A MICRO- AND MACRO ANALYSIS OF RAILWAY EMBANKMENT FAILURES

Master thesis

VERSION
Final report

DATE
28 - 06 - 2024

AUTHOR
D Wilms (Daan)

ProRail
UNIVERSITY OF TWENTE.

Page left blank intentionally

A micro and macro analysis of railway embankment failures

by

Daan Wilms

to obtain the degree of Master of Science
at the University of Twente

Institution:	University of Twente
UT Supervisor:	Dr. Ir. O.A.L. Eikenbroek (Oskar)
Head graduation committee:	Prof. Dr. Ir. E.C. Van Berkum (Eric)
Location:	Faculty of Engineering Technology, Enschede
External parties:	ProRail B.V.
Daily external supervisor:	Drs. Ingram, J. (Jasper)
Programme:	Integrated Civil Engineering - Smart Cities
Document version:	Final report
Project duration:	February 2024 – June 2024

Cover image: railway embankment slip circle visualization (ProRail, 2024)

ProRail

UNIVERSITY OF TWENTE.

Preface

Before you lies the Master Thesis "*A micro and macro analysis of railway embankment failures*". This thesis project is carried out to graduate for the degree of Master of Science in Civil Engineering and Management, with the profile Smart Cities from the Integrated Civil Engineering Systems track. I have executed this research between February and June 2024 at ProRail.

I want to express my gratitude and appreciation to my daily supervisors from ProRail and the University of Twente: Jasper Ingram and Oskar Eikenbroek. During the execution of this research project, several changes in approach and scope were necessary to obtain meaningful results. Without the flexibility and expertise of my supervisors, I could have never finished this thesis. I had the opportunity to meet with these enthusiastic and inspiring individuals more often than initially agreed in the pre-thesis period. I have always left these meetings motivated to continue with the research and investigate the discussed possibilities.

Within the organization of ProRail, I was allowed to work within two groups: the railway embankment project group and the Department of Asset Strategy. I felt like a part of both teams instantly and I very much appreciate the tips and interest that I received from my colleagues.

Lastly, I want to thank my friends and family for their continuous support. Especially my study friends, who made time to brainstorm with me when I got stuck and even proof-read this thesis.

I hope that you enjoy reading my work!

Daan Wilms

Enschede, June 2024

Summary

Purpose – In a changing climate where the transport of individuals and goods is expected to increase, a robust and reliable railway network is essential. In the Netherlands, the railway network is situated on embankments that were mostly built before 1920. With an increase in railway embankment failures, the implementation of maintenance strategies that pre-emptively address issues before they occur. To switch from corrective maintenance to predictive maintenance, an understanding of the causes of railway embankment failures is essential. The purpose of this thesis is therefore to bridge existing knowledge gaps by analysing impact factors that contribute to the occurrence of railway embankment failures on a micro and macro scale. The goal is to contribute to a better understanding of the causes of railway embankment failures to facilitate capacity increase and boost the transition from corrective to predictive maintenance.

Methodology – To find the contribution of different factors to the occurrence of railway embankment failures, four research questions are tackled. The first question aimed to create a short list of impact factors that are potentially influential regarding the failures and can be considered for this research. The second question consisted of data acquisition, pre-processing and a statistical analysis of the available data. For variables, the IQR method and an adaptive Tukey-fence method have been used to set up extremity boundaries. The adaptive Tukey-fence method identifies boundaries that enclose 98% of the values of the variable impact factors. Third, a micro-analysis of specific railway embankment failures is performed. The acquired and pre-processed data is linked to the spatial and temporal component of specific accident reports in the ProRail area Noord-West. Utilizing the boundaries and visualizations the possible relationship between specific embankment failures and the impact factors is investigated. Following the micro-analysis, the impact factors and embankment failures are mapped on a macro scale for the last research question. After visualizing the impact factors and embankment failures, a database is exported that can be subjected to various machine learning methods varying in functionality. The performance of these models is determined based on the R-squared value. For each model, the most impactful factors are highlighted. To find the overall importance of different factors a composite score is obtained by multiplying the importance score by a value based on the performance of the machine learning model. This composite score indicates the importance of an impact factor in explaining the variance in the occurrence of railway embankment failures in the Netherlands.

Results – From a literature review and expert judgement, four areas are investigated to create a long list of impact factors: soil, climate, train and tracks. From each of these areas, at least 1 factor is considered for the analysis. Based on data availability, expert judgement and expected feature correlation, seven impact factors make up the shortlist: precipitation, drought periods, soil saturation, soil strength, train weight, track usage intensity and fractal value D1. When analysed, all variable factors had a right-skewed distribution. For all of these factors, a higher value is considered more detrimental to the railway embankment. Then, the accident reports of the last decades were filtered to find the specific reports that were registered in the case study area for the microanalysis and related to embankment failures. The spatial and temporal components have been linked to the impact factors with the extremity boundaries to visualize whether the occurrence of these failures can be related to the impact factors. From the micro-analysis, no evident correlations have been found between the impact factors and specific occurrences of railway embankment failures. The macro analysis is conducted by first creating a map with the railway tracks in the Netherlands and the occurrences of embankment failures. From the visual representations of the impact factors on a large scale with the embankment failures, no direct relationships are found. The table with impact factors has been subjected to 9 machine learning models. The best-performing model was Gradient Boosting

Regressor, which is able to explain almost 50% of the variance in railway embankment failures in the Netherlands based on the impact factors. The most influential factors for this model and the other best-performing models are soil composition, rainfall and the yearly increase in fractal value D1.

Discussion – This thesis had some limitations and decisions made that potentially influenced the outcome of the analyses. First, the long list of impact factors is based on literature and experts. Actual railway embankment failures have not been inspected. The soil strength and saturation have been estimated based on other datasets. Some data constraints also reduce the acceptability of the outcome of this research. The average daily tonnes per year have been obtained as well as the total number of trains on a track section. This is not as specific as desired for conducting the analyses. With regard to the extremity calculations, the IQR method is used for not-normally distributed datasets, and values are considered extreme when they are in the top 2% of the dataset.

Conclusion – from the analyses performed, it is concluded that the impact of different factors to the occurrence of railway embankment failures can be explained modestly based on the analyses performed in this research. On a micro scale, higher resolution data on more impact factors is considered necessary to predict the time and location of an embankment failure. On the macro scale, the most influential factors for the explanation of the variance in railway embankment failures are soil type, precipitation and fractal analysis value D1. Further research should focus on increasing the number of impact factors to predict railway embankment failures and increase the resolution of the data on these factors.

Samenvatting

Intentie – in een veranderend klimaat waar het vervoer van personen en goederen naar verwachting toeneemt, is een robuust en betrouwbaar spoorweg netwerk essentieel. In Nederland ligt het spoor netwerk op spoordijken, ook wel baanlichamen genoemd. Het grootste deel van deze baanlichamen is gebouwd voor 1920. Het verzakken van deze baanlichamen gebeurt steeds vaker. Het is daarom noodzakelijk om de onderhoudsstrategie van correctief naar voorspellend te schakelen. Zo kan onderhoud worden uitgevoerd voordat de baanlichaam verzakkingen voorkomt. Voor deze transitie is kennis van het verzakken van baanlichamen essentieel. Het doel van dit onderzoek is daarom om bestaande kennisleemtes te overbruggen door verschillende impact factoren te analyseren die bij kunnen dragen aan het ontstaan van baanlichaam verzakkingen op micro- en macro niveau. De intentie is om een bijdrage te leveren aan het begrijpen van de oorzaken van baanlichaam verzakkingen om de capaciteitsgroei en de overgang naar voorspellend onderhoud te faciliteren.

Methode – door middel van vier onderzoeksvragen is de bijdrage van verschillende impact factoren aan het ontstaan van baanlichaam verzakkingen geanalyseerd. Als eerst is een lijst met impact factoren opgesteld, waarna met experts en literatuur een shortlist is gedestilleerd met factoren die mee zijn genomen in het onderzoek. Daarna is data verkregen en voorbewerkt om vraag twee uit te voeren. Hier is een statistische analyse gedaan van de variabele impact factoren, en met de IQR-methode en een adaptieve Tukey fence-methode zijn grenswaarden bepaald die de extremiteit van variabelen aangeven. In het geval van Tukey fence is de grenswaarde gevonden die 98% van de waarden omvat. Als derde is een micro-analyse uitgevoerd met specifieke baanlichaam verzakkingen. De ruimtelijke en tijdelijke component van onregelmatigheidsrapporten in het ProRail gebied Noord-West zijn gekoppeld aan de data van de impact factoren. Met visualisaties en de verkregen grenswaarden wordt de mogelijke relatie tussen baanlichaam verzakkingen en de impact factoren verkend. Als laatste worden de impact factoren en baanlichaam verzakkingen op een macro schaal gekoppeld. Na het visualiseren op een kaart is een database geëxporteerd waar verschillende machine-learning methoden op zijn getest. De prestatie van deze modellen is bepaald gebaseerd op de R-kwadraat waarde. Voor elk model worden de meest invloedrijke factoren aangegeven. Om de algemene impact van de verschillende factoren te bepalen, is een composietscore berekend door de importance score te vermenigvuldigen met een waarde gebaseerd op de prestatie van het machine learning-model.

Resultaten – uit een literatuuronderzoek en het oordeel van experts zijn vier gebieden bepaald voor het vinden van impact factoren: ondergrond, klimaat, trein en rails. Van elk van deze gebieden is ten minste één factor meegenomen voor de analyses. Gebaseerd op data beschikbaarheid, de mening van experts en verwachte correlaties tussen de factoren zijn zeven impactfactoren in de shortlist meegenomen: neerslag, droogte perioden, grondverzadiging, ondergrond sterkte, treingewicht, gebruiksintensiteit van het spoor en fractalwaarde D1. Alle variabele factoren hebben een rechts-scheve verdeling. Voor alle impact factoren is een hogere waarde als meer schadelijk aangenomen voor het baanlichaam. Daarna zijn rapporten van onregelmatigheden uit de laatste decennia gefilterd om specifieke ongevallen te vinden in het studiegebied voor de micro analyse die te maken hebben met baanlichaam verzakkingen. Vervolgens zijn de extremiteitsgrenzen gekoppeld aan de waarden van de impact factoren om mogelijke relaties te kunnen vinden. Vanuit de micro analyse is geen directe evidente relatie gevonden tussen impact factoren en baanlichaam verzakkingen. De macro-analyse is uitgevoerd door eerst visualisaties van de impact factoren te maken op een kaart met het spoornet van Nederland en de hoeveelheid verzakkingsmeldingen. Ook hieruit zijn geen directe relaties gevonden. De resulterende tabel is onderworpen aan negen machine learning methoden. De methode met de beste prestatie is Gradient Boosting Regressor, die 50% van de variatie in

baanlichaam verzakkingen kan verklaren aan de hand van de impact factoren. De belangrijkste factoren die uit dit model, en de andere best presterende methoden komen zijn grondtype, neerslag en de jaarlijkse stijging in fractalwaarde D1.

Discussie – er waren een aantal limitaties en keuzes die een invloed kunnen hebben gehad op het resultaat van de analyses. De impact factoren zijn gebaseerd op deskundigen en literatuur, maar niet op het fysiek bekijken van een verzakking. Bodemsterke en grondsaturatie zijn geschat gebaseerd op andere datasets. Sommige andere data beperkingen verlagen tevens de aannemelijkheid van de resultaten van het onderzoek.

Conclusie – vanuit de analyses in dit onderzoek kan geconcludeerd worden dat tot op zekere schaal het ontstaan van baanlichaam verzakkingen verklaart kan worden aan de hand van de zeven impact factoren die zijn meegenomen in het onderzoek. Op micro-schaal is een hogere resolutie nodig voor het voorspellen van tijd en locatie van een baanlichaam verzakking. Op macro-schaal zijn ondergrondtype, neerslag en fractalwaarde D1 de meest invloedrijke factoren voor het verklaren van variatie in baanlichaam verzakkingen in Nederland. Verder onderzoek zou moeten focussen op het vergroten van het aantal impact factoren die gebruikt wordt om baanlichaam verzakkingen te voorspellen en het verhogen van de data resolutie.

Contents

Table of Tables

Table of Figures

1. Introduction

In this first chapter, an introduction to the research topic is presented. First, some background on the topic of embankment failures is provided, after which the problem context of this research is given. Finally, the research dimensions including the objective, main research question and scope of the research are provided.

1.1. Background

On a regular working day in the Netherlands, almost a million people travel by train (Nederlandse Spoorwegen, 2022). By 2030, the amount of individuals and goods per train is expected to increase by 30% and 50% respectively compared to the current situation (ProRail, 2022b). To facilitate this expected increase in demand, the railway infrastructure must be reliable enough to host more, faster and heavier trains.

The current railway infrastructure is not capable of facilitating this expected increase in demand: in the period between November 15th 2022 and November 15th 2023, a total of 2,684 out of 5,298 failures were registered as issues caused by infrastructure and not by third parties or for instance snow (RijdenDeTreinen, 2023). Infrastructure that complies with the safety norms is essential for a functioning rail transport network. Being the asset owner, ProRail maintains the railway infrastructure to make sure trains can keep driving. The operations executed by ProRail are divided into 8 categories that are indicated in Figure 1.

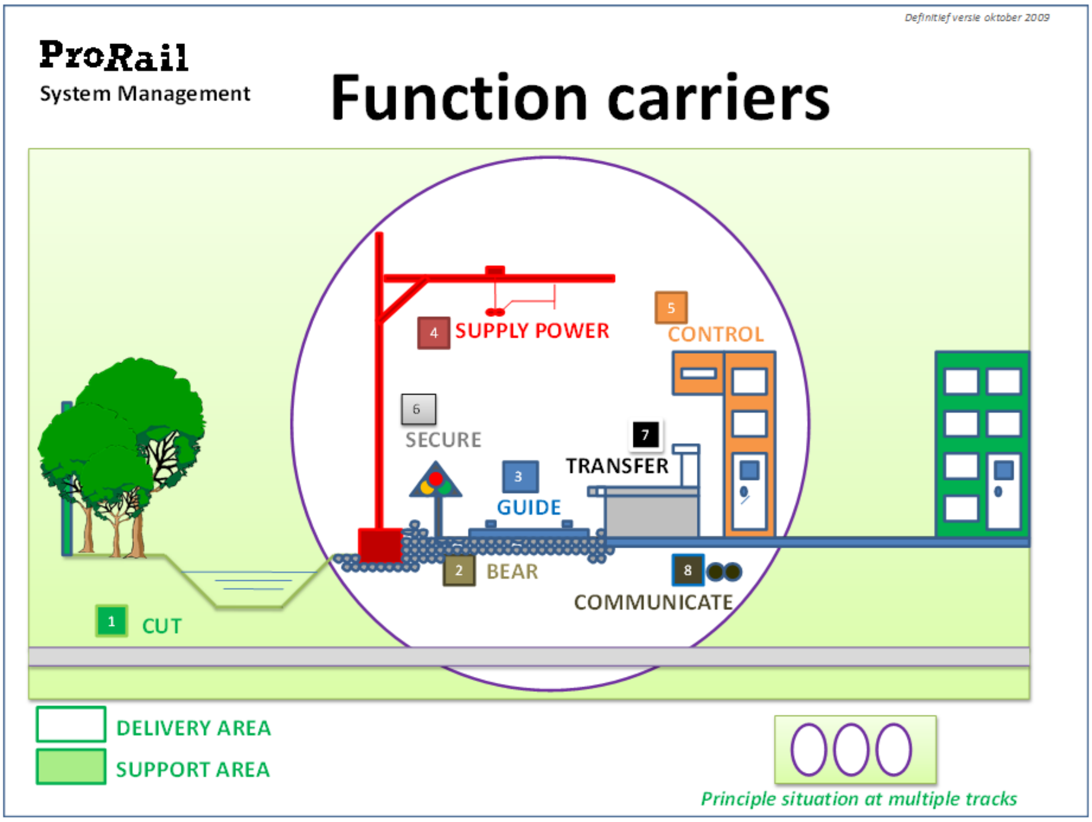


Figure 1: Function carriers of ProRail (Source: (ProRail, 2024)

One specific type of failure that has been given more attention recently is railway embankment failure. In 2022, the Netherlands had a total of 7023 kilometres of railway track length (ProRail, 2022a). The entirety of the Dutch railway network is situated on top of embankments, and as stated by ProRail (2022a) 80% of these embankments were built before 1920. When the soil underneath railway tracks starts sliding, as visualized in Figure 4, hindrance occurs and expensive maintenance activities must be executed quickly. Moreover, lethal accidents such as the major incident in Weesp in 1918 (Figure 2) may happen. This incident is stated to mark the beginning of geotechnical research in the Netherlands. The increasing number of failures in the railway embankments makes the research on this topic more relevant.

Figure 2: Disaster in Weesp, 1918. (SpoorPro, 2018)

The composition of the railway network in the Netherlands is important to understand when considering these embankment failures. Essentially, 5 layers can be distinguished underneath the wheels of the train. The wheels move over rails. These rails are connected to sleepers, usually made of concrete or wood. Figure 3 gives a visual representation of rails, sleepers and their connection. In the figure the next layer is also shown: the ballast bed. The ballast bed makes sure the tracks remain straight, reduces vibrations and aids in rainwater drainage. Underneath the ballast, the railway embankment is situated. Finally, underneath the railway embankment is the surface layer.

Figure 3: Layers underneath train wheels (Verloop, 2019)

The failures in the railway embankment can be categorized into two groups: slip circle failures as indicated in Figure 4 and reduction in bearing capacity. The later happens when one of the first four layers fails. A representation of where the loss of bearing capacity occurs is shown in Figure 5. Slip circle failures have a low occurrence rate with extreme consequences, where the loss of bearing capacity occurs more often. Capacity loss can be addressed more easily with less impact: reduced speed and short-term maintenance are often the solutions for this failure category (ProRail, 2024).

Figure 4: Visual representation of railway embankment slip circles. Source: ProRail

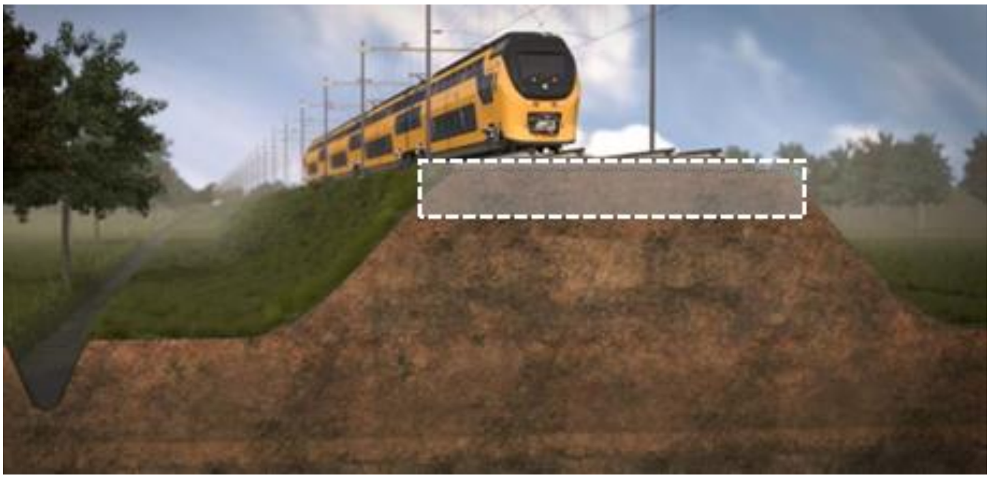


Figure 5: location significant for loss of bearing capacity. Source: ProRail

Currently, a large-scale project in the Netherlands called Landelijke Netwerk Analyse (LNA) is being executed. This collaborative project of ProRail with several engineering firms investigates the strength of the railway embankments and what measures can be taken to ensure their structural integrity

In addition to variables regarding the railway embankment itself, the changing climate with more extremes in drought and precipitation is also expected to impact this type of failure (Lancellotta, 1995). The highest yearly rainfall numbers on 1 day in the region of Zaandam, derived from KNMI (2024) is given in Figure 6. The trendline shows that on average, there is a steady increase in these maxima, and it can be seen that the extremes can increase, such as in 2012. These extremes are possibly related to some recent railway embankment failures, and mitigation or maintenance is expensive and time-consuming, as indicated in the headlines in Figure 7.

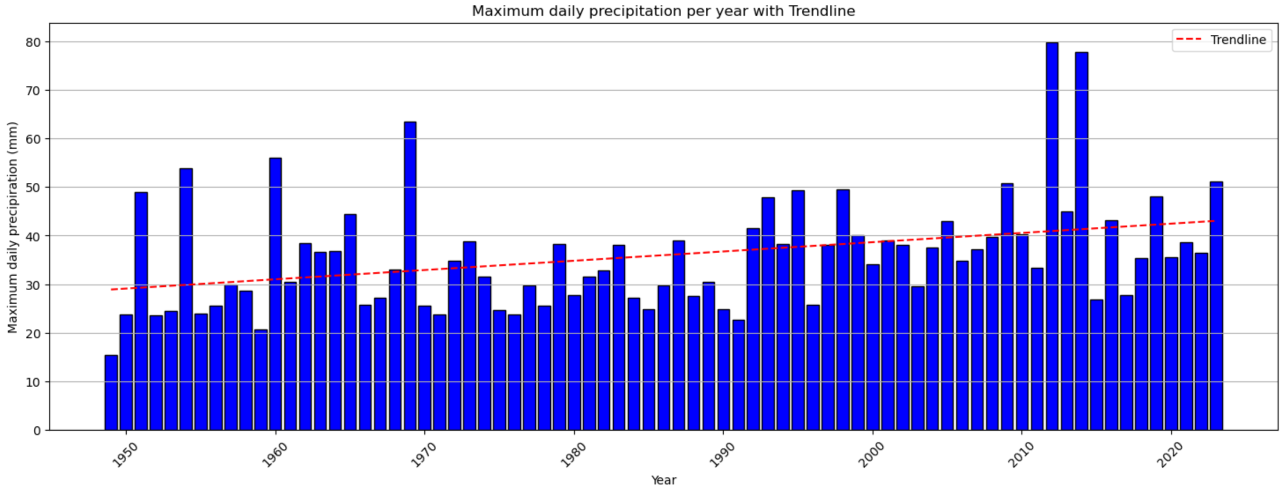


Figure 6: Daily precipitation maxima between 1949 and 2023 including trendline. Source: (KNMI, 2024)

Figure 7: Headlines of embankment failures. Source left: (Lancefield, 2024), source right: (ProRail, 2024)

Current developments in data acquisition and modelling increase the desire to be able to predict where these embankment failures occur. In the ideal scenario, ProRail is able to precisely predict the location and time at which an embankment would fail to be able to plan maintenance activities before failures occur. Data-driven asset management is an upcoming topic which allows asset owners to integrate data into the decision-making process. With the current state-of-the-art, there is still too much uncertainty to integrate this way of asset management on a large scale within the company. The knowledge on embankment failures and the possible additions that computer models and data-driven decisions can have on this topic are hardily investigated.

1.2. Problem context

The railway infrastructure and its use are being monitored to find what segments of the network do not meet the requirements or perform sub-optimally. If a problem occurs that is within the 8 functional carriers of ProRail, corrective maintenance activities are performed. Currently, being able to predict disruptions that cause the network performance to decrease based on the available data is not yet integrated, as the focus is currently on keeping the functionality of the network. Corrective maintenance has been the standard approach for rail maintenance because predictive maintenance is not sufficiently investigated in this sector. The first step for facilitating the shift in maintenance planning is to understand the correlations between the different variables that impact the occurrence of these failures. In the case of railway embankment failures, this investigation is limited to the functional carriers guiding and bearing.

With an embankment failure occurring at a certain place and time, both spatial and temporal factors are expected to be correlated to railway embankment failures. The changing climate, with more temperature and precipitation extremes, is expected to correlate to the occurrence of railway embankment failures. For the asset owner, knowledge of the most influential factors and being able to determine locations that are prone to disruptions enables the planning of maintenance activities before infrastructure failure occurs. This saves time, money and increases safety. In addition, increasing the knowledge on railway embankment failures allows ProRail to make decisions on what locations can withstand changes in capacity-related variables and prioritize maintenance on locations that can not. This way, the demand increase can be facilitated safely, and with less downtime and maintenance costs. Two orders of magnitude are identified for this research. On a micro level, the specific occurrences of railway embankment failures are investigated. These come forward from the specific configuration of impactful factors at the time and location of a failure. Essentially, the micro level explains the chance of an embankment failure based on the configuration of different factors at that time and location. Then, on a macro level, trends in the occurrence of railway embankment failures can be considered, for instance, if a large number of failures occurs near rivers or in the winter. This larger scale considers the impact of different factors on the number of embankment failure occurrences. Both these micro and macro level correlations between factors that potentially impact embankment failures are not known.

In the last decade, ProRail has gathered data on several potentially impactful factors in the rail transport network. Making sense of this data and being able to make use of the databases is desired to make the network robust and reliable. This data can potentially be used to find locations where the infrastructure is likely to fail or where the infrastructure can facilitate an increased number of heavier and faster trains. As stated by ProRail (2024), slip circles are hard to predict whilst having serious consequences for the timetable and passenger safety. More knowledge is already housed regarding the impact of bearing capacity reduction.

With current knowledge and monitoring data, a data-driven maintenance approach seems to increase in likeliness. This predictive maintenance aims to perform maintenance activities before failures occur. Predictive maintenance is advantageous for both the user and the asset owner because timely maintenance activities are cheaper than corrective maintenance and increase safety by mitigating failures. Utilizing the different impact factors as input for machine learning models with predictive capabilities can lead to a further understanding of the influence of different factors on the occurrence of railway embankment failures, and these models can possibly find correlations between these factors more easily than humans. However, because the background knowledge on correlations between different factors is not yet sufficient, a physical model, being a near copy of the real-world situation, is currently not possible. Therefore, this research focuses on combining expert knowledge with a data-driven approach to find the most important factors impacting railway embankment failures to ultimately provide a step in the right direction for data-driven asset management and the capacity expansion of the railway network in the Netherlands.

1.3. Research dimensions

The dimensions of the research are divided into objective, questions and scope.

1.3.1. Research objective

The goal of this research is to find the relationship between rail embankment failures and climate, train and railway variables and embankment parameters. Because of the goal of ProRail to facilitate more, faster and heavier trains with proven safe rideability of the tracks, the goal is aimed at functional carriers bearing and guiding. The knowledge gained within this research will aid in informed decision-making when planning and prioritizing (predictive) maintenance and in the decision-making process on tracks that are fit for increased usage. Based on the problem statement and preliminary investigation, the aim of this research is formulated as follows:

"The objective of this research is to analyse the contribution of embankment parameters, rail transport variables and climate to the variability in railway embankment failure occurrences"

The variety of rail embankment failures is based on the two orders of magnitude presented in the problem context. On a micro-scale, this variety is considered binary, being whether or not an embankment failure has occurred at a certain location and time. On a macro scale, this variety refers to the number of failures in a certain area, finding more generalized correlations between these factors.

1.3.2. Research questions

To reach the objective and solve the problem stated in 1.2, a main research question is formulated:

"To what extent can currently available data on impact factors be used to explain railway embankment failures on a micro and macro scale?"

The main research question will be answered by means of four sub-questions.

Q1: What factors are expected to be the most indicative for railway embankment failure?

Q2: When are factors related to railway embankment failures considered extreme?

Q3: How are the considered impact factors related to specific occurrences of rail embankment failures in the case study area?

Q4: What is the impact of the considered impact factors on the variation in rail embankment failures in the Netherlands?

1.3.3. Scope

This research will focus on railway embankment failures in the Netherlands. The scope of this research has a spatial and a temporal component. The temporal component is defined by the availability of data on rail embankment failures and on the expected indicative variables. Since 2006, rail embankment failures have been registered and stored in a database. The temporal scope of this research is therefore after 01-01-2006 dependent on the variables that are being analysed on their impact.

The spatial component of the research is twofold. The micro analysis only considers a specific case study area, the contract and maintenance area Noord-West, which has some tracks along with the trace between Alkmaar and Amsterdam. Several reasons contributed to the decision of this case study area. First, the size of the area made the amount of data manageable. Second, the trace Alkmaar-Amsterdam is planned to facilitate more trains in the future (ProRail, 2023). The start of the work activities is planned in 2026 and aimed to be finished in 2031. Experts at ProRail also suggested this track due to a known number of failures that are expected to be climate-related, and for which several variables are already measured. Next to the micro analysis, the macro analysis will consider the rail embankment failures on a national scale. All railway lines in the Netherlands are considered for this analysis. In Figure 8, the railway tracks in the Netherlands are depicted. In addition, a magnified image of the micro-analysis study area is given with the geocodes. Amsterdam is part of another ProRail area and therefore not in the case study area of the micro analysis.

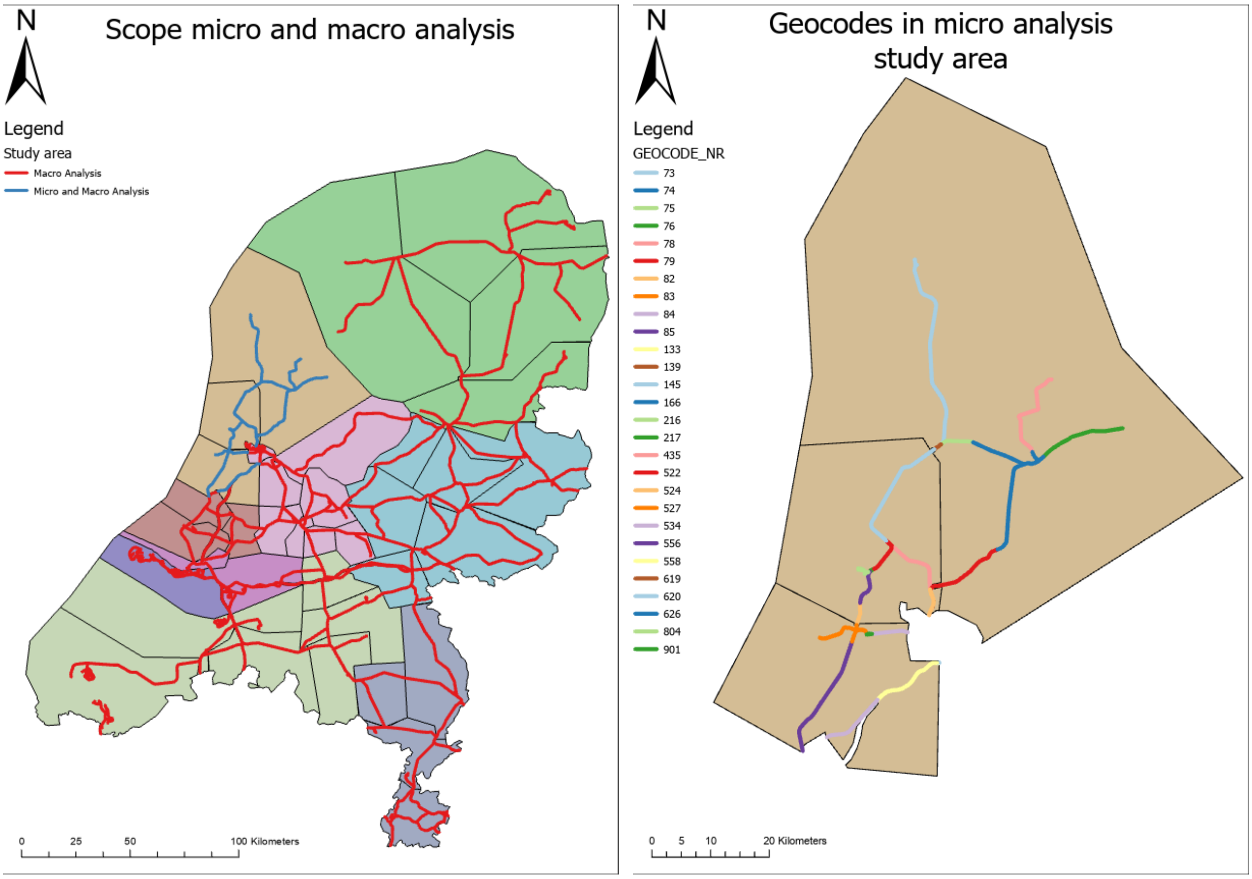


Figure 8: Scope (left) and geocodes in micro analysis study area (right)

Lastly, there are some external factors that can impact the occurrence of railway embankment failures, such as natural disasters and animals. This research considers only variables that can be measured and predicted properly.

The rest of this report is structured as follows: the next chapter outlines the methodology used to answer the main research questions by means of a step-by-step approach to each research question. Then, Chapter 3 shows the results structured per the research question, following the steps indicated in the methodology. The results, method and their shortcomings are discussed in Chapter 4. The answers to the research question are summarized and concluded in Chapter 5. Finally, recommendations for future research are given in the final chapter: Recommendations.

Any additional information that is not included in the main report is given in the Appendices section at the end of this report after the overview of the used literature.

2. Methodology

The four research questions that are introduced in 1.3.2 together serve to answer the main research question and reach the objective. In this chapter, the contribution of these four research questions to the goal is elaborated. In addition, the background as well as the methodology used for answering the four research questions is given. The first research question aims to improve the understanding of railway embankment failures and create a shortlist of impact factors that are considered in the analyses conducted in the rest of the research. The second research question serves as a basis for concluding the micro and macro analysis of the research. A statistical view of data from the shortlist factors is established. In addition, the data is being pre-processed in this section. Then, in research question three, the impact factors of the shortlist are linked to specific railway embankment failure occurrences. For each of these occurrences, the results of the statistical analysis in question two serve as a benchmark to estimate the extremity of the variables at the location and time of the incident. Finally, research question four aims to find broader trends in the occurrence of railway embankment failures and potentially validate findings from the micro-analysis. At the end of this chapter, the research questions and the steps to obtain results and answers are visually represented in a flowchart.

2.1. What factors are expected to be the most indicative for railway embankment failure?

Many variables play a role in the occurrence of a railway embankment failure. This research question provides background information on the occurrence of railway embankment failures. To obtain the answer to this question, literature is combined with ProRail’s safety documents and expert judgement. With the knowledge obtained in the background analysis on embankment failures, a long list of factors that are expected to impact the occurrence and severity of railway embankment failures is created. Then, a shortlist of these impact factors is set up. This shortlist forms the basis for the other research questions and also serves as the hypothesis to be tested.

This research question is divided into 5 steps: a literature study on railway embankment failures is conducted from which a first set of impactful factors is derived. This list is checked and complemented by experts from ProRail. The data availability for each factor is then considered. Then, the expected correlation between these factors is determined. The findings from the previous steps are then combined to create a shortlist of impact factors. These steps are individually detailed in sections 2.1.1 until 2.1.5.

2.1.1. Establish long list of impact factors

A background analysis of railway embankment failures serves as the basis for the long list of impact factors. These failures occur at specific locations and times. Therefore, both spatial and temporal aspects must play a role in their occurrence. Otherwise, heavy rainfall (temporal) would lead to several embankment failures at one moment in multiple parts of the tracks, or one specific soil composition (spatial) would cause embankment failures regardless of the time. The background study considers safety documents from ProRail, literature on soil mechanics and macro stability in dikes, being a comparable failure mechanism. In the results section, each area with factors contributing to the long list of impact factors is investigated.

2.1.2. Expert judgement integration

The factors derived from the literature study are discussed with several experts from ProRail. A civil engineer, data specialist, maintenance engineers and colleagues from the embankment failure programme gave their opinions on the long list, what impact factors must be incorporated and the prioritization of these factors. The findings from the experts also verify the findings in the literature study from a practical perspective and therefore contribute to the establishment of a sound priority list.

2.1.3. Data characteristics

To test whether the selected factors indeed impact the railway embankment failures, the availability and quality of the data are of great importance. Therefore, a data specialist of ProRail has helped with initial confirmation on the data acquisition in general and the obtainability of that data for this research. In addition, factors in the shortlist of which the data is not owned by ProRail have been investigated. For each factor, the following characteristics have been established for the data, as these are important for the analyses.

- Measurement method
- Source
- Variable or characteristic
- Measure frequency

The above characteristics combined with whether the data is available in the first place served as an important aspect of the decision-making process of whether or not to include an impact factor in the shortlist.

2.1.4. Impact factor correlations

When the long list of impact factors is established along with the characteristics of the data, the direct correlation between the impact factors is visualized. This assists in determining the importance of the individual factors. An example of this importance can be the number of passengers in a train and its correlation to the train's weight. The number of passengers adds to the force that is exerted on the rails. The number of passengers is however very variable and the relative extra weight of passengers compared to the weight of the vehicle itself is low. In the case of train weight, the axle loads intrinsically consider the number of passengers. Prioritization of impact factors not only supports hypotheses by visualizing expected relationships but also ensures the manageability of the data that is taken into account. Therefore, this step of visualizing the expected correlation between impact factors is an important last step in the decision-making process leading to the shortlist.

2.1.5. Set up shortlist

Finally, the long list of impact factors is considered with all obtained information to create a prioritized shortlist. The result of this last step is a list of impact factors that are expected to influence the occurrence of railway embankment failures of which a sufficient amount of data is available to link the factors to railway embankment failure data. Testing the shortlist factors both on a micro-scale and a macro scale will test the hypothesis that they impact railway embankment failures in the Netherlands.

2.2. When are factors related to railway embankment failures considered extreme?

The second research question is intended to gain insight into the impact factors and their meaning. The ability to derive insights from the results by understanding the meaning of the values is essential in concluding the results of the micro and macro analysis. The knowledge obtained from this research question therefore aids in drawing substantiated conclusions and determining the extremity and occurrence of certain values. This exploration and interpretation phase consists of four steps. The first step is to obtain the data of shortlist factors and, if needed, clean the datasets. Then, for every variable factor, the distribution is shown in the form of histograms and boxplots. The histograms indicate the skewness of the data, whereas the boxplots more clearly visualize the range and extremes of each variable. Third, the boundaries for considering a variable value as extreme are set up by the Inter Quartile Range (IQR) method. Then, to further specify extremity, an adaptive Tukey's fence method is applied. The final result of this research question serves as a statistical basis for assessing the extremity of variable values by means of two boundaries. Given the spatial and temporal aspects mentioned earlier in the research, these characteristics are preserved and examined through additional visualization techniques such as heatmaps, spatiotemporal plots, and time-series analysis in the analyses following research questions three and four. These methods provide a more comprehensive view of the data, maintaining the integrity of its complexities, such as spatial and temporal dimensions. The following subsections provide a more in-depth view of the steps taken and involved statistics.

2.2.1. Data acquisition and cleaning

To be able to analyse the impact factors of the shortlist, datasets are gathered and pre-processed. Data cleaning includes removing Not a Number (NaN) values, adjusting dates to be in the same format and removing obvious outliers such as negative train loads. The data pre-processing is performed in Excel and Python. In addition, a PostgreSQL database is set up and managed employing pgAdmin4 software. After preprocessing all data, a database with workable datasets is obtained.

2.2.2. Visualization of data

Visualizing data makes it easier to understand and interpret (Srivastava, 2023). This step involves two ways of visualizing the datasets: histograms and boxplots. From a histogram, the skewness of a dataset becomes visible immediately. Because the extremity of values is of interest, the way that a dataset is distributed provides valuable insight into the results from the IQR method described in Dash et al. (2023). To visualize the IQR results, boxplots are also created for each shortlist variable. An explanation of how boxplots serve to better understand the IQR method is given along with the IQR approach. Finally, in this step, basic statistics such as mean, standard deviation, minimum and maximum values are given.

2.2.3. Inter Quartile Range method

The Inter Quartile Range (IQR) method as described by Dash et al. (2023) is used. The boxplot in Figure 9 gives a visual representation of what this method looks like for a normally distributed dataset. The advantages of the IQR method are among others the simplicity and ease of interpretation (Dastjerdy et al., 2023). The IQR is visualized by means of a box and whisker plot, more commonly referred to as a boxplot. A boxplot clearly shows the dispersion of a dataset. For the development of these boxplots with IQR, the first step is to reorder the dataset into ascending order. The value of Quartile 1 (Q1) indicates the boundary under which 25% of the datapoints are found. Then, Quartile 3 (Q3) shows the boundary under which 75% of the datapoints are found. The IQR is the difference between Q1 and Q3. With the IQR, the boundary values at which a value is considered an extreme or potential outlier is calculated with the following formula:

$$\text{Lower bound} = Q1 - k \times IQR$$
$$\text{Upper bound} = Q3 + k \times IQR$$

Traditionally in this formula $k = 1{,}5$

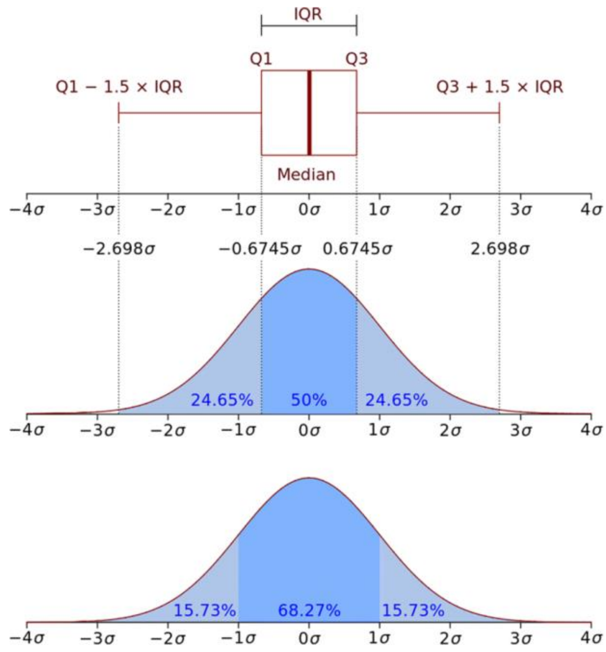


Figure 9: IQR method for outlier detection. Source: (Ibrahim et al., 2021)

2.2.4. Adaptive Tukey Fence

The IQR method provides a basis for further understanding the distribution of the data and a first estimation of when a datapoint can be interpreted as an extreme value. The traditional IQR range from Dash et al. (2023) focuses on finding the values that do no fall within 98% of the range. This method only finds the 98% percentile of the values in normally distributed datasets. To prevent obtaining a large number of outliers due to either right-skewed or left-skewed distributions, the IQR method is extended with an adaptive version of the Tukey fence method proposed by Páez and Boisjoly (2022). Utilizing a computerized model, the value of k for which the number of outliers is equal to 2% of the total datapoints is calculated. This k-value with the IQR provides the upper and lower bounds that can be used to identify extreme values. The following formula represents the method:

$$P(Q1 - K \times IQR \leq N \leq Q3 + K \times IQR) = 0{,}98$$

In this formula, *P* represents the proportion of data points within the Tukey fence bounds, *K* is the variable factor and *N* is the number of data points within the dataset. This way, the occurrence of large quantities of outliers is prevented. A practical translation of this step to the analyses is given by the following example:

If high rainfall based on the IQR boundary seems to be linked to 9/10 accident reports from 1 year, but the frequency of values above the boundary in that year is 80, the impact of rainfall can seem more relevant than it is in reality. If the Tukey fence occurrence of rainfall is apparent in 2/10 accident reports from 1 year but has a frequency of 8, this is a more significant extremity indication.

2.3. How are the considered impact factors related to specific occurrences of rail embankment failures in the case study area?

By utilizing registered disruptions, a micro-scale analysis is executed, in which the embankment failures in the case study contract area Noord-West are linked to the impact factors of the shortlist. The reduction of the scope to a micro-level serves to find specific instances that can be extensively investigated. The aimed result of this research question is obtained by following five steps: determining the number of registered embankment failures, filtering these records to only consider the case study area, linking the impact factors to these records, applying the findings and boundaries from the research question two to the variables and finally investigating these findings.

2.3.1. Determine the number of embankment failures accident records

To gain insight into the amount of railway embankment failures that have been registered, an export of all accident reports is obtained from ProRail. These accident reports are registered by the operation specialist of the switch and report centre. Accident reports are mainly telephonically announced by train drivers. Otherwise, they are received from monitor systems like tunnel installations or the business operation system for energy supply. In most cases, the train traffic controller of the service area for which the accident report is made announces the event. If the failure is related to infrastructure, a so-called RVO, which translates to accident report, is made in SAP, which is an internationally used software for Systems, Applications and Products. Within the accident report, several variables are registered. Some variables that are important for this research include:

- Date and time of announcement
- Facts table containing:
    - Abbreviation of emplacement
    - Abbreviation of object
    - Kilometre indication of accident report (if known)
    - Short description of the accident
- Geocode for which the failure report is registered
- Cause code

There are numerous cause codes, that can vary from copper theft and short-circuits to incorrect operations during maintenance. Within these cause codes, three are possibly related to railway embankment failures. In Table 1, these codes, groups, descriptions in Dutch and the translation are given.

Table 1: cause codes relevant for embankment failures

Code	Omschrijving	Description
207	Onjuiste geometrie/ligging/blinde vering	Incorrect geometry, position, improper support for sleeper by ballast
208	Klapper	Sleeper not on ballast
209	Verzakking / klink / zetting	Prolapsation / latch / settlement

In Table 2, an example of a few of these accident reports is shown with the important variables.

Table 2: example of accident reports

Melddatum en tijd SAP	Fact_RVO_SAP_UI5.Feiten tekst	geocode	Oorzaakcode
22/02/2014 14:07	Vdg-Vdo : Km 4.9 verzakking.	115	209
09/05/2013 18:06	Zvno: Sp-FZ km.72,6(Duitse km)km.111,026(Ned.km)spoorverzakking op de landsgrens,nederlands baan loopt t/m km.111.1	39	207
09/05/2013 16:34	Gvc : Sp-4 slechte ligging bij uiteinde kap.	560	207
03/06/2010 19:44	Gd-Gdg : km-31.0 Deur in geluidswand open. Direct sluiten	105	207
10/03/2008 12:14	Zlw-Ddr km 25.2 verzakking spoor bij onderdoorgang Kilweg trdl rijdt met VR	119	207
15/07/2006 12:07	Gdg achter wls 475B knik in spoor15-7-2006 om 14:00 geaccepteerd door trdl herstel moet vannachtplaatsvinden. Schiftslag weinig hinder wel lstg 40 km	105	209
08/04/2006 19:04	Rtn-Rtd knik in spoor thv km 5	114	209
19/10/2013 12:48	2+ Asn : Wsl 391 A/B geheel uit controle.Herstel volgens WBI NO 694264.	654	208
19/09/2015 15:04	Hlm-Lis: knik in spoor km 20.6 Sp-MH . Vr 40 km/h	85	207
18/10/2014 13:35	Amf-Dld : Hobbel in spoor. Km 19.0 AD spoor.	90	207

The total number of accident reports with a cause code related to the track position is 8662 (January 2006 – March 2024). Using lemmatization and filtering in Excel the table with accident reports are reduced to only include reports that are related to railway embankment failures.

Essentially, the lemmatization formula checks whether the word 'verzak', being Dutch for prolapse, is in the report text. This leaves a total of 2045 reports of railway embankment failures in the Netherlands between January 2006 and March 2024. Figure 10 shows the number of accident reports related to railway embankment failures aggregated both per month and per year. Only full years are taken into account for these figures, therefore the year 2024 is excluded from further analysis. From the figures, it can be stated that the month does not immediately seem to be directly related to embankment failures. In the different years, however, there are some clear differences. 2015 has a relatively low number of accident reports, while from 2018 onward there are a lot of accident reports.

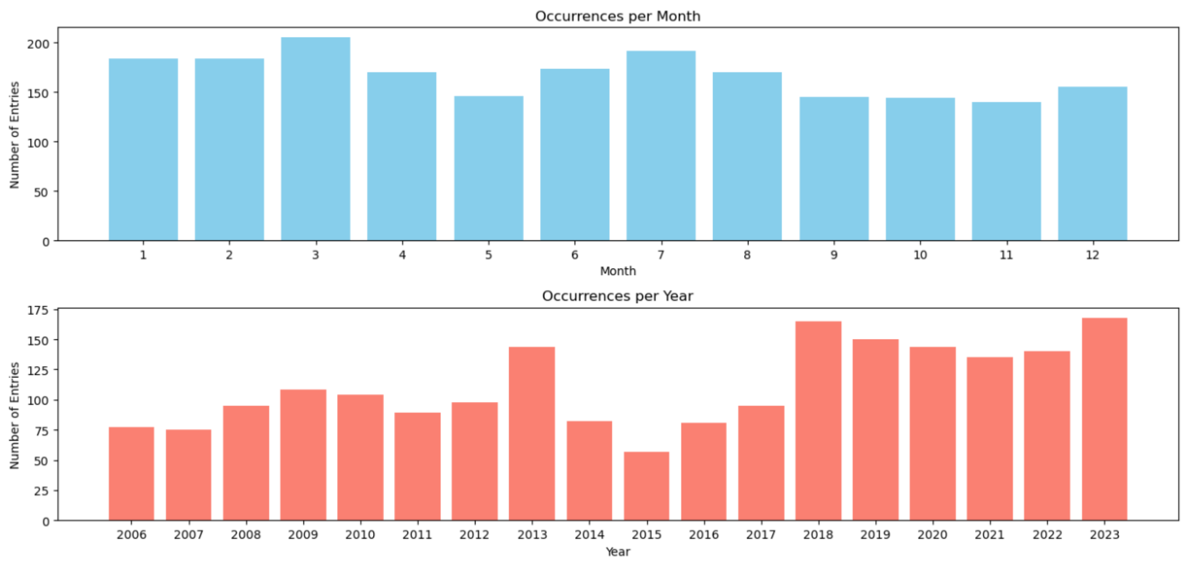


Figure 10: Embankment failures per month (top) and per year (bottom)

2.3.2. Filter records in the case study area

The micro analysis is focused on a specific part of the Netherlands. As stated in the scope demarcation in section 1.3.3, only accident report in the ProRail area Noord-West are considered for the micro analysis. By filtering the table with accident reports to only contain these geocodes, 27 of total 331 geocodes are selected.

The exact location of the disruption is hard to determine. As described in section 2.3.1, a call to announce an irregularity is made by a train driver after which the report is made by the operating expert. The most exact location is usually stated somewhere in the report text. As with the geocodes, the location of the accident, if stated in the report text, is extracted. With the operations in the failure record file executed, the leftover accident reports contain all failures between January 2006 and March 2024 of which the record text specifically states that the prolapsing of the subsurface is the reason for the registration of the failure, and the location of these failures.

2.3.3. Link records to impact factors based on location and date

From the previous steps, a filtered dataset with both spatial and temporal components of accident reports is obtained. This step focuses on linking the impact factors from the shortlist to the accident reports. The database that has been subject to a statistical analysis will be linked to either the spatial or temporal component of the accident report. After the linking process, a table is obtained with all shortlist factors linked to the accident reports.

2.3.4. Apply IQR & Tukey fence boundaries

The final step before concluding the findings is to visualize the extremity of the shortlist variables. To enable fast assessment of the results, conditional formatting is integrated into the resulting table. The following conditional formatting rules are used for all shortlisted variables in the micro-analysis table:

Value > MAX (Tukey Fence boundary ; IQR boundary)
Value > MIN (Tukey Fence boundary ; IQR boundary)
Value > mean + standard deviation
Other outstanding value

Figure 11: conditional formatting rules for micro-analysis table

2.3.5. Micro-analysis results

The final stage of the third research question includes concluding the findings in the micro-analysis. The extremity of the shortlist values at a spatial and temporal scale is analysed. As mentioned in research question one, a railway embankment failure must occur due to at least a combination of factors. Therefore, correlations between impact factors is determined by the occurrence of multiple extreme values at the same failure. The micro-analysis aims to assess specific instances of railway embankment failures. Therefore, all considered failure reports are investigated in the result sections. The final result of this research question is a conclusion of whether or not the shortlist factors are indeed explanatory for the occurrence of railway embankment failures, and why.

2.4. What is the impact of the considered impact factors on the variation in rail embankment failures in the Netherlands?

The final research question aims to find general trends in the occurrence of railway embankment failures in the Netherlands. The macro-analysis allows for establishing knowledge regarding more widespread factors such as climate and trends in track usage intensity. The total number of accident reports regarding rail embankment failures per year is used in this analysis. The plan of action of this macro-analysis consists of four steps. First, the total number of failures in the Netherlands per geocode is visualized on a map. This allows one to relate the shortlist factors to the track branches in the map accordingly. The linkage of the shortlist factors allows for the following step, being the visualization on the macro-scale. Finally, with the visualization, spatial trends can be identified. In addition to the visual inspection, several machine learning models are used to find trends and correlations.

2.4.1. Define failures in the Netherlands

The first step in this analysis is to create an overview of the number of failures in the Netherlands per area. As indicated in Figure 12, the 9 ProRail areas and the geocodes are used as the spatial component for linking factors. If possible, the shortlist factors are linked to the track branches because these are more specific. When linkage to the track branch is not possible, the ProRail areas will be used. Additionally, the ProRail areas can be used to obtain more generalized figures. This makes sure that small segments with many failures do not remain unnoticed. Figure 12 depicts occurrences of railway embankment failures per geocode, uses Natural Breaks (Jenks) for the symbology of failures whereas the right-hand side indicating per ProRail area has the symbology based on the equal interval method. As visible in the images, an appropriate symbology method is important for the interpretation of the results: deep red lines in the left figure show geocodes with failures ranging from 85 total to 275 total.

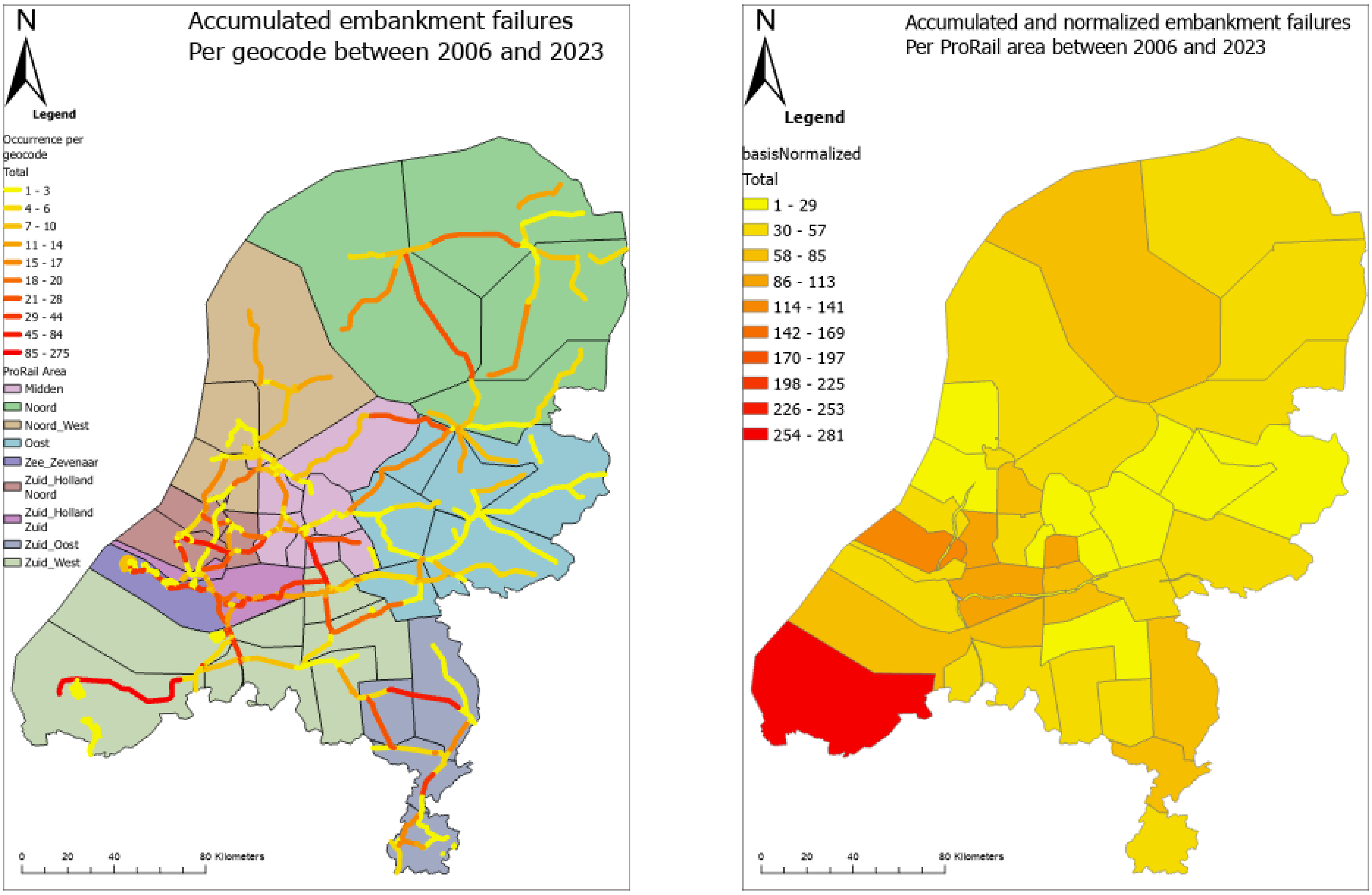


Figure 12: accumulated embankment failures per geocode (left) and per ProRail area (right) between 2006 and 2023

2.4.2. Add shortlist factors to the macro scale

Various mapping tools can be used to link shortlist factors to the map. The main option that is used is a join function, which allows linkage based on common characteristics, such as geocode or ProRail area. If there is no common characteristic between the input layer and the table to join, other tools are used. Especially overlay tools such as Spatial Join and Intersect serve for fast linkage based on spatial characteristics. To make sure that these operations could be performed, all data that is added to a map with a common spatial reference: RD New. This projection is the most commonly used system in the Netherlands (EPSG, 2024). After the shortlist factors are linked to the geocodes and/or ProRail areas, the final two steps of result generation and analysis are executed.

2.4.3. Visualize macro-scale impact factors

The visualization of the shortlist factors serve as a basis to assess the relationship between railway embankment failures and shortlist factors on a macro scale. The figures that are created show the general findings and potentially allow for determining trends and correlations. As stated in 2.4.1, the settings for symbology are important for the interpretation of the results. In addition, to make comparison easier and ensure consistency, a standard layout file is used in which the visualized map layers and their symbology are indicated in the legend.

2.4.4. Machine learning methods

By exporting the linked data from the map, a file is obtained with the shortlist factors and number of accident reports per geocode. This output lends itself for the utilization of machine learning models. With these models, feature engineering is used to find the most influential factors concerning rail embankment failures. Nine machine learning techniques varying in functionality have been used to estimate feature importance: linear regression, ridge regression, lasso regression, elastic net regression, decision trees, random forest, gradient boosting regression, support vector regression and K-nearest neighbours models are set up. A description of these models is given in Appendix A: machine learning models used. Two operations are executed to assess the model performance: a trial feature and visual inspection of the predicted and actual values. The trial feature is equal to the target feature. The machine learning models should be able to almost perfectly predict the outcome when this feature is added, indicating that the models are capable of capturing underlying patterns and relationships. The R-squared and plots of predicted versus actual values are presented and analysed. As stated by Chicco et al. , the R-squared value is informative for model performance and is easily interpretable. An R-squared value of 1 indicates perfect prediction. A negative value means that the model performs worse than a naïve model that always predicts the mean of the target feature.

For each feature and each machine learning method, the importance score is extracted from the model. To assess the overall importance of each feature, these scores are normalized per machine learning method because the magnitude of feature importance differs in linear models and tree models.

To be able to set up a composite score of the features and assess their importance utilizing various approaches, the scores of methods with a positive R-square score have been multiplied by the R-squared value of the machine learning method used. the following formula is used to obtain the overall importance per feature based on the machine learning models R-squared values:

$$I_{overall} = \sum_{i=1}^{n} I_{feat,n} \times R_n^2$$

In this formula, *I* is the feature importance score for machine learning method *n*. Then, the composite scores of each feature are obtained by normalizing $I_{overall}$ for each feature. This value indicates the relative importance of different features for explaining the variance in railway embankment failures for varying machine learning models.

2.4.5. Macro-analysis results

Resulting from the analyses performed in this chapter, the impact of the shortlist factors on the variation in rail embankment failures is derived. The composite scores of seemingly impactful features is highlighted and substantiated based on the literature review, expectations and the functionality of the machine learning models. Moreover, the meaning of the results are elaborated in this section.

Figure 13 shows the steps that are taken for each research questions.

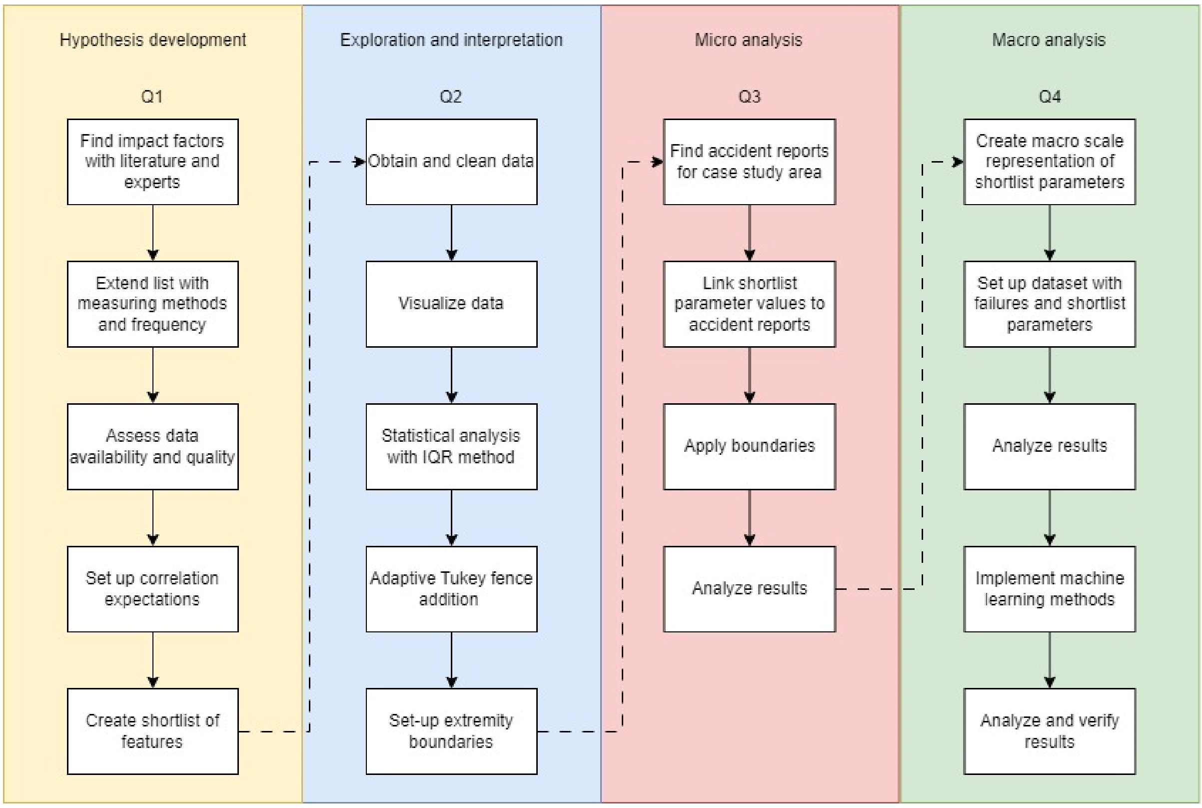


Figure 13: Methodology represented as flowchart

3. Results

In this chapter, the results obtained by following the methodology described in chapter two are given. The methodology broke down the research questions into steps. If these steps generate output, the steps are provided as headings to guide the reader through the establishment of results. The answer to the final research question is given in the conclusion after chapter four has discussed the shortcomings and remarks of this research.

3.1. What factors are expected to be the most indicative for railway embankment failure?

The first question is aimed at creating a list of potentially influential factors regarding the occurrence of embankment failures. From a literature review, four main areas for the occurrence of rail embankment failures are derived: geophysics, climate, train variables and the physical status of the tracks. Each of these main areas is elaborated in section 3.1.1. The expert judgement on the impact factors is integrated in this section since mostly minor additions to the literature and extra insights into the practical side of embankment failures. The end of each area elaboration is complemented by a table providing an overview of the impact factors along with their data characteristics.

3.1.1. Development of long list of impact factors

Derived from geophysics

Rail embankment failure is caused by the sliding of the soil layer underneath the railway tracks. This failure mode appears similar to the macro stability failure mechanism of dikes. The loads in this case are however atop the embankment rather than on the side. Therefore, factors related to this failure mechanism are considered as possible indicators for rail embankment failures. The geotechnical cause of rail embankment failures is a slip circle, as depicted in Figure 14. Three types of slip circle are depicted. The left and middle figures depict the most common slip circles for rail embankment failures. These occur more often because the crest width of an embankment is usually too small for the base circle failure mechanism to occur (RSBB, 2011). The safety factor of the embankment can be calculated using the revised Taylor slope stability proposed by Steward et al. (2011). The first factors that are expected to influence the potential of rail embankment failure are derived from this method. The list of geophysical factors is supplemented with additional elements that can be found in rail embankments such as the prevention measures indicated in Roshan et al. (2022). The list of soil-related factors that are derived from this section are given with the data characteristics as stated in 2.1.3 in Table 3.

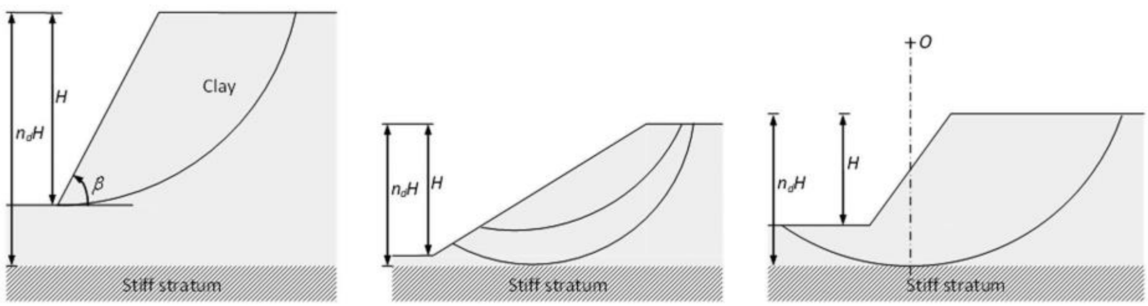


Figure 14: Slip circles by Steward et al. (2011): toe circle (left), slope circle (centre), base circle (right)

First, the soil composition is crucial for determining the strength of the embankment. Parameters such as the internal friction coefficient and weight are all dependent on the material type of the soil. Recently, a large project called Landelijke Netwerk Analyse (LNA) has started (ProRail, 2024). In this project, the stability of the railway embankments is estimated by performing calculations of possible soil compositions. Based on known soil maps, the possible soil compositions of the railway embankments are determined. For various segments, the corresponding safety factors of these possible soil compositions are calculated based on the bishop circle. Secondly, the soil saturation is an important variable in the slip circle calculations. The saturation is dependent not only on the soil type but also on the weather conditions. The saturation of the soil impacts the unit weight of the soil (γ_{dry} or γ_{wet}), being a fundamental parameter within the slip circle calculations (Steward et al., 2011). The soil saturation is partially dependent on the runoff capacity. Several factors impact the runoff capacity of the soil, including the morphology of the soil, cover type of the soil, wind, vegetation and more (Ramke, 2018). Three factors that may be interesting for this research include whether the embankment is in an urban or a non-urban environment, the slope steepness and the vegetation on the embankment. The urban and non-urban aspect is considered interesting as urban areas are expected to have worse runoff capacity (Chen et al., 2007; Li et al.). In addition to the runoff capacity, slope steepness (β) is an important factor that impacts the type of slip circle. Extremely high values of β lead to toe circles as shown in the right slip circle of Figure 14. Finally, next to the impact on runoff capacity, roots of vegetation are expected to have some impact on soil sliding. Finally, the presence of a prevention measure such as such as a sheet pile can increase the strength of the embankment and reduce the chance of failure.

Table 3: Soil-related factors

Factor	Measurement method	Source	Variable of characteristic	Measure frequency
Soil composition	LNA results, soil probe testing	LNA	Characteristic	Project-based
Soil Strength	LNA bishop circle calculations	LNA	Characteristic	Project-based
Soil saturation	Estimation		Variable	-
Presence of earth-retaining structure	Document analysis	ProRail	Characteristic	-
Urban/non-urban area	Estimation	Satellite images	Characteristic	-
Slope	Estimation	Satellite images/height maps	Characteristic	-
Vegetation	Estimation	Satellite images	Characteristic	-

Climate change variables

Climate extremes are an expected cause of the increase in railway embankment failures. Two evident variables that come forward in climate change are the temperature and precipitation extremes. From a geophysical perspective, these two variables influence the soil saturation and consequentially impact the overall strength of the soil. In addition, in an interview with an expert from ProRail (Verloop, 2024) and with TNO, the periods of drought and fluctuations between heavy precipitation and drought are considered possible indicators for rail embankment failure as well. Roshan et al. (2022) state that slope erosion of the embankment is influenced by the duration of a precipitation event. In the Netherlands, however, the duration of precipitation events is not registered (KNMI, 2024). Lastly, forthcoming from the runoff capacity, wind can play a role in the occurrence of a railway embankment failure. In addition to the influence on runoff capacity, Blanco and Lal (2023) state that wind impacts the eroding of the top layer of soil when there is no closed grass cover. Table 4 shows the climate-related factors along with the data characteristics.

Table 4: Climate-related factors

Factor	Measurement method	Source	Variable or Characteristic?	Measure frequency
Precipitation	Rain meter	KNMI	Variable	Daily
Rainfall duration	Time	-	Variable	-
Temperature	Thermometer	KNMI	Variable	Daily
Drought periods	Rain meter	KNMI	Variable	Daily
Fluctuation	Rain meter	KNMI	Variable	Daily
Wind	Wind meter	KNMI	Variable	Daily

Train variables

The external variable load that is executed on top of the rail embankments is exerted by the trains via the tracks. The magnitude of the load is based on several train variables. First and foremost, the weight of the train itself is the largest indicator of the force exerted on top of the rails. The distribution of this force is dependent on the number of axles that the trains have. Then, the speed of the vehicle not only determines the duration that the force is exerted on the rails but is also positively correlated to the axle load (Ciotlaus et al., 2017). Lastly, RSBB (2011) states that the track usage intensity also impacts the possibility of embankment stability problems. Currently, when the bearing capacity of the rails is found to perform below par, a speed restriction is the main intervention applied until maintenance is performed (ProRail, 2021). The train frequency is considered characteristic because the number of trains that are supposed to pass on a trajectory is known in advance. There are exceptions, such as when a segment is used as a detour. This would cause the usage intensity of a part of the railway network to increase. A smaller intensity can also occur when other disruptions cause the segment to become unusable.

Table 5: train-related factors

Factor	Measurement method	Source	Variable or Characteristic	Measure frequency
Number of Axles (combined with speed for time under pressure)	Track sensor	ProRail	Variable	Daily (But the data is aggregated into yearly values)
Train intensity	Timetable and track sensor	ProRail	Characteristic	Daily
Train speed	Track sensor	ProRail	Variable	Daily
Train weight	Track sensor	ProRail	Variable	Daily

Track characteristics

The final category of influence factors is derived from the track characteristics. Significant changes in ballast and rail height can indicate the overall subsidence of embankment sections, as they are placed directly atop the embankment. The distance between the railway tracks is also considered as a possible indicator as, together with rail height, the oblique deformation can be determined. The track geometry is measured by a measurement train. The flowchart in Figure 15 indicates the process from regulations to measurement needs to data in the Daily Operation department of ProRail Asset

Management. The measurement train creates images that can be processed every 0.5 meters and also measures various frequencies in the tracks. Depending on track usage, the measurement train takes 1 to 4 measurements per year on a track.

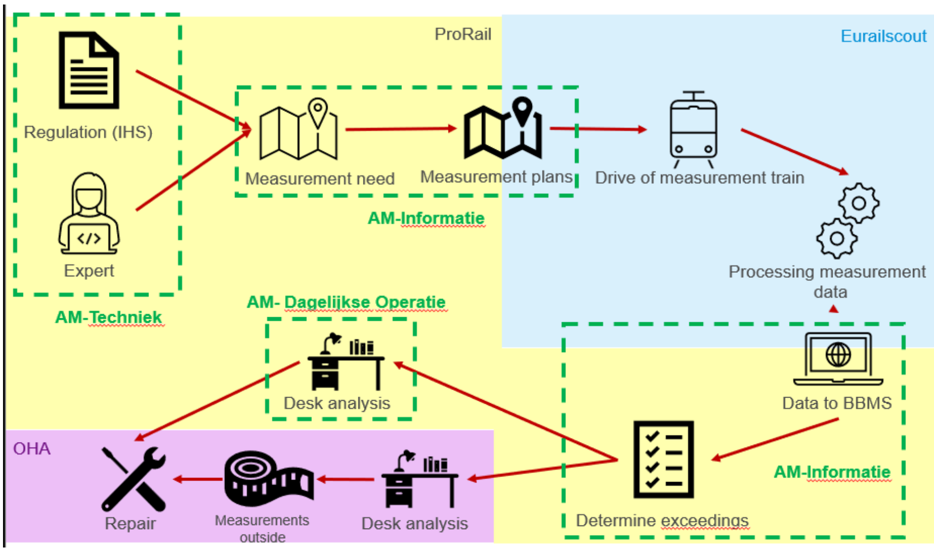


Figure 15: Measurement train flowchart. Source: (De Wit et al., 2022)

The indication of the bearing capacity underneath the rail, consisting of a combination of both the soil and ballast underneath the rails, is currently determined based on the results of a fractal analysis. In a fractal analysis, the height signal found by the measurement train is subject to a mathematical analysis that decomposes frequencies and tracks geometry measurements into wavelength components. These wavelength components are divided into three wavelength ranges based on Austrian Standards (2014):

- Short-wave range between 1 and 3 meters
- Mid-wave range between 3 and 25 meters
- Long-wave range between 25 and 70 meters

Anomalies in the different wavelengths can indicate irregularities and issues. Short-wave anomalies can for instance indicate cracks in the rails and indicate the relation between sleeper and ballast. Middle-wave anomalies indicate irregularities in the ballast itself or the soil layer underneath the tracks. Long-wave anomalies are related to either subsurface or design, such as bridges and hills. RSBB (2011) proposes that the dimensions and settlement levels of the ballast can be indicators for embankment failure. Therefore, middle wavelengths, referred to as the D1 value, can be used to provide insight into the bearing capacity of the ballast and soil layer underneath the tracks. This D1 value is isolated from the overall signal and the fractal dimension is calculated. This fractal dimension is a way to measure complexity, in which a higher value indicates more irregularities and potential subsidence. These patterns, as indicated in Figure 16, are obtained from BBMS as an indication of the bearing capacity. The figure shows two regions of anomalies. In region 1, the fractal dimension of mid-waves (green) is very large. Because the long-wave (blue) also shows some anomalies. This indicates that in region 1, there are potential issues with the subsurface and the ballast. Region 2, on the other hand, does not show large fractal dimensions in the large-wave region. This indicates that only the ballast and layer directly underneath it show potential problems.

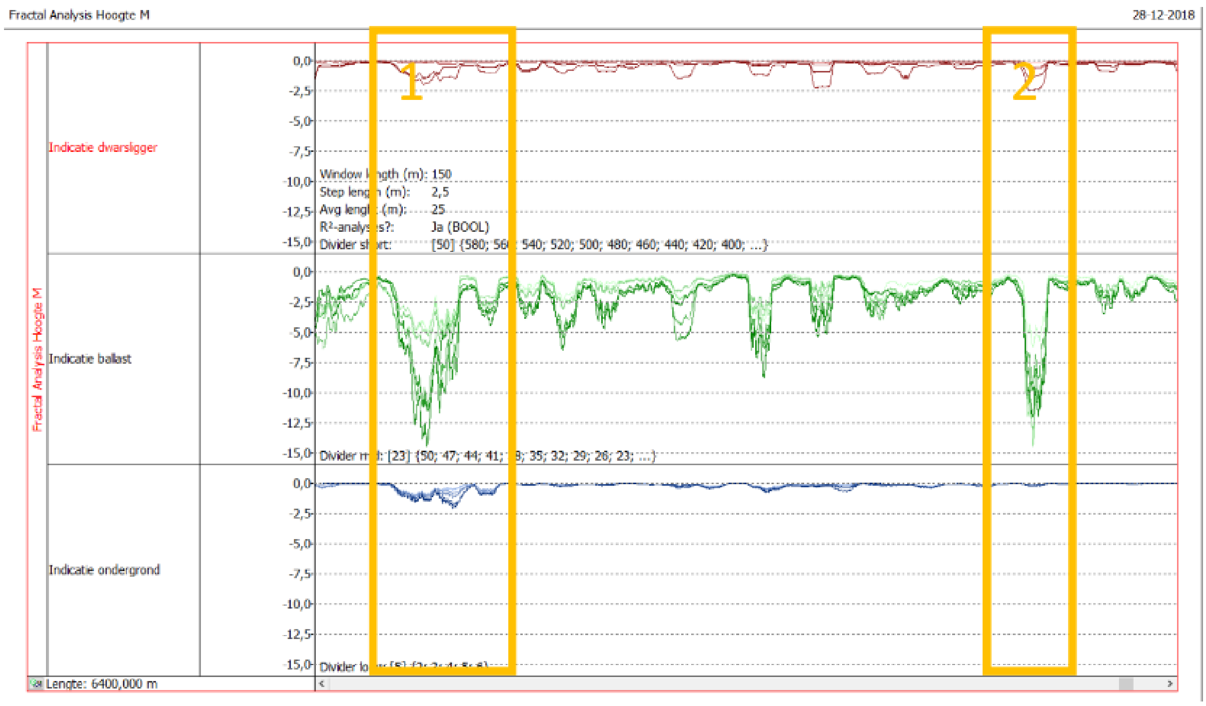


Figure 16: Example of fractal values. Source: (Schuurmans & ProRail, 2019)

Table 6: track-related factors

Factor	Measurement method	Source	Variable or Characteristic?	Measure frequency
Ballast	Measurement train	BBMS	Variable	Half-yearly
Rail height	Measurement train	BBMS	Variable	Half-yearly
Rail degradation value D1	Measurement train and Fractal Analysis	BBMS	Variable	Half-yearly
Distance between tracks	Measurement train	BBMS	Variable	Half-yearly
Sleeper dimensions	Design	ProRail & BBMS	Characteristic	Half-yearly
Sleeper settlement	Measurement train	BBMS	Variable	Half-yearly

3.1.2. Expectation of impact factor correlation

The last step before creating the shortlist of impact factors that will be taken into consideration for the micro and macro analysis is to establish the correlation between the factors from the long list. As stated in the method, this step ensures data manageability and assists in determining the importance of the factors. Figure 17 shows the expected correlation between these factors. This correlation is based on the findings of the previous sections. The occurrence of embankment failure is generalized into two main impact categories: bearing capacity and load. As mentioned earlier, reducing the long list of impact factors into a shortlist containing factors that are considered in the analysis helps keeping the amount of data manageable. Visualizing the expected correlations assists in the decision-making process for the shortlist. In addition to the impact factors and their correlation, the colour of the block indicates in what way the factor will be considered in the micro and macro analysis. Green blocks are directly considered in the micro and macro analysis. Orange blocks are indirectly considered. The temperature, for example, is part of the soil saturation calculation given in B.1. Soil saturation. Red blocks are not considered. Drought periods are not individually in the correlation figure because drought periods are directly derived from precipitation levels. Section 3.1.2 elaborates on the decision-making process.

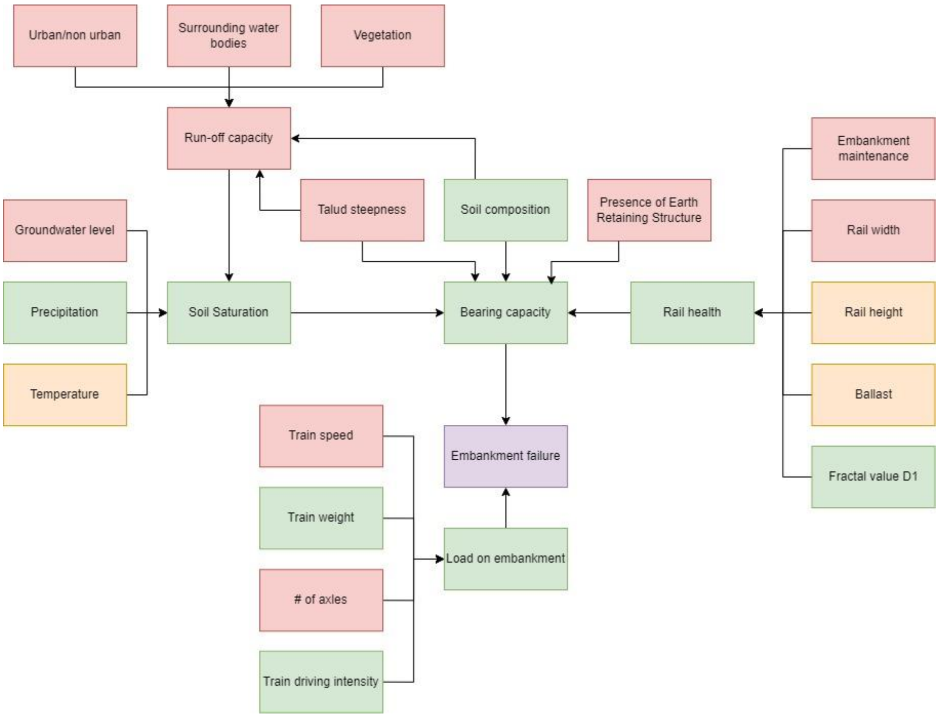


Figure 17: Impact factor correlation for embankment failure

3.1.3. Shortlist selection

The factors that are expected to be the most valuable to consider in this research are selected from a shortlist of impact factors. The selection of these factors is based on several factors: expected impact, expert judgement, data availability and correlations between the factors are the main contributing factors in the decision-making process. Table 7 showcases the shortlist of impact factors. With this shortlist, each of the four impact sections considered in the development of the long list of impact factors has at least one factor taken into consideration in this research. For each factor, an elaboration on the impact is given.

Table 7: Shortlist of impact factors

Impact factors
Precipitation
Drought periods
Soil saturation
Soil strength
Train weight
Track usage intensity
Fractal value D1

Climate change and its impact is the motivation for this research. The precipitation is measured daily by the KNMI, which is a reliable source of weather information in the Netherlands. From the same dataset, drought can be directly derived easily. The soil saturation is related to the weather extremes and is shown by various literature sources to play a large role in the occurrence of slip circles. Because currently the large-scale network analysis of soil bearing capacity is being executed, the soil strength parameter has been advised by experts from SWECO, Royal Haskoning DHV and ProRail to estimate based on the available LNA data and consider for this research.

The load on the embankment is an impact category that is expected to be influenced by four train-related variables. For the train speed, only the maximum allowed driving speed per trajectory was available at the time of the research. Because this does not indicate the actual driving speed of the vehicles, this variable is not considered valuable for this research. The number of axles was too specific, and without the train speed, the overall train weight is a more valuable and more manageable indicator. The track usage intensity is derived from the same department as the train weight. The weight and intensity together are assumed to provide enough information regarding the load exerted by vehicles on top of the embankment.

Finally, experts from ProRail expect the fractal value to be a proper indicator of bearing capacity issues concerning the ballast and the layer directly underneath it. This is a relatively new method that is not yet extensively used. In addition, this value represents all other impact factors in the long list regarding track health in some way. Therefore, this value is taken into consideration for the shortlist.

The other values that have been added to the long list of impact factors are not taken into consideration due to various reasons. The main reason was unavailable or unworkable data. This was the case for most soil and climate factors. Wind, for instance, is measured but the directional component combined with the embankment characteristics to capture the impact of this variable makes it near impossible to integrate wind in the analyses. Another example is the slope of the embankment. This parameter must be derived from satellite maps and height maps. Next to possible measurement errors, creating an overview of the slope of embankments at all locations in the Netherlands or case study area is not possible with the time constraints of this research. Track-related factors other than fractal value D1 are not considered because the measurement train data for those factors could not be acquired in time.

3.2. What is the variation in values of impact factors regarding rail embankment failures?

The resulting shortlist from research question one will first be subject to statistical analysis to become acquainted with the data before performing the micro and macro analysis. As stated in the methodology, four steps will be taken to create the foundation of the analyses from research questions three and four: data acquisition and cleaning of the shortlist factors, analysis of their basic statistics and distribution, IQR outlier analysis and finally an adaptive Tukey fence boundary demarcation.

3.2.1. Acquisition and pre-processing of data

From the shortlist, two variables are not directly available as datasets and need to be derived from other sources first; the soil saturation and soil strength. As mentioned in the results of research question one, soil saturation is influenced by many factors. Taking all impacting factors into account is an exhaustive analysis in itself. Therefore, the soil saturation has been estimated based on the Thornthwaite (1948) potential evapotranspiration formula. The soil strength is derived by performing calculations on the available LNA datasets. Both the Thornthwaite formulae and the soil strength derivation are given in Appendix B. Derivation of soil saturation and soil strength.

After estimating the soil saturation and soil strength, a dataset is available for each shortlist factor. The datasets must be pre-processed to ensure the workability and correctness of the data. The workability includes mainly steps such as setting equal decimal delimiters and date formats. After ensuring the workability the data must be filtered for erroneous data such as measurement mistakes and non-numerical values in numerical columns. This filtering has been performed by plotting histograms of each variable. A clear example of this process is found in the visualization of the LNA safety factor, considering the soil strength with respect to bishop circle sliding. This value can become less than 0 due to for instance missing values or errors in the input. The distribution of the datapoints then becomes distorted due to these negative values. Since a safety factor value in reality cannot become less than 0, any negative value is set to 0. The visualization of the data both with and without these values is given in Figure 18. This step in the pre-processing ensures that during the execution of the analyses, incorrect values do not disturb the output.

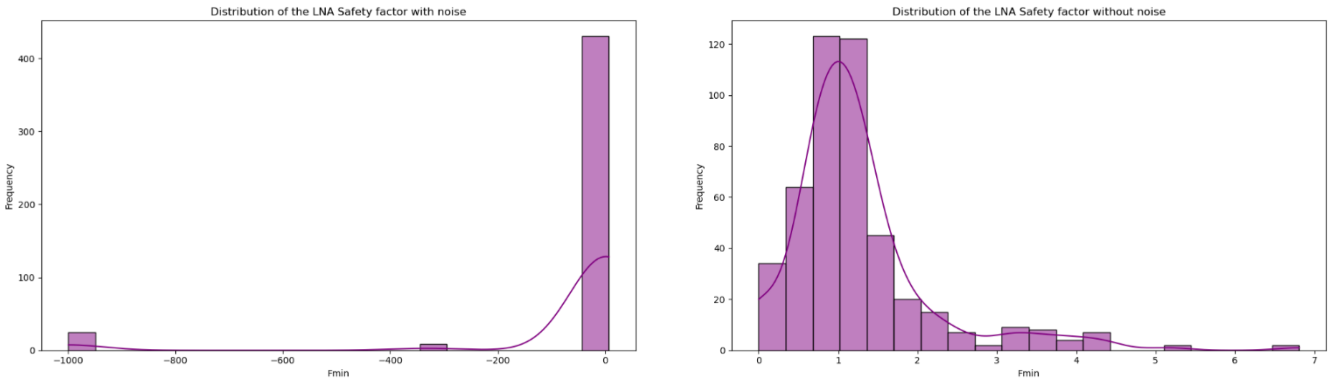


Figure 18: distribution of LNA safety factor with negative values (left) and without negative values(right)

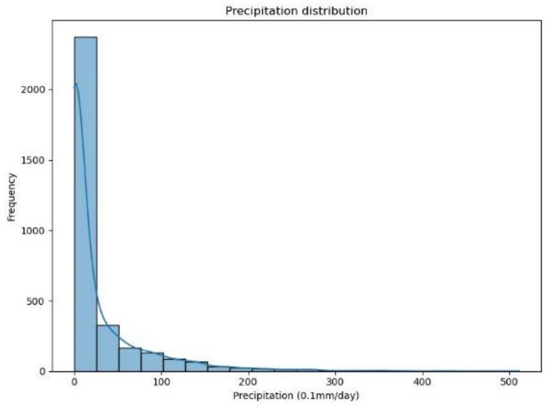
3.2.2. Basic statistics and distribution

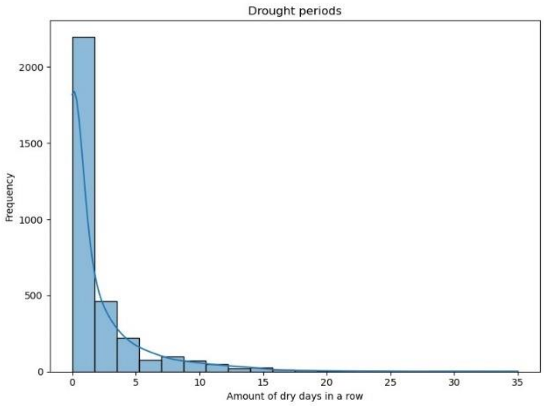
The cleaned histogram in Figure 18 (right) indicates that the distribution of the LNA safety factor resembles a right-skewed distribution. In Figure 19, the distribution for the other shortlist factors is given. For each of these variables, a different dataset forms the basis for the analysis. The above figures, which are derived from the LNA, are taken for the available geocodes in the initial case study contract area Noord-West. The available geocodes are for the trajectory of Alkmaar – Amsterdam. The temperature data, derived from KNMI (2024), is taken from weather station Zaandam in the year range of 2006 – 2023. For this first analysis, the year range 2015-2023 is used to correspond with the other variables. The fractal value D1 was available from 2014 – 2022. The measurement train takes measurements of all rails in the entirety of the Netherlands. To obtain the most significant distribution regarding this value, all measurements are taken into account. This will indicate the most standardized distribution of values and therefore provide the most insightful information with regard to track degeneration. The number of trains using the tracks as well as the tonnes that move over the tracks per day are available for the years 2015-2019, 2022 and 2023. For each track, a value for both the left- and right track is available. These values differ due to the sinusoidal movement induced by the self-stabilizing conic wheels of trains. The largest value is taken because the highest load is expected to be most impactful for embankment failures.

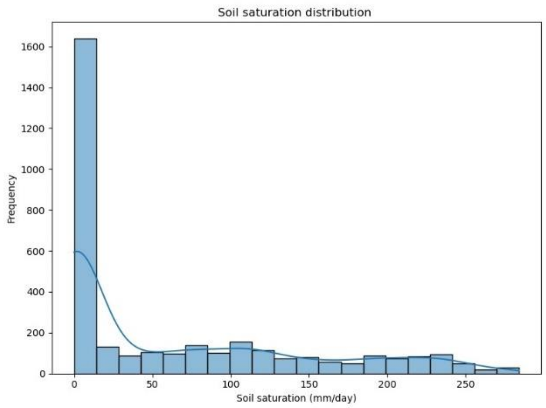
For each variable, the standard statistical values are also calculated. These include number of data points or count, mean, standard deviation, minimum and maximum value. Table 8 showcases these basic statistics for each shortlist variable.

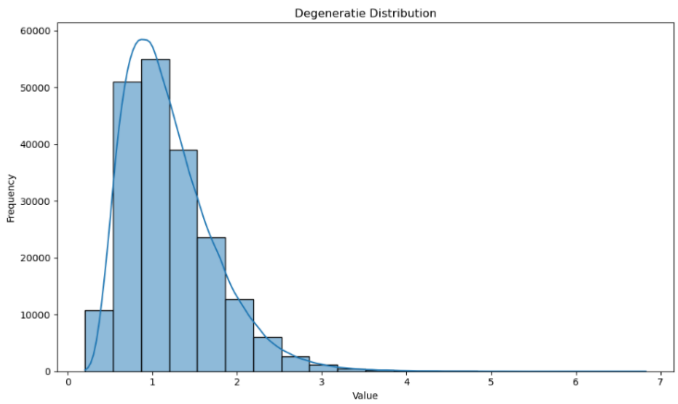
Table 8: basic statistics of shortlist variables

Factor	Count	Mean	s	Min	Max
Precipitation (0,1mm)	3266	27,2	52,6	0	511,0
Drought periods (days)	3266	2,0	3,7	0,0	35,0
Soil saturation (0,1mm)	3266	63,6	56,4	0,0	284
F_{min}	464	1,3	0,9	0,0	6,8
Train weight (tonnes/year)	56630	14224	17235	0	109944
Track usage intensity (trains/year)	56630	15096	17307	0	127903
Fractal value D1	175482	1,2	0,5	0,3	3,4









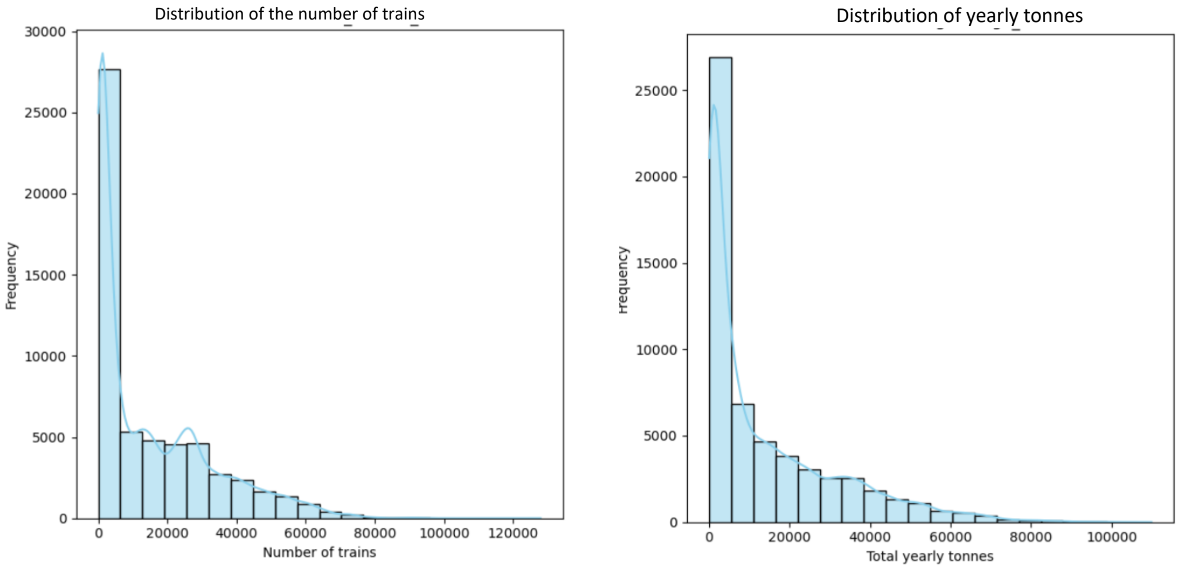


Figure 19: histograms of shortlist impact factors

3.2.3. IQR Extremity calculations

From the histograms created in step two of this research question, it becomes evident that all variables taken into consideration are distributed in a right-skewed manner. For this distribution type, the Inter Quartile Range (IQR) is a suitable method for determining an extremity boundary. This boundary indicates a high value but does not necessarily indicate that a value is outside of the 98th percentile value. In this step, the IQR range and boundaries for all shortlist variable except for F_{min} are set up. Because the motivation of this research is climate change, the precipitation dataset has been used to showcase the steps that are taken to obtain the IQR value.

The boxplots in Figure 20 depict the precipitation distributions of the previous 9 years. 2018 was a relatively dry year, whereas 2023 was a lot wetter. For the IQR method, the values of the entire period between 2015 and 2023 have been taken into account. For all variables in the shortlist, negative values are impossible. The IQR method is however likely to provide a negative lower bound for right-skewed distributions. Because for all variables except for the F_{min}, a high value has a detrimental impact on the rail embankment, the lower bounds are considered insignificant for the analysis. The upper bound provides a value above which the values are considered large, but not necessarily extreme. This boundary serves as a first classification of potentially impactful factors.

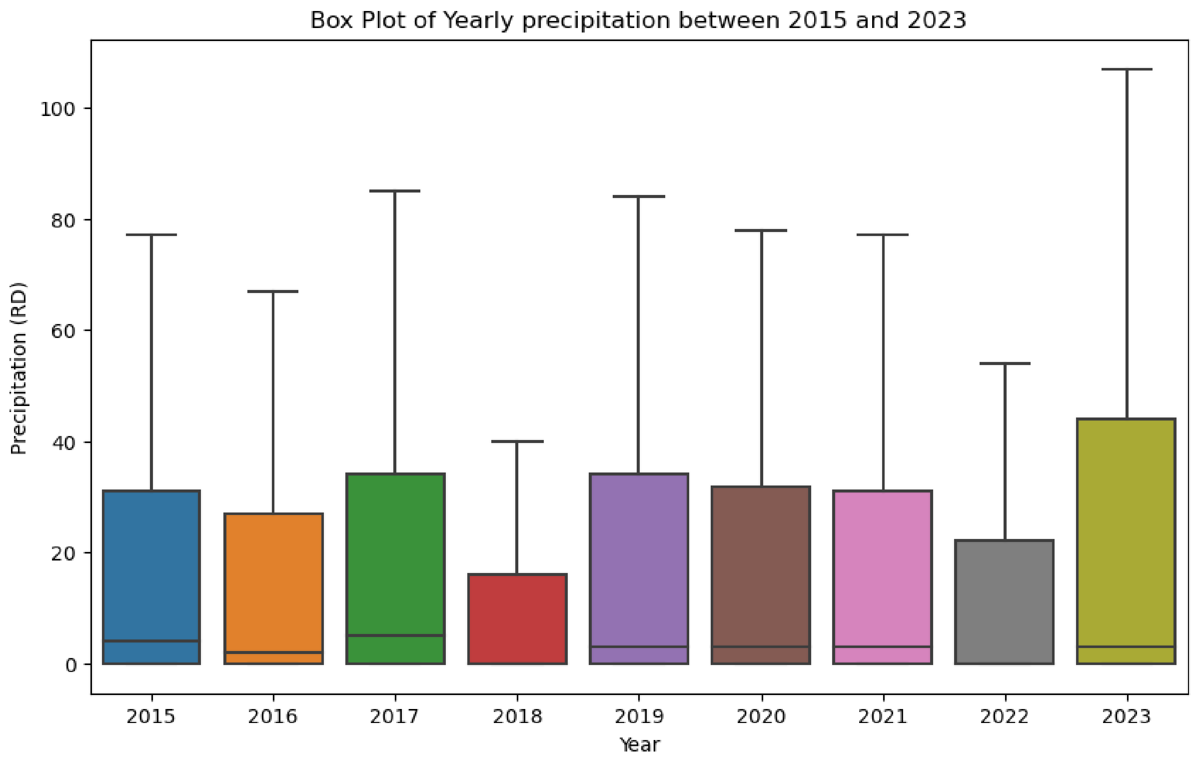


Figure 20: Precipitation between 2015 and 2023

With the traditional IQR method, 408 extreme values occur from 3266 days' worth of data. This would indicate that 12,5% of the values would be considered extreme values. From the visual inspection of the weather data in step one of this research question, measurement errors are not directly identified. In this case, extreme values based on the IQR method would not give meaningful insights, as there are so many occurrences. The adaptive Tukey fence extremity calculations of the next step are aimed at providing a boundary that encapsulates significant extremes. Thereafter, a table showcasing the IQR boundaries, the number of outliers found utilizing the IQR method and the Tukey fence results are given.

3.2.4. Adaptive Tukey fence extremity calculation

An adaptive Tukey fence method has been executed to find the value of K for which 98% of the data points are within the boundaries. In the example case of precipitation, the desired number of extreme values is 65. The model iterates starting from the standard k value of 1,5. If the number of extremes is larger than desired, k is increased by 0,01, and vice versa if the number is too low. For the precipitation a value of k=5,6 yields the desired number of extreme values. The accompanying lower- and upper bounds were -168,30 and 198,30. Because the precipitation was never lower than 0, only the changed upper bound is relevant. The distribution of precipitation values with the boundary lines of both the IQR and the Tukey fence method is given in Figure 21.

Table 9: Statistics of precipitation between 2015 and 2023, including Tukey fence IQR method

Statistic	Value
IQR	30,0
IQR method Lower bound	-45,0
IQR method Upper bound	75,0
IQR number of extreme values	408
Desired amount of extremes	65
Adaptive value of k	5,6
Lower bound	-168,3
Upper bound	198,3
Amount of extreme values	65

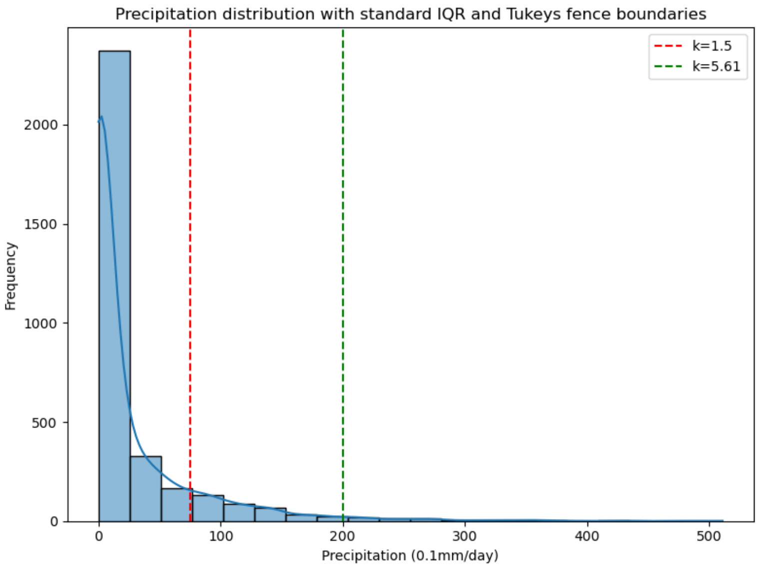


Figure 21: Precipitation distribution with standard IQR and Tukey's fence boundaries

The k-value derived from Tukey's fence method gives an indication of the range in the dataset and how clustered the data points are around the median value. For each shortlist variable of which a high value is detrimental for railway embankments, the k-value that is derived with this method is given below. A high value of k can indicate that a data cluster is found near the median, resulting in a small inter quartile range, or that a large number of values is found deviating far from the median. Therefore, the IQR value and the boundaries of the IQR value are given as well.

Table 10: IQR and Tukey fence boundaries of shortlist variables

Variable	IQR upper bound	k-value	Tukey fence upper bound	Leftover outliers
Precipitation	75,0	5,6	198,3	65
Drought periods	5,0	6,00	6,0	65
Soil saturation	283,7	1,2	250,8	65
Train weight	62847	1,9	70497	1132
Track usage intensity	60591	1,4	58627	1109
Rail degradation	2,5	1,6	2,5	3518

The obtained knowledge on extremes will be used to substantiate findings when the shortlist variables are linked to the failure reports. From a statistical perspective, the 98 percentile values boundaries indicate that more than 1 in 50 failures variable values must be an extreme before the variable is considered significant. For each shortlist variable, an overview of the basic statistical values, the IQR boundaries and the adaptive Tukey fence boundaries is given in Appendix C: statistical results of . The boundaries derived from these analysis provide insight into the meaning of the values obtained when the variables are linked to accident reports in the micro analysis. Additionally, the boundaries created from the IQR and Tukey fence are used to visualize these outcomes which is used to quickly spot correlations or peculiarities.

3.3. How are the considered impact factors related to specific occurrences of rail embankment failures in the contract area Noord-West?

To perform the micro analysis the shortlist factors are linked to the spatiotemporal components of the accident reports in the case study area. By means of the boundaries calculated in research question two, the extremity of the shortlist factors can be visualized. The results are concluded from notable instances in the visual representation of the micro-analysis.

3.3.1. Linkage

The micro-analysis is aimed to take an in-depth look at specific rail embankment failures. The case study area is selected based on the availability of LNA data and the plans of scaling up the train weight, speed and usage at the trajectory of Alkmaar – Amsterdam. As shown in Figure 22, a manageable number of accident reports has been obtained from 2006 onward. A total of 60 embankment failures have been reported in the contract area Noord-West.

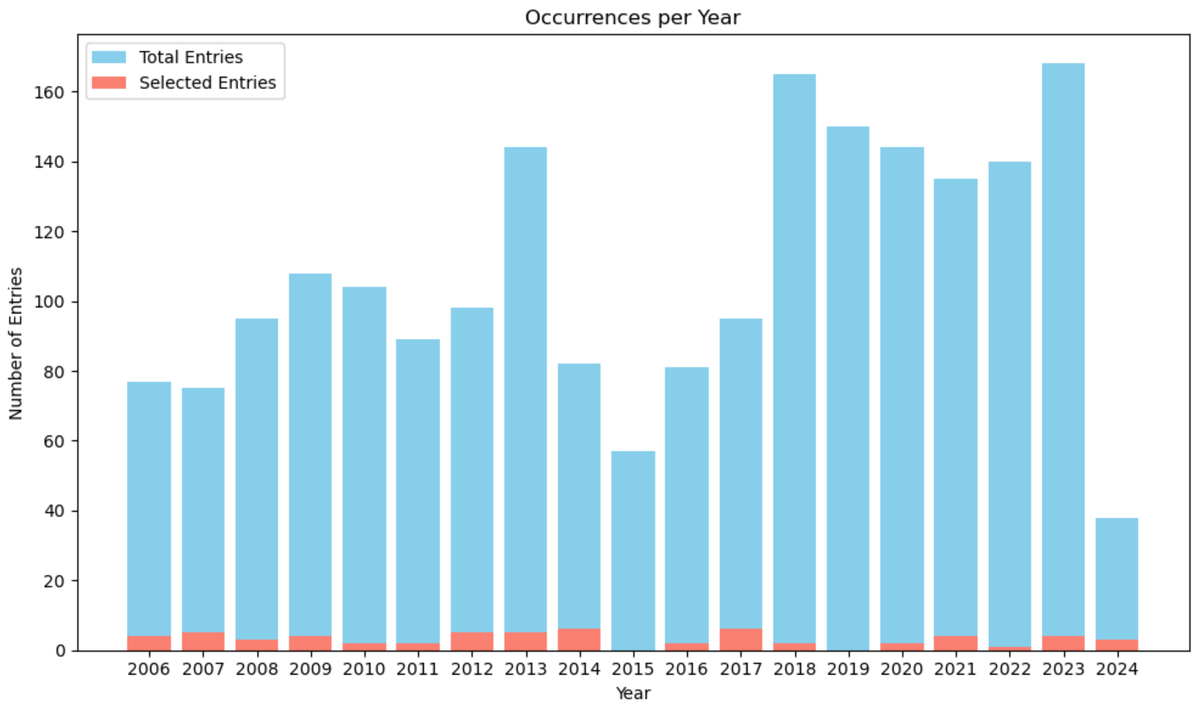


Figure 22: Yearly number of accident reports in the case study area

Each of the shortlist factors must be linked to the accident report in some way. Because a rail embankment failure always occurs at a certain place at a given time, two dimensions can be used for linking factors to the accident report. For each shortlist factor, Table 11 shows the dimension which is used for linking. In the case of the fractal value and the train characteristics, the temporal component indicates the year in which the data was taken. As stated in research question two, the fractal value D1 is available between 2014 and 2022 and the train characteristics in the year range 2015-2019, 2022 and 2023. Therefore, from 2015 onwards, all data was in some way available when extrapolating the train characteristics. From the period 2015-2024, 20 remaining reported railway embankment failures with kilometre indications derived from the description and for which the second phase of the LNA is finished, are investigated individually.

Table 11: Data linking for factors

Factor	Dimension
Precipitation	Temporal
Drought periods	Temporal
Soil saturation	Temporal
Soil strength	Spatial
Train weight	Spatial + Temporal
Track usage intensity	Spatial + Temporal
Fractal value D1	Spatial + Temporal

3.3.2. Apply boundaries

Because climate change is expected to impact the occurrence of railway embankment failures, the soil saturation and precipitation amounts have been linked to the date of each of the accident reports. To assess whether either precipitation extremes or oversaturation of soil is expected to cause the failure of the embankment, a scatterplot is created for each year with the dates of embankment failures shown. An example of the embankment failures plotted in the soil saturation and precipitation in the year 2018, which was determined to be a relatively dry year, is given in Figure 23 below. The other available years are depicted in Appendix C. From these figures, the weather patterns from the days leading up to the railway embankment failure can be investigated. The soil saturation figure shows that there are some peak moments in saturation rather than individual extreme values. For this reason, conclusions taken from solely the boundary values of IQR and Tukey fence is too short sighted and figures are used to substantiate potential findings.

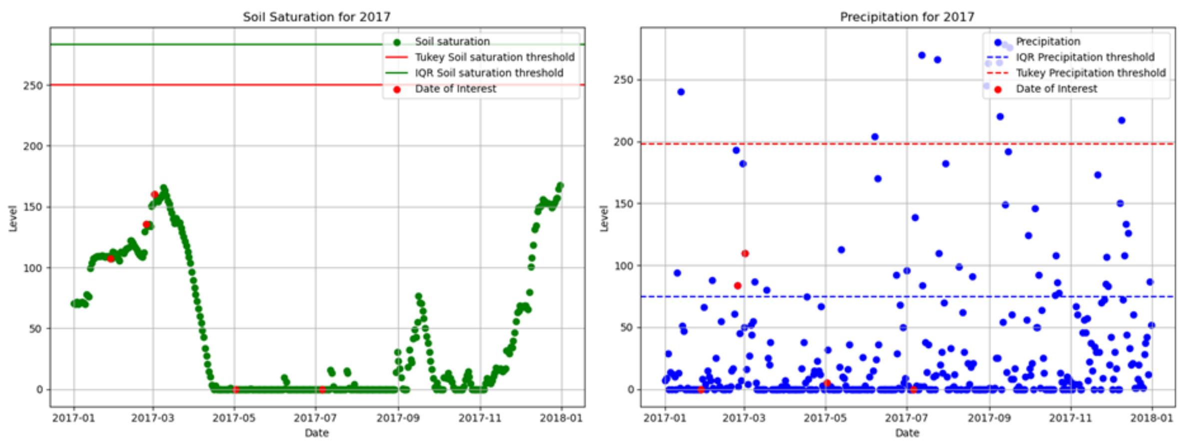


Figure 23: soil saturation (left) and precipitation (right) for the year 2018 with accident report dates

From this figure, it appears that the occurrence of a railway embankment failure is not directly related to precipitation or soil saturation. In continuation of the factor linkage, all accident reports of the study area have been assigned the values for train weight and usage intensity of the track section in which they occurred. In addition, the closest available values for the Bishop Circle safety factor F_{min} and fractal value D1 have been matched with the location and year of the accident report. The colour scales as indicated in Figure 11 are included in this table. Given that the F_{min} values are safety factors, the individual colour coding of “other outstanding result” has been assigned to values below 1. This table includes all shortlist factors that have a spatial characteristic assigned to them.

Table 12: Micro-analysis results with colour codes

Datetime	geocode	Cause code	Kilometer	Fractal value D1	Intensity	Load	Fmin
2016-03-08 15:26:33	588	207	1,5	1,20	1332	2146,318	1,14
2016-11-01 05:15:44	486	207	76	0,96	20965	9732,005	1,65
2017-01-28 14:52:30	525	207	50,8	0,79	34547	28327,2	0,96
2017-02-24 10:18:17	84	208	8,7	2,04	31314	31694,61	1,14
2017-03-02 23:42:26	79	207	22,3	1,11	0	0	1,14
2017-05-02 14:27:01	145	207	56,5	1,35	34548	38554,75	0,73
2017-07-06 06:34:43	556	207	6,5	0,60	437	377,1479	1,14
2018-02-01 15:20:57	84	207	9,3	1,15	53715	51565,92	1,14
2018-02-10 17:52:11	523	209	53,8	1,11	44953	41304,03	0,00
2020-06-03 06:25:28	145	207	56,5	1,56	719	541	0,73
2021-02-10 16:52:19	523	209	177,5	0,70	43	0	0,00
2021-02-23 14:22:11	486	209	77,1	1,04	20218	1297	1,65
2021-05-09 12:50:33	523	209	179,2	1,98	657	558	0,00
2021-11-04 11:03:23	558	209	58,05	1,38	10989	8306	0,77
2022-07-08 13:03:35	78	207	59,1	0,63	4745	3691	0,66
2023-05-11 17:32:17	79	209	3,6	1,47	22562	6148	1,14
2023-07-26 17:03:03	75	207	52,9	1,01	42053	47172	3,69
2023-08-10 12:37:38	75	207	44	0,58	36577	28843	1,24
2023-09-10 17:50:15	75	207	52,3	1,31	24016	18464	1,31

3.3.3. Investigate results `

In the resulting table from the micro-analysis, some results immediately become evident. First and foremost, no values that are above both extremity thresholds of the IQR and the adaptive Tukey fence method are found in the resulting table of the micro-analysis. There is a similarity to be found in the occurrence of extreme values in intensity and load. These two variables were expected to be correlated because an increase in track usage logically results in an increase in weight on top of the tracks. In the other spatial variables, there is no direct correlation to be found.

To assess whether a correlation is found between the temporal climate characteristics and the spatial load characteristics, several figures have been created to visualize the variable configuration found at the time and location of the accident report. An example of the combined results from the fractal value D1, estimated soil strength F_{min} and the soil saturation are plotted in Figure 24 to display the relation between the climate variables and the soil health variables for 2017, which is the year that had the most accident reports registered. All plots of these values between 2016 and 2023 have been investigated and are situated in Appendix D: micro-analysis figures.

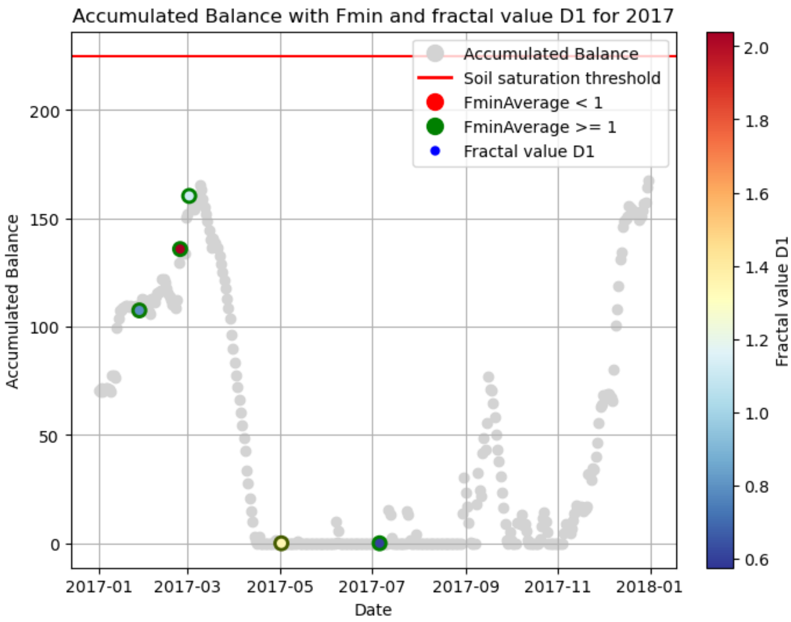


Figure 24: Soil saturation combined with Fractal value D1 and Fmin

As with the resulting table of the micro-analysis, seemingly no obvious correlations are found in the resulting figures combining climate and soil factors. Possible causes that potentially cause the lack of evidence for the relation between the shortlist factors and the occurrence of railway embankment failures are the data resolution, assumptions and estimations and influential factors that are not considered in this research. Each of these possible causes is elaborated below.

Data resolution:

The first expected reason for the absence of correlations is the data resolution. With data resolution the specificity of both the spatial and the temporal components of the data. The train variables, containing track usage intensity and train weight, are aggregated databases that take the average yearly tonnes and total yearly trains for the track branches. Daily, however, the weight and driving frequency already differ. The timetable of Nederlandse Spoorwegen (2024) changes per year, but also changes in frequency occur regularly during the year between weekends and weekdays as well as in holiday periods. Additionally, the Fractal Value D1 is a combined indication variable that is derived from the measurement train. This causes this value to be calculated between 1 and 4 times per year. These have been aggregated to yearly numbers in the dataset that has been used in this research.

Estimation errors:

The second possible cause is errors that may have occurred by including assumptions and estimations within the micro-analysis. As stated in research question two, the values for soil saturation and soil strength are estimations based on several assumptions. The soil saturation, for example, is based solely on the potential evapotranspiration as suggested by Thornthwaite (1948), using the variables temperature, rainfall and sunlight hours to estimate this number. Additionally, the soil strength is based on a generalization of the LNA results. In short, this means that the safety factor F_{min} is based on the chance of a certain soil composition in a track branch multiplied by the lowest safety factor of that soil composition. In B.2. Soil strength, this estimation is elaborated.

Other impactful factors:

Finally, the shortlist is created by a combination of data availability, expert judgement, estimated correlations between factors and expected overall impact. To fully understand the correlations between these impact factors, the entire long list of impact factors should have been matched to the accident reports on a scale as specific as possible. This way, the potential impact of all longlist factors could have been investigated. Due to data and time constraints, however, only the shortlist factors were connected to the accident reports.

3.4. What is the impact of the considered impact factors on the variation in rail embankment failures on the contract areas of ProRail?

The final research question aims to find more generalized correlations between the shortlist factors and the railway embankment failures based on macro-scale representations of the shortlist factors on the railway tracks of the Netherlands. These generalized correlations include for instance increase in usage over the years and average yearly precipitation. By visualizing the values of the shortlist factors the hypotheses of influential factors regarding railway embankment failures are tested on a macro-scale. The shortlist factors are translated to a macro-scale to assess their impact on railway embankment failures. After a visual inspection of the results, several machine learning models have been used to assist in the analyses.

3.4.1. Establish a railway map with shortlist factors

The railway map of the Netherlands is imported, after which the factors from the shortlist are linked to the track branches in several ways. Based on either Kilometering, railway branch number or geocode, integrated operations are executed to obtain feature layers that contain the factors projected on the railway branch. This allows for visual inspection of the importance of each factor regarding railway embankment failures. The shortlist factors have been translated to be suitable for the macro-analysis accordingly. The temporal components are again derived from KNMI weather stations. For each ProRail Area, the precipitation numbers are assumed to be the same for the entire area. A weather station with weather data at least from 2014 until 2023 is selected for each ProRail area. For the macro analysis, the yearly precipitation levels have been linked to all railway track branches that are within a certain ProRail area. The precipitation levels and total number of accident reports are visualized next to each other. To assess differences between the years, the precipitation numbers per year are normalized based on the entire year range. An example of this visualization is given in Figure 25. The other climate factors, days of drought and soil saturation, are not linked to the map for the macro analysis. The yearly precipitation levels are assumed to provide an insight in how wet a year has been, and therefore inherently considers the drought as well. In addition, the soil saturation estimation is heavily dependent on precipitation as well. Taking the average soil saturation per year is almost linearly correlated to the precipitation and therefore not taken into account for the macro analysis.

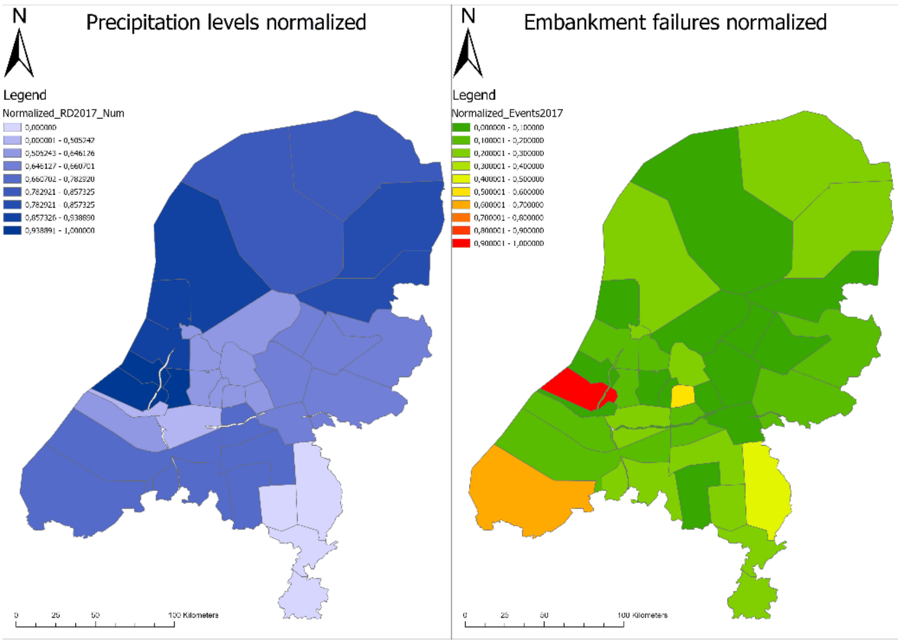


Figure 25: Precipitation and embankment failures in 2017

The spatial factors needed more extensive work to be made suitable for the macro analysis. Unlike the precipitation data, the spatial factors are linked to segments of the railway tracks and are not projected onto the ProRail areas or contract areas. For each shortlist factor with a spatial component, an elaboration is given below to indicate how the values are projected in the macro-analysis.

Soil strength

In research question two, the soil strength is derived from Bishop slip circle calculations performed in the LNA. These national results were not yet available at the time of this research. From the analysis, however, it became clear that the soil type is a highly influential component of the outcome. Therefore, the macro-analysis focuses on finding the soil types underneath the railway tracks for each geocode. These compositions have been derived with the following operations:

First, the soil type map of the Netherlands by TNO and DINOloket (2021) has been imported. This map indicates the main soil type of the first 1,20 meters depth in the Netherlands. The map has over 50 soil types in the Netherlands and the first kilometres of the North Sea. This map has been stripped to only cover the ProRail areas. Then, by performing the Spatial Join operation for each geocode, the soil compositions for each geocode are derived. The soil compositions for each contract area could also be derived because these are linked to the geocodes. Figure 26 displays the soil map of the Netherlands with the track branches and the soil composition of the contract area Utrecht.

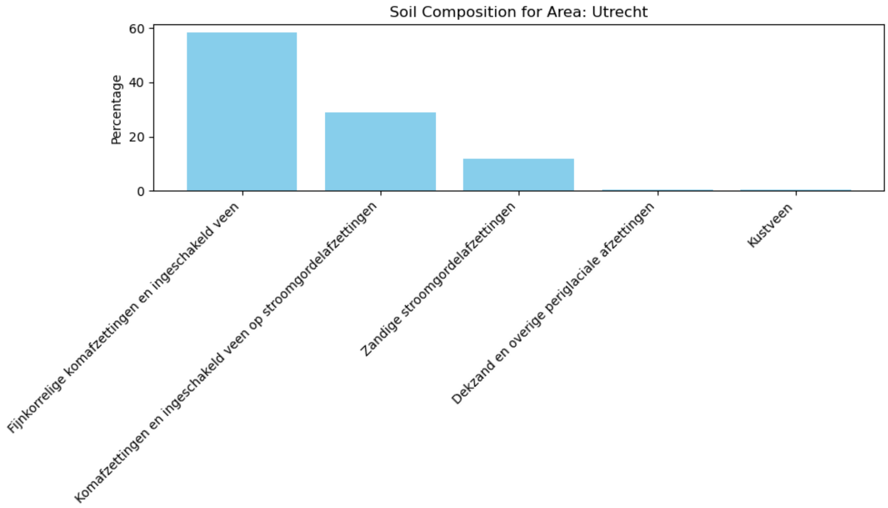


Figure 26: Soil composition and railway tracks in the Netherlands (left) and soil composition of contract area Utrecht (right)

Train weight & track usage intensity

The train weight and track usage intensity are expected to be closely related. The data processing for the macro-analysis for these variables is the same since these values are derived from the same datasets. In the micro-analysis, the kilometre indication from the accident report has been linked to the train weight and usage intensity. For the macro analysis, the values for train weight and track usage intensity have been combined for each geocode. For each geocode, two values are derived for both train weight and track usage: the mean and maximum values over the track branches in the geocode.

- Maximum yearly tonnage on a track branch within the geocode;
- Average yearly tonnage on a track branch within the geocode;
- Maximum yearly intensity on a track branch within the geocode;
- Average yearly intensity on a track branch within the geocode.

Fractal value D1

The final variable that must be processed before the execution of the macro-analysis is the fractal value D1. As with the train variables, various values can be taken into consideration for this analysis. The value for D1 becomes higher every year due to track usage and the age of the rails. Each year that no maintenance activities are performed, more anomalies will therefore occur. When the value decreases between two years, maintenance is assumed to have taken place. Since one value is given per year for this variable, the average increase over all years is determined. This is the same as the worsening per year per track branch. If the value has decreased, thus indicating maintenance, the difference between that year and the year prior has not been considered. With these calculations, the increase in D1 per track branch is obtained. As with the train variables, both the average and maximum values for each geocode are taken. The trend in the fractal value D1 is considered a better indicator then the maximum or average value over the years. Figure 27 displays the values for both variables for each geocode in the Netherlands. There is a large difference in range for the average value and the maximum value. The colours used to indicate extremity in the figures is the same, but the values that correspond to the colours are different. When considering these values, the specificity of this value against the geocodes is very important to note: there is an average and maximum degradation value for each geocode. In 250 geocodes, there are around 23.000 segments that each have an individual value for degradation. Therefore, per geocode, the average and maximum degradation value of about 90 entries is taken. To assess these values further, Table 13 below shows the mean and standard deviation of the average and maximum degradation against the total entries of the fractal value D1 dataset.

Table 13: Mean and standard deviation of generalized datasets for fractal value D1

Value	Mean	Standard deviation
Total entries of fractal value D1	0,1	0,09
Average per geocode	0,1	0,05
Maximum per geocode	0,4	0,2

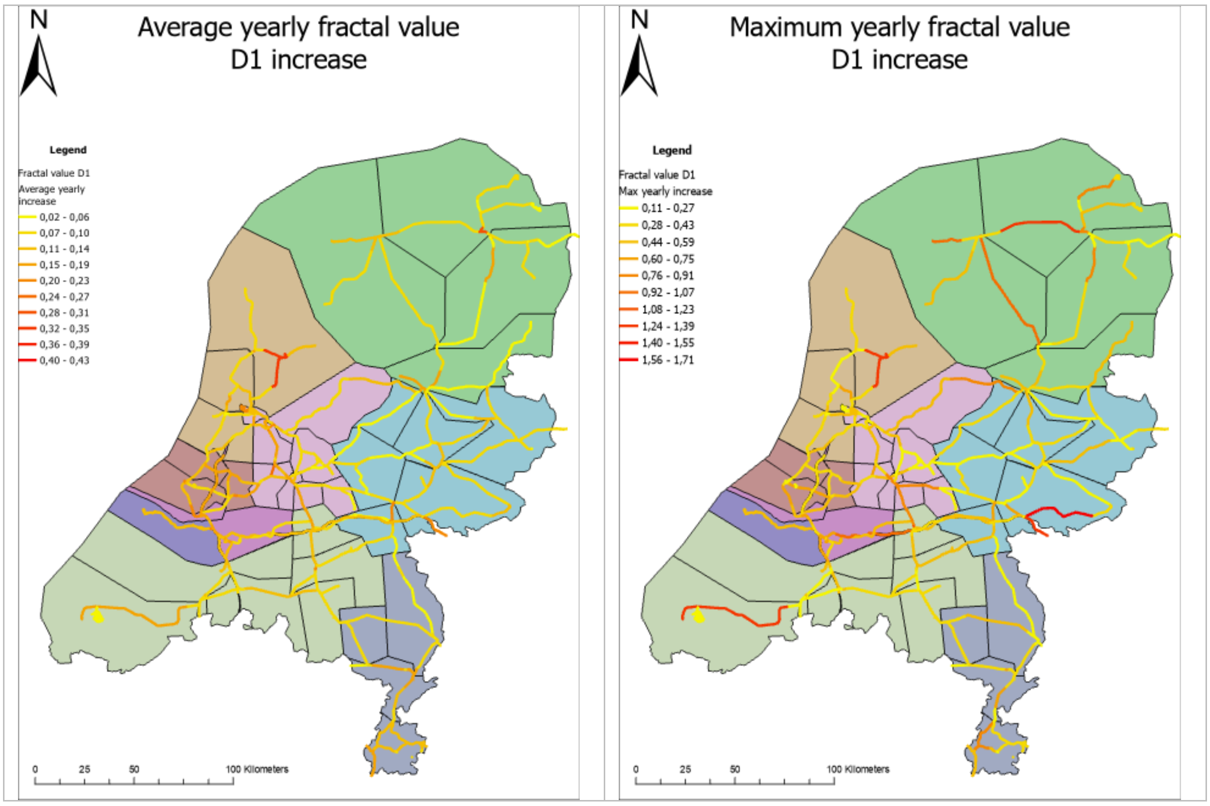


Figure 27: Average and maximum D1 increase per geocode

With the data of each shortlist variable processed to be taken into account for the macro-analysis, the data is visualized and analysed to find potential correlations and test the hypothesis that the value of the shortlist factors impacts the occurrence of embankment failures on a macro scale. The maps with factors that have been analysed are given in Appendix E: visualisations of macro-analysis. From the inspection of the maps, no direct correlation between accident reports and the shortlist factors has been found.

3.4.2. Machine learning methods

The visual inspection of the macro analysis output is followed by the utilization of machine learning models. The target feature which the models try to predict is the number of accident report per geocode per year. These numbers range from 0 to 64 accident reports, with an average of 4,47 accident reports per geocode per year. The features depicted in 3.4.1 are used to predict the values. For each of the nine used machine learning models, the R^2 value is given in Table 14 below. This value indicates how well the machine learning models are able to predict the variability of the target feature.

Table 14: R-squared value of utilized machine learning models

Machine learning method	R^2
Gradient Boosting Regressor	0,49
Random Forest	0,41
SVR	0,27
Linear Regression	0,17
Ridge Regression	0,16
Elastic Net Regression	0,15
Lasso Regression	0,12
Decision Tree	0,10
KNN Regressor	0,08

The Gradient Boosting Regressor machine learning method has the highest R-squared value of 0,49. Figure 28 shows the actual and predicted number of events using this method. The red line indicates a perfect prediction. Some large values are predicted rather accurately, but the highest peak is not explained and several misses can be identified as well.

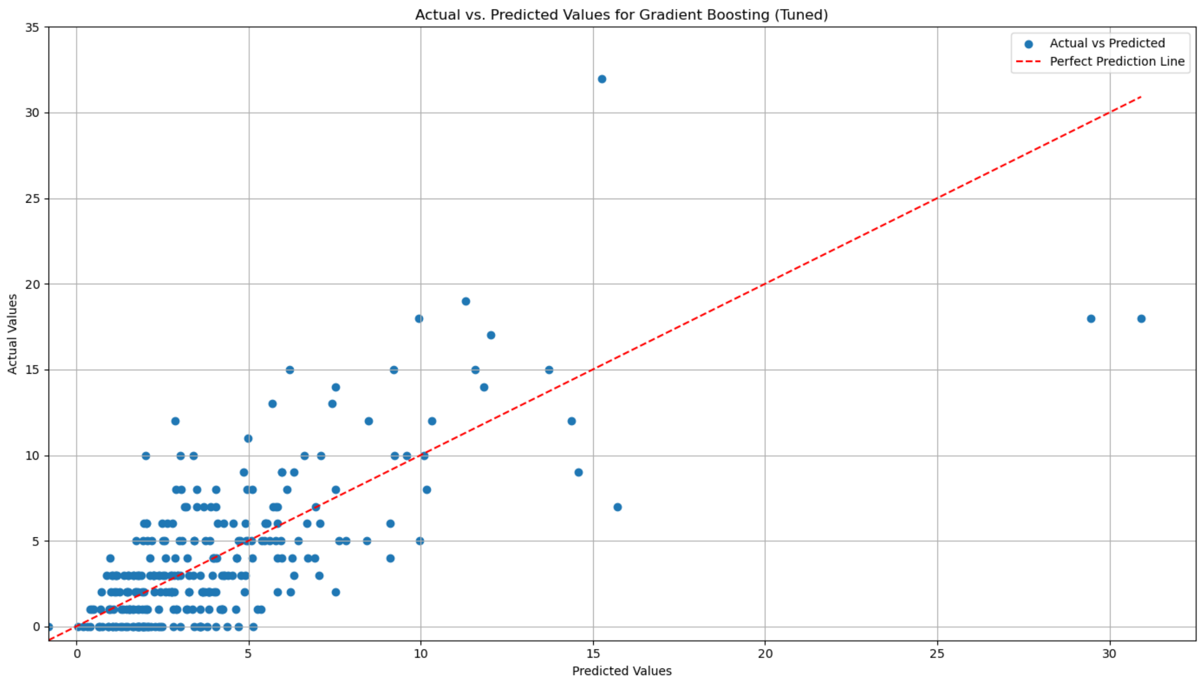


Figure 28: Actual and predicted embankment failures using Gradient Boosting Regressor

Utilizing the trial input feature, all models except for SVR and KNN obtained an R-square value near 1. That would indicate that the variance of the target feature could be fully predicted by means of the input features. The performance KNN also proved to be poor based on visual inspection of the scatterplot. Predictions based on neighbouring data points hardly outperforms a naïve model that always estimates the mean number of accident reports. The number of neighbours considered for the KNN method is the square root of the total number of data points, leading to 42 neighbours. The model is also tested with different numbers of neighbours ranging from 32 to 52, but the best R^2-value is derived with 42 neighbours.

For the shortlist factors the composite scores are given in Table 15. The aggregated soil type is considered the most important feature in this analysis. Since the soil composition is an important factor in the regular calculations of soil strength and slip circles, the large importance of the soil type is not an unexpected outcome. The large contribution of soil type can also potentially be explained due to the large number of features that are aggregated for the final soil score.

Forthcoming from the spatial and temporal component The soil type is expected to be related to a temporal factor. With this input for the machine learning models, rainfall showed to be the second most important feature. The highest fractal value D1 for a geocode in a year is the third most important feature for the prediction of railway embankment failures. In the best performing models, Random Forest and Gradient Boosting Regression, the top three most important factors also consists of soil type, rainfall and fractal value D1.

Table 15: Composite scores of top 10 impactful features for railway embankment failures

Feature	Composite importance score
Soil	0,53
Rainfall	0,17
Max increase of fractal value D1 per year	0,13
Average fractal value D1	0,07
Mean number of trains	0,03
Mean tonnes	0,03
Max tonnes	0,02
Max fractal value D1	0,02
Max number of trains	0,02

In Appendix F: machine learning combined output, the predicted and actual embankment failure numbers are given for each machine learning method. The figures include the R^2 value for each machine learning model and the most important features in that method.

3.4.3. Interpretation of results

The results of this macro analysis give some insights in the effect of the shortlist impact factors on the variance in railway embankment failures**.** The score for Gradient Boosting Regressor indicates that 49% of the variance in the number of embankment failures can be explained with this machine learning method. Some patterns are recognized, but the majority of the variance still can not be explained based on the input features. Forthcoming from the R^2 values of the machine learning models, there is a modest possibility of explaining the variance in railway embankment failures in a geocode based on the impact factors in the shortlist. When attempting to explain the variance in the occurrence of railway embankment failures in a geocode with these machine learning methods, the soil type is responsible for 53% of the model output. Rainfall and the maximum fractal value D1 on a geocode in the year of the occurrence make up 17% and 13% of the model output respectively.

4. Discussion

Before any conclusions can be made based on the results of this research, there are several limitations and decisions that must be mentioned and considered. The discussion is divided into three segments. First, the limitations based on the provided data are touched upon. Forthcoming from this, assumptions made within this research are considered. Finally, general remarks that are expected to have played a role in the outcome of this research are provided. Each section in the discussion follows the structure of the rest of the report.

4.1. Data constraints

This section elaborates on the data constraints that impacted this research. The first data constraints are found in the accident reports. In chapter 2.3.1, the accident reports related to embankment failures are analysed to determine what embankment failures are reported within the case study area. A discrepancy was found in the kilometre indication of the accident report and the indication found in the accident description. To determine the location of the accident, an excel operation has been performed to find the kilometre indication within the accident description. This method is prone to potential failure, for instance when the embankment failure spans a larger section of railway, if the indication is before the letters km or when there is text between km and the location, for instance if the text would state “km near 24.0”. Some accident reports within the case study area could therefore have been wrongly left out of the analysis.

The next data constraints that are considered are based on the long list of impact factors. For the distillation towards a shortlist, several aspects played a role among which the data availability was a key requirement for integration of an impact factor. Although some variables could be estimated, various factors were not measured nor able to be estimated in this research. These factors included vegetation, slope, urban or non-urban area and rainfall duration.

Next to factors that were not measured, there are also factors that were not available during the time of this research. The train speed is measured, but could not be obtained for the case study area other than the maximum allowed driving speed at the different track locations. Because the driving speed differs for each train and drivers can reduce the speed themselves if they are aware of poor embankment conditions, the maximum allowed speed over the track is not a proper indicator for embankment failures. The specificity of the other train variables, intensity and load, was no more precise than yearly aggregated data for each track segment. This included the average daily tonnes over the years, and the total trains that have passed during the year. These yearly aggregations are not specific enough for the railway embankment failures. In addition, there are some years of aggregated data missing. For the years 2020 and 2021, the monthly intensity and average daily tonnes were available. Manual aggregation of this data was however a tedious task, and utilizing monthly data asked for extensive data transformation. A large part of working with the available data, especially the train variables, was this transformation. As discussed in 3.2.1, there were discrepancies in capitalization, column names, table names et cetera. Transforming this data in order to automate a model to link all variables to accident reports has proven to be a time-consuming task.

Then, the data provided by the measurement train is dependent on how often the vehicle drives over the tracks. As stated in De Wit et al. (2022) the measurement train drives 1 to 4 times per year depending on some track variables like usage intensity and known issues. The fractal value D1 has been derived from calculations performed on data from the measurement train. As with the train variables weight and track usage intensity, yearly data was available. The fractal value D1 is therefore based on calculations performed on 1 to 4 measurements per year. This caused the data linkage of the temporal component of the accident report dates to fractal value D1 to be based on the year itself, because the exact date of the measurements was not known.

The LNA is being executed in the time of this research. This causes not all parts of the Netherlands to be analysed. Phase 1, being the estimation of the possible soil compositions of the rail embankments in the Netherlands, is finished. The second phase, including the slip circle calculations is not finished for all areas.

One data type that was too specific rather than too general is the soil map of the Netherlands, which has been utilized for the macro analysis. While the first map that was considered contained over 200 soil types, the geological map obtained from TNO and DINOloket (2021) still had around 40 soil types occurring within the Netherlands. The desired map would have less soil types that were somewhat more general, such as fine sand, middle sand and coarse sand. Now the overall impact of soil type has been aggregated by combining the importance score of all soil types. More ideal would be to be able to find the impact of large amounts of sand or clay and how these correlate with the other factors.

Unfortunately, such a map was not available. In addition, the soil types separately functioned as features in the macro analysis. This is due to the fact that the soil types were linked to the geocodes. This causes the soil type entries to be treated the same way as for instance axle load and fractal value D1.

4.2. Assumptions and estimations

This research project relied on several assumptions and estimations. The research was not feasible without these, but awareness of these shortcomings is very important when using or interpreting the results.

Firstly, in the filtering of the accident reports, the word 'verzak' within the description of accident reports with cause codes related to track position is considered indicative for the railway embankment failures. This assumption potentially left out some accident reports that used another verb and can cause accident reports indicating prolapsing of fences near the tracks or a guardrail to be included. Additionally, any rail embankment failure registered with a different cause code are omitted. Moreover, the operation specialists of the switch and control centre are assumed to not make spelling mistakes.

Then, regarding the extremity of values, the values that do not lie within the 98% range of each factor are considered extreme in the adaptive Tukey fence boundary calculation. This assumption is made for all factors except the soil strength safety factor.

There are two factors that have been estimated based on the available data, being the soil saturation and the soil strength. The soil saturation is assumed to be based solely on the rainfall and evapotranspiration of the soil. For the evapotranspiration of the soil, the Thornthwaite (1948) estimation formula has been used. The rainfall, temperature and sunlight hours from KNMI (2024) have been assumed to be uniform over each of the nine ProRail areas. Finally regarding the soil saturation, based on Shang et al. (2007), the soil is assumed to be completely dry after 16 days without rain. Then, the soil strength is assumed based on the outcome of LNA phase two. The safety factor of the slip circles is assumed to be a proper indication of the soil strength. To obtain one safety factor from the various slip circle options and soil composition scenarios, the chance of a certain soil composition occurring is combined with the lowest safety factor calculated for that segment.

In the macro analysis, some estimations are made to generalize the data for large-scale assessment of the embankment failures in the Netherlands. The extremity of the yearly precipitation is considered by normalizing the values over the years from 2014 to 2023. This way the different years could be visually compared. When generalizing for the large scale assessment, the average value of all years is taken as an indicator for the overall wetness of the area.

Then, to make estimations regarding the train variables in the macro analysis, four generalized values for each geocode have been derived from the dataset to use in the machine learning model. These values that are assumed to properly represent the characteristics and trends of the trains on the tracks. For these values, either the average or maximum value of the weights and intensity per year on a geocode are taken. These are thought to be the best potential indicators on embankment failures.

4.3. Other limiting factors

Some other events that could have been important for this research must be highlighted before concluding this research. First, the reduction in train usage due to COVID-19 is not taken into account in this research. For the years 2020-2021, the axle load and intensity were not available. No extra assumptions have been made regarding for instance registration of embankment failure reports in these periods.

Then, in 2021, there was a large number of registered embankment failures in the northern part of the Netherlands. This area has been prone to earthquakes with several events having a Richter scale indication of 3,0-3,5 in 2018, 2019, 2021 and 2022 (SODM, 2023). The possibility of earthquakes contributing to the occurrence of railway embankment failure is not considered in this research.

The negative impact that beavers and badgers currently have on the health of embankments and potential contributions that these animals have to railway embankment failures in the Netherlands is also not considered.

4.4. General remarks

Next to the limitations there are general remarks on the research method. The long list of impact factors is based on a literature review of railway embankment failures and expert interviews. The research could have been significantly improved by inspecting specific railway embankment failures to make additions to the long list of impact factors and also to verify the results in a later stage of the research. Due to the limited time of this research, visual inspection of actual embankment failures has not been performed.

The method of determining the extremity of the impact factors must be considered. Two boundaries are set up based on the IQR method and an adaptive Tukey fence method. This method has some shortcomings: the IQR method is traditionally used to find boundaries for normally distributed datasets. None of the considered impact factors was normally distributed. The Tukey fence method is set up to find the boundary to identify the 2% most extreme values. This method assumes that the top 2% of a factor indicates an extreme. The extremity of values is also not verified with the experts from ProRail.

With the set up of the machine learning models, it was not possible to make a test set to see how well the model could predict any occurrences of embankment failures that happened in 2024. This is because the measurement train data could not be obtained. In addition, aggregating the precipitation data of the first half year results in a seemingly dry average as usually the wettest months are August and December (KNMI, 2024). Moreover, the decision-making for machine learning methods to use is based on prior knowledge from a Data Science course and the methods that could be integrated easily into the model. An in-depth analysis of the most suitable methods and how to assess their performance is therefore not integrated in this research.

5. Conclusion

The goal of this research is to analyse the contribution of soil parameters, rolling stock variables and climate extremes to rail embankment failures. With the results of this research, a contribution to the development of data-driven asset management is made. To find the current possibility of predicting railway embankment failures based on available data, accident reports of embankment failures in the Netherlands have been analysed on a micro and macro scale. The micro scale served to find the specific cause for the occurrence of an embankment failure on a given place and time. The macro scale analysis aimed to explain the impact of different factors on the number of embankment failures in each geocode in the Netherlands.

The factors that have been used for this analysis were divided into four categories: climate, soil, train and track related factors. For each of these categories, a literature review is performed to set up the a long list of possible impact factors. This list is discussed with experts from ProRail who also supplemented additional factors to the list. A selection of these factors is taken into consideration for the micro and macro analysis based on several factors: the expected relation between the factors, data quality and availability, and expert judgement were the key contributors in the decision-making for the impact factors that are taken into account for the analyses. A shortlist of seven factors has been established with at least one factor from each impact category. The factors that have been considered are precipitation, drought periods, soil saturation, soil strength, train weight, track usage intensity and fractal value D1.

The variation in the variables is an important factor when explaining the occurrence of embankment failures based on these factors. Therefore, extremity boundaries have been set up for each variable in the shortlist by means of the IQR method and an adaptive Tukey fence method. These extremity boundaries helped in the assessment of the variable values of the cases analysed in the micro analysis. For all factors except for the safety factor F_{min} that is used to estimate to the soil strength, a higher value is detrimental for the railway embankment. With these boundaries set up, a micro scale analysis is conducted in which the specific accident reports of ProRail area Noord-West are linked to the shortlist factors. The result is a table of which the values of spatial and temporal factors are shown along with their extremity based on the IQR and Tukey fence boundaries.

From the resulting table of the micro analysis, it can not be concluded what the exact causes were for an embankment failure to occur at a certain place and time. Three possible causes are identified for this result: the impact of factors that have not been considered, the data resolution and possible errors in assumptions and estimations.

In the macro scale analysis more generalized factors such as average yearly precipitation and percentage of a certain soil type underneath a geocode to find trends in the shortlist factors and the occurrence of railway embankment failures. This way, the general impact of factors on the variation in the occurrence of embankment failures is researched. After a visual inspection of the impact factors and embankment failures projected on the Netherlands, nine machine learning methods varying in functionality have been used to find potential underlying correlations. The best performing machine learning method based on the R-squared value is the Random Forest method, which is able to explain 39% of the variance in the number of embankment failures occurrence in the Netherlands based on the impact factors used as input. The most impactful factors based on the findings of this model are the soil composition, precipitation and maximum yearly increase in fractal value D1 on a geocode. Based on a composite score that relates the importance score of a feature to the R-squared value of different methods these three features are also considered the most impactful.

Based on the results of the micro analysis, the visual inspection of impact factors on a macro scale and the utilization of various machine learning models, the answer to the main research question:

"To what extent can currently available data on impact factors be used to explain railway embankment failures on a micro and macro scale?"

Is that the currently available data on impact factors considered in only able to explain embankment failures modestly on a macro scale. For the micro scale analysis, the data resolution must be increased and additional factors need to be considered as well to explain the occurrence of embankment failures on a given place and time. On a macro scale, about half of the variance in embankment failures can be explained based on the impact factors used as input for the machine learning models.

The potential of computerized models to facilitate data-driven asset management is expected to increase when more detailed and more different factors are added to the models. More research is however necessary to facilitate predictive maintenance and data-driven decision making on railway embankments.

6. Recommendations

From the results and limitations of this study, recommendations for future research are found to improve, verify and validate the findings of this thesis. The structure of this chapter is based on the chapters in this research.

The first recommendation is regarding the standardization of the data. Uniformity in the registration of measured data regarding factors such as the names and capitalizations within columns can save time in future research. Additionally, within the accident reports, there are potential options to make it easier to distinguish between railway embankment failures and other accident reports than to lemmatize the cause description. The kilometre indications within these accident reports should be managed such that the extraction location within the accident report registration environment is the same as the indication in the cause description.

Then, regarding the long list of impact factors, there can be more factors that have an impact on the failures of railway embankments. If possible, all factors of the long list should be investigated to find the importance. To enable the analysis of all impact factors, measurements need to be executed on all factors for which there is currently no data available. In addition, the exact soil types of locations with large numbers of embankment failures should be analysed by means of for instance soil probing. Moreover, the soil saturation is recommended to be calculated by means of a more robust method that also considers factors such as soil type, runoff capacity and vegetation.

For performing both the micro- and macro analysis, a lot of time has been spent in the linkage of shortlist factors to the accident reports. For larger scale research on this topic, it is advised to automate the linkage of potentially impactful factors to the accident reports. This way, trends can be analysed on a large scale more efficiently. Moreover, registration of measurement train drive times and what factors are obtained by it can increase the robustness of these analyses. This way, the accident reports can be linked to the corresponding impact factors values obtained at the closest time point. Future research is also encouraged to use more specific data regarding the train variables. By using for instance the daily tonnes rather than yearly average values or the load per axle rather than the train load, the findings can be substantiated better. Additionally, the driving speed of the trains should be included.

The factors that have not been integrated in this research are also recommended to be taken into account for future research. This includes the impact of COVID-19, earthquakes in the northern part of the Netherlands and the long term effects of different types of maintenance on the railway tracks.

Literature

Austrian Standards. (2014). *EN 13848-5. Track geometry quality - Part 5: Geometric quality levels - Plain line, switches and crossings*.

Blanco, H., & Lal, R. (2023). *Soil Conservation and Management*. Springer Nature.

Chen, L., Liu, C., Li, Y., & Wang, G. (2007, 2007/03/01). Impacts of climatic factors on runoff coefficients in source regions of the Huanghe River. *Chinese Geographical Science, 17*(1), 047-055. https://doi.org/10.1007/s11769-007-0047-4

Chicco, D. A.-O., Warrens, M. A.-O. X., & Jurman, G. A.-O. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. (2376-5992 (Electronic)).

Ciotlaus, M., Kollo, G., Moldovan, D., & Muntean, L. (2017, 2017/01/01/). Slope Stability of Railway Embankments. *Procedia Engineering, 181*, 52-59. https://doi.org/https://doi.org/10.1016/j.proeng.2017.02.362

Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023, 2023/03/01/). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal, 6*, 100164. https://doi.org/https://doi.org/10.1016/j.dajour.2023.100164

Dastjerdy, B., Saeidi, A., & Heidarzadeh, S. (2023). Review of Applicable Outlier Detection Methods to Treat Geomechanical Data. *Geotechnics, 3*(2), 375-396.

De Wit, N., Vermeulen, F., & ProRail. (2022). *Summer School Meettreinen*.

EPSG. (2024). *EPSG: 28992; Amersfoort / RD New -- Netherlands - Holland - Dutch*. epsg.io/28992

Ibrahim, E., Shouman, M., Torkey, H., & El-Sayed, A. (2021, 05/01). Handling missing and outliers values by enhanced algorithms for an accurate diabetic classification system. *Multimedia Tools and Applications, 80*, 1-23. https://doi.org/10.1007/s11042-021-10727-0

KNMI. (2024). *Dagwaarden neerslagstations*. https://www.knmi.nl/nederland-nu/klimatologie/monv/reeksen

Lancefield, N. (2024, 4-4-2024). £ 2,8bn to protect railway as record rainfall takes toll. *Metro*.

Lancellotta, R. (1995, 01/01). Geotechnical Engineering.

Li, L. A.-O., Yu, Q., Gao, L., Yu, B., & Lu, Z. The Effect of Urban Land-Use Change on Runoff Water Quality: A Case Study in Hangzhou City. LID - 10.3390/ijerph182010748 [doi] LID - 10748. (1660-4601 (Electronic)).

Nederlandse Spoorwegen. (2022). *Reizigersgedrag 2022*. https://dashboards.nsjaarverslag.nl/reizigersgedrag

Nederlandse Spoorwegen. (2024). *Dienstregeling*. https://www.ns.nl/reisinformatie/dienstregeling

Páez, A., & Boisjoly, G. (2022). Exploratory Data Analysis. In A. Páez & G. Boisjoly (Eds.), *Discrete Choice Analysis with R* (pp. 25-64). Springer International Publishing. https://doi.org/10.1007/978-3-031-20719-8_2

ProRail. (2021). *Instandhoudingsspecificatie spoor, deel 1; Onderhoudswaarden, Interventiewaarden, Onmiddelijke actiewaarden*.

ProRail. (2022a). *Jaarverslag 2022 ProRail*. https://www.prorail.nl/siteassets/homepage/over-ons/documenten/prorail-jaarverslag-2022.pdf

ProRail. (2022b). *Samen Nederland vooruit helpen*. https://www.prorail.nl/

ProRail. (2023). *Tracébesluit PHS Alkmaar – Amsterdam ter inzage*. https://www.prorail.nl/nieuws/tracebesluit-phs-alkmaar--amsterdam-ter-inzage#:~:text=Het%20spoor%20tussen%20Alkmaar%20en%20Amsterdam%20en%20bij%20Heerhugowaard%20wordt,trac%C3%A9besluit%20gepubliceerd%20in%20de%20Staatscourant.

ProRail. (2024). *Onderzoek Baanstabiliteit*. https://www.prorail.nl/programmas/onderzoek-baanstabiliteit

Rahmati, M., Groh, J., Graf, A., Pütz, T., Vanderborght, J., & Vereecken, H. (2020, 2020/01/01). On the impact of increasing drought on the relationship between soil water content and evapotranspiration of a grassland. *Vadose Zone Journal, 19*(1), e20029. https://doi.org/https://doi.org/10.1002/vzj2.20029

Ramke, H.-G. (2018). 8.2 - Collection of Surface Runoff and Drainage of Landfill Top Cover Systems. In R. Cossu & R. Stegmann (Eds.), *Solid Waste Landfilling* (pp. 373-416). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-407721-8.00019-X

RijdenDeTreinen. (2023). *Archief Treinstoringen; 15 november 2022 - 15 november 2023*. https://www.rijdendetreinen.nl/storingen/archief?lines=&reasons=&date_before=15-11-2023&date_after=15-11-2022

Roshan, M. J., A Rashid, A. S., Abdul Wahab, N., Tamassoki, S., Jusoh, S. N., Hezmi, M. A., Nik Daud, N. N., Mohd Apandi, N., & Azmi, M. (2022, 2022/11/01/). Improved methods to prevent railway embankment failure and subgrade degradation: A review. *Transportation Geotechnics, 37*, 100834. https://doi.org/https://doi.org/10.1016/j.trgeo.2022.100834

RSBB. (2011). *The effects of railway traffic on embankment stability*. https://www.networkrail.co.uk/wp-content/uploads/2017/02/RSSB-report-The-effects-of-railway-traffic-on-embankment-stability-1.pdf

Schuurmans, L., & ProRail. (2019). *Fractal analysis als toetsmethode voor het draagvermogen van het baanlichaam*.

Shang, K., Wang, S., Ma, Y., Zhou, Z., Wang, J., Liu, H., & Wang, Y. (2007, 05/30). A scheme for calculating soil moisture content by using routine weather data. *Atmospheric Chemistry and Physics, 7*. https://doi.org/10.5194/acpd-7-7451-2007

SODM. (2023). *Voormalige gaswinning Groningen-gasveld*. M. v. E. z. e. Klimaat. https://www.sodm.nl/sectoren/gaswinning-groningen#:~:text=Met%20een%20totaal%20van%209,op%2012%20april%20bij%20Zijldijk.

SpoorPro. (2018). Dodelijke treinramp Weesp van 1918 herdacht. *SpoorPro*. https://www.spoorpro.nl/materieel/2018/09/14/dodelijke-treinramp-weesp-van-1918-herdacht/

Srivastava, D. (2023, 04/04). An Introduction to Data Visualization Tools and Techniques in Various Domains. *International Journal of Computer Trends and Technology, 71*, 125-130. https://doi.org/10.14445/22312803/IJCTT-V71I4P116

Steward, T., Sivakugan, N., Shukla, S. K., & Das, B. M. (2011, 2011/08/01). Taylor's Slope Stability Charts Revisited. *International Journal of Geomechanics, 11*(4), 348-352. https://doi.org/10.1061/(ASCE)GM.1943-5622.0000093

Suryakanthi, T. (2020, 01/01). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*. *International Journal of Advanced Computer Science and Applications, 11*. https://doi.org/10.14569/IJACSA.2020.0110277

Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review, 38*(1), 55-94. https://doi.org/10.2307/210739

TNO, & DINOloket. (2021). *Geologische kaart 2021*. https://www.dinoloket.nl/ondergrondmodellen/kaart

Verloop, R. (2019). *Introductie techniekveld Baan*

Verloop, R. (2024). *Interview with Remco Verloop, Civil Engineering Specialist* [Interview].

Wickham, H. (2016). Data Analysis. In H. Wickham (Ed.), *ggplot2: Elegant Graphics for Data Analysis* (pp. 189-201). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_9

Appendices

Appendix A: machine learning models used

For this research, nine machine learning models have been used: linear regression, ridge regression, lasso regression, elastic net regression, decision trees, random forest, gradient boosting regression, support vector regression and k nearest neighbours. These models can be divided into two general types: models with direct importance scores for features and models with indirect importance scores for features. The linear and tree-based models have a direct importance score, where either the linear coefficient or the reduction in Gini impurity or information gain in tree models indicates the importance. For the other models, permutation feature importance is used to find the importance factor of the different features. The following sections provide an explanation of the functionality of the used models.

A.1. Linear regression

The first method, linear regression, assumes that there is a linear relationship to be found between the features and the target variable. The functionality of the model is to fit a linear equation. The optimization of the model is performed by minimizing the sum of the squared difference between the observed and the predicted values. The importance factor that is given to each feature in this model is equal to the coefficient in the linear equation.

A.2. Ridge regression

Like linear regression, ridge regression attempts to find a linear relationship between the features and the target variable. Ridge regression incorporates a regularization term, also referred to as an L2 penalty, to prevent overfitting. The regularization term is equal to the magnitude of coefficients to a loss function squared. With using soil types as features, this method is expected to perform much better than linear regression. The importance factor for ridge regression is also the coefficient of the linear function.

A.3. Lasso regression

Least Absolute Shrinkage and Selection Operator Regression, or lasso regression, performs an operation similar to ridge regression. The difference is in the penalty type: lasso regression uses an L1 penalty which is the absolute value of the magnitude of coefficients. Because L1 penalty can set values to zero, feature selection is inherently executed by this model type. For this method as well, the importance value for each feature is equal to the coefficient in the linear function.

A.4. Elastic Net Regression

The last linear regression model, elastic net regression, linearly combines the L1 and L2 penalty derived from the previous two methods to derive a linear relationship for the input features with regard to the target feature. The concept of combining the L1 and L2 penalty is to overcome the limitations of the Ridge and Lasso regression methods. If the penalty L1 accidentally negates the impact of a feature that should not have been neglected, the combined penalty does consider the feature. The feature importance value in Elastic Net Regression is the coefficient in the linear function.

A.5.Decision trees

The first nonlinear function that is used is decision trees. The data is split into subsets based on the feature values. The features are represented by nodes, for which decision rules are represented by branches. The outputs are called leaf nodes. Therefore, by following the decision trees, estimations can be made both regarding classification and regression. The importance of features in this method is related to the impact that they have on the Gini index. The Gini index is probability index for a random instance to be misclassified when chosen randomly (Suryakanthi, 2020).

A.6. Random forest

By creating a multitude of decision trees and merging the results, the random forest machine learning algorithm attempts to reduce the possibility of overfitting and increase the accuracy of the overall model. The importance function is a composite Gini index influence of the feature. The output of the random forest algorithm is expected to outperform decision trees.

A.7. Gradient boosting regression

Gradient boosting regression is also an algorithm that makes use of decision trees. Multiple models are built, in which errors by one model are corrected by the next model. A loss function is incorporated and optimized by adding weak learners from decision trees. While random forest reduces overfitting, gradient boosting regression is rather prone to overfitting.

A.8. Support vector regression

This regression method uses the support vector machines principle to find a hyperplane that can be used to fit the data within certain thresholds. Support vector machines finds an optimal line or hyperplane in which the distance between data points with different classifications is as large as possible. By combining multiple support vector machines, this classification method can be used for regression problems. The importance function for each feature in this algorithm is based on permutation feature importance, which is a feature selection method that adds the features in a randomized order to the dataset and finds the improvement caused by the addition of each feature.

A.9. K – nearest neighbours

Finally, k-Nearest Neighbours makes predictions on the outcome of a feature combination based on the closest training examples within the feature space. This makes k-nearest neighbours non parametric. The output is dependent on the chosen number K, being the number of closest training examples used. Usually, the square root of the number of feature entries is taken for K.

While KNN is often used for classification, averaging the values of the neighbours allows for the usage in regression problems. As with SVR, the importance function is dependent on permutation feature importance.

Appendix B. Derivation of soil saturation and soil strength

B.1. Soil saturation

From the literature review of research question one, it has been concluded that soil saturation is dependent on many variables and soil characteristics. The variables for the Thornthwaite formula include average monthly temperature, monthly sun hours and daily precipitation numbers. These values are derived from KNMI. This estimation neglects factors like vegetation and wind for the runoff capacity. Since rail embankments are generally above ground, adjacent water bodies and groundwater tables can be omitted for the estimation of soil saturation. The Thornthwaite formula provides an estimation of the evapotranspiration in the soil rather than a direct indication of the soil saturation. Rahmati et al. (2020) state evapotranspiration is the main cause of soil desaturation. Therefore, the precipitation and potential evapotranspiration combined are assumed as an estimation for the soil saturation level. Shang et al. (2007) state that after 16 days of drought, the soil is completely unsaturated. Combining the above factors, the saturation level of the soil is calculated in millimetres on a certain day with the following statement:

$$PET_{Thorn} = \begin{cases} 0, & \text{If T} < 0 \\ 16 \cdot \dfrac{N}{360} \cdot \left(\dfrac{10 \cdot T}{I}\right)^{a}, & \text{If } 0 \leq \text{T} \leq 26 \\ \dfrac{N}{360} \cdot (-415.85 + 305{,}332.24 \cdot T - 0.43 \cdot T^2), & \text{If T} > 26 \end{cases}$$

$$I = \sum_{Jan}^{Dec} \left(\frac{Max[0, T_m]}{5}\right)^{1{,}514}$$

$$a = (6.75 \cdot 10^{-7} \cdot \text{I}^3) - (7.71 \cdot 10^{-5} \cdot \text{I}^2) + (0.01792 \cdot \text{I}) + (0.49239)$$

$$Sat_{dayY} = \begin{cases} 0 & \text{If } \text{D}_{\text{dry}} \geq 16 \\ \displaystyle\sum_{i=1}^{Y_i} RD_i - PET_{Thorn,i} & \text{If } \text{D}_{\text{dry}} < 16 \end{cases}$$

In the following formula:

N = sunlight duration in hours (based on KNMI)

T = average daily air temperature

I = heat index

T_m = average monthly temperature

RD = precipitation

D_{dry} = consecutive number of days with no precipitation

B.2. Soil strength

In the Landelijke Netwerk Analyse (LNA), for all railway embankments in the Netherlands, the subsurface has been investigated. From this investigation, several possible soil profiles have been determined for each railway segment. Each of these profiles has an occurrence probability. For each profile and each segment, calculations have been executed to find the safety factor regarding critical bishop circles in various scenarios. The LNA consists of three phases. In phase 1, a rough estimation is made of the safety factors and soil profiles. From this worst-case scenario, any segment that has a safety factor above 1 is safe to such an extent that it does not have to be investigated further for phase 2. During the execution of this research, phase 1 of the LNA is finished and phase 2 is being set up. Since the LNA is a large project, multiple companies have been involved to obtain results of this analysis. In this Appendix, the process of diluting these soil profiles and chances into a single value that can be linked to the location of a rail embankment failure is explained.

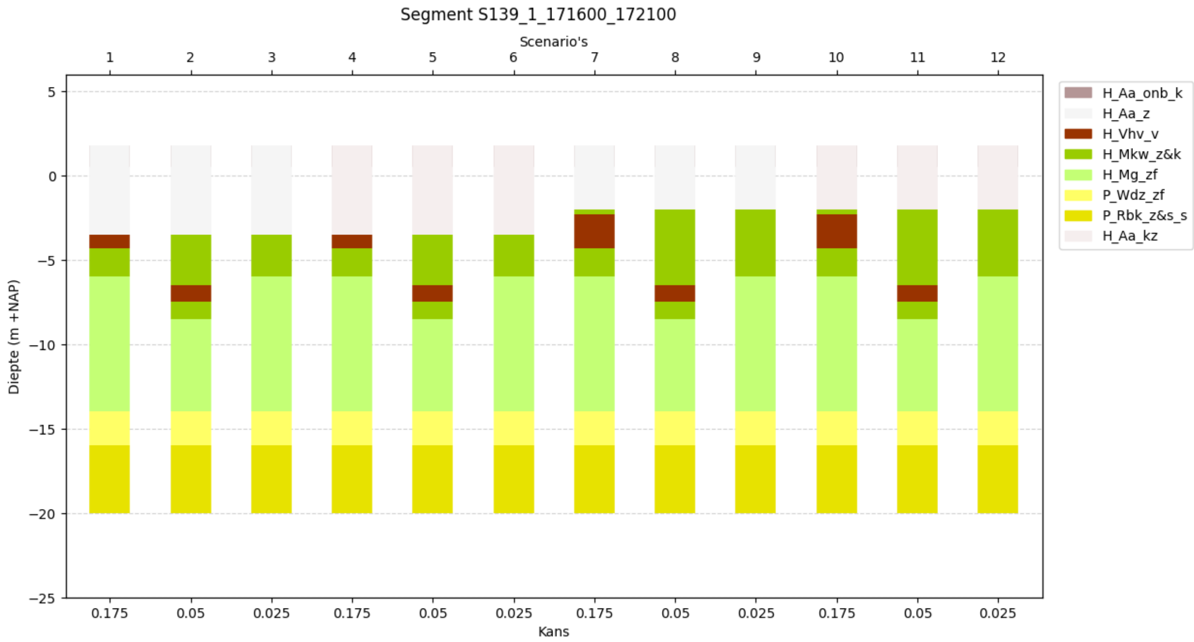
B.2.1. Soil profiles:

In the case study area, Royal Haskoning DHV provided ProRail with the soil profiles and results of the geo-stability calculations. Only GeoCode 620, which includes Amsterdam Centraal, is calculated by a different company. The results for phase 1 calculations of the other GeoCodes in the case study area have been distributed for this thesis. For each GeoCode, three folders are obtained. The first folder contains the input for soil profile estimations. The second folder contains the output values for the soil profile estimations. The output folder contains two files; the first is a table with the segment IDs, soil profile IDs and probability for each soil profile. An example of such a segment is as follows:

segment_id	soilprofile_id	probability
S139_1_171600_172100	S139_1_171600_172100_1D1	0,175
S139_1_171600_172100	S139_1_171600_172100_1D2	0,05
S139_1_171600_172100	S139_1_171600_172100_1D3	0,025
S139_1_171600_172100	S139_1_171600_172100_1D4	0,175
S139_1_171600_172100	S139_1_171600_172100_1D5	0,05
S139_1_171600_172100	S139_1_171600_172100_1D6	0,025
S139_1_171600_172100	S139_1_171600_172100_1D7	0,175
S139_1_171600_172100	S139_1_171600_172100_1D8	0,05
S139_1_171600_172100	S139_1_171600_172100_1D9	0,025
S139_1_171600_172100	S139_1_171600_172100_1D10	0,175
S139_1_171600_172100	S139_1_171600_172100_1D11	0,05
S139_1_171600_172100	S139_1_171600_172100_1D12	0,025

This segment_id is called S139_1_171600_172100. This is read as: GeoCode 139, from kilometer 171,600 until 172,100. For this segment_id, twelve alternatives for the soil profiles are found. They have the same name as the segment_id, followed by the alternative number (e.g. D1 or D7). Each profile has an occurrence probability. This is found in the right-most column. A probability of 0,175 corresponds to an occurrence probability of 17,5%.

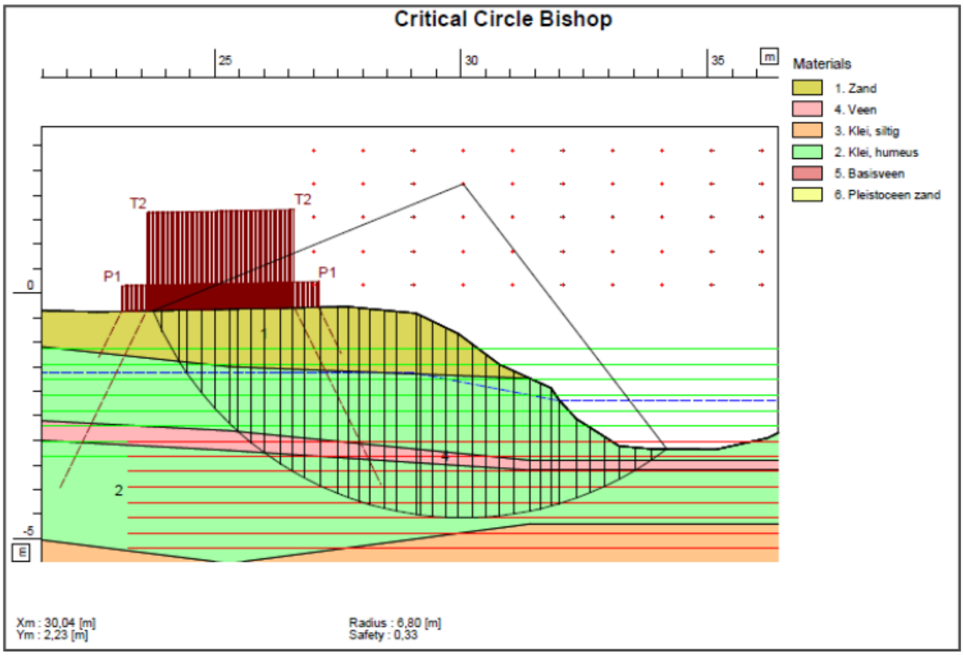
Then, for each segment, a graph is created that visualizes the soil compositions of these profiles. The graph that corresponds to the above segment_id is given below. In addition, a table can be found with the exact numbers of each profile. An example of the first soilprofile (D1) is given in the table below.



soilprofile_id	bottom_level	soil_name
S139_1_171600_172100_1D1	0.5	mv
S139_1_171600_172100_1D1	1.8	H_Aa_onb_k
S139_1_171600_172100_1D1	-3.5	H_Aa_z
S139_1_171600_172100_1D1	-4.3	H_Vhv_v
S139_1_171600_172100_1D1	-6.0	H_Mkw_z&k
S139_1_171600_172100_1D1	-14.0	H_Mg_zf
S139_1_171600_172100_1D1	-16.0	P_Wdz_zf
S139_1_171600_172100_1D1	-20.0	P_Rbk_z&s_s

B2.2. Slip circle safety factors

With the soil profiles for each segment, calculations have been executed to find the safety factors regarding embankment stability. The third folder within each Geocode map contains the calculation results. Internal loads, phreatic levels, creek water height (if applicable) height geometries and more input values are used to calculate the safety factor. Multiple calculations are executed to obtain the safety factors with various load cases. The load cases consider for instance the left- or right hand side of the embankment, and the type of slip circle. An example of a critical bishop circle on the right hand side in a midpoint circle failure scenario is given in the figure below. As with the probabilities of soil profile occurrence, each of the load case has an occurrence probability.



B2.3. Integration in data analysis

The various soil profiles and calculations need to be reduced into one number in order to make them useful for the data analysis. Therefore,

The first step is to aggregate the safety factors of the various safety calculations. Because this analysis focuses on rail embankment failures that have occurred, the lowest safety factor has been used for each soil profile. In some cases, the safety factor of one of these soil profiles was -999. Because this would be an extreme outlier for the statistical analysis and a safety factor below zero is not possible, the safety factor in these cases is set to 0.

With each soil profile having a safety factor, the next step is to aggregate the various soil profiles for a segment. The number of possible soil profiles is different for each segment. Therefore, the aggregated safety factor of each segment is calculated with the following formula:

$$F_{segment} = \sum_{i=1}^{N} P_i \times F_{min,i}$$

In which *N* is the number of soil profiles of the segment, P is the occurrence probability of the soil profile and F_{min} is the lowest safety factor of the calculations. If the lowest safety factor of all possible soil profiles was taken in the analysis, a profile with 0,5% occurrence probability and a safety factor of 0 would cause the entire segment to be considered unsafe. Therefore, the probability and minimal safety factor of each soil profile are taken for the aggregated safety factor.

Appendix C: statistical results of variables

C.1. Temperature variables

	Rain (0,1mm/day)	Daily average Temperature (0,1°C)	Estimated soil saturation (0.1mm)
Mean	27.20	113,5	63,62
Standard Deviation	52,62	56.36	80,80
Min	0.00	-65,00	0.00
Max	511,0	295,00	284,06
Median	2	111,00	13,83
IQR	30,00	89,00	25,66
IQR lower bound	-45,00	-62,5	-170,19
IQR upper bound	75,00	293,0	283,71
# of extremes (IQR)	408/3266	2/3266	2/3266
Tukey fence K value	5,61	0,91	1,21
Tukey fence lower bound	-168,30	-10,87	-137,31
Tukey fence upper bound	198,30	241,87	250,80
# of extremes (TF)	65/3266	64/3266	65/3266

C.2. Train characteristics

Note that only datasets have been derived with average daily tonnes and track usage intensity the following 2015-2019, 2022 and 2023.

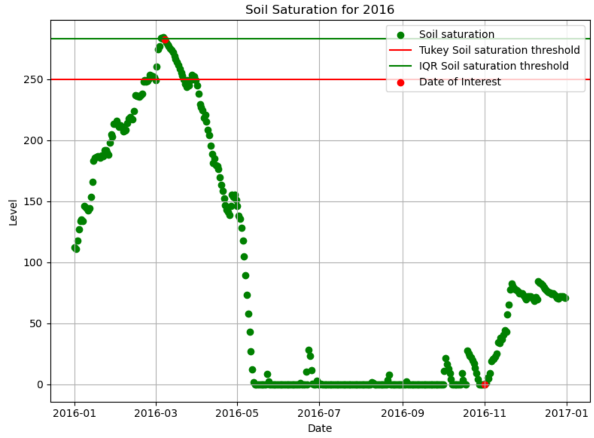
	Daily tonnes	Usage intensity
Count	56630	56630
Mean	14224	15096
Standard Deviation	17235	17307
Min	0	0
Max	109944	127903
Median	6273	7078
IQR	21235	24550
IQR lower bound	-39171	-35645
IQR upper bound	62847	62555
# of extremes (IQR)	1105	841
Tukey fence K value	1,88	1,42
Tukey fence lower bound	-38746	-33681
Tukey fence upper bound	62421	60591
# of extremes (TF)	1132	1109

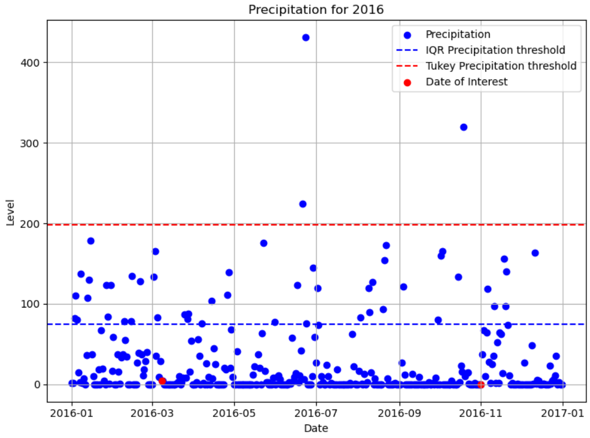
C.3. Fractal value D1 and yearly increase D1

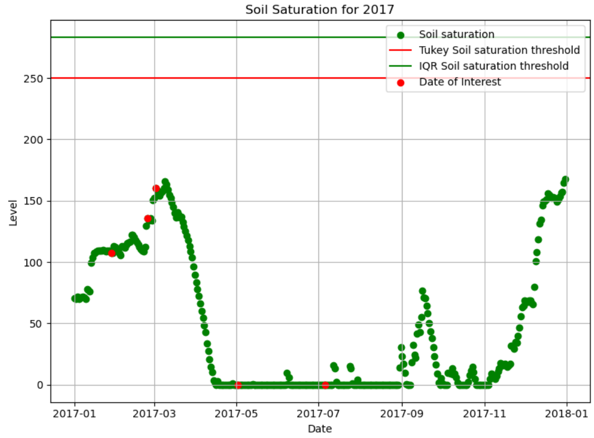
	Average fractal value D1	Increase in D1 (per geocode)
count	175482	903
mean	1,23	0,45
std	0,52	0,25
min	0,30	0
0,25	0,84	0,29
0,5	0,11	0,40
0,75	0,15	0,53
max	3,35	1,71
IQR	0,67	0,24
Lower bound	-0,17	-0,07
Upper bound	2,53	0,89
outliers	4048	51
Tukey K value	1,58	3,22
Tukey Lower bound	-0,26	-0,46
Tukey Upper bound	2,54	1,28
Tukey Outliers	3518	22

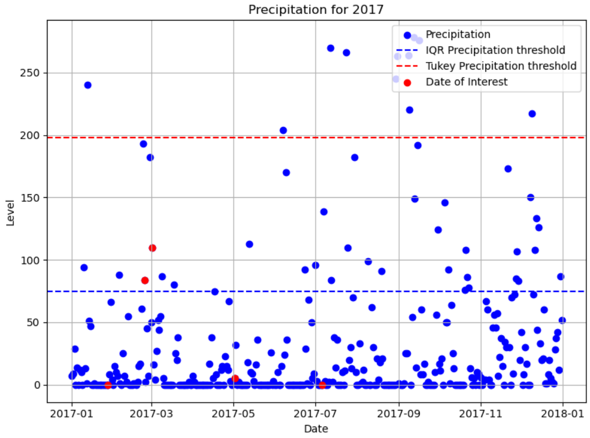
Appendix D: micro-analysis figures

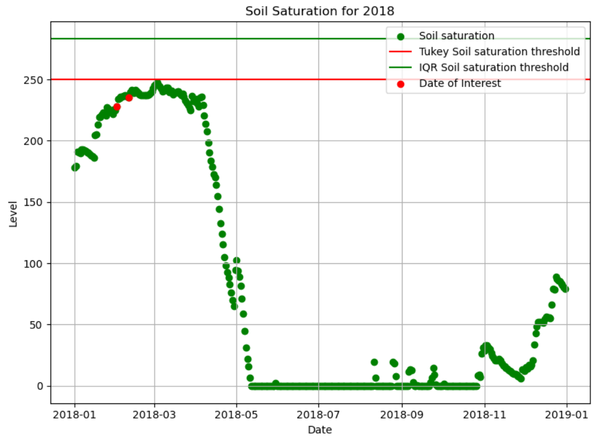
D.1. Rainfall, soil saturation and embankment failure

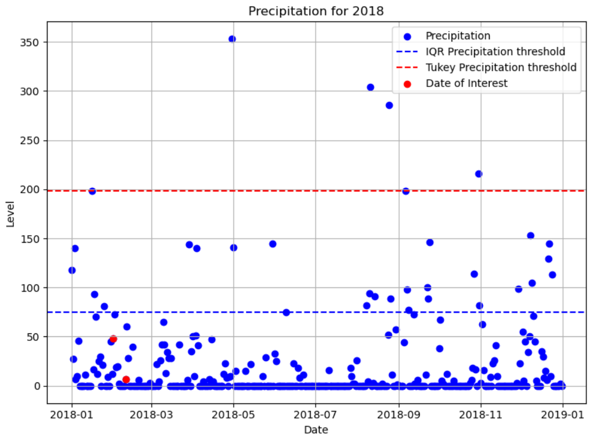


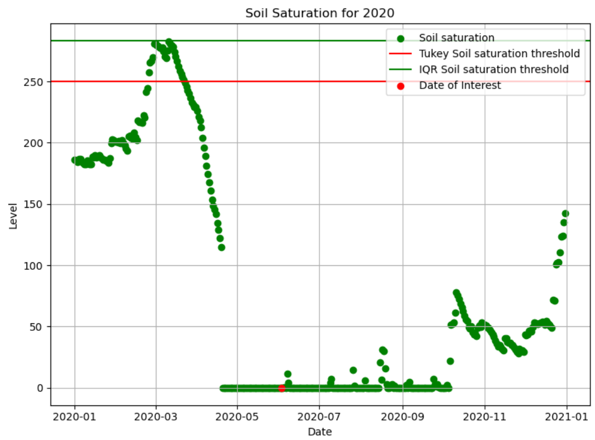


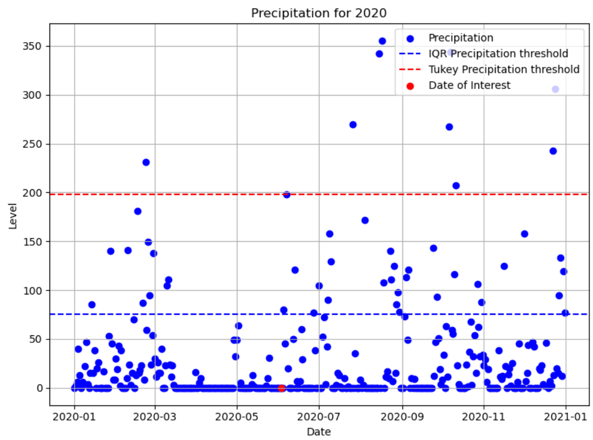


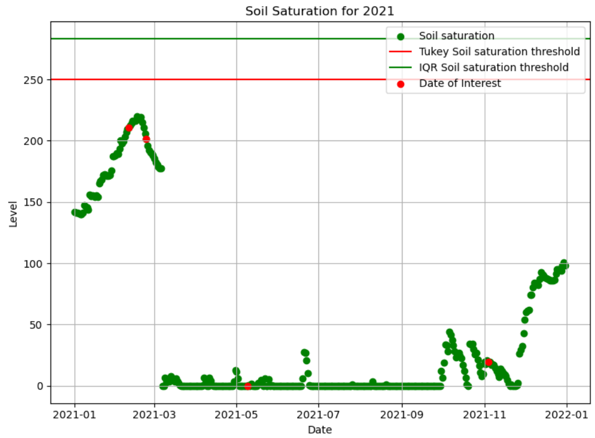


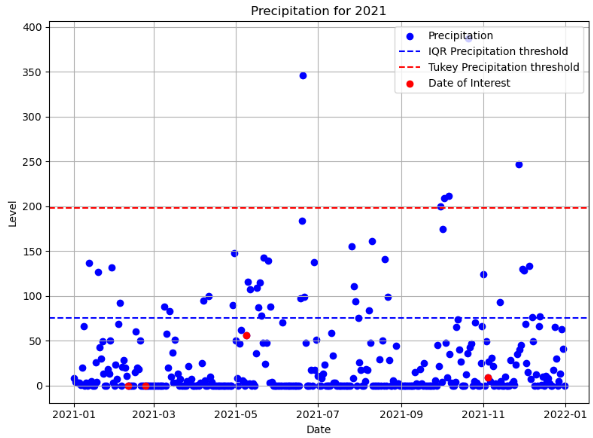


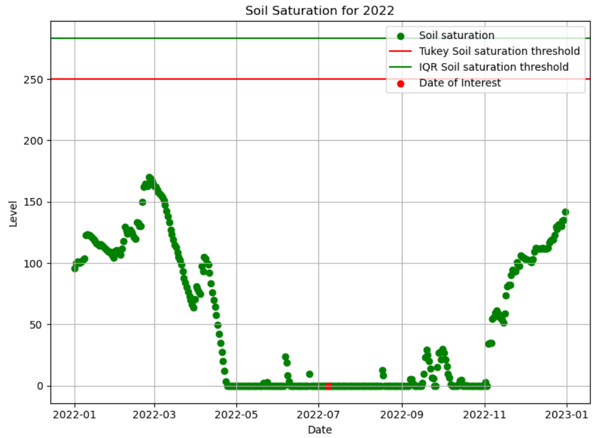


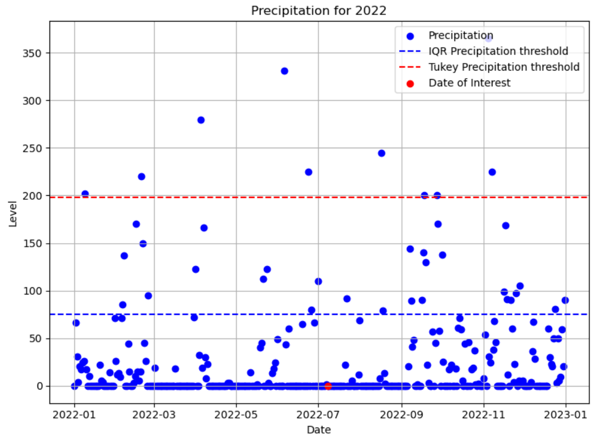

Soil Saturation for 2020
Soil saturation
Tukey Soil saturation threshold
IQR Soil saturation threshold
Date of Interest
Level
Date

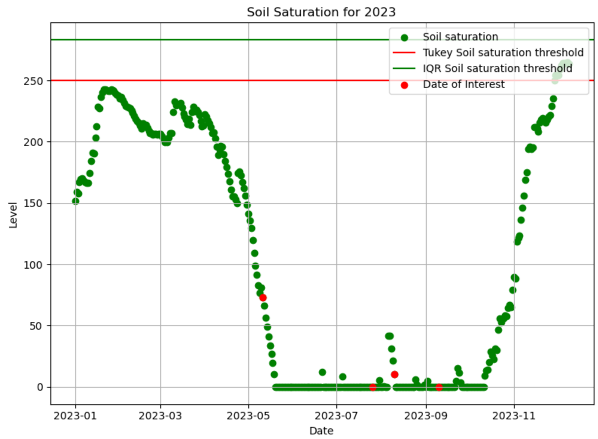

Precipitation for 2020
Precipitation
IQR Precipitation threshold
Tukey Precipitation threshold
Date of Interest
Level
Date

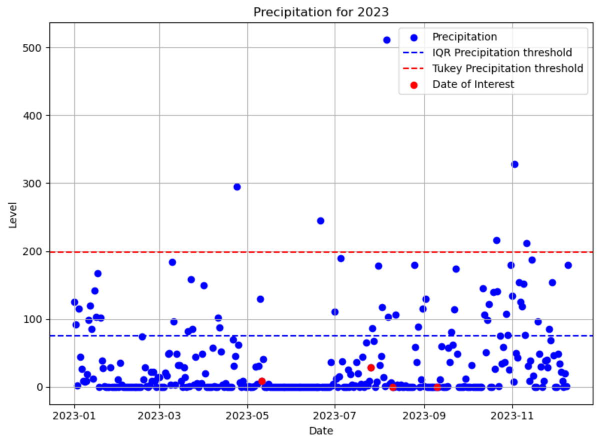

Soil Saturation for 2021
Soil saturation
Tukey Soil saturation threshold
IQR Soil saturation threshold
Date of Interest
Level
Date


Precipitation for 2021
Precipitation
IQR Precipitation threshold
Tukey Precipitation threshold
Date of Interest
Level
Date

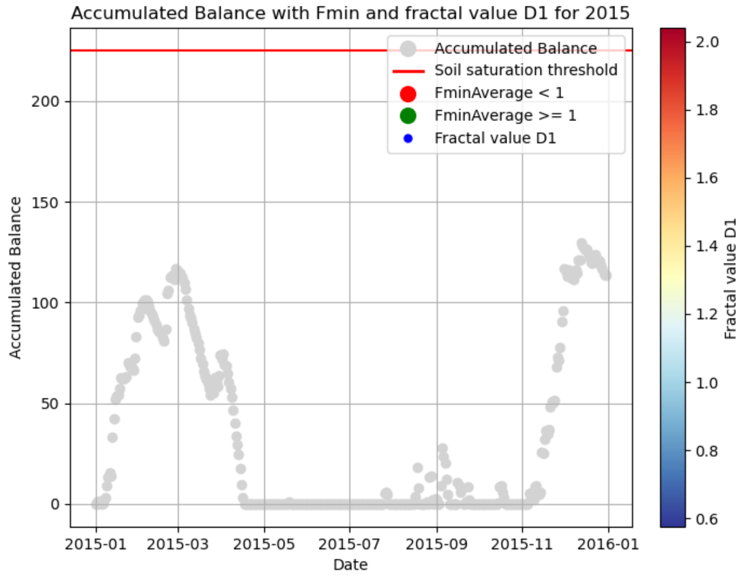

Soil Saturation for 2022
Soil saturation
Tukey Soil saturation threshold
IQR Soil saturation threshold
Date of Interest
Level
Date

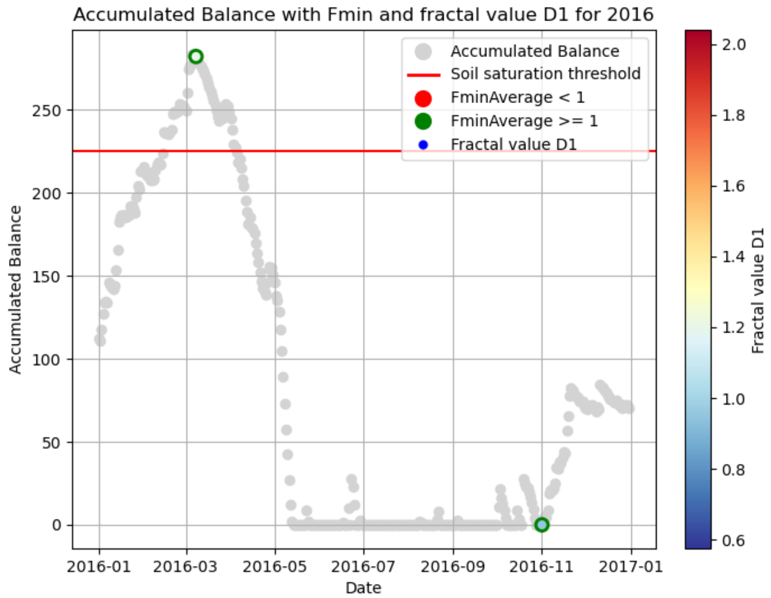

Precipitation for 2022
Precipitation
IQR Precipitation threshold
Tukey Precipitation threshold
Date of Interest
Level
Date

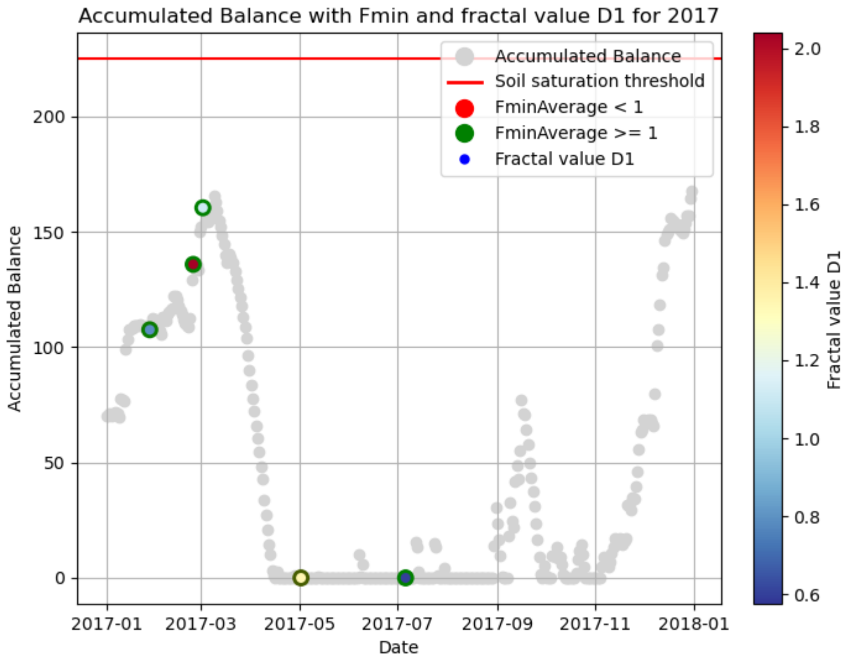

Soil Saturation for 2023
Soil saturation
Tukey Soil saturation threshold
IQR Soil saturation threshold
Date of Interest
Level
Date

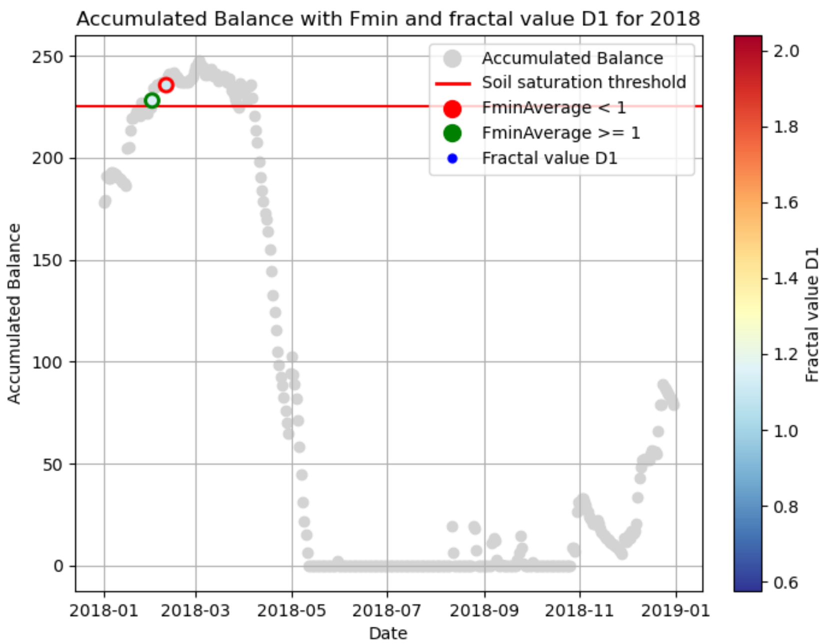

Precipitation for 2023
Precipitation
IQR Precipitation threshold
Tukey Precipitation threshold
Date of Interest
Level
Date

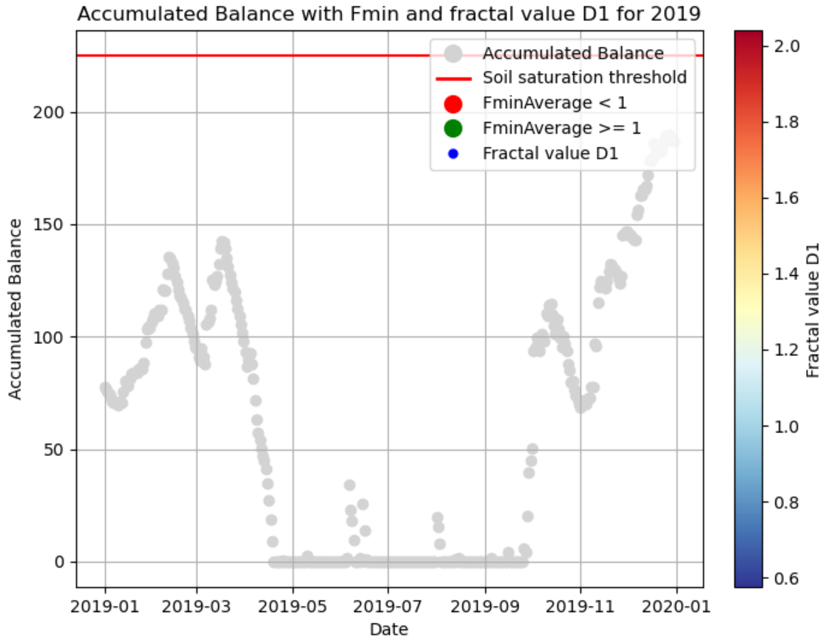
D.2. Fractal value, Fmin and soil saturation



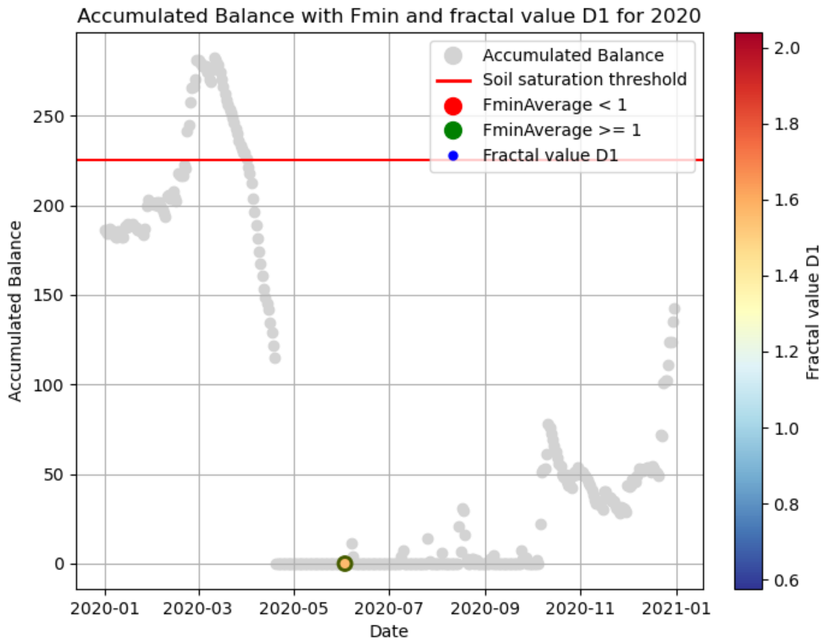



Accumulated Balance with Fmin and fractal value D1 for 2017
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1
0
50
100
150
200
2017-01
2017-03
2017-05
2017-07
2017-09
2017-11
2018-01
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

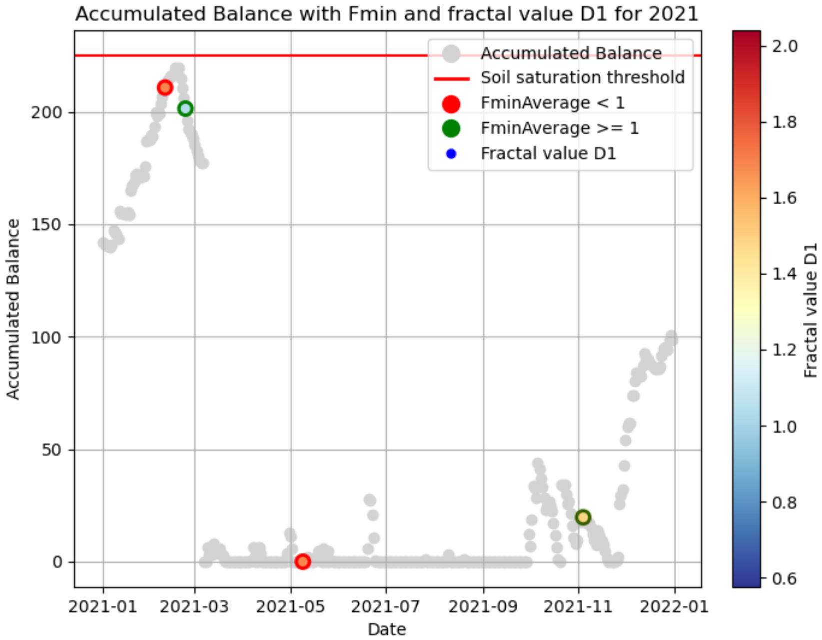

Accumulated Balance with Fmin and fractal value D1 for 2018
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1
0
50
100
150
200
250
2018-01
2018-03
2018-05
2018-07
2018-09
2018-11
2019-01
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

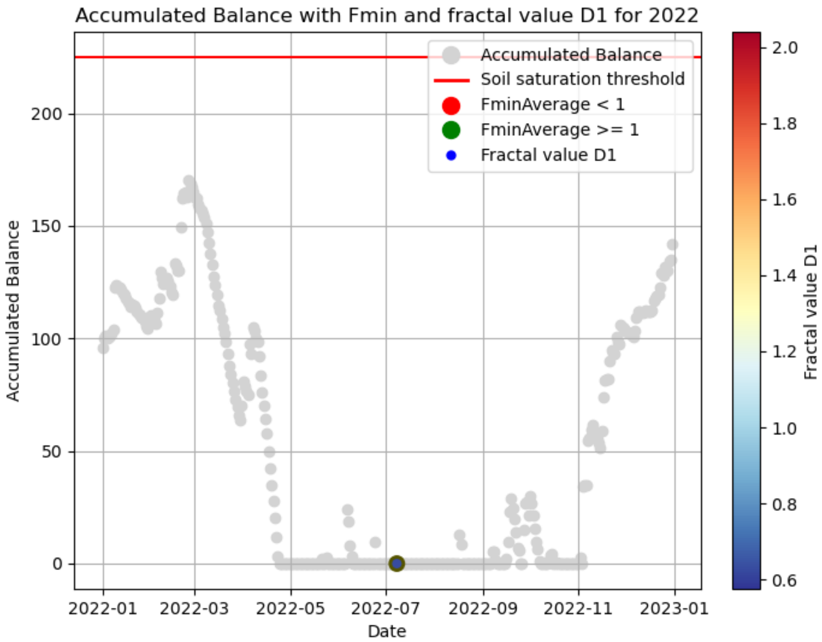


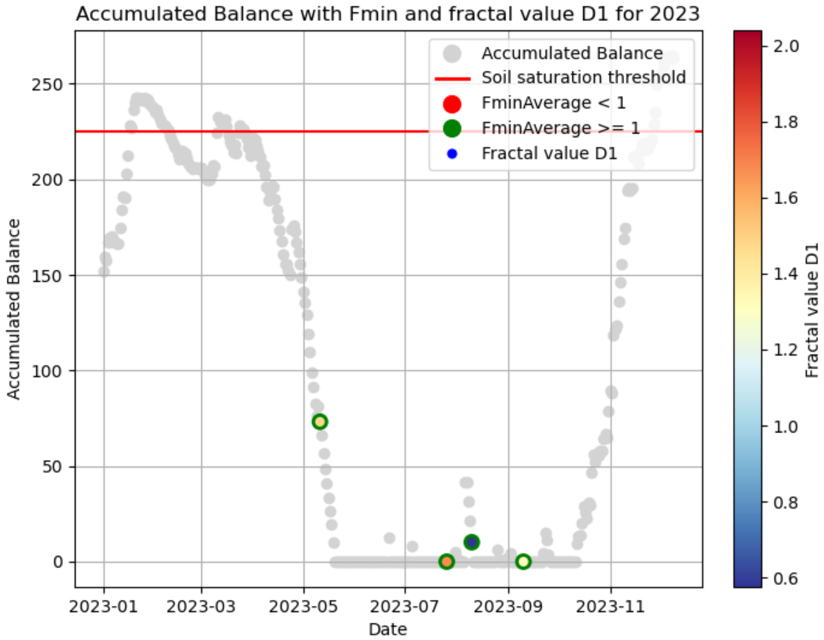
Accumulated Balance with Fmin and fractal value D1 for 2019
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1



Accumulated Balance with Fmin and fractal value D1 for 2020
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1


Accumulated Balance with Fmin and fractal value D1 for 2021
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1
0
50
100
150
200
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
2022-01
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

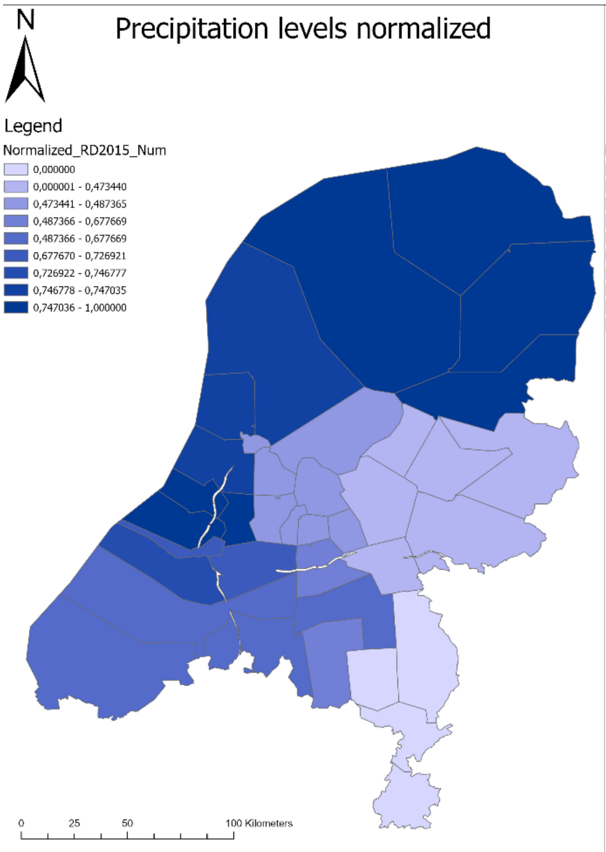

Accumulated Balance with Fmin and fractal value D1 for 2022
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1
0
50
100
150
200
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

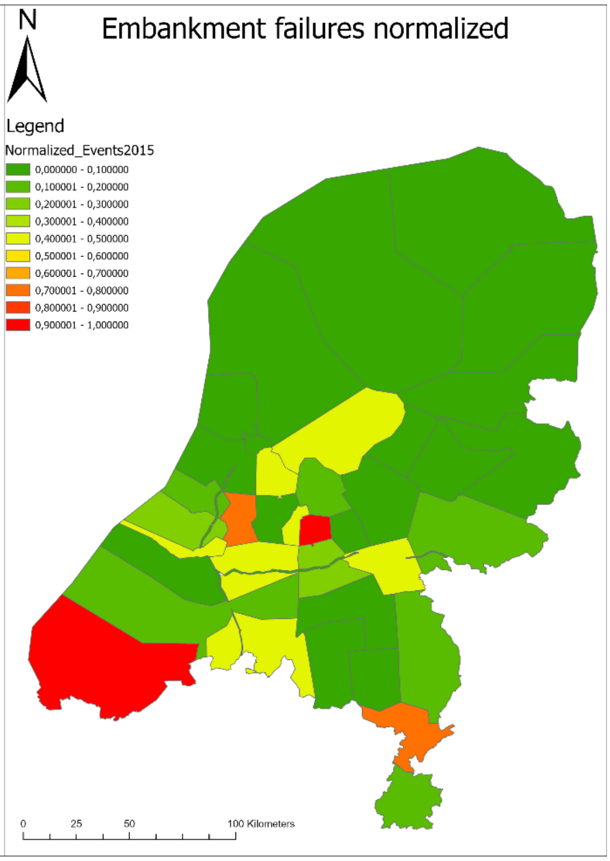


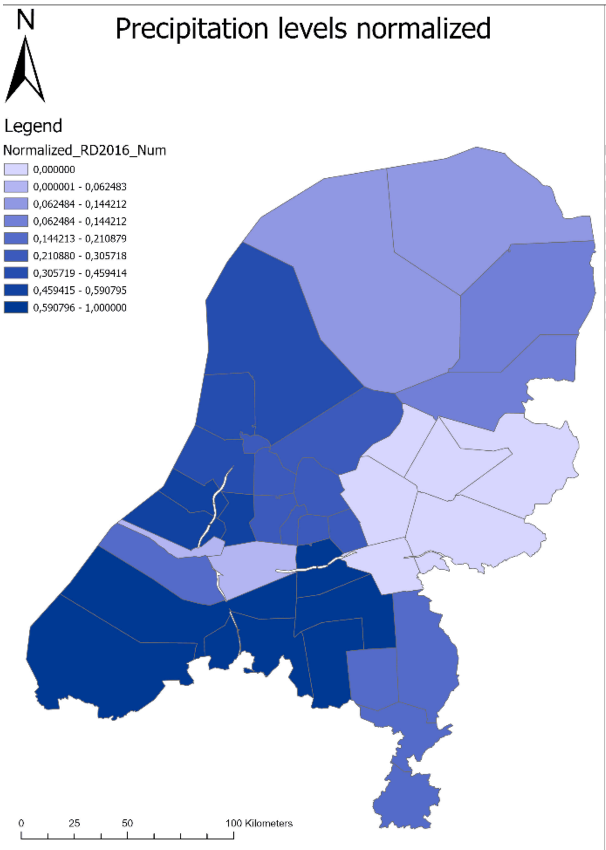
Accumulated Balance with Fmin and fractal value D1 for 2023
Accumulated Balance
Soil saturation threshold
FminAverage < 1
FminAverage >= 1
Fractal value D1
Accumulated Balance
Date
Fractal value D1
0
50
100
150
200
250
2023-01
2023-03
2023-05
2023-07
2023-09
2023-11
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

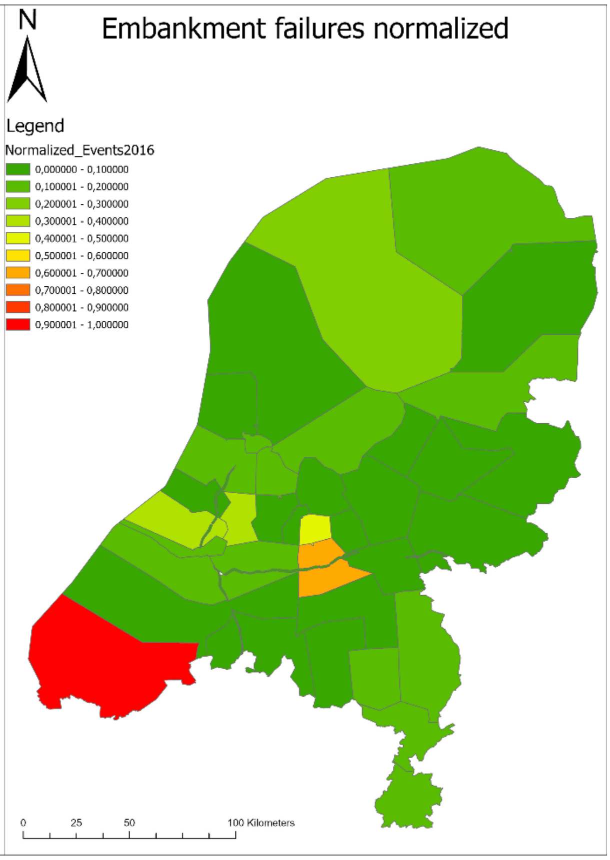
Appendix E: visualisations of macro-analysis

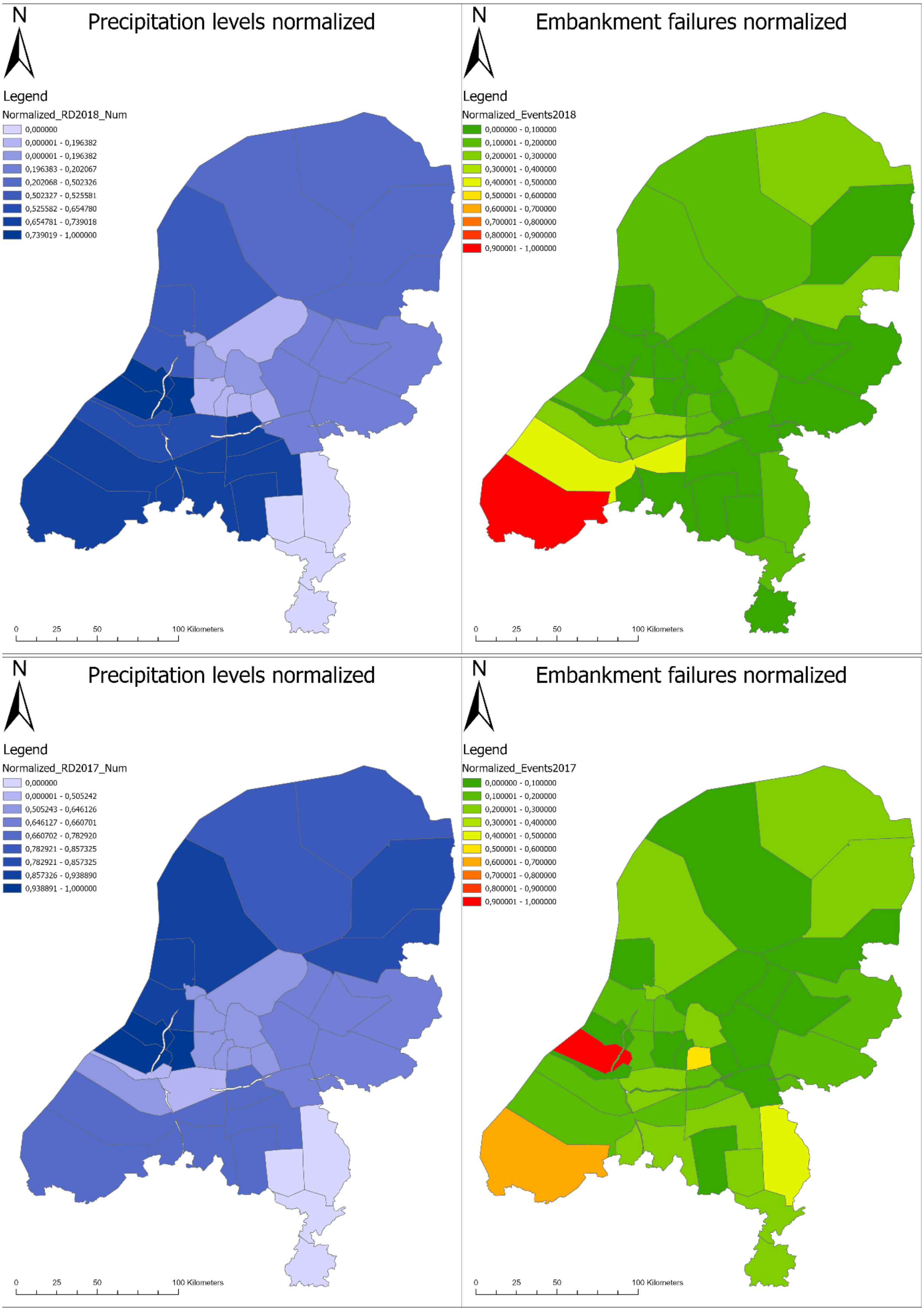
E.1. Rainfall and embankment failures

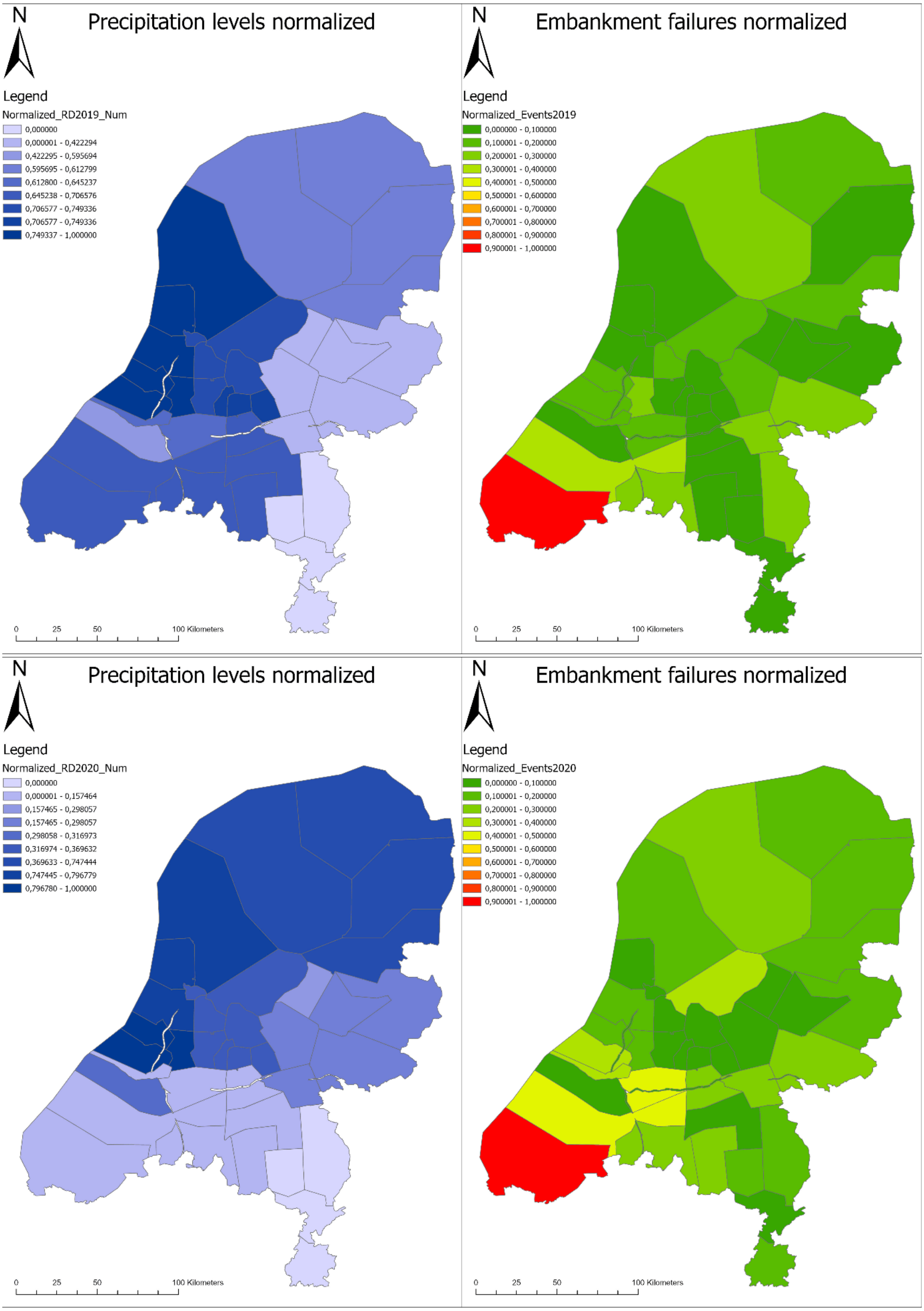


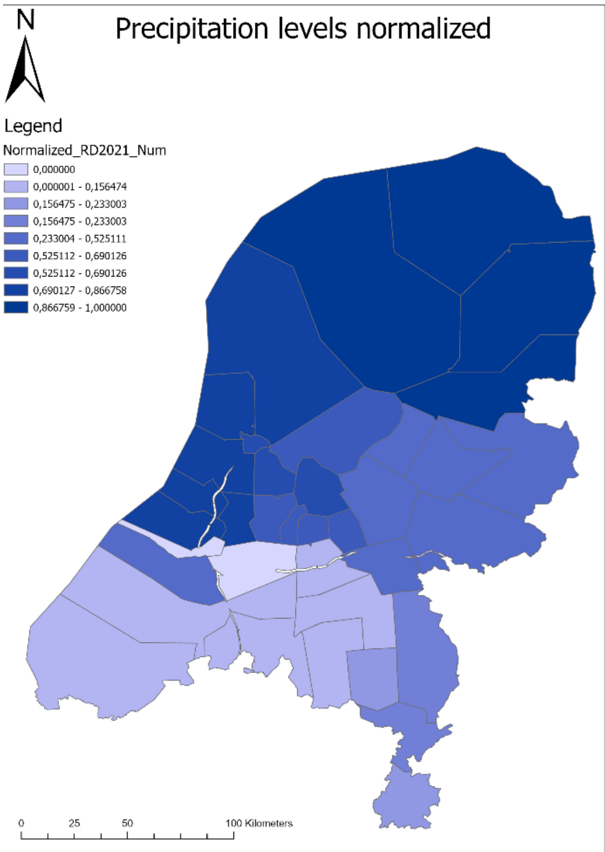


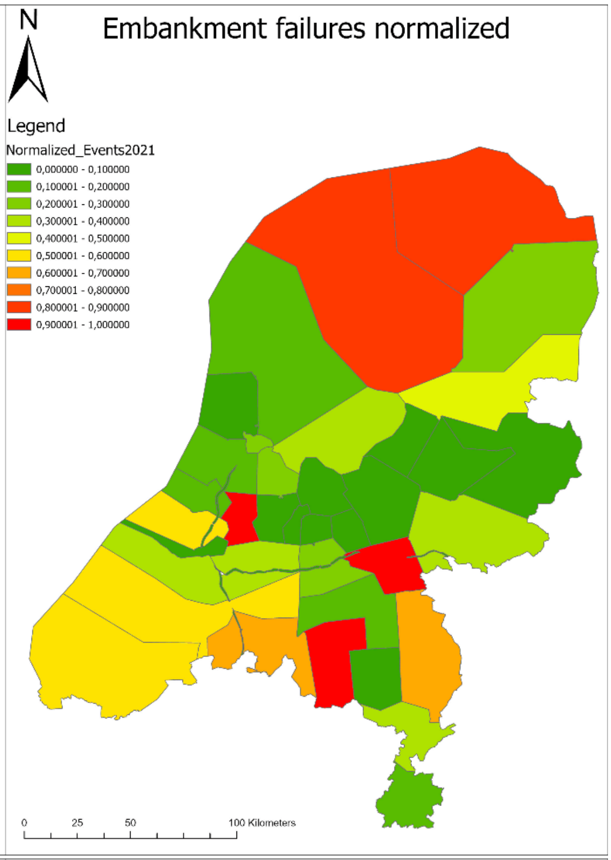


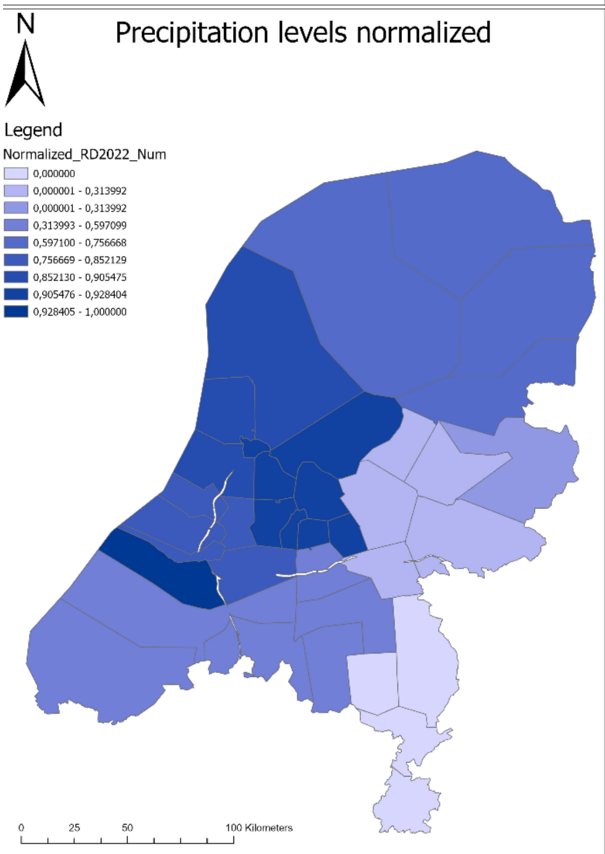


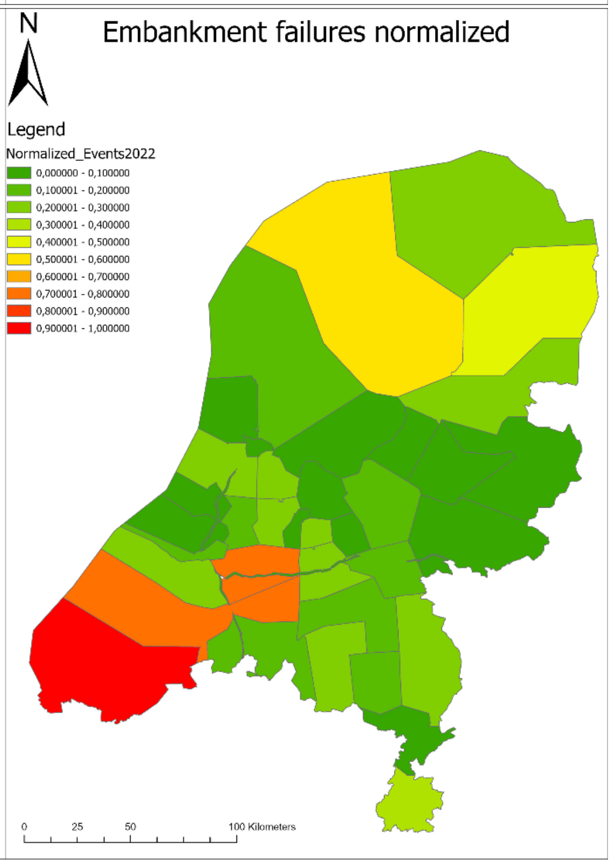

N
Precipitation levels normalized
Legend
Normalized_RD2018_Num
0,000000
0,000001 - 0,196382
0,000001 - 0,196382
0,196383 - 0,202067
0,202068 - 0,502326
0,502327 - 0,525581
0,525582 - 0,654780
0,654781 - 0,739018
0,739019 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2018
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers
N
Precipitation levels normalized
Legend
Normalized_RD2017_Num
0,000000
0,000001 - 0,505242
0,505243 - 0,646126
0,646127 - 0,660701
0,660702 - 0,782920
0,782921 - 0,857325
0,782921 - 0,857325
0,857326 - 0,938890
0,938891 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2017
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers

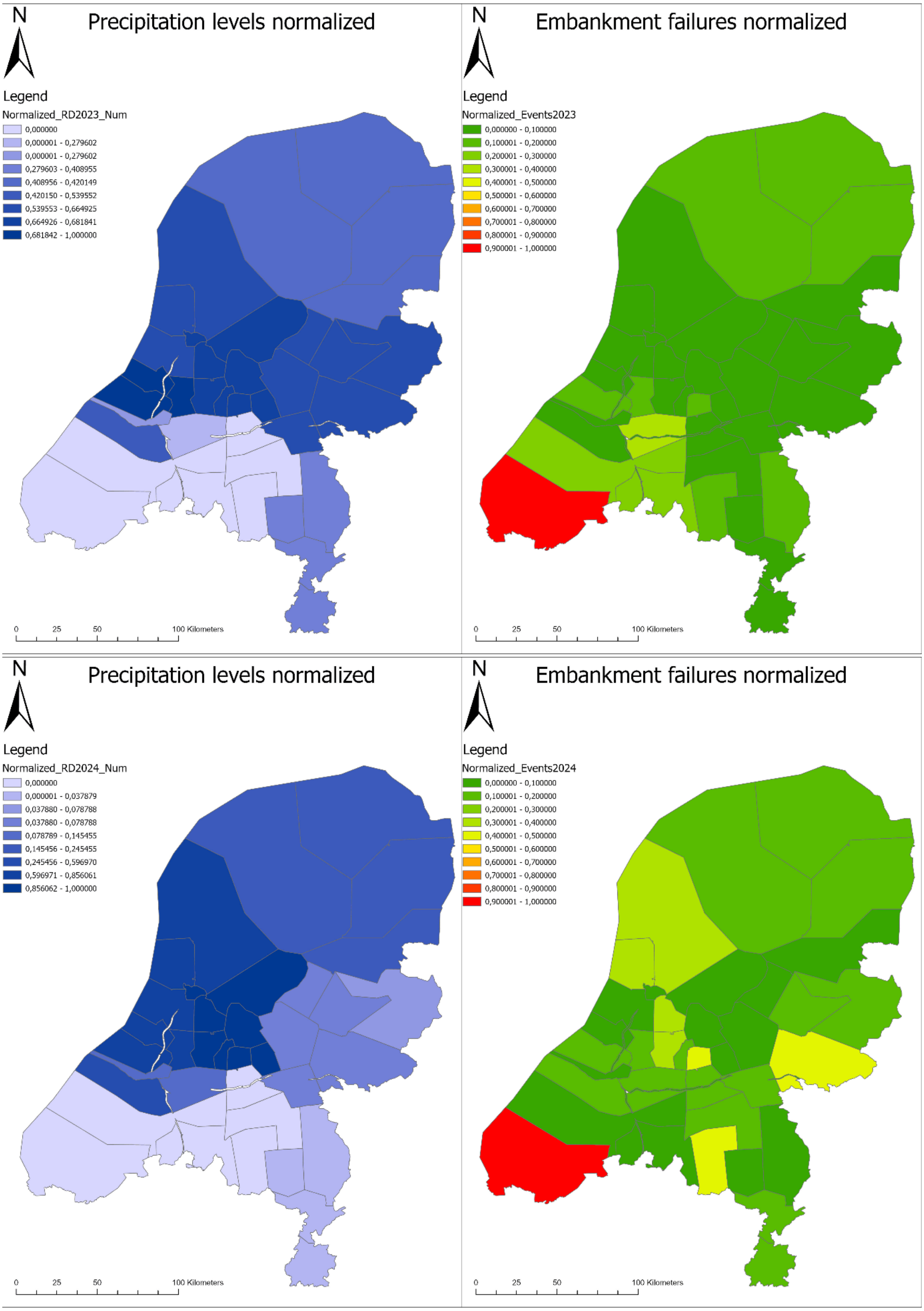

N
Precipitation levels normalized
Legend
Normalized_RD2019_Num
0,000000
0,000001 - 0,422294
0,422295 - 0,595694
0,595695 - 0,612799
0,612800 - 0,645237
0,645238 - 0,706576
0,706577 - 0,749336
0,706577 - 0,749336
0,749337 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2019
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers
N
Precipitation levels normalized
Legend
Normalized_RD2020_Num
0,000000
0,000001 - 0,157464
0,157465 - 0,298057
0,157465 - 0,298057
0,298058 - 0,316973
0,316974 - 0,369632
0,369633 - 0,747444
0,747445 - 0,796779
0,796780 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2020
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers


N
Precipitation levels normalized
Legend
Normalized_RD2021_Num
0,000000
0,000001 - 0,156474
0,156475 - 0,233003
0,156475 - 0,233003
0,233004 - 0,525111
0,525112 - 0,690126
0,525112 - 0,690126
0,690127 - 0,866758
0,866759 - 1,000000
0
25
50
100 Kilometers

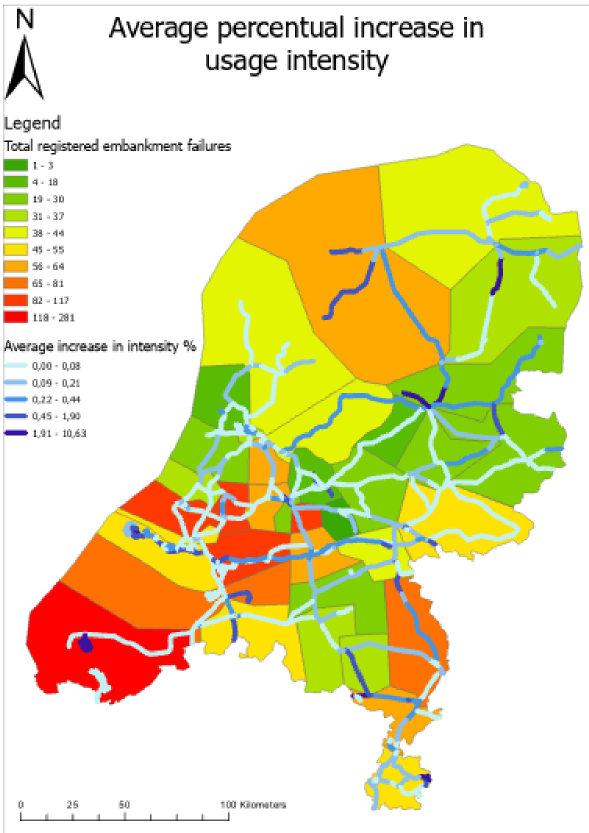

N
Embankment failures normalized
Legend
Normalized_Events2021
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0
25
50
100 Kilometers

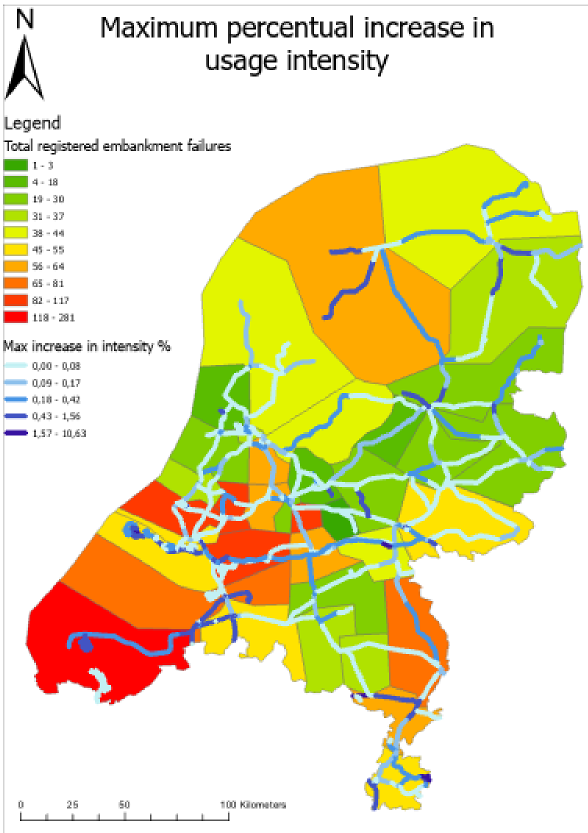

N
Precipitation levels normalized
Legend
Normalized_RD2022_Num
0,000000
0,000001 - 0,313992
0,000001 - 0,313992
0,313993 - 0,597099
0,597100 - 0,756668
0,756669 - 0,852129
0,852130 - 0,905475
0,905476 - 0,928404
0,928405 - 1,000000
0
25
50
100 Kilometers

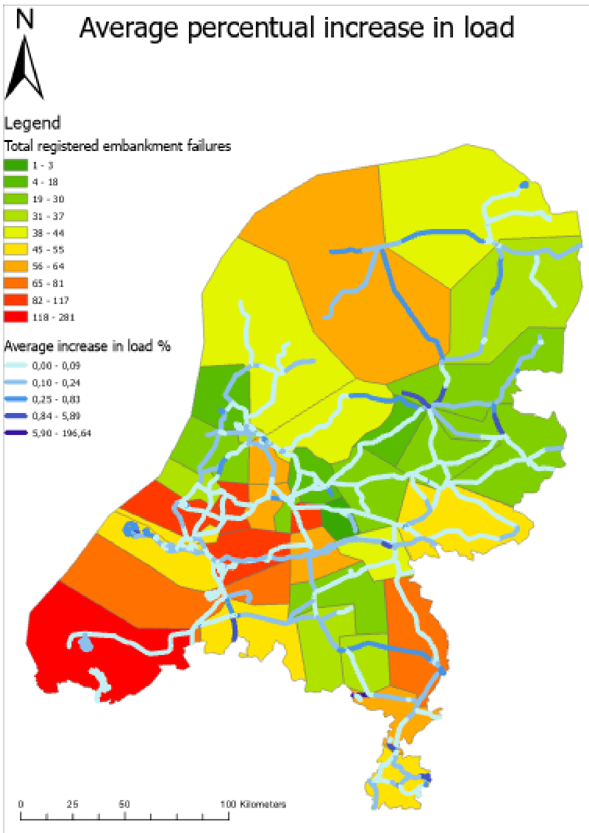

N
Embankment failures normalized
Legend
Normalized_Events2022
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0
25
50
100 Kilometers

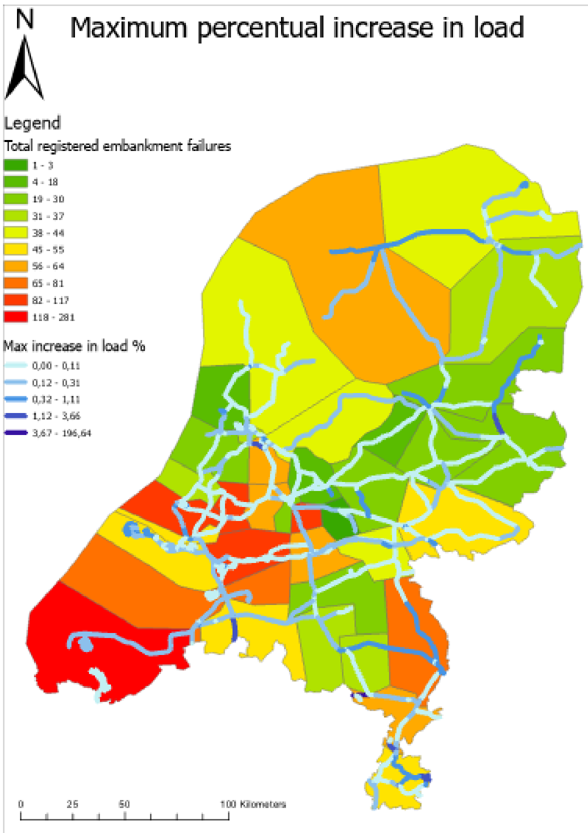

N
Precipitation levels normalized
Legend
Normalized_RD2023_Num
0,000000
0,000001 - 0,279602
0,000001 - 0,279602
0,279603 - 0,408955
0,408956 - 0,420149
0,420150 - 0,539552
0,539553 - 0,664925
0,664926 - 0,681841
0,681842 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2023
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers
N
Precipitation levels normalized
Legend
Normalized_RD2024_Num
0,000000
0,000001 - 0,037879
0,037880 - 0,078788
0,037880 - 0,078788
0,078789 - 0,145455
0,145456 - 0,245455
0,245456 - 0,596970
0,596971 - 0,856061
0,856062 - 1,000000
0 25 50 100 Kilometers
N
Embankment failures normalized
Legend
Normalized_Events2024
0,000000 - 0,100000
0,100001 - 0,200000
0,200001 - 0,300000
0,300001 - 0,400000
0,400001 - 0,500000
0,500001 - 0,600000
0,600001 - 0,700000
0,700001 - 0,800000
0,800001 - 0,900000
0,900001 - 1,000000
0 25 50 100 Kilometers

E.2. Train characteristics on national scale

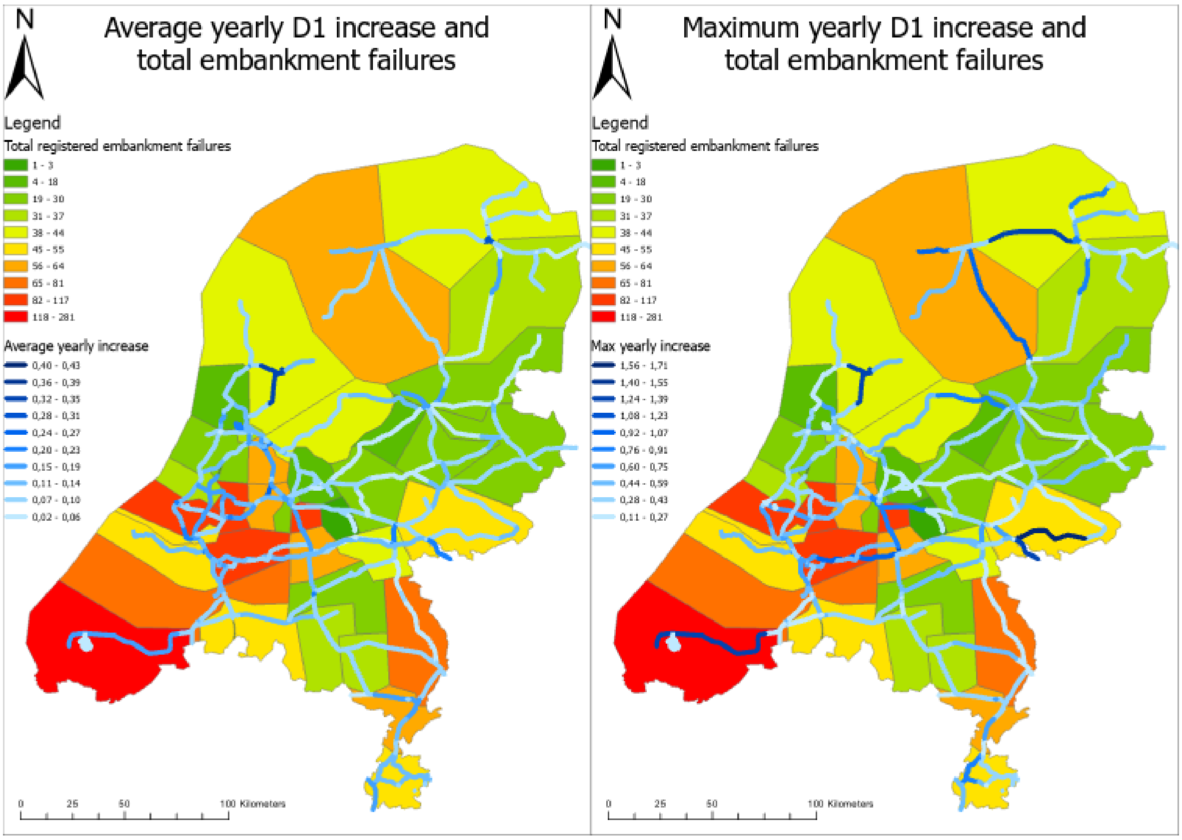





N
Average percentual increase in load
Legend
Total registered embankment failures
1 - 3
4 - 18
19 - 30
31 - 37
38 - 44
45 - 55
56 - 64
65 - 81
82 - 117
118 - 281
Average increase in load %
0,00 - 0,09
0,10 - 0,24
0,25 - 0,83
0,84 - 5,89
5,90 - 196,64
0
25
50
100 Kilometers

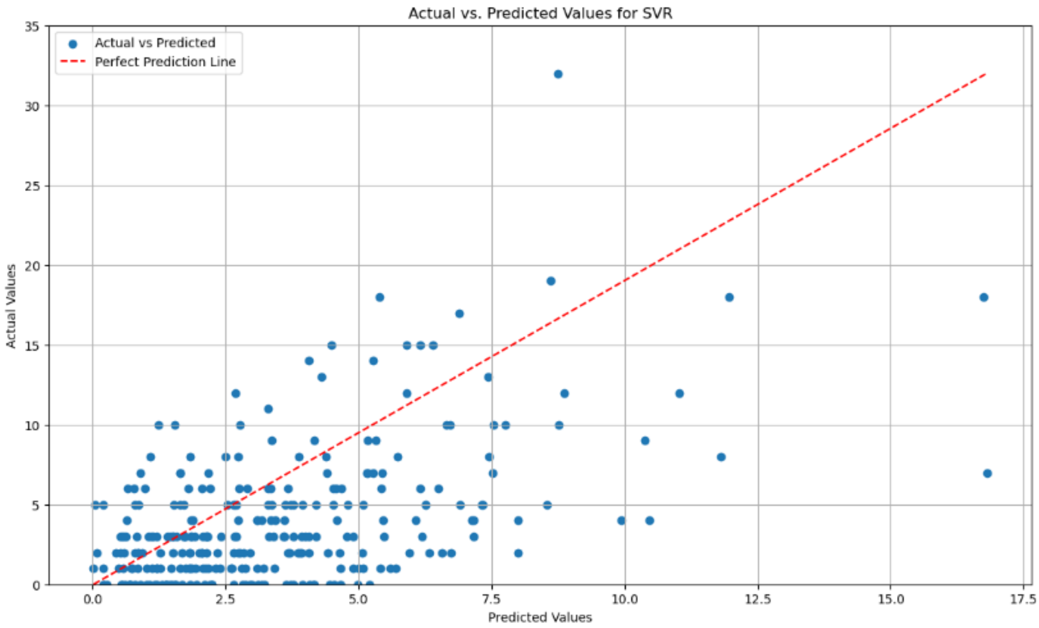

N
Maximum percentual increase in load
Legend
Total registered embankment failures
1 - 3
4 - 18
19 - 30
31 - 37
38 - 44
45 - 55
56 - 64
65 - 81
82 - 117
118 - 281
Max increase in load %
0,00 - 0,11
0,12 - 0,31
0,32 - 1,11
1,12 - 3,66
3,67 - 196,64
0
25
50
100 Kilometers

E.3. track characteristics

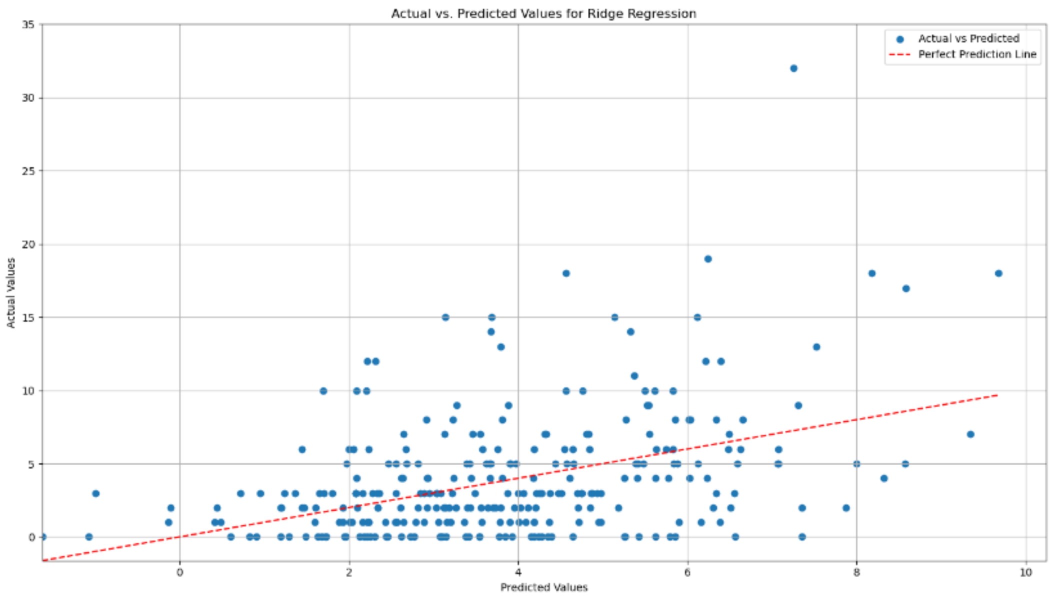


Appendix F: machine learning combined output

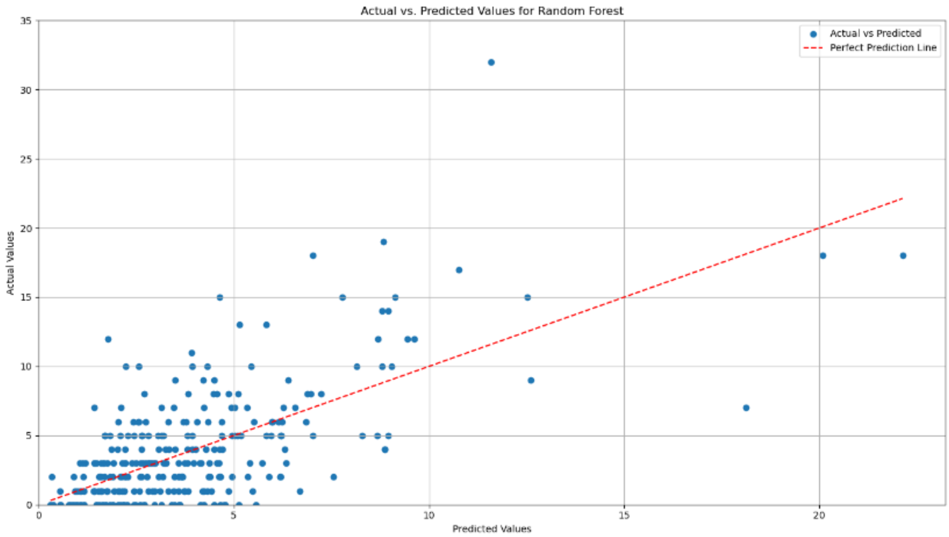
```
R-squared on Test Set for SVR: 0.2659
```



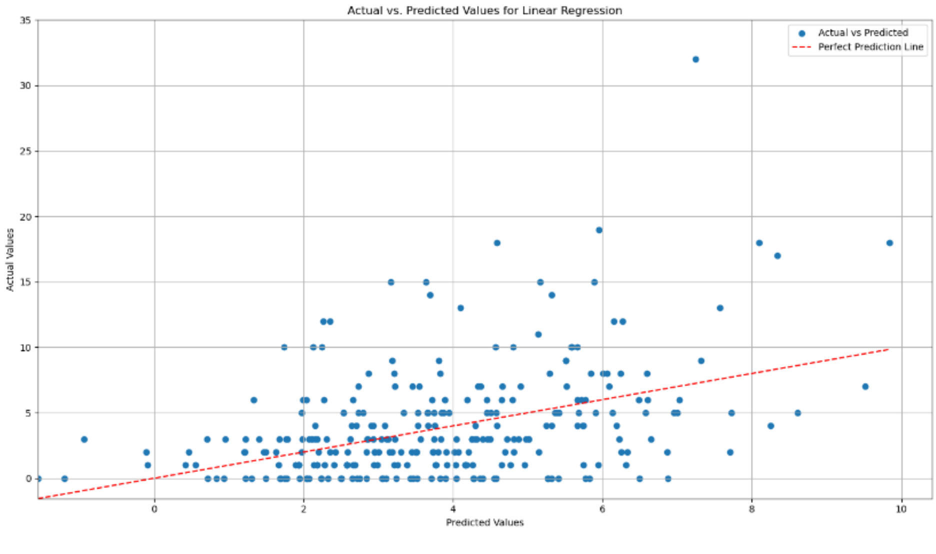

```
R-squared on Test Set for Ridge Regression: 0.1624
Ridge Regression Coefficients:
D1avg             1.902451e+00
DegStatMaxYear    9.020993e-01
D1Max             4.060092e-01
Rainfall          1.649940e-01
MeanTon           3.812221e-05
MeanNo            3.605856e-05
MaxNo             1.867961e-05
MaxTon            7.395490e-06
Soil              5.827836e-10
dtype: float64
```

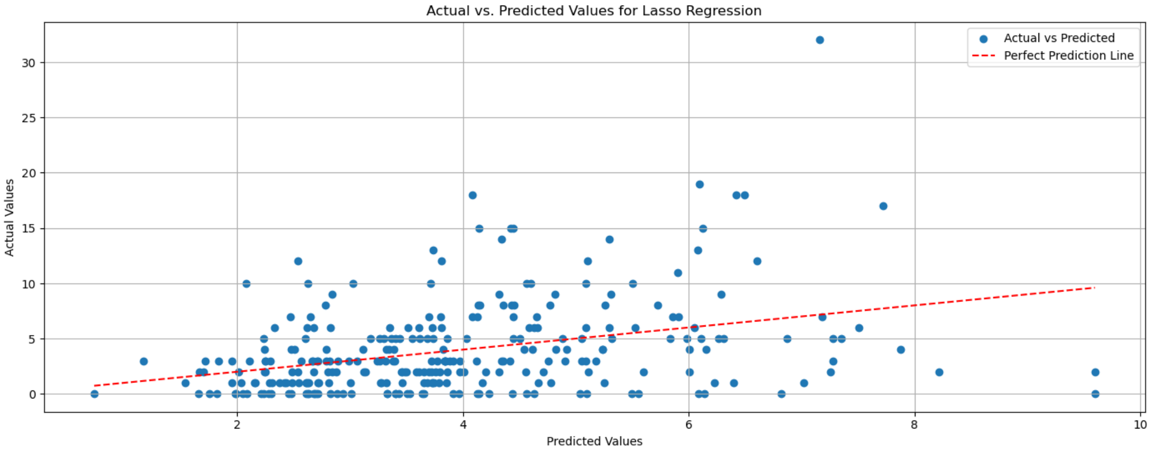


```
R-squared on Test Set for Random Forest: 0.4148
Random Forest Feature Importances:
Soil            0.470021
Rainfall        0.176073
DegStatMaxYear  0.113080
MeanTon         0.054139
MaxTon          0.050555
MeanNo          0.050373
MaxNo           0.037371
D1Max           0.026093
D1avg           0.022295
dtype: float64
```



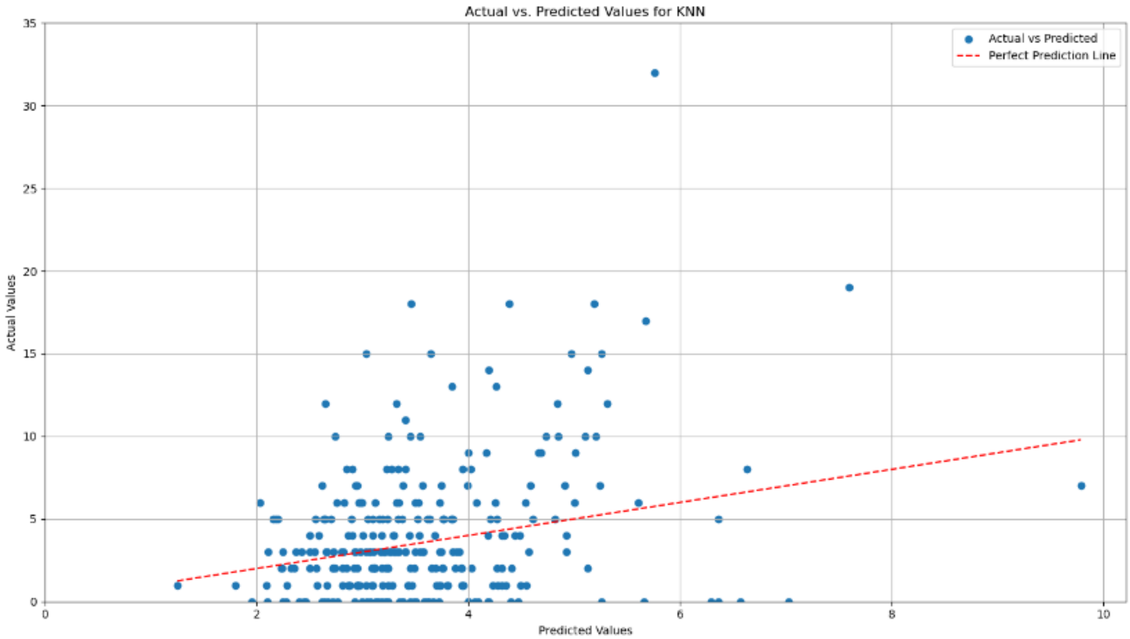

```
R-squared on Test Set for Linear Regression: 0.1667
Linear Regression Coefficients:
Soil            2.612528e+09
D1avg           3.134364e+00
DegStatMaxYear  9.063606e-01
D1Max           2.759838e-01
Rainfall        1.622257e-01
MeanTon         3.376603e-05
MeanNo          2.655387e-05
MaxNo           2.279878e-05
MaxTon          9.600073e-06
dtype: float64
```



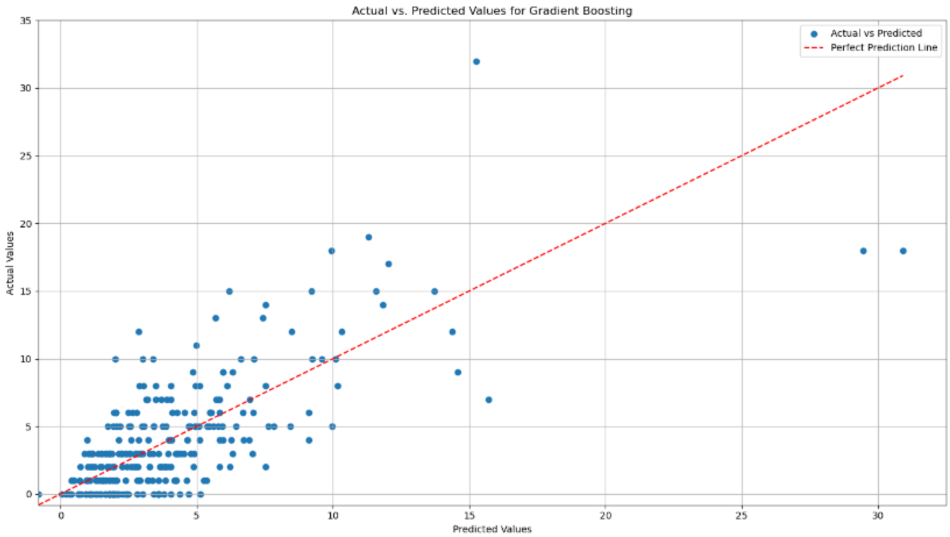


R²: 0.1206
Top 10 Feature Importances:
Soil: 0.7116
Rainfall: 0.2882
MeanNo: 0.0001
MaxTon: 0.0000
MeanTon: 0.0000
MaxNo: 0.0000
D1Max: 0.0000
D1avg: 0.0000
DegStatMaxYear: 0.0000

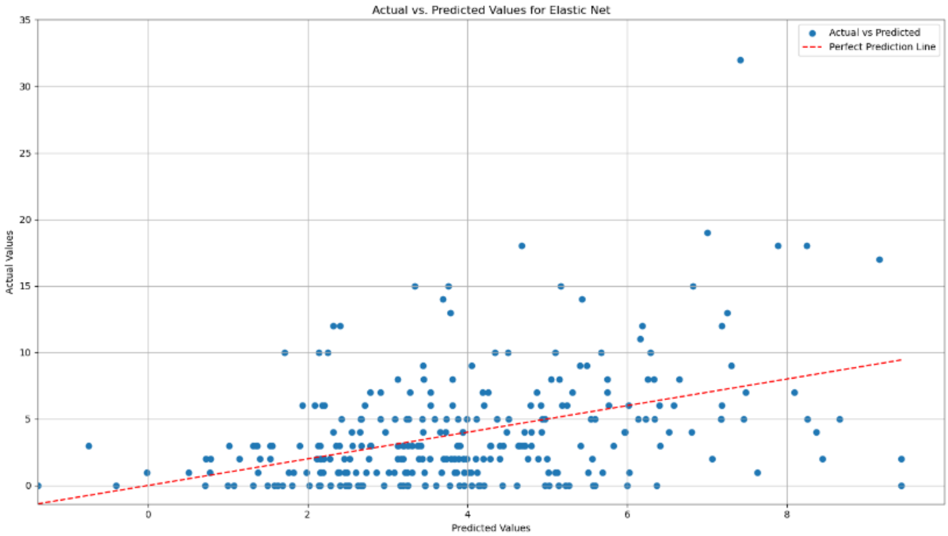
```
R-squared on Test Set for KNN: 0.0924
```



```
R-squared on Test Set for Gradient Boosting: 0.4882
Gradient Boosting Feature Importances:
Soil              0.446078
Rainfall          0.200429
DegStatMaxYear    0.140803
MeanNo            0.054816
MeanTon           0.045961
MaxTon            0.041364
MaxNo             0.038860
D1Max             0.016673
D1avg             0.015014
dtype: float64
```




```
R-squared on Test Set for Elastic Net: 0.1494
Elastic Net Feature Importances:
Soil              1.099307
DegStatMaxYear    0.737800
Rainfall          0.167097
MeanNo            0.000061
MeanTon           0.000040
MaxNo             0.000006
MaxTon            0.000003
D1Max             0.000000
D1avg             0.000000
dtype: float64
```



```
R-squared on Test Set for Decision Tree: 0.1049
Decision Tree Feature Importances:
Soil                0.554419
Rainfall            0.286480
DegStatMaxYear      0.071373
D1Max               0.039852
D1avg               0.016556
MaxTon              0.013735
MaxNo               0.011607
MeanNo              0.005275
MeanTon             0.000704
dtype: float64
```