0	562	217	61	21	861
L	32	35	29	7	103
M	2	1	4	11	18
E	0	4	1	13	18
Sum	596	257	95	52	1000

Balanced Multinom

Predicted \ Inspected	0	L	M	E	Sum
0	413	109	17	8	547
L	106	59	23	2	190
M	47	55	25	12	139
E	30	34	30	30	124
Sum	596	257	95	52	1000

9.4 Properties of environmental predictors at decisegment level

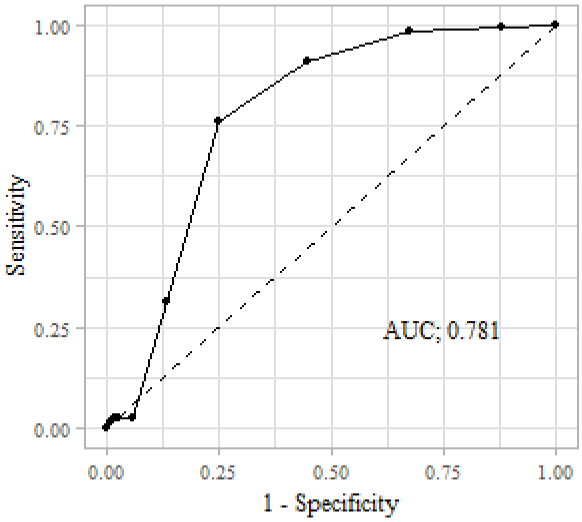
Decisegments; Generated by Rstudio model g5f_ENV_Histogram

Factor	Unit	5th percentile	Mean	95th percentile
Age	Years	2	5,36	9
Tree cover	Percentage	0	6,85	70
Mean days of frost	Days/year	41	55	64
Mean hours of rain	Hours/year	125	141	160
Heavy traffic	MVT/etm[17]	124.100	409.262	977.835
Light traffic	MVT/etm[17]	1.320.570	3.968.151	10.545.945
Leverage area	Percentage	0	10	100

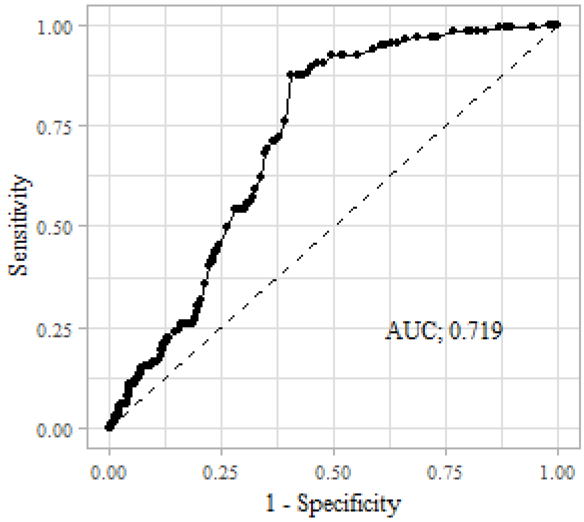
		Class true	Class false	
Has parallel road	Boolean	15%	85%	
Tree cover	Boolean	11%	89%	
Leverage area	Boolean	11%	89%	

9.5 ROC curves of linear regression predicting inspection results

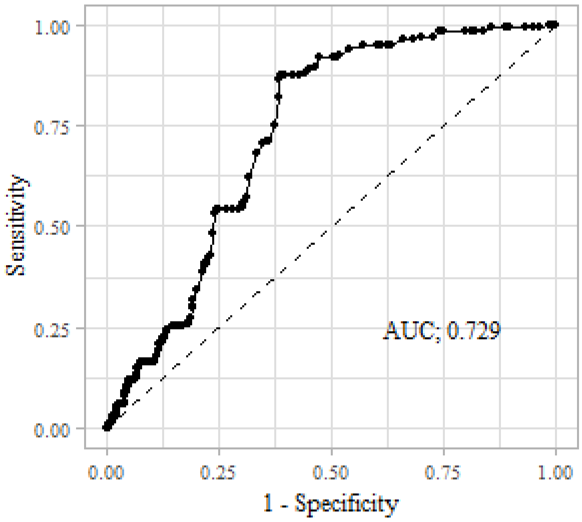
A; Only age (0-hypothesis)



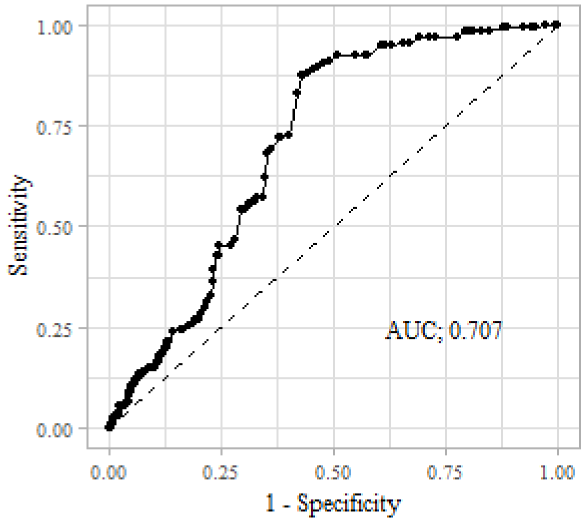
B; All but HasParRoad



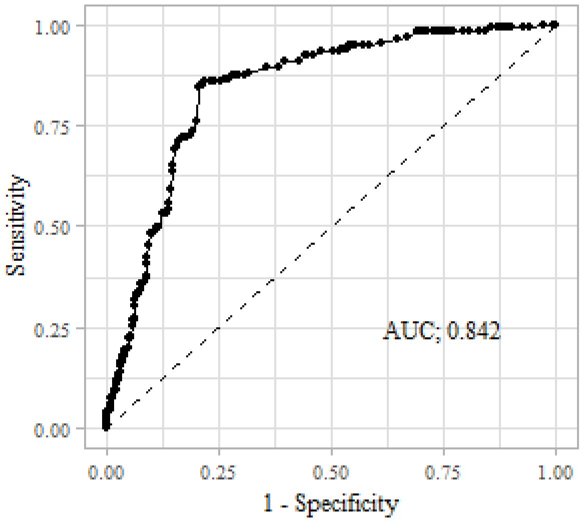
C; All but HasParRoad and HeavyTraffic



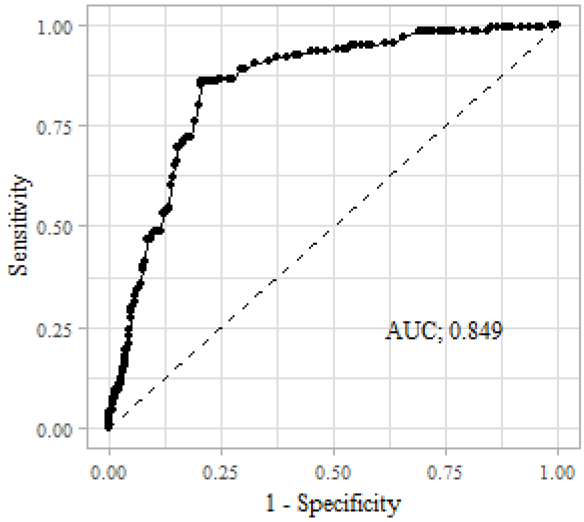
D; All but HasParRoad and traffic



E; Weather as function of age

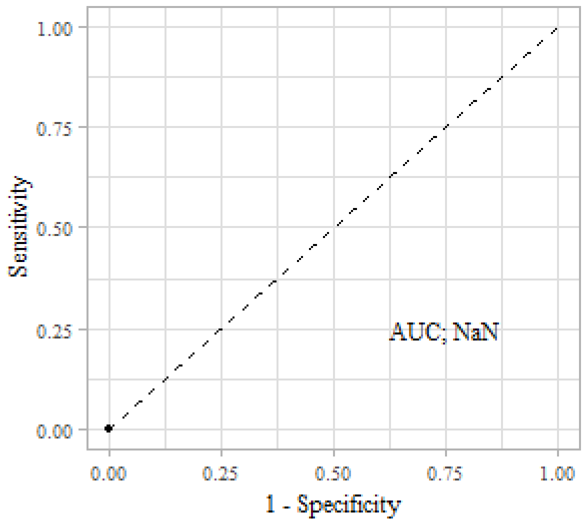


F; Traffic and weather as function of age

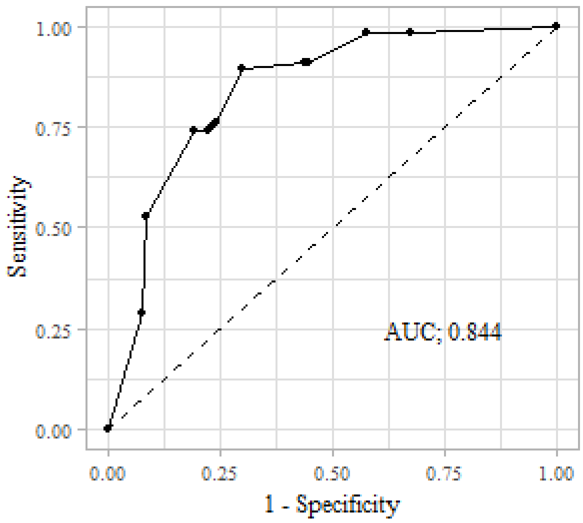


9.6 ROC curves of RandomForest predicting inspection results

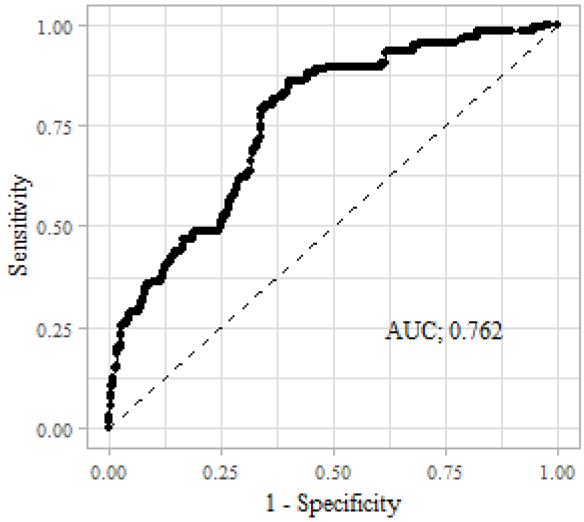
A; Only age



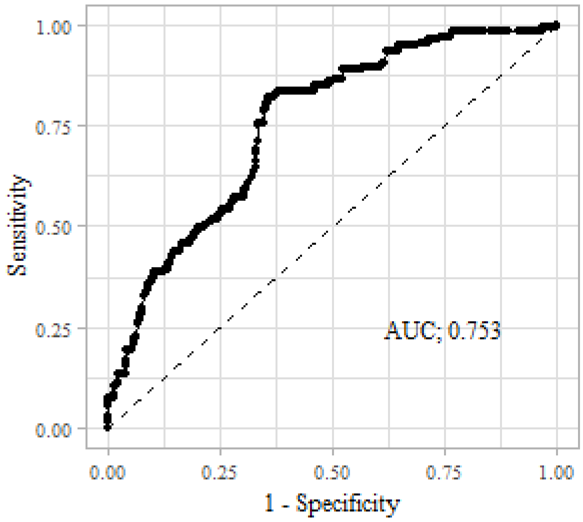
B; Only age at stratified sampling



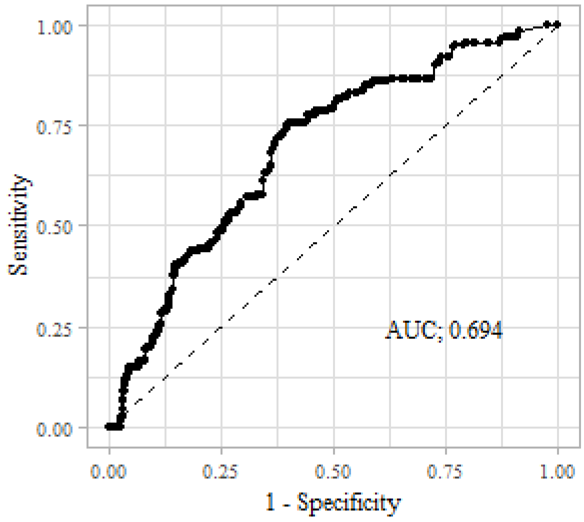
C.2; All predictors, excl. HasParRoad



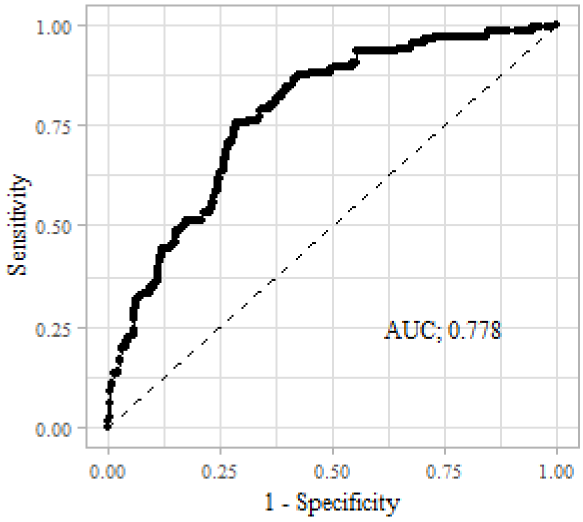
D; Excluding HasParRoad and TreeCover



E; Without stratified sampling



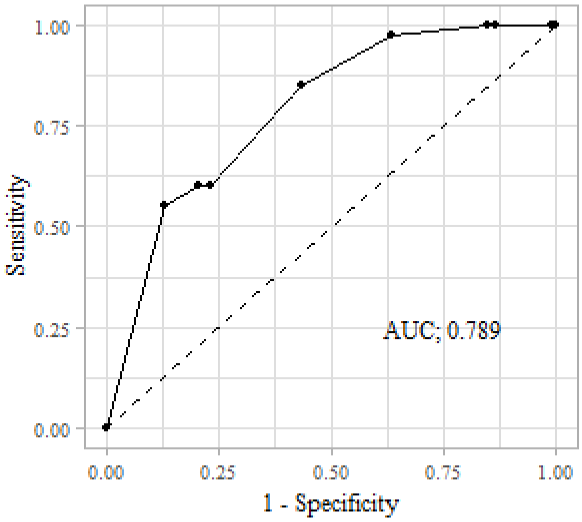
F; Optimizing Mtry -> Mtry=1



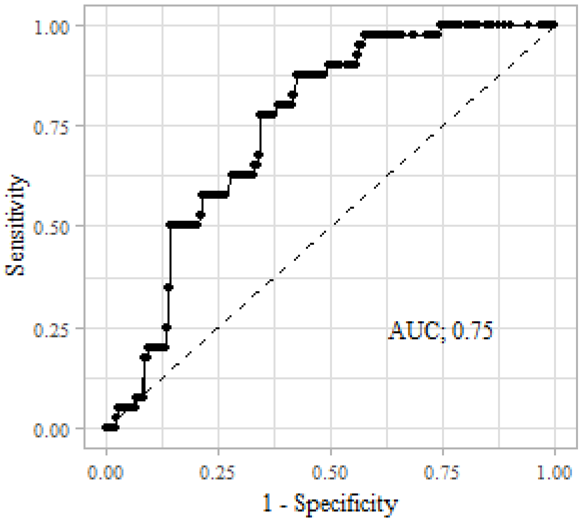
9.7 ROC curves of RandomForest predicting measurement result

ML applied to predict roughness per decisegment, ‘roughness to severity’ model used to vary prediction threshold and ‘decisegment severities to road segment EoL’ model used to predict EoL

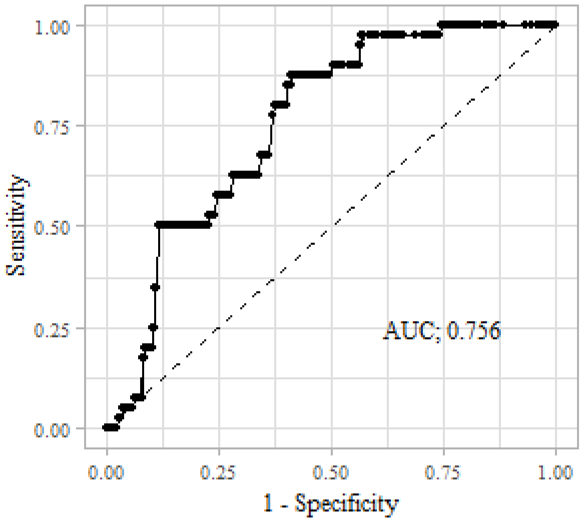
A; Only age



E; Optimized model



G; Optimized model, Ntree=1000



10 References

Conference papers

Bochove, G. v. (2014). *Self Healing Asfalt*. Paper presented at the CROW Infradagen 2014, Ermelo. https://www.crow.nl/bijeenkomsten-en-congressen/crow-infradagen-2014.

Nagelhout, M., Wennink, P. M., & Gerritsen, W. (2004). *Detection of ravelling*. Paper presented at the SURF 2004, Fifth Symposium on Pavement Surface Characteristics, Toronto, Ontario, Canada.

Ng, A. Y., & Jordan, M. I. (2002). *On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes.* Paper presented at the Advances in neural information processing systems

Ooijen, V. W., & Bol, V. D. (2004). *High-Speed Measurement of Raveling on Porous Asphalt*. Paper presented at the SURF 2004, Fifth Symposium on Pavement Surface Characteristics, Toronto, Ontario, Canada.

Woodward, D., Millar, P., & McQuaid, G. (2014). *Use of 3D modelling techniques to better understand road surface textures*. Paper presented at the International Safer Roads Conference, 4th, 2014, Cheltenham, United Kingdom.

Laws, norms and standards

Besluit geluidhinder, (2006)

CROW. (2011a). *Handboek visuele inspectie 2011*. Ede: CROW.Personal library

CROW. (2011b). *Handleiding globale visuele inspectie 2011*. Ede: CROW.Personal library

CROW. (2011c). *Wegbeheer 2011*.Personal library, #Appendix II; standard intervention classes

Handleiding toelatingsprocedure en ringonderzoek High Speed Road Profiler (2017)

NEN. (1998). Geometrical Product Specifications (GPS) — Surface texture: Profile method; Surfaces having stratified functional properties *Part 2: Height characterization using the linear material ratio curve* (Vol. NEN-EN-ISO 13565-2:1998 en): Internationale Organisatie voor Standaardisatie [ISO]

NEN. (2002). Karakterisering van de oppervlaktetextuur van wegdekken met behulp van textuurprofielen - Deel 3: Specificatie en classificatie van profielmeters (Vol. NEN-ISO 13473-3:2002 en): ISO

NEN. (2004). Karakterisering van de textuur van bestratingen met oppervlakprofielen - Deel 1: Gemiddelde profieldiepte (Vol. ISO 13473-1:1997 en)

Online databases

CBS. (2017a). Length of roads; Per province. http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70806ned&D1=8&D2=0,5-16&D3=l&HDR=G2&STB=T,G1&VW=T.

CBS. (2017b). Provinciebegrotingen; baten en lasten per functie 2005-2016. http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70932NED&D1=0,16&D2=1&D3=0,10&D4=l&HDR=T,G1,G3&STB=G2&VW=T.

Provincie Gelderland. (2017b). Geoserver provincie Gelderland. geoserver.prvgld.nl/geoserver.

Government policies

Provincie Gelderland. (2008). *Actieplan geluid Gelderland 2008-2012*.

Provincie Gelderland. (2017a). *Begroting 2017; Onderhoud kapitaalgoederen*.

Provincie Gelderland. (2018). *Begroting 2018; Onderhoud kapitaalgoederen*.

Peer-reviewd reports

Arnberg, P. W., Burke, M. W., Magnusson, G., Oberholtzer, R., Råhs, K., & Sjögren, L. (1991). *The Laser RST: Current Status*. Retrieved from

Hagos, E. T. (2008). *De Invloed van Veroudering op de Eigenschappen van het Bindmiddel in Zeer Open Asfaltbeton*. Retrieved from

Huurman, M. (2008). *Lifetime Optimisation Tool, LOT, Main Report*. Retrieved from

Izevbekhai, B. I., Watson, M., Clyne, T., & Wong, M. S. (2014). *Hot Mix Asphalt Surface Characteristics*. Retrieved from #Goes into porous and non-porous asphalt surface characteristics in combination with profilometry

Kersten, H. T. M., Bobbink, E., & Reinink, H. F. (2014). *SMA-NL 8G+; Steenmastiekasfalt akoestisch geoptimaliseerd*. Retrieved from

Rasmussen, R. O., Sohaney, R., Wiegand, P., & Harrington, D. (2011). *Measuring and Analyzing Pavement Texture*. Retrieved from

Rijkswaterstaat. (2007). *Advies dunne geluidsreducerende deklagen op niet-autosnelwegen*. Retrieved from

Tsai, Y., & Wang, Z. (2015). *Development of an Asphalt Pavement Raveling Detection Algorithm Using Emerging 3D Laser Technology and Macrotexture Analysis*. Retrieved from

van den Bol-de Jong, M. E., Bouman, F. G. M., van Ooijen, W. H., & Verra, N. (2003). *Kalibratie ARAN-lasermetingen met betrekking tot rafeling van ZOAB*. Retrieved from #Research setting up a predictive model on ZOAB pavement based on visual inspections of 18 mile post segments

van Reisen, F., Erkens, S., v.d. Ven, M., Voskuilen, J., & Hofman, R. (2008). *Verbetering levensduur tweelaags ZOAB* (DVS-2008-020). Retrieved from

Journal articles

CROW. (2016). *Besprekingsverslag 15e vergadering platform wegmetingen.* Paper presented at the Platform wegmetingen, kennisplatform CROW

Henning, T. F. P., & Roux, D. C. (2012). A probabilistic approach for modelling deterioration of asphalt surfaces. *Journal of the South African Institution of Civil Engineering, 54*(2), 36-44.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography, 17*(2), 145-151.

Mo, L., Huurman, M., Wu, S., & Molenaar, A. A. A. (2009). Ravelling investigation of porous asphalt concrete based on fatigue characteristics of bitumen–stone adhesion and mortar. *Materials & Design, 30*(1), 170-179. doi:10.1016/j.matdes.2008.04.031.

Opara, K. R., Skakuj, M., & Stöckner, M. (2016). Factors affecting raveling of motorway pavements—A field experiment with new additives to the deicing brine. *Construction and Building Materials, 113*, 174-187. doi:10.1016/j.conbuildmat.2016.03.039.

Park, S. H., Goo, J. M., & Jo, C.-H. (2004). Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean Journal of Radiology, 5*(1), 11-18.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C (2nd ed.): the art of scientific computing*: Cambridge University Press

Qiao, Y., Flintsch, G., Dawson, A., & Parry, T. (2013). Examining effects of climatic factors on flexible pavement performance and service life. *Transportation Research Record: Journal of the Transportation Research Board*(2349), 100-107. #Several possible environmental factors described

Ueckermann, A., Wang, D., Oeser, M., & Steinauer, B. (2015). Calculation of skid resistance from texture measurements. *Journal of Traffic and Transportation Engineering (English Edition), 2*(1), 3-16. doi:10.1016/j.jtte.2015.01.001.

Zhang, Y., & Leng, Z. (2017). Quantification of bituminous mortar ageing and its application in ravelling evaluation of porous asphalt wearing courses. *Materials & Design, 119*, 1-11. doi:10.1016/j.matdes.2017.01.052.

Books

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer

Werkgroep Stille wegdekken. (2010). *Stille wegdekken*. Ede: CROW