Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation UVEK
Département fédéral de l'environnement, des transport, de l'énergie et de la communication DETEC
Dipartimento federale dell'ambiente, dei trasporti, dell'energia e delle communication DATEC

Bundesamt für Strassen
Office fédéral des routes
Ufficio federale delle Strade

Removing the unexplained variability in road condition indicator values (COEUS)

Identifizierung der Ursachen für unerklärliche Variabilität in Strassenzustandsdaten

Suppression de la variabilité inexpliquée des valeurs des indicateurs d'état des routes

Berner Fachhochschule Architektur, Holz und Bau (AHB)
Prof. Dr. Aybike Öngel
Dr. Moises Bueno

ETH Zürich
Dr. Marcelo Galleguillos Torres
Prof. Dr. Bryan Tyrone Adey

Research project TRU_20_02B_01 at the request of the Federal Roads Office FEDRO
December 2025 | 1815

Removing the unexplained variability in road condition indicator values (COEUS)

Identifizierung der Ursachen für unerklärliche Variabilität in Strassenzustandsdaten

Suppression de la variabilité inexpliquée des valeurs des indicateurs d'état des routes

Berner Fachhochschule Architektur, Holz und Bau (AHB)
Prof. Dr. Aybike Öngel
Dr. Moises Bueno

ETH Zürich
Dr. Marcelo Galleguillos Torres
Prof. Dr. Bryan Tyrone Adey

Research project TRU_20_02B_01 at the request of the Federal Roads Office FEDRO
December 2025 | 1815

Imprint

Research centre and project team

Project management
Aybike Öngel, BFH-AHB

Members
Moises Bueno, BFH-AHB
Marcelo Galleguillos Torres, ETH Zürich
Bryan Tyrone Adey, ETH Zürich

Leading expert commission

TRU_20_02 Grunddaten für Zustandsprognosen

Advisory commission

President
Erik Bühlmann

Members
Stefano Bradanini
Robert Braber
Mehdi Ould Henia
Martin Horat
Yvan Ramel
Frank Schiffmann

Co-financing of the research project

ETH Zürich

Applicant

Arbeitsgruppe Trassee und Umwelt (TRU)

Source

The document can be downloaded free of charge from
https://www.mobilityplatform.ch/

December 2025

Table of contents

List of illustrations

List of tables

Summary

Switzerland has a dense road network, including national, cantonal, and communal/private roads. National roads, despite making up only 3% of the network, handle around 40% of traffic and 62.9% of goods transport. The management and maintenance of roads are distributed among the Swiss Federal Roads Office (FEDRO), cantonal authorities, and municipalities.

Roads deteriorate over time due to aging, weather conditions, and usage factors like salt in winter or terrain shifts. To maintain an optimal service level, road conditions must be assessed periodically. Swiss road managers follow strict guidelines to ensure accuracy, comprehensibility, repeatability, and reproducibility in pavement assessments. Accuracy refers to minimizing measurement errors, comprehensibility ensures clear data interpretation, and repeatability/reproducibility analyse measurement consistency under different conditions.

Pavement Management Systems (PMS) are strategic tools used by road authorities to plan, prioritize, and optimize maintenance and rehabilitation interventions across road networks. Their goal is to maintain pavements at desired service levels while managing costs over the long term. The efficiency of PMS relies heavily on the accuracy, completeness, and consistency of condition data. Influential factors include the quality of location referencing, historical data consistency, comprehensive spatial and temporal coverage, and alignment with evolving infrastructure management practices. Errors in data acquisition or processing can result in flawed treatment recommendations and misallocation of maintenance budgets.

Network-level data collection typically favors high-speed methods, such as profilometers and inertial profilers, due to their speed, reduced labor demands, and minimal traffic disruption. However, these methods can introduce data variability caused by equipment calibration differences, lateral positioning, environmental conditions, and seasonal fluctuations. Operators and service providers may apply different measurement methodologies, further impacting consistency. As a result, rigorous equipment validation and repeatability studies are essential to ensure data quality and enable meaningful comparisons across time and regions.

Recent research has explored the application of machine learning techniques to model pavement deterioration and forecast performance. Advanced frameworks now integrate methods such as BorutaShap for feature selection, Bayesian Neural Networks (BNNs) for probabilistic modeling, and SHAP (SHapley Additive Explanations) values for model interpretation. These approaches enhance the ability to capture complex interactions among pavement design, traffic, environment, and maintenance history. However, studies highlight that the largest source of uncertainty in these models is poor data quality, underscoring the need for improved data governance and validation procedures.

Visual inspections remain a valuable assessment method, especially in settings without access to automated systems. Their reliability varies depending on rater experience, training, and the pavement condition being evaluated. Inspections are generally more accurate for pavements in good condition, while subjective interpretation introduces more errors at the boundary between fair and poor. Training programs and calibration exercises have been shown to significantly improve both accuracy and consistency. Monte Carlo simulations and error transition matrices have been used to quantify the impact of visual inspection errors on condition indices, particularly at moderate and high distress severity levels.

Surface roughness measurements, often quantified through the International Roughness Index (IRI), are widely used in network evaluations. However, IRI values can vary significantly depending on profiler type, lane position, traffic volume, and pavement type. Studies show that single test runs may not adequately capture condition variability, especially on state routes. Comparative analyses of profiler data against reference equipment highlight the need for periodic calibration, multiple repetitions, and statistical evaluation to ensure meaningful interpretation. Notably, geographic and seasonal influences also contribute to measurement variation.

Skid resistance, an essential indicator for traffic safety, is influenced by surface texture, aggregate properties, weather, and traffic loading. Techniques such as the GripTester and dynamic friction testers (DFT) are used alongside laser scanning systems to measure surface texture and friction. Seasonal variation is a major factor; for example, summer polishing tends to reduces skid resistance. Long-term studies have linked friction trends to changes in texture depth, rainfall patterns, and heavy vehicle traffic. Recent innovations include the development of statistical models to estimate dry-season friction levels from wet-season data.

Statistical methods play a central role in managing variability and error in pavement data. Tools such as Measurement System Analysis (MSA), Gage Repeatability & Reproducibility (GR&R), and latent Markov models allow researchers to identify, quantify, and reduce error sources. MSA has been widely adopted in other industries but is increasingly applied in transportation research to evaluate the reliability of measurement systems. Hybrid methodologies combining experimental design, simulation, and advanced probabilistic modeling offer deeper insights into the influence of construction quality, traffic loading, and environmental exposure on pavement deterioration.

Finally, the integration of modern analytics with traditional engineering practices enhances the precision and reliability of pavement condition assessments. While automated and machine learning-based methods show great promise, their success depends on rigorous data management, consistent measurement practices, and well-calibrated systems. As road networks age and budgets tighten, the ability to make data-driven, confident decisions on maintenance priorities will remain essential. Continued investment in measurement technology, staff training, and cross-jurisdictional data standardization will support more resilient and cost-effective pavement infrastructure management worldwide.

The assessment of road condition in Switzerland is governed by a set of national norms, each detailing specific procedures and requirements for evaluating the performance of road surfaces. Although each part of the monitoring process is covered by a dedicated norm, there are frequent overlaps and cases where multiple norms apply to the same task. These norms are particularly relevant after construction or rehabilitation phases, as well as during the operational life of road networks. The norm VSS 40 525 for example, defines the acceptance criteria for road surfaces upon completion of works, focusing on four characteristics: local irregularities, longitudinal evenness, transversal evenness, and surface friction. Irregularities or distresses are evaluated as described in SN 640 516-7, with specific acceptance thresholds for different pavement layers. Longitudinal evenness is measured based on the road's longitudinal profile, with two methods allowed by VSS 40 517: one based on angular variation (W and standard deviation Sw), and another based on waveband analysis (NBO), which examines short, medium, and long surface waves. The second method is more comprehensive and better adapted to detect multi-scale surface irregularities. Transversal evenness is assessed by measuring rut depth and the theoretical water depth within the ruts, following VSS 40518. While rut depth itself does not have specific acceptance thresholds, the water depth does, especially for safety purposes on high-speed roads. For instance, on roads with speed limits over 80 km/h, the maximum permissible theoretical water depth is 4.0 mm. Surface friction, another critical parameter, is evaluated using dynamic and static methods according to VSS 40 512. Dynamic systems include the Skiddometer BV 8 and the SKM device, while the pendulum method (SRT) serves as the static alternative. These tests take seasonal conditions into account and are typically conducted 3–6 months after the road is open to traffic, with speed-specific thresholds applied.

Pavement condition can be summarized using four indices defined in VSS 40 925b: I0 (surface damage), I2 (longitudinal evenness), I3 (transversal evenness), and I4 (surface friction). Each index is scored from 0 (very good) to 5 (very poor), based on one or multiple parameters. The selection of which parameters to evaluate and how often depends on the road's function, traffic load, and management methodology. At the network level, these indices are used in accordance with SN 640 900 and stored in management systems defined by VSS 40 944 and VSS 40 904. Project-level assessments follow norms VSS 40 730 and VSS 40 925b. While most indicators can be assessed year-round as long as the pavement is dry, temperature conditions must be observed when evaluating surface friction and load-bearing capacity.

Surface damage assessments differ between bituminous and concrete surfaces. For asphalt pavements, evaluation includes smoothness, deformation, structural damage, and repairs. For concrete, factors like material loss, joint degradation, vertical displacements, and cracking are considered. This work was focused on bituminous surfaces that correspond to highways. Traditional methods rely on inspectors walking or driving along the segment, although many companies now use high-resolution imaging mounted on vehicles, which are then analysed by trained professionals—or increasingly, by artificial intelligence systems. Surface damage is characterized by its extent (scale 0–3) and severity (scale 1–3), which are combined using a matrix to yield a score from 0 to 9. In summary assessments, parameters are grouped, scored, weighted, and combined into the I0 index. Standard segment lengths are 50 meters for full carriageways and 100 meters for single-lane roads. The general rule is to report the worst severity level per group and a representative extent. For project-level detailed assessments, both dimensions are recorded for each individual parameter.

Longitudinal evenness (I2) is evaluated through several methods, with the most direct being the measurement of angle changes (W) between adjacent 1-meter sections of the road surface. The standard deviation Sw of these W values is used to derive the index, as shown in a graph within VSS 40517. Alternative methods include the International Roughness Index (IRI), which estimates user ride comfort based on simulated vertical motion of a vehicle suspension system; the NBO waveband method, which analyses energy distribution across different wave frequencies; and the weighted longitudinal profile (BLP), which balances amplitude contributions from different wavelengths. These methods help isolate periodic or isolated defects. However, the norm only provides a direct transformation of Sw into I2 values and does not specify how to convert results from other techniques into the I2 scale.

Transversal evenness (I3) is based on rut depth measurements and the associated water accumulation potential in those ruts. Ruts can be measured manually with a ruler or string, and it must be indicated which method is used. Values from both left and right wheel paths are required, and multiple measurements must be taken over the evaluation segment. When evaluating road networks, only one rut may be assessed. The spacing between transversal profile points must not exceed 10 cm, and the vertical precision of measurements should be 0.5 mm for dynamic and 1.0 mm for static systems. A minimum of 10 measurements per road segment is recommended, and both statistical summaries and extreme values should be documented. When both ruts are measured, the worst value is used for I3.

In practical terms, both I2 and I3 measurements are conducted using advanced laser profilometers integrated into survey vehicles. These technologies ensure high measurement precision and speed, essential for large-scale road network assessments. The systems are capable of accurately scanning the road surface and subsequently computing standardized parameters for designated sections (e.g. wheel paths) in accordance with the relevant specifications. Despite the existence of traditional manual methods, automated systems provide greater efficiency, reliability, and reproducibility, and are now standard in most monitoring operations.

Overall, the Swiss approach to road condition assessment is highly structured and standardized, ensuring consistency across projects and regions. The integration of norms with advanced measurement systems and increasing digital workflows supports data-driven decision-making in road maintenance and rehabilitation planning.

Switzerland's national road network is managed through two divisions: West and East. Division West oversees the regional branches Estavayer-le-Lac and Thun, while Division East manages Zofingen, Winterthur, and Bellinzona. Regional branches are named "*Filialen*" in Switzerland. Each *Filiale* is responsible for measuring road condition indicators within its designated area. The data collected is stored in the TRA-Trassee application, which supports road infrastructure management and planning. The data undergoes verification before integration into the system, though resource limitations affect quality control. Major companies involved in road condition measurement include Schniering GmbH (now part of TÜV Rheinland) and Infralab SA, which employ advanced high-speed measurement techniques. Data adjustments are conducted by Geologix AG before final storage.

In this study the analysis is focused on two methods to explore the complex and large datasets. On one side, visualizations play a key role in data analysis, helping identify inconsistencies and trends over time in a simple and intuitive way. Various statistical methods, such as histograms and scatter plots, aid in assessing data distribution, detecting outliers, and ensuring consistency in road condition indicators. On the other hand, a more complex approach like factor analysis is also a tool that can be considered a critical part of pavement condition assessment. In this study, weather data from MeteoSwiss was integrated, matching relevant meteorological stations to each *Filiale*. The combination of weather data and measurement factors allowed exploring more in detail the importance when explaining the variability in the datasets. To enhance prediction accuracy, advanced machine learning models—including AdaBoost, CatBoost, LightGBM, Random Forest, and XGBoost—are optimized using Bayesian Optimization (BO). These ensemble learning methods help capture complex relationships in pavement condition data. The model performance is assessed using R^2 and RMSE metrics to determine the most accurate predictions. SHAP (Shapley Additive Explanations) and permutations techniques provide interpretability for the predictive models, allowing insight into the influence of different variables on road condition variability. The quantification of the importance of features such as climate, intervention history, and measurement methods ensure a data-driven approach to road infrastructure planning and maintenance.

Coming to the results of the visualization process, it could be observed for example for the regional branch Estavayer-le-lac that on Highway N1, for stretches with no major intervention post-2000, most indicators showed a general pattern of deterioration, as expected. However, some unexpected improvements in certain indicators (notably I2 and I4) raised concerns about measurement reliability. Periods between 2009 and 2013, and 2013 and 2017, showed anomalies where different indicators presented conflicting trends—for example, I2 and I4 remained stable or improved, while I0 and I3 indicated degradation. This highlights a lack of synchronization between indicators, which should ideally follow a similar aging trajectory in the absence of interventions.

In segments that experienced major interventions, improvements were sometimes visible but not always consistent across all indicators. For example, in the stretch intervened between 2009 and 2013, I0 clearly reflected the intervention, while I2 showed an unexpected lack of response, and I3/I4 revealed mixed or muted trends. Similarly, in the post-2013 intervention stretch, only I2 and I4 reflected the improvement as expected, while I0 remained stable and failed to register the intervention's impact.

On Highway N9, sections without reported interventions also showed conflicting trends. Improvements in indicators like I0 and I4 between 2013 and 2017—without any registered works—suggest that undocumented maintenance or errors in the construction date database might have occurred. In segments with confirmed interventions, indicators like I0 and I4 tended to behave as expected, showing clear improvements followed by gradual deterioration. However, I2 and I3 again demonstrated inconsistencies, with some improvements occurring prior to interventions and, in some cases, further improvement years after.

In general, the results from the visualization exploration indicate that the variations observed in the different indicators over time are inconsistent with expected patterns. Rather than a gradual and predictable deterioration in road conditions from one survey to the next, the observed changes appear almost random. This issue is most pronounced in indicators I0, I2, and I3, whereas indicator I4 exhibits slightly more stability over time. Such inconsistencies raise concerns regarding the reliability and accuracy of the collected data. A particularly intriguing observation is the apparent disconnection between the different indicators. Despite measuring distinct aspects of pavement conditions, all indicators should theoretically align in reflecting a continuous deterioration trend or a marked improvement following an intervention. However, the analysis suggests that this is not the case, further highlighting possible flaws in the measurement process. Additionally, variability in measurements increases significantly between successive surveys, even when the same company is responsible for data collection. This inconsistency suggests that external factors, measurement techniques, or equipment calibration issues could be affecting the results.

Another unexpected finding is the disparity in road conditions within the same stretch of pavement. While it is reasonable to assume that some sections of a road may deteriorate at different rates due to varying traffic loads or environmental influences, a certain level of consistency is still expected. The fact that some segments appear to be in significantly better condition than others raise further concerns about the accuracy of the measurement process. When interventions occur between two survey periods, their impact is sometimes evident in certain indicators but absent in others. This lack of uniformity is difficult to explain and points to potential deficiencies in data collection and processing. The involvement of multiple companies in the measurement process seems to be another contributing factor to data inconsistencies, with the extent of the issue depending on the specific indicator being analysed. The visualization analysis also revealed considerable discrepancies in the quantity and consistency of data available across different *Filialen*. These missing records hinder the ability to track pavement conditions over time, limiting the effectiveness of long-term infrastructure planning and maintenance strategies.

The variable importance analysis conducted for each indicator provided further insights into the key factors influencing data variability.

For the case of surface damage, the analysis is based on paired measurements from consecutive survey years on 100-meter road segments. The primary variable of interest—Grade Difference—captures the change in I0 values between two surveys and serves as the dependent variable in the modeling process. A variety of independent variables were considered to understand this variability. These included the measuring speeds in both survey years, the difference between them, whether an intervention occurred between surveys, the time elapsed between measurements, whether the surveying organization changed, and several climatic factors, such as temperature, wind speed, and relative humidity differences between the two surveys.

The descriptive analysis showed that the Grade Difference was generally symmetrically distributed, with a slight positive skew—suggesting a natural tendency for gradual surface deterioration. Measuring speeds exhibited bimodal and trimodal distributions, indicating varying operational settings during surveys. Notably, about 48% of the road sections were measured by different organizations in consecutive years, and only 13% underwent documented interventions. Climate-related variables, although centered around zero, exhibited substantial extremes in some cases, particularly temperature differences of up to ±20°C.

A correlation matrix revealed important insights: there was a moderate positive correlation (0.59) between measuring speed in the second year (MS2) and Grade Difference, suggesting that faster measurement speeds in the second survey year tend to result in higher reported surface degradation. Conversely, the presence of an intervention showed a negative correlation (-0.32) with Grade Difference, aligning with the expectation that interventions typically lead to improved conditions. Strong correlations among measuring speed variables confirmed the internal consistency of speed-related variables. Organizational changes showed moderate correlations with speed and grade differences, pointing to possible procedural or calibration differences between surveying entities. Nevertheless, for about twenty years, the measuring methods have undergone major changes. Originally, the visual survey was conducted in a vehicle driving at low speed (or sometimes by foot) where an operator had to define the type of distresses. Years after, the operator remotely assessed the condition of the road surface from the office by analysing high-resolution images captured in the field. Nowadays, most companies assess surface damage by using advanced imaging systems (*i.e.* 3D lasers) that are able to *detect* automatically and classify the distresses according to VSS 40 925. These images are recorded continuously at a fixed measurement interval. Consequently, while variable measurement speed may appear to have no practical effect on data quality, it may, in fact, hide the influence of methodological changes in visual survey techniques.

To model and predict the Grade Difference, five ensemble learning algorithms were employed: CatBoost, XGBoost, Random Forest, LightGBM, and AdaBoost. Initial results showed CatBoost and XGBoost as the top-performing models, each achieving an R^2 of 0.689. After applying Bayesian Optimization (BO) to fine-tune hyperparameters, CatBoost further improved to $R^2 = 0.693$ with the lowest RMSE of 0.422, making it the most accurate model for this dataset. These results indicate that optimized ensemble models can explain nearly 70% of the variance in surface damage measurements, with high predictive reliability.

To interpret the model results, SHAP (SHapley Additive Explanations) values were calculated for the CatBoost model. These allowed the identification of feature importance and the direction of each variable's impact on the model output. Again, measuring speed emerged as the most influential factor, strongly linked to positive changes in grade differences. This implies that higher speeds during surveys tend to yield higher degradation scores, potentially due to reduced resolution or noise sensitivity in data acquisition. Temperature differences were the second most impactful variable, with both extreme increases and decreases affecting the measurement outcomes. This demonstrates the sensitivity of surface condition assessments to climatic variation.

Organizational changes ranked third, with clear evidence that a change in surveying organization often resulted in greater measurement differences—likely due to differences in equipment calibration or subjectivity of assessment protocols. The Comparison Period showed a non-linear effect. Both short and long intervals between measurements tended to result in smaller grade differences, possibly due to stabilizing conditions over time or limited observable changes in very short windows. Wind speed and humidity differences had smaller, more nuanced effects, while intervention presence was linked to lower grade differences—as expected. However, due to the limited number of intervention cases in the dataset, conclusions regarding their precise effect remain tentative and warrant further investigation with a more balanced dataset.

In summary, for indicator I0, the data studied with the different assessment methods shows that the most critical factor affecting measurement discrepancies among the analysed variables is measuring speed, followed by temperature and differences between measurement companies. This model accounts for approximately 70% of the observed variability, leaving 30% unexplained.

For longitudinal evenness, the analysis aims to understand how external factors—such as interventions, environmental conditions, and organizational differences—influence the variability in I2 measurements, defined through the Grade Difference: the change in indicator value between two survey years on the same 100-meter road segment. To model this variability, six input variables were selected. The dependent variable is the grade difference between two surveys. The independent variables include: Intervention, a binary variable indicating whether maintenance or rehabilitation was carried out between surveys, test year length, the time interval between the two survey, project difference, indicating whether the same or different contractor performed the surveys, temperature difference, dew point difference, and relative humidity difference between the two measurement periods. After cleaning and filtering the dataset to ensure weather and contractor data were available, a total of 95,439 rows remained.

A visualization of the input variables revealed several trends. Most road segments did not experience an intervention between surveys, and the time gap was generally four to five years. Importantly, temperature and humidity differences between surveys were often substantial, highlighting that weather variability could play a role in measurement inconsistencies. As for contractor changes, there was a slight majority of road segments surveyed by the same organization across both years.

Correlation analysis showed generally low relationships between the independent variables and the grade difference. However, one notable exception was the Intervention variable, which consistently displayed the highest (though modest) negative correlation with grade difference, averaging around -0.3. This indicates that segments which underwent interventions tended to show reduced I2 values in the subsequent survey—consistent with improved pavement evenness after maintenance. This result held true both at the level of individual *Filialen* and in the aggregated dataset, reinforcing the practical relevance of interventions in explaining longitudinal evenness improvements. Despite identifying this pattern, the overall predictive power of machine learning models remained limited. Several models were tested, including CatBoost, LightGBM, XGBoost, Random Forest, and ensemble learning methods. The best-performing model, CatBoost, achieved an R^2 of 0.35, meaning only 35% of the variance in grade differences could be explained. Hyperparameter tuning—such as increasing tree depth and the number of estimators—did not yield significantly better results. Alternative methods, including neural networks, also failed to outperform the tree-based models. This moderate model performance suggests that a significant portion of the variability in longitudinal evenness cannot be captured using the analysed variables alone. The high degree of unexplained variance could stem from random measurement error, unmeasured variables (e.g., traffic loading, material type, or pavement layer conditions), or inherent noise in how the I2 indicator is derived. This aligns with earlier observations from exploratory analysis, which highlighted inconsistency in measurement behavior even across similar segments. Nevertheless, the 35% of explained variability still offers useful insight.

A feature importance analysis confirmed that "Intervention" was consistently the most influential variable across all models which is expected (though most roads did not experience intervention). This underlines its central role in the evolution of longitudinal evenness. Interventions generally lead to smoother surfaces and lower I2 scores, validating the indicator's responsiveness to road maintenance. Weather-related variables also played a noteworthy role. Relative humidity difference and temperature difference emerged as relevant contributors to I2 variability. These findings suggest that environmental conditions at the time of measurement can influence the results, either directly (e.g., surface expansion or contraction) or indirectly through measurement equipment sensitivity. Therefore, greater attention should be paid to the climatic context in which surveys are conducted, especially when comparing results across years. In contrast, contractor differences (project difference) showed relatively lower importance, indicating that measurement consistency between organizations may be higher than initially expected, at least for this indicator. Still, its inclusion remains important for quality control, especially when considering multi-year monitoring programs.

In summary, for the indicator I2, the available models could only explain 35% of the variability, with interventions being the most influential factor, followed by weather parameters such as humidity and temperature.

For the case of the transversal evenness, the study analyses how different external and procedural factors influence the variation in this indicator across repeated road surveys. Again, the primary variable of interest is the Grade Difference—defined as the change in I3 scores between two consecutive surveys on the same 100-meter road segment. Seven independent variables were used in this analysis: Intervention, a binary variable indicating if maintenance occurred between the two surveys, company, a binary variable showing whether the same or different contractor conducted both measurements, temperature difference, wind direction difference, wind speed difference, visibility difference, and relative humidity difference, all representing climatic conditions at the time of measurement.

Initial exploration through variable distribution plots provided useful insights. The Grade Difference was roughly symmetrically distributed, suggesting that typical surface deterioration was frequently offset by condition improvements due to interventions. As expected, most road segments showed no intervention between surveys, while about half were surveyed by different companies, a factor that may introduce inconsistencies in measurement due to equipment or methodology differences. Weather variables—particularly temperature and humidity—showed normal distributions centered around zero, though extreme differences (e.g., over 10°C or 20°C in temperature) were common. This highlighted the potential impact of environmental variability on the survey results.

A correlation heatmap was created to assess relationships between the independent and dependent variables. Overall, correlation values were low, which suggests that I3 measurement differences result from a complex interaction of weakly related variables. Among all factors, the change of company between surveys had the highest correlation with grade difference, indicating that organizational practices and technologies may significantly influence the measurement of transversal evenness. Although climatic variables showed some inter-correlation (e.g., between temperature and humidity), they were not so high as to suggest redundancy, making them valid inputs for modeling.

To explore predictive relationships, various machine learning models were applied. The best performance was obtained using the LightGBM model, which achieved an R^2 of 0.4958 and an RMSE of 0.441, indicating that around 50% of the variability in grade differences could be explained using the selected variables. This is a moderately strong result and suggests that while environmental and procedural variables do influence measurement outcomes, a significant portion of the variance remains unexplained—likely due to unmeasured variables such as traffic loading, pavement structure, material properties, or measurement system calibration. Notably, increasing model complexity (e.g., deeper trees or more estimators) during parameter tuning did not improve performance, and in some cases, reduced model effectiveness. Other advanced techniques, including neural networks, were also tested but failed to significantly outperform the tree-based models. These findings reinforce the idea that randomness and noise, as well as unrecorded variables, limit the ability of predictive models to fully capture the behavior of Indicator I3.

Despite these limitations, the variable importance analysis provided meaningful insight. Company difference emerged as the most important factor influencing grade differences. This strongly suggests that organizational variation, whether due to equipment, calibration, or operational procedures has a substantial effect on the consistency of transversal evenness assessments. This confirms the need for standardized protocols across surveying entities to ensure data comparability over time. Intervention was the second most influential factor. This aligns with expectations, as maintenance activities typically lead to significant surface improvements, reducing rut depth and improving evenness. In the dataset, interventions were linked to negative grade differences, reflecting improved conditions in the second survey. This highlights the sensitivity of Indicator I3 to road maintenance efforts and reinforces its value as a reliable marker of pavement performance over time. Climatic variables, particularly temperature and relative humidity, also contributed meaningfully to the model, though their individual effects were weaker than those of intervention or company difference. Their influence likely results from the physical effects of temperature fluctuations on pavement materials (e.g., expansion, contraction, softening) and how moisture or visibility conditions may affect data capture during measurement runs. However, the impact of each single weather variable was less pronounced, and their combined effect likely explains more of the measurement variability.

Resuming, the analysis for indicator I3 highlights the strong impact of the measurement company on variability, with an explained variance of around 50%. This finding underscores the need for greater standardization in measurement methodologies. Additionally, interventions and weather conditions contribute to variations in indicator I3.

Surface friction data was collected using the RK-SKM 80 test vehicle, with a standard measurement frequency of every 100 meters and an operating speed of 80 km/h. To ensure spatial consistency across multiple survey years, despite natural fluctuations in measurement starting points, the data was standardized using a nearest-neighbor approach, aligning it to fixed 100-meter segments from a defined zero marker. The analysis concentrated on the center lane and used 1-kilometer intervals as the basic analysis unit by aggregating ten consecutive measurements. The study covered 12 major highways (N1–N16, excluding four), encompassing 474 kilometers of road. For each road section, survey years were selected to ensure repeated measurements at consistent locations across time, yielding a dataset of 1,322 valid road samples. Each sample contained data from two consecutive surveys, and when missing values were detected within the 1-km interval, the sample was excluded to preserve data quality.

To measure the extent of variability in friction measurements, a novel metric called the Average Difference (AD) was introduced. This metric calculates the average change in friction values between two survey years within a 1-kilometer segment. The distribution of AD followed a slightly negatively skewed Gaussian pattern, suggesting a general trend toward improvement in surface friction values—often due to interventions—over time.

Factors potentially influencing AD were grouped into three categories: Measuring information: including whether an intervention occurred (binary), the project difference or change of inspection company between years, and test year length indicating the time between surveys. Another category was measuring conditions: such as measurement speeds in each year, their difference, and the standard deviations of speed during testing. Finally, weather conditions were also used: including temperature difference and wind speed difference between the two survey periods.

The correlation analysis showed low direct correlations between AD and individual variables, indicating that no single factor dominantly explains friction variability, and suggesting a multi-factorial influence. However, the presence of interventions and the change of contractor between years emerged as variables of interest. Machine learning models were then used to investigate the data. A CatBoost regression model demonstrated outstanding performance, achieving an R^2 of 0.9186 and an RMSE of 0.362, explaining more than 92% of the variability in the AD. This high level of predictive accuracy highlighted that the recorded variables effectively capture the key influences on surface friction changes. Notably, attempts to further optimize the model through parameter tuning did not result in significant performance gains.

To move beyond correlation and investigate causal relationships, a structural causal inference framework was applied using the DoWhy library. A causal graph was constructed based on domain knowledge and SHAP (SHapley Additive exPlanations) analysis results, mapping out the hypothesized causal relationships between the input variables and AD. This allowed the computation of Average Treatment Effects (ATEs) for each variable. The causal analysis revealed that interventions had the most significant positive causal effect on friction improvement, with an ATE of 0.7193. This means that, on average, interventions resulted in a 71.93% greater friction value difference compared to segments with no interventions—strongly supporting the effectiveness of maintenance. In contrast, several factors, including project difference, measurement speed, measurement speed difference, and temperature difference, were found to have negative causal effects, although the interpretation of their directionality is less straightforward due to their continuous nature. The SHAP analysis also confirmed temperature difference and intervention as the most influential features in explaining AD. Temperature differences may affect pavement texture or the responsiveness of the measuring system, while the importance of interventions aligns with their physical impact on road surface quality. The SHAP beeswarm plot further revealed that test year length and standard deviation of measurement speed also influence the AD, particularly through outlier values. Finally, robustness checks were conducted on the causal inference results using three techniques: introducing a random common cause, placebo treatment, and data subset validation. All tests confirmed the stability and reliability of the estimated causal effects, strengthening confidence in the findings.

Conversely, indicator I4 exhibits the highest degree of explainability, with up to 91% of the variability accounted for. Temperature is the most significant influencing factor, followed closely by wind speed, interventions, and measuring speed.

The analysis of differences between regional branches or *filialen* focused on Indicators I3 (transversal evenness) and I0 (surface damage) revealed notable disparities in data quality, quantity, and variability drivers. Some *filialen* maintained more complete and consistent datasets across survey years and indicators, while others showed data gaps or missing survey years. These inconsistencies raise questions about whether the data was never collected or simply not uploaded, which hinders the ability to generate reliable nationwide insights. Additionally, the primary factors influencing indicator variability differ by *filiale*, possibly due to regional differences in geography, climate, or traffic loads.

For Indicator I0, similar patterns were observed. While measuring speed is generally a dominant factor across *filialen*, in *Filiale 2* (Thun), weather-related variables had a greater impact. This again may reflect regional environmental differences or procedural variations in how measurements are carried out. Furthermore, the proportion of explained variability also varied across *filialen*, indicating that some regions have more unexplained noise in their data than others. These findings are important for harmonizing data collection standards and ensuring consistent evaluation practices nationwide, especially if the goal is to develop general rules or update national-level pavement condition norms.

Based on these findings, several recommendations are proposed to improve the reliability and usability of the collected data.

First, the systematic visualization of data should be enhanced to facilitate quick assessments of pavement conditions and data consistency. By developing an automated visualization system that updates with new data, stakeholders such as road managers and measurement companies would have easier access to crucial information. This approach would allow for the rapid identification of potential measurement issues, as well as missing data that could compromise long-term assessments. Second, weather parameters should be systematically measured at the same time as pavement indicators. The use of portable weather stations, which are now both affordable and reliable, would provide essential data on temperature, humidity, and other environmental conditions that may impact measurement results. Additionally, survey scheduling should be optimized to ensure that measurements are conducted under conditions similar to those of previous surveys. This would help to reduce inconsistencies caused by external environmental factors. Third, to mitigate the impact of different companies on measurement variability, contractual agreements should ensure that the same company is responsible for at least two consecutive survey years. This approach would reduce inconsistencies arising from methodological differences between companies. Furthermore, all companies conducting measurements should be required to regularly calibrate their equipment and conduct comparative assessments to ensure uniformity in data collection processes. Fourth, the process of recording interventions should be more systematic and comprehensive. The current inconsistencies in intervention data make it difficult to assess their true impact on pavement conditions. Both major and minor interventions should be documented in detail, including information on the type of intervention, the extent of repairs, and materials used. This would enhance the ability to correlate intervention activities with subsequent pavement condition changes. Fifth, the scope of data analysis should be expanded to incorporate additional variables that could help explain measurement variability. In particular, the inclusion of detailed traffic data—distinguishing between heavy trucks and passenger vehicles—would provide a more accurate picture of pavement stress and degradation patterns. Additionally, integrating material properties as well as topographic and soil parameters into the analysis could offer further insights into how geographical and geological conditions influence pavement deterioration regarding the type of road surface. Finally, investment in the further development of the TRA-Trassee application is strongly recommended. Enhancing its functionality to integrate all pavement-related data, including visualization tools and analytical capabilities, would significantly improve its usability. Making the system more accessible to managers, measurement companies, researchers, and policymakers would facilitate better decision-making and ensure that critical data is effectively utilized.

In conclusion, the findings of this study highlight significant inconsistencies in road condition measurements, underscoring the need for improved data accuracy and reliability. Implementing the proposed recommendations—including enhanced data visualization, systematic weather measurements, standardized company practices, detailed intervention documentation, expanded data analysis, and the development of TRA-Trassee will contribute to more effective pavement condition monitoring and long-term infrastructure planning. These improvements will ultimately ensure the sustainability and efficiency of Switzerland's national road network.

Zusammenfassung

Die Schweiz verfügt über ein dichtes Strassennetz mit nationalen, kantonalen und kommunalen/privaten Strassen. Obwohl die Nationalstraßen nur 3 % des Netzes ausmachen, wickeln sie rund 40 % des Verkehrs und 62,9 % des Güterverkehrs ab. Die Verwaltung und der Unterhalt der Strassen sind zwischen dem Bundesamt für Strassen (ASTRA), den kantonalen Behörden und den Gemeinden aufgeteilt.

Der Zustand der Strassen verschlechtert sich mit der Zeit aufgrund von Alterung, Witterungsbedingungen und Nutzungsfaktoren wie Salz im Winter oder Geländeveränderungen. Um ein optimales Serviceniveau aufrechtzuerhalten, muss der Strassenzustand regelmässig bewertet werden. Schweizer Strassenverwalter befolgen strenge Richtlinien, um die Genauigkeit, Nachvollziehbarkeit, Wiederholbarkeit und Reproduzierbarkeit bei der Bewertung von Strassenbelägen sicherzustellen. Die Genauigkeit bezieht sich auf die Minimierung von Messfehlern, die Verständlichkeit gewährleistet eine klare Dateninterpretation, und die Wiederholbarkeit/Reproduzierbarkeit analysiert die Konsistenz der Messungen unter verschiedenen Bedingungen.

Pavement Management Systems (PMS) sind strategische Instrumente, die von Strassenbaubehörden zur Planung, Priorisierung und Optimierung von Instandhaltungs- und Sanierungsmassnahmen in Strassennetzen eingesetzt werden. Ihr Ziel ist es, die Beläge auf dem gewünschten Leistungsniveau zu halten und gleichzeitig die Kosten langfristig zu kontrollieren. Die Effizienz von PMS hängt stark von der Genauigkeit, Vollständigkeit und Konsistenz der Zustandsdaten ab. Zu den Einflussfaktoren gehören die Qualität der Standortreferenzierung, die Konsistenz der historischen Daten, die umfassende räumliche und zeitliche Abdeckung und die Anpassung an die sich entwickelnden Praktiken des Infrastrukturmanagements. Fehler bei der Datenerfassung oder -verarbeitung können zu fehlerhaften Behandlungsempfehlungen und einer Fehlallokation von Instandhaltungsbudgets führen.

Bei der Datenerfassung auf Netzebene werden in der Regel Hochgeschwindigkeitsmethoden wie Profilometer und Trägheitsprofiler bevorzugt, da sie schnell sind, weniger Arbeit erfordern und den Verkehr nur minimal beeinträchtigen. Diese Methoden können jedoch zu Datenschwankungen führen, die durch Unterschiede in der Gerätekalibrierung, seitliche Positionierung, Umweltbedingungen und saisonale Schwankungen verursacht werden. Betreiber und Dienstleister können unterschiedliche Messmethoden anwenden, was die Konsistenz weiter beeinträchtigt. Daher sind strenge Gerätevalidierungen und Wiederholbarkeitsstudien unerlässlich, um die Datenqualität zu gewährleisten und aussagekräftige Vergleiche über Zeit und Regionen hinweg zu ermöglichen.

Jüngste Forschungsarbeiten haben die Anwendung von Techniken des maschinellen Lernens zur Modellierung der Verschlechterung von Strassenbelägen und zur Leistungsprognose untersucht. Fortgeschrittene Rahmenwerke integrieren nun Methoden wie BorutaShap für die Merkmalsauswahl, Bayes'sche Neuronale Netze (BNNs) für die probabilistische Modellierung und SHAP-Werte für die Modellinterpretation. Diese

Ansätze verbessern die Fähigkeit, komplexe Wechselwirkungen zwischen Belagsdesign, Verkehr, Umwelt und Instandhaltungshistorie zu erfassen. Studien zeigen jedoch, dass die grösste Unsicherheitsquelle bei diesen Modellen die schlechte Datenqualität ist, was die Notwendigkeit einer verbesserten Datenverwaltung und Validierungsverfahren unterstreicht.

Visuelle Inspektionen sind nach wie vor eine wertvolle Bewertungsmethode, vor allem in Umgebungen ohne Zugang zu automatischen Systemen. Ihre Zuverlässigkeit variiert je nach Erfahrung und Ausbildung der Prüfer sowie dem zu bewertenden Belagszustand. Inspektionen sind im Allgemeinen genauer für Beläge in gutem Zustand, während die subjektive Interpretation an der Grenze zwischen „gut“ und „schlecht“ zu mehr Fehlern führt. Es hat sich gezeigt, dass Schulungsprogramme und Kalibrierungsübungen sowohl die Genauigkeit als auch die Konsistenz erheblich verbessern können. Monte-Carlo-Simulationen und Fehlerübergangsmatrizen wurden eingesetzt, um die Auswirkungen von Fehlern bei der visuellen Inspektion auf die Zustandsindizes zu quantifizieren, insbesondere bei mittleren und hohen Schadensschweregrade.

Messungen der Oberflächenrauheit, die häufig mit dem Internationalen Rauheitsindex (IRI) quantifiziert werden, sind bei Netzbewertungen weit verbreitet. Die IRI-Werte können jedoch je nach Profiltyp, Fahrbahnposition, Verkehrsaufkommen und Belagstyp erheblich variieren. Studien zeigen, dass einzelne Testläufe die Variabilität des Zustands nicht angemessen erfassen können, insbesondere auf Staatsstrassen. Vergleichende Analysen von Profilerdaten mit Referenzgeräten machen deutlich, dass eine regelmässige Kalibrierung, mehrere Wiederholungen und eine statistische Auswertung erforderlich sind, um eine sinnvolle Interpretation zu gewährleisten. Auch geografische und saisonale Einflüsse tragen zu Messschwankungen bei.

Die Griffigkeit, ein wichtiger Indikator für die Verkehrssicherheit, wird durch die Oberflächenbeschaffenheit, die Eigenschaften der Zuschlagstoffe, das Wetter und die Verkehrsbelastung beeinflusst. Techniken wie der GripTester und dynamische Reibungstester (DFT) werden neben Laserscanning-Systemen zur Messung der Oberflächenbeschaffenheit und Reibung eingesetzt. Ein wichtiger Einflussfaktor sind jahreszeitliche Schwankungen. Beispielsweise wird die Grifigkeit durch Polieren im Sommer verringert. Langfristige Studien haben Reibungstrends mit Veränderungen der Texturtiefe, Niederschlagsmustern und starkem Fahrzeugverkehr in Verbindung gebracht. Zu den jüngsten Innovationen gehört die Entwicklung statistischer Modelle zur Schätzung der Reibungswerte in der Trockenzeit anhand von Daten aus der Regenzeit.

Statistische Methoden spielen eine zentrale Rolle beim Umgang mit der Variabilität und den Fehlern in Belagsdaten. Werkzeuge wie die Messsystemanalyse (MSA), die Wiederholbarkeit und Reproduzierbarkeit von Messgeräten (GR&R) und latente Markov-Modelle ermöglichen es Forschern, Fehlerquellen zu identifizieren, zu quantifizieren und zu reduzieren. MSA ist in anderen Branchen weit verbreitet, wird aber zunehmend auch in der Verkehrsforschung eingesetzt, um die Zuverlässigkeit von Messsystemen zu bewerten. Hybride Methoden, die Versuchsplanung, Simulation und fortschrittliche probabilistische Modellierung kombinieren, bieten tiefere Einblicke in den Einfluss von Bauqualität, Verkehrsbelastung und Umwelteinflüssen auf die Verschlechterung von Strassenbelägen.

Schliesslich erhöht die Integration moderner Analysemethoden mit traditionellen technischen Verfahren die Präzision und Zuverlässigkeit der Zustandsbewertung von Strassenbelägen. Während automatisierte und auf maschinellem Lernen basierende Methoden vielversprechend sind, hängt ihr Erfolg von einem rigorosen Datenmanagement, konsistenten Messverfahren und gut kalibrierten Systemen ab. Angesichts alternder Strassennetze und knapper werdender Budgets wird die Fähigkeit, datengestützte und zuverlässige Entscheidungen über Instandhaltungsprioritäten zu treffen, weiterhin von entscheidender Bedeutung sein. Kontinuierliche Investitionen in Messtechnik, Mitarbeiterschulungen und die länderübergreifende Standardisierung von Daten werden weltweit zu einem widerstandsfähigeren und kostengünstigeren Management der Strasseninfrastruktur beitragen.

Die Bewertung des Strassenzustands in der Schweiz wird durch eine Reihe von nationalen Normen geregelt, die jeweils spezifische Verfahren und Anforderungen für die Bewertung der Leistungsfähigkeit von Strassenbelägen enthalten. Obwohl jeder Teil des Überwachungsprozesses durch eine eigene Norm abgedeckt ist, gibt es häufig Überschneidungen und Fälle, in denen mehrere Normen für dieselbe Aufgabe gelten. Diese Normen sind insbesondere nach Bau- oder Sanierungsphasen sowie während der Betriebsdauer von Strassennetzen relevant. Die Norm VSS 40 525 beispielsweise definiert die Abnahmekriterien für Strassenoberflächen nach Abschluss der Arbeiten und konzentriert sich auf vier Eigenschaften: Einzelunebenheiten, Längsebenheit, Querebenheit und Oberflächenreibung. Einzelunebenheiten werden bewertet, wie in SN 640 516-7 beschrieben, mit spezifischen Akzeptanzschwellen für verschiedene Belagsschichten. Die Ebenheit in Längsrichtung wird anhand des Längsprofils der Strasse gemessen, wobei die VSS 40 517 zwei Methoden zulässt: eine auf der Grundlage der Winkelvariation (W und Standardabweichung Sw) und eine andere auf der Grundlage der Wellenbandanalyse (NBO), bei der kurze, mittlere und lange Oberflächenwellen untersucht werden. Die zweite Methode ist umfassender und besser geeignet, um mehrstufige Oberflächenunregelmässigkeiten zu erkennen. Die transversale Ebenheit wird durch die Messung der Spurrinnentiefe und der theoretischen Wassertiefe innerhalb der Spurrinnen in Anlehnung an die VSS 40 518 bewertet. Für die Spurrinnentiefe selbst gibt es keine spezifischen Akzeptanzschwellen, wohl aber für die Wassertiefe, insbesondere aus Sicherheitsgründen auf Hochgeschwindigkeitsstrassen. Auf Strassen mit Geschwindigkeitsbegrenzungen über 80 km/h beträgt die maximal zulässige theoretische Wassertiefe 4,0 mm. Die Oberflächenreibung, ein weiterer kritischer Parameter, wird mit dynamischen und statischen Methoden nach VSS 40 512 bewertet. Zu den dynamischen Systemen gehören das Skiddometer BV 8 und das SKM-Gerät, während die Pendelmethode (SRT) als statische Alternative dient. Diese Tests berücksichtigen die jahreszeitlichen Bedingungen und werden in der Regel 3-6 Monate nach der Verkehrsfreigabe der Strasse durchgeführt, wobei geschwindigkeitsspezifische Schwellenwerte angewendet werden.

Der Zustand der Fahrbahn kann anhand von vier in der VSS 40 925b definierten Indizes zusammengefasst werden: I0 (Oberflächenschäden), I2 (Längsebenheit), I3 (Querebenheit) und I4 (Oberflächenreibung). Jeder Index wird von 0 (sehr gut) bis 5 (sehr schlecht) bewertet, basierend auf einem oder mehreren Parametern. Die Auswahl der zu bewertenden Parameter und die Häufigkeit der Bewertung hängen von der

Funktion der Strasse, der Verkehrsbelastung und der Verwaltungsmethodik ab. Auf Netzebene werden diese Indizes gemäss SN 640 900 verwendet und in Managementsystemen gemäss VSS 40 944 und VSS 40904 gespeichert. Die Bewertung auf Projektebene erfolgt nach den Normen VSS 40 730 und VSS 40 925b. Während die meisten Indikatoren das ganze Jahr über bewertet werden können, solange der Belag trocken ist, müssen bei der Bewertung der Oberflächenreibung und der Tragfähigkeit die Temperaturbedingungen beachtet werden.

Die Bewertung von Oberflächenschäden unterscheidet sich bei bituminösen und Betonoberflächen. Bei Asphaltbelägen umfasst die Bewertung Glätte, Verformung, strukturelle Schäden und Reparaturen. Bei Beton werden Faktoren wie Materialverlust, Fugenverschlechterung, vertikale Verschiebungen und Rissbildung berücksichtigt. In dieser Arbeit haben wir uns auf bituminöse Oberflächen konzentriert, die den Autobahnen entsprechen. Herkömmliche Methoden beruhen auf Inspektoren, die den Abschnitt zu Fuss oder mit dem Auto abfahren, obwohl viele Unternehmen inzwischen hochauflösende, auf Fahrzeugen montierte Bilder verwenden, die dann von geschulten Fachleuten - oder zunehmend auch von Systemen mit künstlicher Intelligenz - analysiert werden. Oberflächenschäden werden durch ihr Ausmass (Skala 0-3) und ihren Schweregrad (Skala 1-3) charakterisiert, die mithilfe einer Matrix kombiniert werden, um eine Bewertung von 0 bis 9 zu erhalten. Bei der zusammenfassenden Bewertung werden die Parameter gruppiert, bewertet, gewichtet und zum I0-Index zusammengefasst. Die Standardabschnittslängen betragen 50 Meter für vollwertige Fahrbahnen und 100 Meter für einspurige Strassen. In der Regel wird der schlechteste Schadensschweregrad pro Gruppe und ein repräsentativer Umfang angegeben. Für detaillierte Bewertungen auf Projektebene werden beide Dimensionen für jeden einzelnen Parameter erfasst.

Die Längsebenheit (I2) wird mit verschiedenen Methoden bewertet, wobei die direkteste die Messung der Winkeländerungen (W) zwischen benachbarten 1-Meter-Abschnitten der Fahrbahnoberfläche ist. Die Standardabweichung Sw dieser W-Werte wird zur Ableitung des Index verwendet, wie in einer Grafik in der VSS 40517 dargestellt. Zu den alternativen Methoden gehören der Internationale Rauheitsindex (IRI), der den Fahrkomfort für den Benutzer auf der Grundlage der simulierten vertikalen Bewegung eines Fahrzeugaufhängungssystems schätzt, die NBO-Wellenbandmethode, die die Energieverteilung über verschiedene Wellenfrequenzen analysiert, und das gewichtete Längsprofil (BLP), das die Amplitudenbeiträge verschiedener Wellenlängen ausgleicht. Diese Methoden helfen bei der Isolierung periodischer oder isolierter Fehler. Die Norm bietet jedoch nur eine direkte Umwandlung von Sw in I2-Werte und gibt nicht an, wie die Ergebnisse anderer Verfahren in die I2-Skala umgerechnet werden können.

Die Querebenheit (I3) basiert auf Messungen der Spurrinnentiefe und dem damit verbundenen Wasseransammlungspotenzial in diesen Spurrinnen. Die Spurrinnen können manuell mit einem Lineal oder einer Schnur gemessen werden, wobei anzugeben ist, welche Methode verwendet wird. Es werden Werte von der linken und rechten Radspur benötigt, und es müssen mehrere Messungen über den Bewertungsabschnitt durchgeführt werden. Bei der Bewertung von Strassennetzen darf nur eine Spurrille bewertet werden. Der Abstand zwischen den Querprofilpunkten darf 10 cm nicht überschreiten, und die vertikale Genauigkeit der Messungen sollte 0,5 mm für dynamische und 1,0 mm für statische Systeme betragen. Es werden mindestens 10 Messungen pro Strassenabschnitt empfohlen, wobei sowohl statistische Zusammenfassungen als auch Extremwerte dokumentiert werden sollten. Wenn beide Spurrinnen gemessen werden, wird der schlechteste Wert für I3 verwendet.

In der Praxis werden sowohl I2- als auch I3-Messungen mit modernen Laserprofilometern durchgeführt, die in Vermessungsfahrzeuge integriert sind. Diese Technologien gewährleisten eine hohe Messgenauigkeit und -geschwindigkeit, die für gross angelegte Bewertungen des Strassennetzes unerlässlich sind. Die Systeme sind in der Lage, die Strassenoberfläche präzise zu scannen und anschliessend standardisierte Parameter für definierte Abschnitte (z. B. Radspuren) gemäss den entsprechenden Vorgaben zu berechnen. Trotz der traditionellen manuellen Methoden bieten automatisierte Systeme eine grössere Effizienz, Zuverlässigkeit und Reproduzierbarkeit und sind heute Standard bei den meisten Überwachungsmassnahmen. Insgesamt ist der Schweizer Ansatz zur Strassenzustandsbewertung stark strukturiert und standardisiert, was die Konsistenz über Projekte und Regionen hinweg gewährleistet. Die Integration von Normen mit modernen Messsystemen und zunehmenden digitalen Arbeitsabläufen unterstützt die datengestützte Entscheidungsfindung bei der Planung von Strassenerhaltung und -sanierung.

Das schweizerische Nationalstrassennetz wird in zwei Bereiche unterteilt: West und Ost. Die Division West betreut die regionalen Filialen Estavayer-le-Lac und Thun, die Division Ost die Filialen Zofingen, Winterthur und Bellinzona. Die regionalen Niederlassungen werden in der Schweiz „Filialen" genannt. Jede Filiale ist für die Messung der Strassenzustandsindikatoren in dem ihr zugewiesenen Gebiet verantwortlich. Die gesammelten Daten werden in der Anwendung TRA-Trassee gespeichert, die die Verwaltung und Planung der Strasseninfrastruktur unterstützt. Die Daten werden vor der Integration in das System überprüft, wobei die Qualitätskontrolle durch begrenzte Ressourcen beeinträchtigt wird. Zu den wichtigsten Unternehmen, die sich mit der Messung des Strassenzustands befassen, gehören die Schniering GmbH (jetzt Teil des TÜV Rheinland) und Infralab SA, die fortschrittliche Hochgeschwindigkeitsmessverfahren einsetzen. Die Daten werden von Geologix AG vor der endgültigen Speicherung bereinigt. In dieser Studie konzentrieren wir unsere Analyse auf zwei Methoden, um die komplexen und grossen Datensätze zu untersuchen.

Einerseits spielen Visualisierungen eine Schlüsselrolle bei der Datenanalyse, da sie helfen, Unstimmigkeiten und Trends im Zeitverlauf auf einfache und intuitive Weise zu erkennen. Verschiedene statistische Methoden wie Histogramme und Streudiagramme helfen bei der Bewertung der Datenverteilung, der Erkennung von Ausreissern und der Sicherstellung der Konsistenz der Strassenzustandsindikatoren. Auf der anderen Seite ist auch ein komplexerer Ansatz wie die Faktorenanalyse ein Werkzeug, das als wichtiger Bestandteil der Strassenzustandsbewertung angesehen werden kann. In dieser Studie integrieren wir Wetterdaten von MeteoSchweiz, indem wir jeder Filiale relevante meteorologische Stationen zuordnen. Die Kombination von Wetterdaten und Messfaktoren ermöglichte es uns, die Bedeutung für die Erklärung der Variabilität in den Datensätzen genauer zu untersuchen. Um die Vorhersagegenauigkeit zu verbessern, werden fortschrittliche Modelle des maschinellen Lernens - darunter

AdaBoost, CatBoost, LightGBM, Random Forest und XGBoost - mithilfe der Bayes'schen Optimierung (BO) optimiert. Diese Ensemble-Lernmethoden helfen bei der Erfassung komplexer Zusammenhänge in den Daten zum Belagszustand. Die Modellleistung wird anhand von R^2- und RMSE-Metriken bewertet, um die genauesten Vorhersagen zu ermitteln. SHAP (Shapley Additive Explanations) und Permutationsverfahren ermöglichen die Interpretation der Vorhersagemodelle und geben Aufschluss über den Einfluss der verschiedenen Variablen auf die Variabilität des Strassenzustands. Die Quantifizierung der Bedeutung von Merkmalen wie Klima, Interventionshistorie und Messmethoden gewährleistet einen datengesteuerten Ansatz für die Planung und Instandhaltung der Strasseninfrastruktur.

Was die Ergebnisse des Visualisierungsprozesses betrifft, so konnten wir beispielsweise für den Regionalzweig Estavayer-le-lac feststellen, dass die meisten Indikatoren auf der Autobahn N1 für Abschnitte ohne grössere Eingriffe nach 2000 erwartungsgemäss ein allgemeines Verschlechterungsmuster aufwiesen. Einige unerwartete Verbesserungen bei bestimmten Indikatoren (vor allem I2 und I4) gaben jedoch Anlass zu Bedenken hinsichtlich der Zuverlässigkeit der Messungen. In den Zeiträumen zwischen 2009 und 2013 sowie 2013 und 2017 traten Anomalien auf, bei denen verschiedene Indikatoren widersprüchliche Trends aufwiesen - so blieben beispielsweise I2 und I4 stabil oder verbesserten sich, während I0 und I3 eine Verschlechterung anzeigten. Dies verdeutlicht die fehlende Synchronisation zwischen den Indikatoren, die sich ohne Eingriffe idealerweise ähnlich entwickeln sollten.

In Abschnitten, in denen grössere Eingriffe vorgenommen wurden, waren Verbesserungen manchmal sichtbar, aber nicht immer gleichmässig bei allen Indikatoren. In dem Abschnitt, in dem zwischen 2009 und 2013 interveniert wurde, spiegelte beispielsweise I0 die Intervention deutlich wider, während I2 eine unerwartet schwache Reaktion zeigte und I3/I4 gemischte oder gedämpfte Trends erkennen liessen. In ähnlicher Weise spiegelten im Interventionsabschnitt nach 2013 nur I2 und I4 die erwartete Verbesserung wider, während I0 stabil blieb und die Auswirkungen der Intervention nicht erfasste.

Auf der Autobahn N9 zeigten Abschnitte ohne gemeldete Eingriffe ebenfalls widersprüchliche Trends. Verbesserungen bei Indikatoren wie I0 und I4 zwischen 2013 und 2017 - ohne registrierte Arbeiten - deuten darauf hin, dass nicht dokumentierte Wartungsarbeiten oder Fehler in der Baudatenbank aufgetreten sein könnten. In Segmenten mit bestätigten Eingriffen verhielten sich Indikatoren wie I0 und I4 tendenziell wie erwartet und zeigten deutliche Verbesserungen, gefolgt von einer allmählichen Verschlechterung. Die Indikatoren I2 und I3 wiesen jedoch erneut Unstimmigkeiten auf, wobei einige Verbesserungen vor den Eingriffen und in einigen Fällen weitere Verbesserungen Jahre danach auftraten.

Im Allgemeinen deuten die Ergebnisse der Visualisierungsuntersuchung darauf hin, dass die bei den verschiedenen Indikatoren im Laufe der Zeit beobachteten Schwankungen nicht mit den erwarteten Mustern übereinstimmen. Anstatt einer allmählichen und vorhersehbaren Verschlechterung des Strassenzustands von einer Erhebung zur nächsten erscheinen die beobachteten Veränderungen fast zufällig. Dieses Problem ist bei den Indikatoren I0, I2 und I3 am stärksten ausgeprägt, während der Indikator I4 im Laufe der Zeit eine etwas grössere Stabilität aufweist. Solche Unstimmigkeiten geben Anlass zu Bedenken hinsichtlich der Zuverlässigkeit und Genauigkeit der erhobenen Daten. Eine besonders auffällige Beobachtung ist die offensichtliche Trennung zwischen den verschiedenen Indikatoren. Obwohl sie unterschiedliche Aspekte des Belagszustands messen, sollten alle Indikatoren theoretisch übereinstimmen und einen kontinuierlichen Verschlechterungstrend oder eine deutliche Verbesserung nach einer Massnahme widerspiegeln. Die Analyse deutet jedoch darauf hin, dass dies nicht der Fall ist, was auf mögliche Fehler im Messverfahren hinweist. Darüber hinaus nimmt die Variabilität der Messungen zwischen aufeinanderfolgenden Erhebungen erheblich zu, selbst wenn dasselbe Unternehmen für die Datenerhebung verantwortlich ist. Diese Inkonsistenz deutet darauf hin, dass externe Faktoren, Messverfahren oder Probleme bei der Gerätekalibrierung die Ergebnisse beeinflussen könnten.

Ein weiterer unerwarteter Befund ist der unterschiedliche Strassenzustand innerhalb ein und desselben Strassenabschnitts. Zwar kann man davon ausgehen, dass sich einige Abschnitte einer Strasse aufgrund unterschiedlicher Verkehrsbelastung oder Umwelteinflüsse unterschiedlich schnell verschlechtern, aber dennoch sollte ein gewisses Mass an Konsistenz vorhanden sein. Die Tatsache, dass einige Abschnitte in einem deutlich besseren Zustand zu sein scheinen als andere, gibt Anlass zu weiteren Bedenken hinsichtlich der Genauigkeit des Messverfahrens. Wenn zwischen zwei Erhebungszeiträumen Eingriffe vorgenommen werden, sind deren Auswirkungen manchmal bei bestimmten Indikatoren erkennbar, bei anderen jedoch nicht. Dieser Mangel an Einheitlichkeit ist schwer zu erklären und deutet auf mögliche Mängel bei der Datenerhebung und -verarbeitung hin. Die Beteiligung mehrerer Unternehmen am Messverfahren scheint ein weiterer Faktor zu sein, der zu Dateninkonsistenzen beiträgt, wobei das Ausmass des Problems von dem jeweiligen analysierten Indikator abhängt. Die Visualisierungsanalyse zeigte auch erhebliche Unterschiede in der Menge und Konsistenz der verfügbaren Daten in den verschiedenen Filialen. Diese fehlenden Datensätze erschweren die Verfolgung des Belagszustands im Zeitverlauf und schränken die Effektivität der langfristigen Infrastrukturplanung und der Unterhaltsstrategien ein.

Die für jeden Indikator durchgeführte Analyse der Variablenbedeutung lieferte weitere Erkenntnisse über die Schlüsselfaktoren, die die Datenvariabilität beeinflussen.

Im Falle von Oberflächenschäden basiert die Analyse auf gepaarten Messungen aus aufeinanderfolgenden Erhebungsjahren auf 100-Meter-Strassenabschnitten. Die primäre Variable von Interesse - die Graddifferenz - erfasst die Veränderung der I0-Werte zwischen zwei Erhebungen und dient als abhängige Variable im Modellierungsprozess. Um diese Variabilität zu verstehen, wurde eine Reihe von unabhängigen Variablen berücksichtigt. Dazu gehörten die Messgeschwindigkeiten in beiden Erhebungsjahren, die Differenz zwischen ihnen, ob zwischen den Erhebungen ein Eingriff stattfand, die zwischen den Messungen verstrichene Zeit, ob die Erhebungsorganisation wechselte und verschiedene klimatische Faktoren wie Temperatur, Windgeschwindigkeit und relative Luftfeuchtigkeit zwischen den beiden Erhebungen.

Die deskriptive Analyse zeigte, dass die Neigungsdifferenz im Allgemeinen symmetrisch verteilt war, mit einer leichten positiven Schräglage, was auf eine natürliche Tendenz zur allmählichen Verschlechterung der Oberfläche hindeutet. Die

Messgeschwindigkeiten wiesen bimodale und trimodale Verteilungen auf, was auf unterschiedliche Betriebseinstellungen während der Messungen hinweist. Bemerkenswert ist, dass etwa 48 % der Strassenabschnitte in aufeinanderfolgenden Jahren von verschiedenen Organisationen gemessen wurden und dass nur bei 13 % dokumentierte Eingriffe vorgenommen wurden. Die klimabezogenen Variablen lagen zwar um den Nullpunkt herum, wiesen aber in einigen Fällen erhebliche Extreme auf, insbesondere Temperaturunterschiede von bis zu ±20°C.

Eine Korrelationsmatrix lieferte wichtige Erkenntnisse: Es zeigte sich eine mässige positive Korrelation (0,59) zwischen der Messgeschwindigkeit im zweiten Erhebungsjahr (MS2) und der Graddifferenz, was darauf hindeutet, dass höhere Messgeschwindigkeiten im zweiten Erhebungsjahr tendenziell zu einer stärkeren gemeldeten Oberflächenschädigung führen. Umgekehrt zeigte das Vorhandensein einer Intervention eine negative Korrelation (-0,32) mit der Höhendifferenz, was mit der Erwartung übereinstimmt, dass Interventionen in der Regel zu einer Verbesserung der Bedingungen führen. Starke Korrelationen zwischen den Variablen zur Messung der Geschwindigkeit bestätigten die interne Konsistenz der geschwindigkeitsbezogenen Variablen. Organisatorische Änderungen zeigten moderate Korrelationen mit Geschwindigkeits- und Notenunterschieden, was auf mögliche Verfahrens- oder Kalibrierungsunterschiede zwischen den Erhebungsstellen hindeutet. Dennoch haben sich die Messmethoden in den letzten zwanzig Jahren erheblich weiterentwickelt. Ursprünglich wurde die visuelle Zustandserhebung bei niedriger Fahrgeschwindigkeit (oder teilweise zu Fuss) durchgeführt, wobei ein Operator die Schadenmerkmale direkt vor Ort definieren musste. Einige Jahre später erfolgte die Bewertung des Fahrbahnzustands durch den Operator aus dem Büro, indem hochauflösende Bilder aus dem Feld analysiert wurden. Heutzutage erfassen die meisten Unternehmen Oberflächenschäden mithilfe fortschrittlicher Bildgebungssysteme (z. B. 3D-Lasersysteme), die in der Lage sind, Schäden automatisch zu erkennen und gemäss SN 640 925 zu klassifizieren. Diese Bilddaten werden kontinuierlich in festen Messintervallen aufgezeichnet. Folglich mag eine variable Messgeschwindigkeit zunächst keinen offensichtlichen Einfluss auf die Datenqualität haben, kann jedoch tatsächlich die Auswirkungen methodischer Veränderungen in der visuellen Zustandserhebung überdecken.

Zur Modellierung und Vorhersage des Notenunterschieds wurden fünf Ensemble-Lernalgorithmen eingesetzt: CatBoost, XGBoost, Random Forest, LightGBM und AdaBoost. Die ersten Ergebnisse zeigten CatBoost und XGBoost als die leistungsfähigsten Modelle, die jeweils ein R^2 von 0,689 erreichten. Nach Anwendung der Bayes'schen Optimierung (BO) zur Feinabstimmung der Hyperparameter verbesserte sich CatBoost weiter auf $R^2 = 0{,}693$ mit dem niedrigsten RMSE von 0,422 und war damit das genaueste Modell für diesen Datensatz. Diese Ergebnisse zeigen, dass optimierte Ensemble-Modelle fast 70 % der Varianz in den Messungen der Oberflächenbeschädigung erklären können, und zwar mit hoher Vorhersagezuverlässigkeit.

Zur Interpretation der Modellergebnisse wurden SHAP-Werte (SHapley Additive Explanations) für das CatBoost-Modell berechnet. Diese ermöglichten die Identifizierung der Bedeutung von Merkmalen und der Richtung des Einflusses jeder Variable auf das Modellergebnis. Die Messgeschwindigkeit erwies sich als der einflussreichste Faktor, der stark mit positiven Veränderungen der Notenunterschiede verbunden ist. Dies bedeutet, dass höhere Geschwindigkeiten während der Erhebungen tendenziell zu höheren Degradationswerten führen, was möglicherweise auf eine geringere Auflösung oder Rauschempfindlichkeit bei der Datenerfassung zurückzuführen ist. Temperaturunterschiede waren die zweitwichtigste Variable, wobei sich sowohl extreme Anstiege als auch Abfälle auf die Messergebnisse auswirkten. Dies zeigt, wie empfindlich die Bewertung des Oberflächenzustands auf klimatische Schwankungen reagiert. Organisatorische Veränderungen standen an dritter Stelle, wobei es deutliche Hinweise darauf gab, dass ein Wechsel der Vermessungsorganisation häufig zu grösseren Messunterschieden führte - wahrscheinlich aufgrund von Unterschieden bei der Kalibrierung der Geräte oder bei den Bewertungsprotokollen. Der Vergleichszeitraum zeigte einen nicht-linearen Effekt. Sowohl kurze als auch lange Intervalle zwischen den Messungen führten tendenziell zu geringeren Notenunterschieden, was möglicherweise darauf zurückzuführen ist, dass sich die Bedingungen im Laufe der Zeit stabilisieren oder dass in sehr kurzen Zeitfenstern nur begrenzte Veränderungen zu beobachten sind. Unterschiede in der Windgeschwindigkeit und der Luftfeuchtigkeit hatten kleinere, differenziertere Auswirkungen, während die Anwesenheit von Interventionen erwartungsgemäss mit geringeren Notenunterschieden verbunden war. Aufgrund der begrenzten Anzahl von Interventionsfällen im Datensatz bleiben die Schlussfolgerungen hinsichtlich ihrer genauen Auswirkungen jedoch vorläufig und bedürfen weiterer Untersuchungen mit einem ausgewogeneren Datensatz.

Zusammenfassend lässt sich für den Indikator I0 sagen, dass die Daten aus den verschiedenen Bewertungsmethoden zeigen, dass die Messgeschwindigkeit der wichtigste Faktor ist, der die Messdiskrepanz zwischen den analysierten Variablen beeinflusst, gefolgt von der Temperatur und den Unterschieden zwischen den Messunternehmen. Das Modell erklärt etwa 70 % der beobachteten Variabilität, während 30 % unerklärt bleiben.

Für die Längsschnittgleichmässigkeit zielt die Analyse darauf ab, zu verstehen, wie externe Faktoren - wie Interventionen, Umweltbedingungen und organisatorische Unterschiede - die Variabilität der I2-Messungen beeinflussen, die durch die Höhendifferenz definiert sind: die Veränderung des Indikatorwerts zwischen zwei Erhebungsjahren auf demselben 100-Meter-Strassenabschnitt. Um diese Variabilität zu modellieren, wurden sechs Eingangsvariablen ausgewählt. Die abhängige Variable ist der Höhenunterschied zwischen zwei Erhebungen. Zu den unabhängigen Variablen gehören: Intervention, eine binäre Variable, die angibt, ob zwischen den Erhebungen Instandhaltungs- oder Sanierungsmassnahmen durchgeführt wurden, Länge des Erhebungsjahres, das Zeitintervall zwischen den beiden Erhebungen, Projektdifferenz, die angibt, ob derselbe oder ein anderer Auftragnehmer die Erhebungen durchgeführt hat, Temperaturdifferenz, Taupunktdifferenz und relative Luftfeuchtigkeitsdifferenz zwischen den beiden Messperioden. Nach Bereinigung und Filterung des Datensatzes, um sicherzustellen, dass Wetter- und Auftragnehmerdaten verfügbar waren, blieben insgesamt 95 439 Zeilen übrig.

Eine Visualisierung der Eingangsvariablen ergab mehrere Trends. An den meisten Strassenabschnitten wurde zwischen den Erhebungen nicht eingegriffen, und der zeitliche Abstand betrug im Allgemeinen vier bis fünf Jahre. Wichtig ist, dass die

Temperatur- und Feuchtigkeitsunterschiede zwischen den Erhebungen oft beträchtlich waren, was darauf hinweist, dass Wetterschwankungen eine Rolle bei Messungsinkonsistenzen spielen könnten. Was den Wechsel des Auftragnehmers betrifft, so gab es eine leichte Mehrheit von Strassenabschnitten, die in beiden Jahren von der gleichen Organisation untersucht wurden.

Die Korrelationsanalyse ergab im Allgemeinen nur geringe Zusammenhänge zwischen den unabhängigen Variablen und dem Höhenunterschied. Eine bemerkenswerte Ausnahme bildete jedoch die Interventionsvariable, die durchweg die höchste (wenn auch bescheidene) negative Korrelation mit dem Höhenunterschied aufwies, im Durchschnitt etwa -0,3. Dies deutet darauf hin, dass die Abschnitte, in denen Massnahmen durchgeführt wurden, bei der anschliessenden Erhebung tendenziell geringere I2-Werte aufwiesen, was mit der verbesserten Ebenheit der Fahrbahn nach der Instandhaltung übereinstimmt. Dieses Ergebnis gilt sowohl für die einzelnen Filialen als auch für den aggregierten Datensatz, was die praktische Relevanz von Massnahmen zur Erklärung von Verbesserungen der Ebenheit im Längsschnitt unterstreicht. Trotz der Identifizierung dieses Musters blieb die Vorhersagekraft der maschinellen Lernmodelle insgesamt begrenzt. Es wurden mehrere Modelle getestet, darunter CatBoost, LightGBM, XGBoost, Random Forest und Ensemble-Lernmethoden. Das beste Modell, CatBoost, erreichte ein R^2 von 0,35, was bedeutet, dass nur 35 % der Varianz der Notenunterschiede erklärt werden konnten. Die Anpassung der Hyperparameter, wie z. B. die Erhöhung der Baumtiefe und der Anzahl der Schätzer, führte nicht zu signifikant besseren Ergebnissen. Alternative Methoden, einschliesslich neuronaler Netze, konnten die baumbasierten Modelle ebenfalls nicht übertreffen. Diese mässige Modellleistung deutet darauf hin, dass ein erheblicher Teil der Variabilität der longitudinalen Ebenheit nicht allein mit den verfügbaren Variablen erfasst werden kann. Das hohe Mass an unerklärter Varianz könnte auf zufällige Messfehler, nicht gemessene Variablen (z. B. Verkehrsbelastung, Baumaterial oder Belagsbedingungen) oder inhärentes Rauschen bei der Ableitung des I2-Indikators zurückzuführen sein. Dies deckt sich mit früheren Beobachtungen aus der explorativen Analyse, die eine Inkonsistenz des Messverhaltens selbst in ähnlichen Abschnitten aufzeigte. Nichtsdestotrotz bieten die 35 % der erklärten Variabilität nützliche Einblicke.

Eine Analyse der Merkmalsbedeutung bestätigte, dass die Variable ‚Intervention' erwartungsgemäss über alle Modelle hinweg die einflussreichste war (obwohl die meisten Strassen keine Intervention erfahren hatten). Dies unterstreicht ihre zentrale Rolle bei der Entwicklung der Längsebenheit. Interventionen führen im Allgemeinen zu glatteren Oberflächen und niedrigeren I2-Werten, was bestätigt, dass der Indikator auf die Strasseninstandhaltung reagiert. Auch wetterbedingte Variablen spielten eine bemerkenswerte Rolle. Der Unterschied in der relativen Luftfeuchtigkeit und die Temperaturdifferenz erwiesen sich als wichtige Faktoren für die I2-Variabilität. Diese Ergebnisse deuten darauf hin, dass die Umweltbedingungen zum Zeitpunkt der Messung die Ergebnisse beeinflussen können, entweder direkt (z. B. Ausdehnung oder Zusammenziehen der Oberfläche) oder indirekt durch die Empfindlichkeit der Messgeräte. Daher sollte dem klimatischen Kontext, in dem die Erhebungen durchgeführt werden, grössere Aufmerksamkeit gewidmet werden, insbesondere beim Vergleich der Ergebnisse zwischen den Jahren. Im Gegensatz dazu waren die Unterschiede zwischen den Auftragnehmern (Projektunterschiede) von relativ geringerer Bedeutung, was darauf hindeutet, dass die Konsistenz der Messungen zwischen den Organisationen zumindest bei diesem Indikator höher sein könnte als ursprünglich erwartet. Dennoch bleibt die Einbeziehung dieses Indikators für die Qualitätskontrolle wichtig, insbesondere wenn mehrjährige Überwachungsprogramme in Betracht gezogen werden.

Zusammenfassend lässt sich sagen, dass die verfügbaren Modelle für den Indikator I2 nur 35 % der Variabilität erklären konnten, wobei die Eingriffe der einflussreichste Faktor waren, gefolgt von Wetterparametern wie Feuchtigkeit und Temperatur.

Für die transversale Ebenheit wird in der Studie analysiert, wie verschiedene externe und verfahrenstechnische Faktoren die Variation dieses Indikators bei wiederholten Strassenerhebungen beeinflussen. Auch hier ist die primäre Variable von Interesse die Höhendifferenz - definiert als die Veränderung der I3-Werte zwischen zwei aufeinanderfolgenden Erhebungen auf demselben 100-Meter-Strassenabschnitt. Sieben unabhängige Variablen wurden in dieser Analyse verwendet: Intervention, eine binäre Variable, die angibt, ob zwischen den beiden Erhebungen Instandhaltungsarbeiten stattgefunden haben, Unternehmen, eine binäre Variable, die angibt, ob derselbe oder ein anderer Auftragnehmer beide Messungen durchgeführt hat, Temperaturdifferenz, Windrichtungsdifferenz, Windgeschwindigkeitsdifferenz, Sichtbarkeitsdifferenz und relative Luftfeuchtigkeitsdifferenz, die alle die klimatischen Bedingungen zum Zeitpunkt der Messung darstellen.

Eine erste Untersuchung anhand von Variablenverteilungsdiagrammen lieferte nützliche Erkenntnisse. Die Neigungsdifferenz war ungefähr symmetrisch verteilt, was darauf hindeutet, dass die typische Oberflächenverschlechterung häufig durch Zustandsverbesserungen aufgrund von Eingriffen ausgeglichen wurde. Wie zu erwarten war, gab es auf den meisten Strassenabschnitten zwischen den Erhebungen keine Eingriffe, während etwa die Hälfte von verschiedenen Unternehmen vermessen wurde, ein Faktor, der aufgrund von Unterschieden in der Ausrüstung oder Methodik zu Inkonsistenzen bei den Messungen führen kann. Die Wettervariablen - insbesondere Temperatur und Luftfeuchtigkeit - wiesen Normalverteilungen auf, die um den Wert Null zentriert waren, obwohl extreme Unterschiede (z. B. Temperaturunterschiede von mehr als 10 °C oder 20 °C) häufig vorkamen. Dies verdeutlicht die potenziellen Auswirkungen der Umweltvariabilität auf die Umfrageergebnisse.

Zur Bewertung der Beziehungen zwischen den unabhängigen und abhängigen Variablen wurde eine Korrelations-Heatmap erstellt. Insgesamt waren die Korrelationswerte niedrig, was darauf hindeutet, dass die Unterschiede in der I3-Messung auf eine komplexe Interaktion von schwach miteinander verbundenen Variablen zurückzuführen sind. Von allen Faktoren wies der Wechsel des Unternehmens zwischen den Erhebungen die höchste Korrelation mit der Bewertungsdifferenz auf, was darauf hindeutet, dass organisatorische Verfahren und Technologien die Messung der transversalen Ebenheit erheblich beeinflussen können. Die klimatischen Variablen wiesen zwar eine gewisse Interkorrelation auf (z. B. zwischen Temperatur und Luftfeuchtigkeit), doch war diese nicht stark ausgeprägt, dass sie auf eine Redundanz hindeutet, sodass sie als Input für die Modellierung geeignet sind.

Zur Untersuchung der prädiktiven Beziehungen wurden verschiedene Modelle des maschinellen Lernens angewandt. Die beste Leistung wurde mit dem LightGBM-Modell erzielt, das ein R^2 von 0,4958 und einen RMSE von 0,441 erreichte, was bedeutet, dass etwa 50 % der Variabilität der Notenunterschiede mit den ausgewählten Variablen erklärt werden konnten. Dies ist ein mässig starkes Ergebnis und deutet darauf hin, dass Umwelt- und Verfahrensvariablen zwar die Messergebnisse beeinflussen, ein erheblicher Teil der Varianz jedoch unerklärt bleibt - wahrscheinlich aufgrund von nicht gemessenen Variablen wie Verkehrsbelastung, Fahrbahnstruktur, Materialeigenschaften oder Kalibrierung des Messsystems. Insbesondere die Erhöhung der Modellkomplexität (z. B. tiefere Bäume oder mehr Schätzer) während der Parameterabstimmung führte nicht zu einer Verbesserung der Leistung und in einigen Fällen sogar zu einer Verringerung der Modelleffektivität. Andere fortschrittliche Techniken, einschliesslich neuronaler Netze, wurden ebenfalls getestet, konnten aber die baumbasierten Modelle nicht wesentlich übertreffen. Diese Ergebnisse bekräftigen den Gedanken, dass Zufälligkeit und Rauschen sowie nicht erfasste Variablen die Fähigkeit von Vorhersagemodellen einschränken, das Verhalten von Indikator I3 vollständig zu erfassen.

Trotz dieser Einschränkungen lieferte die Analyse der Bedeutung der Variablen aussagekräftige Erkenntnisse. Unternehmensunterschiede erwiesen sich als der wichtigste Faktor, der die Unterschiede in der Bewertung beeinflusste. Dies deutet stark darauf hin, dass organisatorische Unterschiede, sei es aufgrund von Ausrüstung, Kalibrierung oder Betriebsverfahren, einen erheblichen Einfluss auf die Konsistenz der Bewertungen der transversalen Ebenheit haben. Dies bestätigt den Bedarf an standardisierten Protokollen zwischen den Vermessungsstellen, um die Vergleichbarkeit der Daten im Laufe der Zeit zu gewährleisten. Interventionen waren der zweitwichtigste Einflussfaktor. Dies entspricht den Erwartungen, da Instandhaltungsmassnahmen in der Regel zu erheblichen Verbesserungen der Oberfläche führen, die Spurrinnentiefe verringern und die Ebenheit verbessern. Im Datensatz waren die Eingriffe mit negativen Notenunterschieden verbunden, was die verbesserten Bedingungen bei der zweiten Erhebung widerspiegelt. Dies zeigt, wie empfindlich der Indikator I3 auf Instandhaltungsmassnahmen reagiert, und unterstreicht seinen Wert als verlässlicher Indikator für die Leistungsfähigkeit der Fahrbahn im Laufe der Zeit. Klimavariablen, insbesondere die Temperatur und die relative Luftfeuchtigkeit, leisteten ebenfalls einen bedeutenden Beitrag zum Modell, wenngleich ihre individuellen Auswirkungen schwächer waren als die der Intervention oder der Unternehmensunterschiede. Ihr Einfluss resultiert wahrscheinlich aus den physikalischen Effekten von Temperaturschwankungen auf Belagsmaterialien (z. B. Ausdehnung, Zusammenziehen, Erweichung) und daraus, wie Feuchtigkeit oder Sichtbedingungen die Datenerfassung während der Messläufe beeinflussen können. Der Einfluss jeder einzelnen Wettervariablen war jedoch weniger ausgeprägt, und ihr kombinierter Effekt erklärt wahrscheinlich einen grösseren Teil der Messvariabilität.

Die Analyse des Indikators I3 zeigt den starken Einfluss des Messunternehmens auf die Variabilität, mit einer erklärten Varianz von etwa 50 %. Dieses Ergebnis unterstreicht die Notwendigkeit einer stärkeren Standardisierung der Messmethoden. Darüber hinaus tragen Eingriffe und Wetterbedingungen zu den Schwankungen bei Indikator I3 bei.

Die Daten zur Oberflächenreibung wurden mit dem Messfahrzeug RK-SKM 80 mit einer Standardmessfrequenz von allen 100 Metern und einer Betriebsgeschwindigkeit von 80 km/h erhoben. Um trotz natürlicher Schwankungen der Messstartpunkte eine räumliche Konsistenz über mehrere Erhebungsjahre hinweg zu gewährleisten, wurden die Daten mit Hilfe eines Nearest-Neighbor-Ansatzes standardisiert, indem sie auf feste 100-Meter-Segmente von einer definierten Nullmarke aus ausgerichtet wurden. Die Analyse konzentrierte sich auf die mittlere Fahrspur und verwendete 1-Kilometer-Intervalle als grundlegende Analyseeinheit, indem zehn aufeinanderfolgende Messungen zusammengefasst wurden. Die Studie erstreckte sich auf 12 Hauptverkehrsstrassen (N1-N16, mit Ausnahme von vier) mit einer Länge von 474 Kilometern. Für jeden Strassenabschnitt wurden Erhebungsjahre ausgewählt, um wiederholte Messungen an gleichbleibenden Orten im Laufe der Zeit zu gewährleisten, was einen Datensatz von 1.322 gültigen Strassenproben ergab. Jede Stichprobe enthielt Daten aus zwei aufeinanderfolgenden Erhebungen, und wenn innerhalb des 1-km-Intervalls fehlende Werte festgestellt wurden, wurde die Stichprobe ausgeschlossen, um die Datenqualität zu erhalten.

Um das Ausmass der Variabilität bei den Reibungsmessungen zu messen, wurde eine neue Metrik eingeführt, die durchschnittliche Differenz (AD). Diese Metrik berechnet die durchschnittliche Veränderung der Reibungswerte zwischen zwei Erhebungsjahren innerhalb eines 1-Kilometer-Segments. Die Verteilung der AD folgte einem leicht negativ verzerrten Gaussschen Muster, was auf eine allgemeine Tendenz zur Verbesserung der Oberflächenreibungswerte - oft aufgrund von Eingriffen - im Laufe der Zeit hindeutet.

Die Faktoren, die AD möglicherweise beeinflussen, wurden in drei Kategorien eingeteilt: Messinformationen: einschliesslich der Frage, ob eine Intervention stattgefunden hat (binär), die Projektdifferenz oder der Wechsel der Inspektionsfirma zwischen den Jahren und die Länge des Testjahres, die die Zeit zwischen den Erhebungen angibt. Eine weitere Kategorie waren die Messbedingungen: z. B. die Messgeschwindigkeiten in jedem Jahr, ihre Differenz und die Standardabweichungen der Geschwindigkeit während der Prüfung. Schliesslich wurden auch die Wetterbedingungen herangezogen, einschliesslich der Temperaturdifferenz und der Windgeschwindigkeitsdifferenz zwischen den beiden Erhebungszeiträumen.

Die Korrelationsanalyse ergab geringe direkte Korrelationen zwischen AD und einzelnen Variablen, was darauf hindeutet, dass kein einzelner Faktor die Reibungsvariabilität überwiegend erklärt, was auf einen multifaktoriellen Einfluss hindeutet. Allerdings erwiesen sich das Vorhandensein von Interventionen und der Wechsel des Vertragspartners zwischen den Jahren als interessante Variablen. Zur Untersuchung der Daten wurden dann Modelle des maschinellen Lernens eingesetzt. Ein CatBoost-Regressionsmodell zeigte mit einem R^2 von 0,9186 und einem RMSE von 0,362 eine hervorragende Leistung und erklärte mehr als 92 % der Variabilität in der AD. Dieser hohe Grad an Vorhersagegenauigkeit zeigt, dass die erfassten Variablen die wichtigsten Einflüsse auf die Veränderungen der Oberflächenreibung effektiv erfassen. Versuche, das Modell durch Parameterabstimmung weiter zu optimieren, führten nicht zu einer signifikanten Leistungssteigerung.

Um über die Korrelation hinauszugehen und kausale Beziehungen zu untersuchen, wurde mit der DoWhy-Bibliothek ein strukturelles kausales Inferenzsystem angewendet. Auf der Grundlage des Domänenwissens und der Ergebnisse der SHAP (SHapley Additive exPlanations)-Analyse wurde ein Kausaldiagramm erstellt, das die angenommenen kausalen Beziehungen zwischen den Inputvariablen und AD abbildet. Dies ermöglichte die Berechnung der durchschnittlichen Behandlungseffekte (ATEs) für jede Variable. Die Kausalanalyse ergab, dass die Interventionen den signifikantesten positiven kausalen Effekt auf die Reibungsverbesserung hatten, mit einem ATE von 0,7193. Das bedeutet, dass die Massnahmen im Durchschnitt zu einem 71,93 % grösseren Unterschied zwischen den Reibungswerten im Vergleich zu Abschnitten ohne Massnahmen führten, was die Wirksamkeit der Instandhaltung deutlich unterstreicht. Im Gegensatz dazu wurden bei mehreren Faktoren, einschliesslich Projektdifferenz, Messgeschwindigkeit, Messgeschwindigkeitsdifferenz und Temperaturdifferenz, negative kausale Auswirkungen festgestellt, obwohl die Interpretation ihrer Richtungsabhängigkeit aufgrund ihrer kontinuierlichen Natur weniger eindeutig ist. Die SHAP-Analyse bestätigte ausserdem, dass Temperaturunterschied und Intervention die einflussreichsten Merkmale zur Erklärung von AD sind. Temperaturunterschiede können sich auf die Fahrbahnbeschaffenheit oder das Ansprechverhalten des Messsystems auswirken, während die Bedeutung von Eingriffen mit ihren physischen Auswirkungen auf die Strassenqualität übereinstimmt. Der SHAP-Bienenschwarmplot zeigte ausserdem, dass die Länge des Testjahres und die Standardabweichung der Messgeschwindigkeit ebenfalls die AD beeinflussen, insbesondere durch Ausreisserwerte. Schliesslich wurden die Ergebnisse der Kausalschlüsse mit drei Techniken auf ihre Robustheit hin überprüft: Einführung einer zufälligen gemeinsamen Ursache, Placebo-Behandlung und Validierung von Datenuntergruppen. Alle Tests bestätigten die Stabilität und Zuverlässigkeit der geschätzten kausalen Effekte, was das Vertrauen in die Ergebnisse stärkt.

Umgekehrt weist Indikator I4 den höchsten Grad an Erklärbarkeit auf, wobei bis zu 91 % der Variabilität erklärt werden können. Die Temperatur ist der wichtigste Einflussfaktor, dicht gefolgt von der Windgeschwindigkeit, den Eingriffen und der Messgeschwindigkeit.

Die Analyse der Unterschiede zwischen den regionalen Niederlassungen oder Filialen konzentrierte sich auf die Indikatoren I3 (transversale Ebenheit) und I0 (Oberflächenschäden) und ergab bemerkenswerte Unterschiede in Bezug auf Datenqualität, -quantität und Variabilitätsfaktoren. Einige Filialen verfügten über vollständigere und konsistentere Datensätze für alle Erhebungsjahre und Indikatoren, während andere, Datenlücken oder fehlende Erhebungsjahre aufwiesen. Diese Unstimmigkeiten werfen die Frage auf, ob die Daten nie erhoben oder einfach nicht hochgeladen wurden, was die Möglichkeit erschwert, zuverlässige landesweite Erkenntnisse zu gewinnen. Darüber hinaus unterscheiden sich die Hauptfaktoren, die die Variabilität der Indikatoren beeinflussen, je nach Filiale, was möglicherweise auf regionale Unterschiede in Bezug auf Geografie, Klima oder Verkehrsbelastung zurückzuführen ist.

Bei Indikator I0 wurden ähnliche Muster beobachtet. Während die Messgeschwindigkeit im Allgemeinen in allen Filialen ein dominierender Faktor ist, hatten in Filiale 2 (Thun) wetterbedingte Variablen einen grösseren Einfluss. Auch dies könnte auf regionale Umweltunterschiede oder verfahrenstechnische Unterschiede bei der Durchführung von Messungen zurückzuführen sein. Darüber hinaus variierte auch der Anteil der erklärten Variabilität zwischen den Filialen, was darauf hindeutet, dass einige Regionen mehr unerklärtes Rauschen in ihren Daten haben als andere. Diese Ergebnisse sind wichtig für die Harmonisierung der Datenerhebungsstandards und die Sicherstellung landesweit einheitlicher Bewertungspraktiken, insbesondere wenn das Ziel darin besteht, allgemeine Regeln zu entwickeln oder die Normen für den Belagszustand auf nationaler Ebene zu aktualisieren.

Auf der Grundlage dieser Ergebnisse werden mehrere Empfehlungen zur Verbesserung der Zuverlässigkeit und Verwendbarkeit der erhobenen Daten vorgeschlagen.

Erstens sollte die systematische Visualisierung von Daten verbessert werden, um eine schnelle Bewertung des Belagszustands und der Datenkonsistenz zu ermöglichen. Durch die Entwicklung eines automatisierten Visualisierungssystems, das mit neuen Daten aktualisiert wird, hätten Interessenvertreter wie Strassenverwalter und Messunternehmen leichteren Zugang zu wichtigen Informationen. Dieser Ansatz würde es ermöglichen, potenzielle Messprobleme sowie fehlende Daten, die langfristige Bewertungen beeinträchtigen könnten, schnell zu erkennen. Zweitens sollten Wetterparameter systematisch zur gleichen Zeit wie die Belagsindikatoren gemessen werden. Der Einsatz von tragbaren Wetterstationen, die heute sowohl erschwinglich als auch zuverlässig sind, würde wichtige Daten zu Temperatur, Luftfeuchtigkeit und anderen Umweltbedingungen liefern, die die Messergebnisse beeinflussen können. Ausserdem sollte die Planung der Erhebungen optimiert werden, um sicherzustellen, dass die Messungen unter ähnlichen Bedingungen wie bei früheren Erhebungen durchgeführt werden. Dies würde dazu beitragen, durch externe Umweltfaktoren verursachte Unstimmigkeiten zu verringern. Drittens sollten vertragliche Vereinbarungen sicherstellen, dass ein und dasselbe Unternehmen für mindestens zwei aufeinanderfolgende Erhebungsjahre verantwortlich ist, um die Auswirkungen verschiedener Unternehmen auf die Messvariabilität zu verringern. Dieser Ansatz würde Unstimmigkeiten, die sich aus methodischen Unterschieden zwischen Unternehmen ergeben, verringern. Ausserdem sollten alle Unternehmen, die Messungen durchführen, verpflichtet werden, ihre Geräte regelmässig zu kalibrieren und vergleichende Bewertungen vorzunehmen, um die Einheitlichkeit der Datenerhebungsverfahren zu gewährleisten. Viertens sollte der Prozess der Erfassung von Interventionen systematischer und umfassender sein. Die derzeitige Uneinheitlichkeit der Interventionsdaten erschwert die Bewertung ihrer tatsächlichen Auswirkungen auf den Belagszustand. Sowohl grössere als auch kleinere Eingriffe sollten detailliert dokumentiert werden, einschliesslich Informationen über die Art des Eingriffs, den Umfang der Reparaturen und die verwendeten Materialien. Dies würde die Möglichkeit verbessern, die Massnahmen mit den nachfolgenden Veränderungen des Belagszustands zu korrelieren. Fünftens sollte der Umfang der Datenanalyse erweitert werden, um zusätzliche Variablen einzubeziehen, die zur Erklärung der Messungsvariabilität beitragen könnten. Insbesondere die Einbeziehung detaillierter Verkehrsdaten - mit einer Unterscheidung zwischen schweren Lastkraftwagen und Personenkraftwagen - würde ein genaueres Bild der Belagsbelastung und der Verschlechterungsmuster liefern. Darüber hinaus könnte die Einbeziehung von Materialeigenschaften sowie topografischen und bodenkundlichen Parametern in die Analyse weitere Erkenntnisse darüber liefern, wie geografische und geologische Bedingungen die Verschlechterung von Strassenbelägen in Abhängigkeit von der Belagsart beeinflussen. Schliesslich wird eine Investition in die Weiterentwicklung der TRA-Trassee-Anwendung dringend empfohlen. Eine Erweiterung der Funktionalität zur Integration aller Belagsdaten, einschliesslich Visualisierungstools und analytischer Fähigkeiten, würde die Benutzerfreundlichkeit deutlich verbessern. Eine bessere Zugänglichkeit des Systems für Manager, Messunternehmen, Forscher und politische Entscheidungsträger würde eine bessere Entscheidungsfindung ermöglichen und sicherstellen, dass wichtige Daten effektiv genutzt werden.

Zusammenfassend lässt sich sagen, dass die Ergebnisse dieser Studie erhebliche Unstimmigkeiten bei der Messung des Strassenzustands aufzeigen und die Notwendigkeit einer verbesserten Genauigkeit und Zuverlässigkeit der Daten unterstreichen. Die Umsetzung der vorgeschlagenen Empfehlungen - einschliesslich verbesserter Datenvisualisierung, systematischer Wettermessungen, standardisierter Unternehmenspraktiken, detaillierter Eingriffsdokumentation, erweiterter Datenanalyse und der Entwicklung von TRA-Trassee - wird zu einer effektiveren Überwachung des Strassenzustands und einer langfristigen Infrastrukturplanung beitragen. Diese Verbesserungen werden letztlich die Nachhaltigkeit und Effizienz des Schweizer Nationalstrassennetzes sicherstellen.

Résumé

La Suisse dispose d'un réseau routier dense, composé de routes nationales, cantonales et communales/privées. Les routes nationales, bien qu'elles ne représentent que 3 % du réseau, assurent environ 40 % du trafic et 62,9 % du transport de marchandises. La gestion et l'entretien des routes sont répartis entre l'Office fédéral des routes (OFROU), les autorités cantonales et les municipalités.

Les routes se détériorent au fil du temps en raison du vieillissement, des conditions météorologiques et des facteurs d'utilisation tels que le sel en hiver ou les changements de terrain. Pour maintenir un niveau de service optimal, l'état des routes doit être évalué périodiquement. Les gestionnaires des routes suisses suivent des directives strictes pour garantir la précision, la compréhensibilité, la répétabilité et la reproductibilité des évaluations des chaussées. La précision consiste à minimiser les erreurs de mesure, la compréhensibilité assure une interprétation claire des données et la répétabilité/reproductibilité analyse la cohérence des mesures dans des conditions différentes.

Les systèmes de gestion des chaussées (SGC) sont des outils stratégiques utilisés par les autorités routières pour planifier, hiérarchiser et optimiser les interventions d'entretien et de réhabilitation sur les réseaux routiers. Leur objectif est de maintenir les chaussées aux niveaux de service souhaités tout en gérant les coûts à long terme. L'efficacité du PMS dépend fortement de la précision, de l'exhaustivité et de la cohérence des données sur l'état des chaussées. Les facteurs influents sont notamment la qualité du référencement, la cohérence des données historiques, la couverture spatiale et temporelle complète et l'alignement sur les pratiques de gestion des infrastructures en constante évolution. Les erreurs d'acquisition ou de traitement des données peuvent entraîner des recommandations de traitement erronées et une mauvaise affectation des budgets d'entretien.

La collecte de données au niveau du réseau privilégie généralement les méthodes à grande vitesse, telles que les profilomètres et les profileurs inertiels, en raison de leur rapidité, de la réduction des besoins en main-d'œuvre et de l'interruption minimale du trafic. Cependant, ces méthodes peuvent introduire une variabilité des données due aux différences d'étalonnage des équipements, au positionnement latéral, aux conditions environnementales et aux fluctuations saisonnières. Les opérateurs et les fournisseurs de services peuvent appliquer des méthodologies de mesure différentes, ce qui a un impact supplémentaire sur la cohérence. Par conséquent, des études rigoureuses de validation et de répétabilité des équipements sont essentielles pour garantir la qualité des données et permettre des comparaisons significatives dans le temps et entre les régions.

Des recherches récentes ont exploré l'application de techniques d'apprentissage automatique pour modéliser la détérioration des chaussées et prévoir les performances. Des cadres avancés intègrent désormais des méthodes telles que BorutaShap pour la sélection des caractéristiques, les réseaux neuronaux bayésiens (BNN) pour la modélisation probabiliste et les valeurs SHAP pour l'interprétation des modèles. Ces

approches améliorent la capacité à saisir les interactions complexes entre la conception de la chaussée, le trafic, l'environnement et l'historique de l'entretien. Toutefois, les études soulignent que la plus grande source d'incertitude dans ces modèles est la mauvaise qualité des données, ce qui met en évidence la nécessité d'améliorer la gouvernance des données et les procédures de validation.

Les inspections visuelles restent une méthode d'évaluation précieuse, en particulier dans les environnements qui n'ont pas accès à des systèmes automatisés. Leur fiabilité varie en fonction de l'expérience et de la formation de l'évaluateur, ainsi que de l'état de la chaussée évaluée. Les inspections sont généralement plus précises pour les chaussées en bon état, tandis que l'interprétation subjective introduit plus d'erreurs à la limite entre passable et mauvais. Il a été démontré que les programmes de formation et les exercices d'étalonnage améliorent considérablement la précision et la cohérence. Des simulations de Monte Carlo et des matrices de transition des erreurs ont été utilisées pour quantifier l'impact des erreurs d'inspection visuelle sur les indices d'état, en particulier à des niveaux modérés et élevés de gravité des dégradations.

Les mesures de rugosité de surface, souvent quantifiées par l'indice de rugosité international (IRI), sont largement utilisées dans les évaluations de réseaux. Toutefois, les valeurs de l'IRI peuvent varier considérablement en fonction du type de profileur, de la position de la voie, du volume de trafic et du type de chaussée. Des études montrent que des essais uniques peuvent ne pas refléter de manière adéquate la variabilité de l'état, en particulier sur les routes nationales. Les analyses comparatives des données des profileurs par rapport à l'équipement de référence soulignent la nécessité d'un étalonnage périodique, de répétitions multiples et d'une évaluation statistique pour garantir une interprétation pertinente. Les influences géographiques et saisonnières contribuent également à la variation des mesures.

L'adhérence, un indicateur essentiel de la sécurité routière, est influencée par la texture de la surface, les propriétés des agrégats, les conditions météorologiques et la charge du trafic. Des techniques telles que le GripTester et les testeurs de frottement dynamique (DFT) sont utilisées parallèlement aux systèmes de balayage laser pour mesurer la texture et le frottement de la surface. Les variations saisonnières sont un facteur important ; par exemple, le polissage estival réduit l'adhérence. Des études à long terme ont permis d'établir un lien entre les tendances en matière de frottement et l'évolution de la profondeur de la texture, des précipitations et de la circulation des véhicules lourds. Parmi les innovations récentes, on peut citer le développement de modèles statistiques permettant d'estimer les niveaux de frottement en saison sèche à partir des données de la saison humide.

Les méthodes statistiques jouent un rôle central dans la gestion de la variabilité et des erreurs dans les données relatives aux chaussées. Des outils tels que l'analyse des systèmes de mesure (MSA), la répétabilité et la reproductibilité des mesures (GR&R) et les modèles de Markov latents permettent aux chercheurs d'identifier, de quantifier et de réduire les sources d'erreur. La MSA a été largement adoptée dans d'autres industries, mais elle est de plus en plus appliquée dans la recherche sur les transports pour évaluer la fiabilité des systèmes de mesure. Les méthodologies hybrides combinant la conception expérimentale, la simulation et la modélisation probabiliste avancée permettent de mieux comprendre l'influence de la qualité de la construction, de la charge du trafic et de l'exposition à l'environnement sur la détérioration des chaussées.

Enfin, l'intégration de l'analyse moderne aux pratiques d'ingénierie traditionnelles améliore la précision et la fiabilité des évaluations de l'état des chaussées. Si les méthodes automatisées et basées sur l'apprentissage automatique sont très prometteuses, leur succès dépend d'une gestion rigoureuse des données, de pratiques de mesure cohérentes et de systèmes bien calibrés. À mesure que les réseaux routiers vieillissent et que les budgets se resserrent, la capacité à prendre des décisions fondées sur des données et sûres sur les priorités d'entretien restera essentielle. L'investissement continu dans la technologie de mesure, la formation du personnel et la normalisation des données entre les différentes juridictions permettront une gestion plus résiliente et plus rentable des infrastructures de chaussées dans le monde entier.

L'évaluation de l'état des routes en Suisse est régie par un ensemble de normes nationales, chacune détaillant des procédures et des exigences spécifiques pour l'évaluation de la performance des revêtements routiers. Bien que chaque partie du processus de surveillance soit couverte par une norme spécifique, il y a souvent des chevauchements et des cas où plusieurs normes s'appliquent à la même tâche. Ces normes sont particulièrement pertinentes après les phases de construction ou de réhabilitation, ainsi que pendant la durée de vie opérationnelle des réseaux routiers. La norme VSS 40 525, par exemple, définit les critères d'acceptation des revêtements routiers à l'issue des travaux, en se concentrant sur quatre éléments : les déformations locales, l'uni longitudinal, l'uni transversal et le frottement de la surface. Les déformations locales sont évaluées comme décrit dans la norme SN 640 516-7, avec des seuils d'acceptation spécifiques pour les différentes couches de la chaussée. L'uni longitudinal est mesuré sur la base du profil longitudinal de la route, avec deux méthodes autorisées par la norme VSS 40 517 : l'une basée sur la variation angulaire, et l'autre basée sur l'analyse de la bande d'ondes, qui examine les ondes de surface courtes, moyennes et longues. La seconde méthode est plus complète et mieux adaptée à la détection des irrégularités de surface à plusieurs échelles. L'uni transversal est évalué en mesurant la profondeur des ornières et la profondeur théorique de l'eau dans les ornières, conformément à la norme VSS 40 518. Si la profondeur de l'ornière n'est pas soumise à des seuils d'acceptation spécifiques, la profondeur de l'eau l'est, en particulier pour des raisons de sécurité sur les routes à grande vitesse. Par exemple, sur les routes où la vitesse est limitée à plus de 80 km/h, la profondeur d'eau théorique maximale autorisée est de 4,0 mm. Le frottement de surface, autre paramètre critique, est évalué à l'aide de méthodes dynamiques et statiques conformément à la norme VSS 40512. Les systèmes dynamiques comprennent le Skiddometer BV 8 et l'appareil SKM, tandis que la méthode du pendule sert d'alternative statique. Ces tests prennent en compte les conditions saisonnières et sont généralement effectués 3 à 6 mois après l'ouverture de la route à la circulation, en appliquant des seuils de vitesse spécifiques.

L'état de la chaussée peut être résumé à l'aide de quatre indices définis dans la norme VSS 40 925b : I0 (endommagement de la surface), I2 (uni longitudinal), I3 (uni transversal) et I4 (frottement de surface). Chaque indice est noté de 0 (très bon) à 5 (très mauvais), sur la base d'un ou de plusieurs paramètres. Le choix des paramètres à évaluer et leur fréquence dépendent de la fonction de la route, de la charge de trafic et de la méthodologie de gestion. Au niveau du réseau, ces indices sont utilisés conformément à la norme SN 640 900 et stockés dans des systèmes de gestion définis par les normes VSS 40 944 et VSS 40 904. Les évaluations au niveau du projet suivent les normes VSS 40 730 et VSS 40 925b. Si la plupart des indicateurs peuvent être évalués tout au long de l'année pour autant que la chaussée soit sèche, les conditions de température doivent être respectées lors de l'évaluation du frottement de surface et de la capacité de charge.

L'évaluation des dommages de surface diffère selon qu'il s'agit d'une surface bitumineuse ou d'une surface en béton. Pour les chaussées en asphalte, l'évaluation porte sur l'uni, la déformation, les dommages structurels et les réparations. Pour le béton, des facteurs tels que la perte de matériau, la dégradation des joints, les déplacements verticaux et les fissures sont pris en compte. Dans ce travail, nous nous sommes concentrés sur les surfaces bitumineuses qui correspondent aux autoroutes. Les méthodes traditionnelles reposent sur des inspecteurs qui marchent ou conduisent le long du segment, bien que de nombreuses entreprises utilisent désormais des images à haute résolution montées sur des véhicules, qui sont ensuite analysées par des professionnels qualifiés ou, de plus en plus, par des systèmes d'intelligence artificielle. Les dommages de surface sont caractérisés par leur étendue (échelle 0-3) et leur gravité (échelle 1-3), qui sont combinées à l'aide d'une matrice pour donner une note de 0 à 9. Dans les évaluations sommaires, les paramètres sont regroupés, notés, pondérés et combinés dans l'indice I0. La longueur standard des segments est de 50 mètres pour les routes à chaussée unique et de 100 mètres pour les routes à voie unique. La règle générale est d'indiquer le niveau de gravité le plus défavorable par groupe et une étendue représentative. Pour les évaluations détaillées au niveau du projet, les deux dimensions sont enregistrées pour chaque paramètre individuel.

L'uni longitudinal (I2) est évalué par plusieurs méthodes, la plus directe étant la mesure des variations d'angle (W) entre des sections adjacentes d'un mètre de la surface de la route. L'écart-type Sw de ces valeurs W est utilisé pour calculer l'indice, comme le montre un graphique dans la norme VSS 40517. Parmi les autres méthodes, citons l'indice international de rugosité (IRI), qui évalue le confort de conduite de l'usager sur la base d'un mouvement vertical simulé du système de suspension d'un véhicule ; la méthode de la bande d'ondes NBO, qui analyse la distribution de l'énergie à travers différentes fréquences d'ondes ; et le profil longitudinal pondéré (BLP), qui équilibre les contributions d'amplitude provenant de différentes longueurs d'onde. Ces méthodes permettent d'isoler les défauts périodiques ou isolés. Cependant, la norme ne fournit qu'une transformation directe de Sw en valeurs I2 et ne précise pas comment convertir les résultats d'autres techniques dans l'échelle I2.

La régularité transversale (I3) est basée sur les mesures de la profondeur des ornières et le potentiel d'accumulation d'eau dans ces ornières. Les ornières peuvent être mesurées manuellement à l'aide d'une règle ou d'une ficelle, et il convient d'indiquer la méthode utilisée. Les valeurs des passages de roues gauche et droite sont nécessaires, et plusieurs mesures doivent être prises sur le segment d'évaluation. Lors de l'évaluation des réseaux routiers, une seule ornière peut être évaluée. L'espacement entre les points du profil transversal ne doit pas dépasser 10 cm et la précision verticale des mesures doit être de 0,5 mm pour les systèmes dynamiques et de 1,0 mm pour les systèmes statiques. Un minimum de 10 mesures par segment de route est recommandé, et les résumés statistiques ainsi que les valeurs extrêmes doivent être documentés. Lorsque les deux ornières sont mesurées, la valeur la plus mauvaise est utilisée pour I3.

En pratique, les mesures I2 et I3 sont effectuées à l'aide de profilomètres laser avancés intégrés dans les véhicules d'enquête. Ces technologies garantissent une précision et une vitesse de mesure élevées, essentielles pour les évaluations du réseau routier à grande échelle. Les systèmes sont capables de scanner avec précision la surface de la chaussée et de calculer ensuite les paramètres normalisés pour des sections désignées (par exemple les bandes de roulement), conformément aux spécifications en vigueur. Malgré l'existence de méthodes manuelles traditionnelles, les systèmes automatisés offrent une efficacité, une fiabilité et une reproductibilité accrues, et sont désormais la norme dans la plupart des opérations de surveillance.

Dans l'ensemble, l'approche suisse de l'évaluation de l'état des routes est très structurée et normalisée, ce qui garantit la cohérence entre les projets et les régions. L'intégration des normes avec des systèmes de mesure avancés et l'augmentation des flux de travail numériques soutiennent la prise de décision basée sur les données dans la planification de l'entretien et de la réhabilitation des routes.

Le réseau routier national de la Suisse est géré par deux divisions : Ouest et Est. La division Ouest supervise les antennes régionales d'Estavayer-le-Lac et de Thoune, tandis que la division Est gère Zofingue, Winterthour et Bellinzone. Les antennes régionales sont appelées « filialen » en Suisse. Chaque Filiale est responsable de la mesure des indicateurs de l'état des routes dans la zone qui lui est attribuée. Les données collectées sont stockées dans l'application TRA-Trassee, qui soutient la gestion et la planification des infrastructures routières. Les données sont vérifiées avant d'être intégrées dans le système, mais les ressources limitées affectent le contrôle de la qualité. Les principales entreprises impliquées dans la mesure de l'état des routes sont Schniering GmbH (qui fait maintenant partie de TÜV Rheinland) et Infralab SA, qui utilisent des techniques avancées de mesure à grande vitesse. Les ajustements de données sont effectués par Geologix AG avant le stockage final. Dans cette étude, nous concentrons notre analyse sur deux méthodes pour explorer les ensembles de données complexes et volumineux.

D'une part, les visualisations jouent un rôle clé dans l'analyse des données, en aidant à identifier les incohérences et les tendances dans le temps d'une manière simple et intuitive. Diverses méthodes statistiques, telles que les histogrammes et les diagrammes de dispersion, permettent d'évaluer la distribution des données, de détecter les valeurs aberrantes et d'assurer la cohérence des indicateurs de l'état des routes. D'autre part, une approche plus complexe comme l'analyse factorielle est également un outil qui peut être considéré comme un élément essentiel de l'évaluation de l'état des chaussées. Dans cette étude, nous intégrons les données météorologiques de MétéoSuisse, en associant les stations météorologiques pertinentes à chaque Filiale. La combinaison des données météorologiques et des facteurs de mesure nous a permis d'explorer plus en détail l'importance de l'explication de la variabilité dans les ensembles de données. Pour améliorer la précision des prédictions, des modèles avancés d'apprentissage automatique - notamment AdaBoost, CatBoost, LightGBM, Random Forest et XGBoost -

sont optimisés à l'aide de l'optimisation bayésienne (BO). Ces méthodes d'apprentissage d'ensemble permettent de saisir les relations complexes dans les données sur l'état des chaussées. La performance du modèle est évaluée à l'aide des mesures R^2 et RMSE afin de déterminer les prédictions les plus précises. Les techniques SHAP (Shapley Additive Explanations) et permutations permettent d'interpréter les modèles prédictifs et de comprendre l'influence des différentes variables sur la variabilité de l'état des routes. La quantification de l'importance de caractéristiques telles que le climat, l'historique des interventions et les méthodes de mesure garantit une approche de la planification et de l'entretien des infrastructures routières fondée sur les données.

En ce qui concerne les résultats du processus de visualisation, nous avons pu observer, par exemple pour la direction régionale d'Estavayer-le-lac, que sur l'autoroute N1, pour les tronçons n'ayant fait l'objet d'aucune intervention majeure après 2000, la plupart des indicateurs ont montré une tendance générale à la détérioration, comme on pouvait s'y attendre. Toutefois, certaines améliorations inattendues de certains indicateurs (notamment I2 et I4) ont suscité des inquiétudes quant à la fiabilité des mesures. Les périodes entre 2009 et 2013, et 2013 et 2017, ont montré des anomalies où différents indicateurs présentaient des tendances contradictoires - par exemple, I2 et I4 sont restés stables ou se sont améliorés, tandis que I0 et I3 ont indiqué une dégradation. Cela met en évidence un manque de synchronisation entre les indicateurs, qui devraient idéalement suivre une trajectoire de vieillissement similaire en l'absence d'interventions.

Dans les segments ayant fait l'objet d'interventions majeures, les améliorations étaient parfois visibles, mais pas toujours cohérentes pour tous les indicateurs. Par exemple, dans le tronçon ayant fait l'objet d'une intervention entre 2009 et 2013, I0 reflétait clairement l'intervention, tandis que I2 montrait une absence de réponse inattendue et que I3/I4 révélaient des tendances mixtes ou atténuées. De même, sur le tronçon ayant fait l'objet d'une intervention après 2013, seuls I2 et I4 ont reflété l'amélioration attendue, tandis que I0 est resté stable et n'a pas enregistré l'impact de l'intervention.

Sur l'autoroute N9, les tronçons sans interventions signalées ont également montré des tendances contradictoires. Les améliorations d'indicateurs tels que I0 et I4 entre 2013 et 2017 - sans travaux enregistrés - suggèrent qu'une maintenance non documentée ou des erreurs dans la base de données des dates de construction ont pu se produire. Dans les segments où les interventions ont été confirmées, les indicateurs tels que I0 et I4 ont eu tendance à se comporter comme prévu, montrant de nettes améliorations suivies d'une détérioration progressive. Cependant, les indicateurs I2 et I3 ont de nouveau montré des incohérences, avec certaines améliorations survenues avant les interventions et, dans certains cas, des années plus tard.

En général, les résultats de l'exploration visuelle indiquent que les variations observées dans les différents indicateurs au fil du temps ne correspondent pas aux modèles attendus. Plutôt qu'une détérioration progressive et prévisible de l'état des routes d'une enquête à l'autre, les changements observés semblent presque aléatoires. Ce problème est plus prononcé pour les indicateurs I0, I2 et I3, tandis que l'indicateur I4 présente une stabilité légèrement supérieure dans le temps. De telles incohérences soulèvent des inquiétudes quant à la fiabilité et à l'exactitude des données collectées. La déconnexion apparente entre les différents indicateurs est une observation particulièrement intrigante. Bien qu'ils mesurent des aspects distincts de l'état des chaussées, tous les indicateurs devraient théoriquement s'aligner en reflétant une tendance continue à la détérioration ou une amélioration marquée à la suite d'une intervention. Cependant, l'analyse suggère que ce n'est pas le cas, ce qui met encore plus en évidence les failles possibles dans le processus de mesure. En outre, la variabilité des mesures augmente considérablement entre les enquêtes successives, même lorsque la même entreprise est responsable de la collecte des données. Cette incohérence suggère que des facteurs externes, des techniques de mesure ou des problèmes d'étalonnage de l'équipement pourraient affecter les résultats.

Une autre constatation inattendue est la disparité de l'état des routes à l'intérieur d'un même tronçon de chaussée. Bien qu'il soit raisonnable de supposer que certains tronçons d'une route se détériorent à des rythmes différents en raison de la variation des charges de trafic ou des influences environnementales, on s'attend à un certain niveau de cohérence. Le fait que certains tronçons semblent être en bien meilleur état que d'autres soulève d'autres inquiétudes quant à la précision du processus de mesure. Lorsque des interventions ont lieu entre deux périodes d'enquête, leur impact est parfois évident pour certains indicateurs, mais absent pour d'autres. Ce manque d'uniformité est difficile à expliquer et indique des déficiences potentielles dans la collecte et le traitement des données. L'implication de plusieurs entreprises dans le processus de mesure semble être un autre facteur contribuant aux incohérences des données, l'ampleur du problème dépendant de l'indicateur spécifique analysé. L'analyse de visualisation a également révélé des écarts considérables dans la quantité et la cohérence des données disponibles dans les différentes filiales. Ces données manquantes empêchent de suivre l'état des chaussées dans le temps, ce qui limite l'efficacité des stratégies de planification et d'entretien des infrastructures à long terme.

L'analyse de l'importance des variables réalisée pour chaque indicateur a permis de mieux comprendre les facteurs clés qui influencent la variabilité des données.

Dans le cas des dommages de surface, l'analyse est basée sur des mesures appariées effectuées au cours d'années d'enquête consécutives sur des segments de route de 100 mètres. La principale variable d'intérêt - la différence de niveau - reflète le changement des valeurs I0 entre deux enquêtes et sert de variable dépendante dans le processus de modélisation. Diverses variables indépendantes ont été prises en compte pour comprendre cette variabilité. Celles-ci comprennent les vitesses de mesure au cours des deux années d'enquête, la différence entre elles, l'existence ou non d'une intervention entre les enquêtes, le temps écoulé entre les mesures, le changement d'organisation de l'enquête et plusieurs facteurs climatiques, tels que les différences de température, de vitesse du vent et d'humidité relative entre les deux enquêtes.

L'analyse descriptive a montré que la différence de niveau était généralement distribuée de manière symétrique, avec une légère asymétrie positive, ce qui suggère une tendance naturelle à la détérioration progressive de la surface. Les vitesses mesurées présentaient des distributions bimodales et trimodales, indiquant des paramètres opérationnels variables au cours des enquêtes. Notamment, environ 48 % des tronçons routiers ont été mesurés par différentes organisations au cours d'années consécutives, et seulement 13 % ont fait l'objet d'interventions documentées. Les variables liées au climat, bien que centrées autour de zéro, présentaient des extrêmes importants dans certains cas, en particulier des différences de température allant jusqu'à ±20°C.

Une matrice de corrélation a révélé des informations importantes : il existe une corrélation positive modérée (0,59) entre l'EM2 et la différence de niveau, ce qui suggère que des vitesses de mesure plus rapides au cours de la deuxième année d'enquête tendent à se traduire par une dégradation de surface plus importante. Inversement, la présence d'une intervention a montré une corrélation négative (-0,32) avec l'écart de niveau, ce qui correspond à l'attente selon laquelle les interventions conduisent généralement à une amélioration des conditions. Les fortes corrélations entre les variables de mesure de la vitesse ont confirmé la cohérence interne des variables liées à la vitesse. Les changements organisationnels ont montré des corrélations modérées avec les différences de vitesse et de niveau, ce qui indique d'éventuelles différences de procédure ou de calibrage entre les entités chargées des enquêtes. Néanmoins, depuis environ vingt ans, les méthodes de mesure ont connu d'importantes évolutions. À l'origine, les inspections visuelles étaient effectuées à bord d'un véhicule circulant à faible vitesse (ou parfois à pied), l'opérateur devant identifier les types de dégradations. Quelques années plus tard, l'opérateur évaluait à distance l'état de la surface routière depuis un bureau, en analysant des images haute résolution capturées sur le terrain. Aujourd'hui, la plupart des entreprises évaluent les dommages de surface à l'aide de systèmes d'imagerie avancés (notamment des lasers 3D) capables de détecter et de classer automatiquement les dégradations selon la norme VSS 40 925. Ces images sont enregistrées en continu à des intervalles de mesure fixes. Par conséquent, bien qu'une vitesse de mesure variable puisse sembler sans incidence pratique sur la qualité des données, elle peut en réalité dissimuler l'influence des évolutions méthodologiques des techniques d'inspection visuelle.

Pour modéliser et prédire la différence de niveau, cinq algorithmes d'apprentissage d'ensemble ont été utilisés : CatBoost, XGBoost, Random Forest, LightGBM et AdaBoost. Les premiers résultats ont montré que CatBoost et XGBoost étaient les modèles les plus performants, chacun atteignant un R^2 de 0,689. Après avoir appliqué l'optimisation bayésienne (BO) pour affiner les hyperparamètres, CatBoost s'est encore amélioré pour atteindre R^2 = 0,693 avec le RMSE le plus bas de 0,422, ce qui en fait le modèle le plus précis pour cet ensemble de données. Ces résultats indiquent que les modèles d'ensemble optimisés peuvent expliquer près de 70 % de la variance des mesures de dommages de surface, avec une grande fiabilité prédictive.

Pour interpréter les résultats du modèle, les valeurs SHAP (SHapley Additive Explanations) ont été calculées pour le modèle CatBoost. Elles ont permis d'identifier l'importance des caractéristiques et la direction de l'impact de chaque variable sur les résultats du modèle. La vitesse de mesure est apparue comme le facteur le plus influent, fortement lié à des changements positifs dans les différences de niveau. Cela implique que des vitesses plus élevées pendant les enquêtes tendent à produire des scores de dégradation plus élevés, potentiellement en raison d'une résolution réduite ou d'une sensibilité au bruit lors de l'acquisition des données. Les différences de température ont été la deuxième variable la plus influente, les résultats des mesures étant affectés à la fois par des augmentations et des diminutions extrêmes. Cela démontre la sensibilité des évaluations de l'état de surface aux variations climatiques. Les changements organisationnels arrivent en troisième position, avec des preuves évidentes qu'un changement d'organisme d'enquête entraîne souvent des différences de mesure plus importantes, probablement dues à des différences dans l'étalonnage de l'équipement ou de la subjectivité des protocoles d'évaluation. La période de comparaison a eu un effet non linéaire. Les intervalles courts et longs entre les mesures ont tendance à entraîner des différences de niveau plus faibles, probablement en raison de la stabilisation des conditions dans le temps ou de changements observables limités dans des fenêtres très courtes. Les différences de vitesse du vent et d'humidité ont eu des effets plus petits et plus nuancés, tandis que la présence d'une intervention était liée à des différences de notes plus faibles, comme on s'y attendait. Cependant, en raison du nombre limité de cas d'intervention dans l'ensemble de données, les conclusions concernant leur effet précis restent provisoires et méritent d'être approfondies avec un ensemble de données plus équilibré.

En résumé, pour l'indicateur I0, les données analysées à l'aide des différentes méthodes d'évaluation montrent que le facteur le plus important affectant dans les écarts de mesure entre les variables étudiées est de mesure est la vitesse de mesure, suivie de la température et des différences entre les sociétés de mesure. Ce modèle explique environ 70 % de la variabilité observée, laissant 30 % inexpliqués.

Pour la régularité longitudinale, l'analyse vise à comprendre comment les facteurs externes - tels que les interventions, les conditions environnementales et les différences organisationnelles - influencent la variabilité des mesures de l'I2, définie par la différence de niveau : le changement de la valeur de l'indicateur entre deux années d'enquête sur le même segment de route de 100 mètres. Pour modéliser cette variabilité, six variables d'entrée ont été sélectionnées. La variable dépendante est la différence de niveau entre deux enquêtes. Les variables indépendantes sont les suivantes l'intervention, une variable binaire indiquant si des travaux d'entretien ou de réhabilitation ont été effectués entre les relevés, la durée de l'année d'essai, l'intervalle de temps entre les deux relevés, la différence de projet, indiquant si les relevés ont été effectués par le même entrepreneur ou par un entrepreneur différent, la différence de température, la différence de point de rosée et la différence d'humidité relative entre les deux périodes de mesure. Après avoir nettoyé et filtré l'ensemble des données pour s'assurer que les données météorologiques et les données sur les entrepreneurs étaient disponibles, il restait un total de 95 439 lignes.

Une visualisation des variables d'entrée a révélé plusieurs tendances. La plupart des tronçons routiers n'ont pas fait l'objet d'une intervention entre les enquêtes, et l'intervalle de temps était généralement de quatre à cinq ans. Il est important de noter que les différences de température et d'humidité entre les enquêtes étaient souvent importantes, ce qui montre que la variabilité météorologique peut jouer un rôle dans les incohérences de mesure. En ce qui concerne les changements d'entrepreneur, il y avait une légère majorité de segments de route étudiés par la même organisation au cours des deux années.

L'analyse des corrélations a montré des relations généralement faibles entre les variables indépendantes et la différence de niveau. Toutefois, une exception notable est la variable Intervention, qui a constamment affiché la corrélation négative la plus élevée (bien que modeste) avec la différence de niveau, avec une moyenne d'environ -0,3. Cela indique que les segments ayant fait l'objet d'interventions ont eu tendance à présenter des valeurs I2 réduites lors de l'enquête suivante, ce qui est cohérent avec l'amélioration de l'uni de la chaussée après l'entretien. Ce résultat s'est vérifié à la fois au niveau de chaque filialen et dans l'ensemble des données agrégées, ce qui renforce la pertinence pratique des interventions pour expliquer les améliorations longitudinales de l'uni. Malgré l'identification de cette tendance, le pouvoir prédictif global des modèles d'apprentissage automatique est resté limité. Plusieurs modèles ont été testés, notamment CatBoost, LightGBM, XGBoost, Random Forest et des méthodes d'apprentissage d'ensemble. Le modèle le plus performant, CatBoost, a obtenu un R^2 de 0,35, ce qui signifie que seuls 35 % de la variance des différences de notes ont pu être expliqués. L'ajustement des hyperparamètres, comme l'augmentation de la profondeur des arbres et du nombre d'estimateurs, n'a pas donné de résultats significativement meilleurs. Les méthodes alternatives, y compris les réseaux neuronaux, n'ont pas non plus été plus performantes que les modèles à base d'arbres. Cette performance modérée des modèles suggère qu'une part importante de la variabilité de l'homogénéité longitudinale ne peut pas être capturée en utilisant uniquement les variables disponibles. Le degré élevé de variance inexpliquée pourrait provenir d'une erreur de mesure aléatoire, de variables non mesurées (par exemple, la charge de trafic, les matériaux de construction ou les conditions de la couche de chaussée), ou d'un bruit inhérent à la manière dont l'indicateur I2 est dérivé. Ceci est en accord avec les observations précédentes de l'analyse exploratoire, qui a mis en évidence des incohérences dans le comportement des mesures, même sur des segments similaires. Néanmoins, les 35 % de variabilité expliquée offrent toujours un aperçu utile.

Une analyse de l'importance des caractéristiques a confirmé que l'« intervention » était systématiquement la variable la plus influente dans tous les modèles, ce qui était attendu (bien que la majorité des routes n'aient pas fait l'objet d'une intervention). Cela souligne son rôle central dans l'évolution de l'uni longitudinal. Les interventions conduisent généralement à des surfaces plus lisses et à des scores I2 plus faibles, ce qui valide la réactivité de l'indicateur à l'entretien des routes. Les variables météorologiques ont également joué un rôle important. La différence d'humidité relative et la différence de température sont apparues comme des facteurs importants de la variabilité de l'I2. Ces résultats suggèrent que les conditions environnementales au moment de la mesure peuvent influencer les résultats, soit directement (par exemple, expansion ou contraction de la surface), soit indirectement par le biais de la sensibilité de l'équipement de mesure. Par conséquent, il convient d'accorder une plus grande attention au contexte climatique dans lequel les enquêtes sont menées, en particulier lors de la comparaison des résultats d'une année à l'autre. En revanche, les différences entre les entrepreneurs (différences entre les projets) ont une importance relativement moindre, ce qui indique que la cohérence des mesures entre les organisations est peut-être plus élevée que prévu, du moins pour cet indicateur. Néanmoins, son inclusion reste importante pour le contrôle de la qualité, en particulier lorsque l'on envisage des programmes de surveillance pluriannuels.

En résumé, pour l'indicateur I2, les modèles disponibles n'ont pu expliquer que 35% de la variabilité, les interventions étant le facteur le plus influent, suivi par les paramètres météorologiques tels que l'humidité et la température.

Dans le cas de l'uni transversal, l'étude analyse comment différents facteurs externes et procéduraux influencent la variation de cet indicateur à travers des enquêtes routières répétées. Là encore, la principale variable d'intérêt est la différence de niveau, définie comme la variation des scores I3 entre deux enquêtes consécutives sur le même segment de route de 100 mètres. Sept variables indépendantes ont été utilisées dans cette analyse : Intervention, une variable binaire indiquant si l'entretien a eu lieu entre les deux enquêtes, entreprise, une variable binaire indiquant si le même entrepreneur ou un entrepreneur différent a effectué les deux mesures, différence de température, différence de direction du vent, différence de vitesse du vent, différence de visibilité et différence d'humidité relative, toutes représentant les conditions climatiques au moment de la mesure.

L'exploration initiale à l'aide de diagrammes de distribution des variables a fourni des informations utiles. La différence de niveau était distribuée de manière à peu près symétrique, ce qui suggère que la détérioration typique de la surface était fréquemment compensée par des améliorations de l'état dues à des interventions. Comme prévu, la plupart des segments de route n'ont subi aucune intervention entre les relevés, tandis qu'environ la moitié ont été relevés par des entreprises différentes, un facteur qui peut introduire des incohérences dans les mesures en raison des différences d'équipement ou de méthodologie. Les variables météorologiques - en particulier la température et l'humidité - présentaient des distributions normales centrées sur zéro, bien que des différences extrêmes (par exemple, plus de 10°C ou 20°C de température) aient été courantes. Cela a mis en évidence l'impact potentiel de la variabilité environnementale sur les résultats de l'enquête.

Une carte thermique de corrélation a été créée pour évaluer les relations entre les variables indépendantes et dépendantes. Dans l'ensemble, les valeurs de corrélation étaient faibles, ce qui suggère que les différences de mesure de l'I3 résultent d'une interaction complexe de variables faiblement liées. Parmi tous les facteurs, le changement d'entreprise entre les enquêtes présentait la corrélation la plus élevée avec la différence de note, ce qui indique que les pratiques et technologies organisationnelles peuvent influencer de manière significative la mesure de l'homogénéité transversale. Bien que les variables climatiques présentent une certaine inter-corrélation (par exemple, entre la température et l'humidité), elles ne sont pas assez élevées pour suggérer une redondance, ce qui en fait des données valables pour la modélisation.

Pour explorer les relations prédictives, divers modèles d'apprentissage automatique ont été appliqués. La meilleure performance a été obtenue avec le modèle LightGBM, qui a obtenu un R^2 de 0,4958 et une RMSE de 0,441, ce qui indique qu'environ 50 % de la variabilité des différences de notes peut être expliquée à l'aide des variables sélectionnées. Ce résultat est modérément fort et suggère que si les variables environnementales et procédurales influencent effectivement les résultats des mesures, une part importante de la variance reste inexpliquée, probablement en raison de variables non mesurées telles que la charge de trafic, la structure de la chaussée, les propriétés des matériaux ou l'étalonnage du système de mesure. Notamment, l'augmentation de la complexité du modèle (par exemple, arbres plus profonds ou plus d'estimateurs) pendant le réglage des paramètres n'a pas amélioré la performance et, dans certains cas, a réduit l'efficacité du modèle. D'autres techniques avancées, y compris les réseaux neuronaux, ont également été testées mais n'ont pas réussi à surpasser de manière significative les modèles à base d'arbres. Ces résultats renforcent l'idée que le hasard et le bruit, ainsi que les variables non enregistrées, limitent la capacité des modèles prédictifs à saisir pleinement le comportement de l'indicateur I3.

Malgré ces limites, l'analyse de l'importance des variables a fourni des informations significatives. La différence entre les entreprises est apparue comme le facteur le plus important influençant les différences de notes. Cela suggère fortement que les variations organisationnelles, qu'elles soient dues à l'équipement, à l'étalonnage ou aux procédures opérationnelles, ont un effet substantiel sur la cohérence des évaluations de la planéité transversale. Cela confirme la nécessité de protocoles standardisés entre les entités d'enquête pour garantir la comparabilité des données dans le temps. L'intervention est le deuxième facteur le plus influent. Cela correspond aux attentes, car les activités d'entretien conduisent généralement à des améliorations significatives de la surface, réduisant la profondeur des ornières et améliorant la planéité. Dans l'ensemble des données, les interventions ont été liées à des différences de niveau négatives, reflétant l'amélioration des conditions lors de la deuxième enquête. Cela met en évidence la sensibilité de l'indicateur I3 aux efforts d'entretien des routes et renforce sa valeur en tant que marqueur fiable de la performance des chaussées au fil du temps.

Les variables climatiques, en particulier la température et l'humidité relative, ont également contribué de manière significative au modèle, bien que leurs effets individuels aient été plus faibles que ceux de l'intervention ou de la différence entre les entreprises. Leur influence résulte probablement des effets physiques des fluctuations de température sur les matériaux de la chaussée (expansion, contraction, ramollissement) et de la manière dont l'humidité ou les conditions de visibilité peuvent affecter la saisie des données pendant les campagnes de mesure. Cependant, l'impact de chaque variable météorologique individuelle était moins prononcé, et leur effet combiné explique probablement une plus grande partie de la variabilité des mesures.

L'analyse de l'indicateur I3 met en évidence le fort impact de la société de mesure sur la variabilité, avec une variance expliquée d'environ 50 %. Ce résultat souligne la nécessité d'une plus grande standardisation des méthodologies de mesure. En outre, les interventions et les conditions météorologiques contribuent aux variations de l'indicateur I3.

Les données sur le frottement de surface ont été collectées à l'aide du véhicule d'essai RK-SKM 80, avec une fréquence de mesure standard tous les 100 mètres et une vitesse de fonctionnement de 80 km/h. Afin d'assurer la cohérence spatiale sur plusieurs années d'enquête, malgré les fluctuations naturelles des points de départ des mesures, les données ont été normalisées à l'aide d'une approche du plus proche voisin, en les alignant sur des segments fixes de 100 mètres à partir d'un marqueur zéro défini. L'analyse s'est concentrée sur la voie centrale et a utilisé des intervalles d'un kilomètre comme unité d'analyse de base en agrégeant dix mesures consécutives. L'étude a porté sur 12 autoroutes principales (N1-N16, à l'exception de quatre), soit 474 kilomètres de route. Pour chaque tronçon routier, les années d'enquête ont été sélectionnées pour garantir des mesures répétées à des endroits cohérents dans le temps, ce qui a permis d'obtenir un ensemble de données de 1 322 échantillons routiers valides. Chaque échantillon contenait des données provenant de deux enquêtes consécutives, et lorsque des valeurs manquantes étaient détectées dans l'intervalle d'un kilomètre, l'échantillon était exclu afin de préserver la qualité des données.

Pour mesurer l'étendue de la variabilité des mesures de frottement, une nouvelle métrique appelée Différence Moyenne (DM) a été introduite. Cette mesure calcule le changement moyen des valeurs de frottement entre deux années d'enquête sur un segment d'un kilomètre. La distribution de l'écart moyen a suivi une courbe gaussienne légèrement négative, suggérant une tendance générale à l'amélioration des valeurs de frottement de surface - souvent due à des interventions - au fil du temps.

Les facteurs susceptibles d'influencer l'AD ont été regroupés en trois catégories : Mesure de l'information : y compris la question de savoir si une intervention a eu lieu (binaire), la différence de projet ou le changement d'entreprise d'inspection entre les années, et la durée de l'année d'essai, qui indique le temps écoulé entre les enquêtes. Une autre catégorie était celle des conditions de mesure : les vitesses mesurées chaque année, leur différence et les écarts types de la vitesse pendant les tests. Enfin, nous avons également utilisé les conditions météorologiques, notamment la différence de température et la différence de vitesse du vent entre les deux périodes d'enquête.

L'analyse de corrélation a montré de faibles corrélations directes entre l'AD et les variables individuelles, ce qui indique qu'aucun facteur unique n'explique de manière dominante la variabilité du frottement et suggère une influence multifactorielle. Cependant, la présence d'interventions et le changement d'entrepreneur entre les années sont apparus comme des variables intéressantes. Des modèles d'apprentissage automatique ont ensuite été utilisés pour étudier les données. Un modèle de régression CatBoost a démontré une performance exceptionnelle, atteignant un R^2 de 0,9186 et un RMSE de 0,362, expliquant plus de 92% de la variabilité de l'AD. Ce niveau élevé de précision prédictive montre que les variables enregistrées capturent efficacement les influences clés sur les changements de frottement de surface. Notamment, les tentatives d'optimisation du modèle par le réglage des paramètres n'ont pas permis d'obtenir des gains de performance significatifs.

Pour aller au-delà de la corrélation et étudier les relations causales, un cadre d'inférence causale structurelle a été appliqué à l'aide de la bibliothèque DoWhy. Un graphe causal a été construit sur la base de la connaissance du domaine et des résultats de l'analyse SHAP (SHapley Additive exPlanations), cartographiant les relations causales supposées entre les variables d'entrée et l'AD. Cela a permis de calculer les effets moyens du traitement (ATE) pour chaque variable. L'analyse causale a révélé que les interventions avaient l'effet causal positif le plus significatif sur l'amélioration de la friction, avec un ATE de 0,7193. Cela signifie qu'en moyenne, les interventions ont entraîné une différence de valeur de frottement supérieure de 71,93 % par rapport aux segments n'ayant fait l'objet d'aucune intervention, ce qui confirme l'efficacité de la maintenance. En revanche, plusieurs facteurs, dont la différence de projet, la vitesse de mesure, la différence de vitesse de mesure et la différence de température, se sont révélés avoir des effets causaux négatifs, bien que l'interprétation de leur directionnalité soit moins évidente en raison de leur nature continue. L'analyse SHAP a également confirmé que la différence de température et l'intervention étaient les caractéristiques les plus influentes pour expliquer l'AD. Les différences de température peuvent affecter la texture de la chaussée ou la réactivité du système de mesure, tandis que l'importance des interventions s'aligne sur leur impact physique sur la qualité de la surface de la route. Le diagramme SHAP beeswarm a en outre révélé que la durée de l'année d'essai et l'écart-type de la vitesse de mesure influencent également l'AD, en particulier par le biais de valeurs aberrantes. Enfin, des contrôles de robustesse ont été effectués sur les résultats de l'inférence causale à l'aide de trois techniques : introduction d'une cause commune aléatoire, traitement par placebo et validation de sous-ensembles de données. Tous les tests ont confirmé la stabilité et la fiabilité des effets causaux estimés, renforçant ainsi la confiance dans les résultats.

A l'inverse, l'indicateur I4 présente le plus haut degré d'explicabilité, avec jusqu'à 91% de la variabilité expliquée. La température est le facteur d'influence le plus important, suivi de près par la vitesse du vent, les interventions et la vitesse de mesure.

L'analyse des différences entre les branches régionales ou filialen axées sur les indicateurs I3 (régularité transversale) et I0 (dégâts de surface) a révélé des disparités notables dans la qualité et la quantité des données, ainsi que dans les facteurs de variabilité. Certaines filiales ont conservé des ensembles de données plus complets et plus cohérents pour l'ensemble des années d'enquête et des indicateurs, tandis que d'autres présentaient des lacunes dans les données ou des années d'enquête manquantes. Ces incohérences soulèvent la question de savoir si les données n'ont jamais été collectées ou si elles n'ont tout simplement pas été téléchargées, ce qui entrave la capacité à générer des informations fiables à l'échelle nationale. En outre, les principaux facteurs influençant la variabilité des indicateurs diffèrent d'une filiale à l'autre, probablement en raison des différences régionales en matière de géographie, de climat ou de charge de trafic.

Pour l'indicateur I0, des tendances similaires ont été observées. Alors que la vitesse de mesure est généralement un facteur dominant dans toutes les filialen, dans la Filiale 2 (Thun), les variables liées aux conditions météorologiques ont eu un impact plus important. Là encore, cela peut refléter des différences environnementales régionales ou des variations procédurales dans la manière dont les mesures sont effectuées. En outre, la proportion de variabilité expliquée varie également d'une filialen à l'autre, ce qui indique que certaines régions ont plus de bruit inexpliqué dans leurs données que d'autres. Ces résultats sont importants pour harmoniser les normes de collecte de données et garantir des pratiques d'évaluation cohérentes à l'échelle nationale, en particulier si l'objectif est de développer des règles générales ou de mettre à jour les normes d'état des chaussées au niveau national.

Sur la base de ces résultats, plusieurs recommandations sont proposées pour améliorer la fiabilité et l'utilisation des données collectées.

Tout d'abord, la visualisation systématique des données devrait être améliorée pour faciliter l'évaluation rapide de l'état des chaussées et la cohérence des données. En développant un système de visualisation automatisé qui se met à jour avec les nouvelles données, les parties prenantes telles que les gestionnaires de routes et les sociétés de mesure auraient plus facilement accès à des informations cruciales. Cette approche permettrait d'identifier rapidement les problèmes de mesure potentiels, ainsi que les données manquantes qui pourraient compromettre les évaluations à long terme. Deuxièmement, les paramètres météorologiques devraient être systématiquement mesurés en même temps que les indicateurs de chaussée. L'utilisation de stations météorologiques portables, qui sont désormais à la fois abordables et fiables, fournirait des données essentielles sur la température, l'humidité et d'autres conditions environnementales susceptibles d'influer sur les résultats des mesures. En outre, la programmation des enquêtes devrait être optimisée afin de garantir que les mesures sont effectuées dans des conditions similaires à celles des enquêtes précédentes. Cela permettrait de réduire les incohérences dues à des facteurs environnementaux externes. Troisièmement, pour atténuer l'impact des différentes entreprises sur la variabilité des mesures, les accords contractuels doivent garantir que la même entreprise est responsable d'au moins deux années d'enquête consécutives. Cette approche permettrait de réduire les incohérences dues aux différences méthodologiques entre les entreprises. En outre, toutes les entreprises effectuant des mesures devraient être tenues d'étalonner régulièrement leur équipement et de procéder à des évaluations comparatives afin de garantir l'uniformité des processus de collecte des données. Quatrièmement, le processus d'enregistrement des interventions devrait être plus systématique et plus complet. Les incohérences actuelles dans les données relatives aux interventions rendent difficile l'évaluation de leur impact réel sur l'état des chaussées. Les interventions, qu'elles soient majeures ou mineures, devraient être documentées en détail, avec des informations sur le type d'intervention, l'étendue des réparations et les matériaux utilisés. Cela permettrait d'améliorer la capacité à corréler les activités d'intervention avec les changements ultérieurs de l'état de la chaussée. Cinquièmement, la portée de l'analyse des données devrait être élargie afin d'intégrer des variables supplémentaires susceptibles d'expliquer la variabilité des mesures. En particulier, l'inclusion de données détaillées sur le trafic - en faisant la distinction entre les poids lourds et les véhicules de tourisme - permettrait d'obtenir une image plus précise de la tension exercée sur la chaussée et des schémas de dégradation. En outre, l’intégration des propriétés des matériaux ainsi que des paramètres topographiques et géotechniques dans l’analyse pourrait apporter des éclairages supplémentaires sur l’influence des conditions géographiques et géologiques sur la dégradation des chaussées en fonction du type de revêtement. Enfin, il est fortement recommandé d'investir dans le développement de l'application TRA-Trassee. L'amélioration de sa fonctionnalité afin d'intégrer toutes les données relatives aux chaussées, y compris les outils de visualisation et les capacités analytiques, améliorerait considérablement sa convivialité. Rendre le système plus accessible aux gestionnaires, aux sociétés de mesure, aux chercheurs et aux décideurs politiques faciliterait la prise de meilleures décisions et garantirait une utilisation efficace des données critiques.

En conclusion, les résultats de cette étude mettent en évidence des incohérences significatives dans les mesures de l'état des routes, soulignant la nécessité d'améliorer la précision et la fiabilité des données. La mise en œuvre des recommandations proposées

- notamment une meilleure visualisation des données, des mesures météorologiques systématiques, des pratiques d'entreprise normalisées, une documentation détaillée des interventions, une analyse élargie des données et le développement de TRA-Trassee - contribuera à une surveillance plus efficace de l'état des chaussées et à une planification à long terme de l'infrastructure. Ces améliorations permettront en fin de compte d'assurer la durabilité et l'efficacité du réseau routier national de la Suisse.

1 Introduction

The road network in Switzerland is dense and consist of 83274 km in total. From those, 2255 km corresponds to national roads, 17278 km cantonal roads and 63742 km of other roads (communal or private roads) (Federal Office of Statistics). National roads are in particular important considering that despite that they constitute only 3% of the total Swiss total network, they account for around 40% of the traffic. Furthermore, around 62.9% of goods traffic in handled via national roads corresponding to a total of 27,2-billion-ton kilometres (FEDRO, 2020).

The planning, construction, maintenance and management of the road network has different responsible depending on the type of roads. For the case of national roads, the Swiss Federal Roads Office (FEDRO/ASTRA/OFROU) is responsible, for cantonal routes, each canton has their own responsible office, and for smaller routes, the municipalities are responsible.

A road network as any infrastructure is expected to deteriorate over time because of normal aging, weather exposure and various stressors like for example the use of salt in winter or deformations of the terrain caused by landslides, flooding or other geological phenomena. In order to maintain a road network that ensures an adequate level of service at an optimal cost, its condition must be evaluated periodically so problems can be detected in time and measures can be planned and implemented.

Swiss road managers, or companies mandated by them, diligently adhere to standards and guidelines when assessing the condition of their road assets. This assessment data plays a crucial role in forecasting asset deterioration and devising optimal intervention strategies. Given the significance of this information, it must meet specific criteria including accuracy, comprehensibility, repeatability, and reproducibility.

Accuracy in data measurement refers to how closely measured values align with the true or actual values of the quantities being assessed. It signifies the degree of correctness in measurements, with minimal error. Factors impacting accuracy encompass calibration, unit selection, and sources of error like systematic biases, random variations, and external conditions. Ensuring accuracy is vital across scientific, engineering, and research domains for reliable data analysis. Uncertainty quantification is often associated with accuracy, providing a range within which true values likely reside. Accurate measurements are fundamental for informed decision-making.

Comprehensibility in data measurement refers to the ease with which individuals can fully grasp and make sense of the data, including its context, significance, and implications. It goes beyond mere understanding to encompass a deeper level of insight. Achieving comprehensibility involves not only presenting data clearly but also providing relevant context, explanations, and interpretations. It ensures that data users can draw meaningful conclusions and make informed decisions based on the information presented. Whether in scientific research, business analytics, or public policy, enhancing comprehensibility is vital for harnessing the true value of data, enabling informed choices, and driving meaningful actions based on data-driven insights.

Repeatability, as determined through variance analysis, hinges on measurement precision under identical conditions. Precision, though related, differs from accuracy as it gauges the consistency of repeated measurements. This involves consistent procedures, measurement systems, operators, operating conditions, and locations, while measurements are replicated on the same objects within a short timeframe.

Reproducibility, on the other hand, is contingent on measurement precision under diverse conditions, encompassing different locations, operators, measuring systems, but maintaining the same or similar objects (Pereira et al., 2016).

Additionally, inherent uncertainties may stem from material properties, measurement devices and their inherent randomness. These variations are introduced by the laws of physics. Mitigating repeatability variation entails enhancing the physical measurement process, while addressing reproducibility variation involves refining procedures.

In this report it is first presented a literature review exploring the main sources of variability that are present in the measurement of indicators of pavement condition. It is specifically focused on the indicators surface damage (I0), longitudinal evenness (I2), transversal evenness (I3) and surface friction (I4). It is then presented in detail the normative in place in Switzerland. After that the methods used in this study are described. Following comes the results of the initial visual exploratory analysis separated by regional branches (named "*filialen*" in Switzerland) and the results of the factor importance analysis separated by indicator. Finally, a discussion of the results is presented and a series of recommendations are proposed in order to reduce the variability in the data and improve the quality and usability of the dataset in the long term.

2 Literature review

2.1 Pavement condition assessment

A Pavement Management System (PMS) can be defined as a set of tools designed to assist in the management of a road network. Its main objective is to determine the optimal maintenance and rehabilitation strategies in order to maintain the pavements of these roads at a specific level of service (Ferreira et al., 2011). The quality of management procedures, as noted in the literature (NCHRP 401, 2009), is influenced by several key factors. These factors include the outsourcing of pavement data collection, the quality of location reference data, the consistency of historical data, the spatial and temporal coverage of the network, as well as new requirements resulting from evolving business practices. It is therefore of utmost importance to ensure the quality of pavement condition data used in the Pavement Management Systems (PMS). The availability of accurate and consistent data over time plays a key role in the modelling necessary to develop effective multi-year maintenance plans and work programmes. It is recognised that there are variations in measurements of condition variables over time, and this variability can affect both the estimation of road deterioration and the accuracy of future condition information (Ruotoistenmäki et al., 2006). Therefore, it is essential to minimise errors in the data, as these can have a significant impact on treatment recommendations and budget requirements (NCHRP 401, 2009).

Network-level data collection involves the collection of large volumes of information about the condition of roads. In this respect, road authorities tend to prefer high-speed data collection techniques, as they are less time-consuming compared to project-level approaches. Although these techniques may be less accurate, their advantage is that they do not disrupt traffic flow and minimise risks for data collection teams. This becomes even more relevant when multiple service providers are employed for data collection, as they may use different equipment, methodologies or innovative technologies. Therefore, rigorous verification and validation of the accuracy and precision of the devices used becomes even more critical.

Variation in the data recorded by high-speed profilometers in the same year arises due to a number of reasons, including the conditions present at the time of measurement, seasonal fluctuations in the profile, the accuracy of the measuring equipment and the method used to process the data, as well as variations in the lateral position of the measuring vehicle. When multiple profilometers of the same type are used, differences in equipment and variations in lateral position caused by different operators also contribute to variability (Ruotoistenmäki et al., 2006).

In recent years, there has been a growing interest in machine learning-based models for predicting pavement performance. These models have the ability to capture intricate relationships but several challenges have hindered their practical application like for example a lack of effective feature selection prior to model development, the complexity of interpreting black-box models, and the omission of uncertainty considerations. To address these limitations, Yao et al. 2022 have created a novel framework for modeling pavement performance evolution incorporating the BorutaShap method for feature selection, the Bayesian neural network (BNN) for constructing models and quantifying uncertainties, and the SHapley Additive exPlanations (SHAP) approach for interpreting the models. It was identified that poor data quality is the primary source of uncertainty in the models. But the model interpretation allows to shed light on the key factors influencing pavement performance.

In a study that examines performance measures for both flexible and rigid surfaced pavements, along with the corresponding threshold values (Corey-Lay, 2014) report the following findings based on a survey that was distributed to members of a Joint Technical Committee on Pavements, with responses received from 14 out of 20 participating states. States commonly utilized rutting and cracking as indicators of flexible pavement performance, with IRI ranking as the third most frequently employed measure, applied consistently to both flexible and rigid pavements. Rutting assessments varied depending on the number and types of sensors used, spanning from five-point sensors to line sensors to three-dimensional cameras. The adoption of cracking as a performance measure required a collaborative effort to establish precise definitions, measurement methods, and threshold criteria. The survey's focus on rigid pavements was largely confined to jointed plain concrete, as 12 out of the 14 responding states indicated that the majority of their rigid pavements fell into this category. Further work is needed to develop a faulting measure, as its accuracy depends on the spacing between consecutive traces. Definitions, methods, and threshold criteria are also essential components yet to be established for other performance measures relating to rigid pavements, including patching, cracked slabs, and damaged joints.

Jia et al. 2016 did an evaluation to explore the impact of data variability on maintenance planning. In the context of Tennessee's current pavement management system, the dominant factors influencing maintenance planning were the variability in roughness and distress severity level. In contrast, the variability in distress extent had a minor effect on maintenance planning, and the variability in rut depth did not significantly influence it.

Evdorides et al. 2013 introduces a methodology for modeling pavement resilience and offers insights into the factors affecting the accuracy of pavement performance predictions. In their work the authors presents a methodology that accomplishes two objectives: (a) it employs the Monte Carlo data generation technique to construct appropriate performance models, and (b) it enables an analysis of how variations in model parameters affect the predicted performance. The analysis of the results reveals that, among the examined data sets, the variability in traffic loading has a significant impact on pavement performance. In contrast, the variability in pavement strength is of lesser importance, while the initial pavement roughness has a negligible effect on pavement performance.

Multiple researchers have focused on the determination of the error associated with measurements made using one or more pieces of equipment. Humplick (1992), for example, studied methods to tackle the issues of measurement errors related to measuring equipment and measurement locations. He classified measurement error analysis into methods examining existing processes and identifying the presence/absence of errors (e.g., statistical comparison and correlation analysis) and methods determining causes of errors (e.g., experimental and quasi-experimental techniques, analysis of variance (ANOVA), and covariance). The first group includes analysis techniques relying on mean and standard deviation calculations. These techniques have been used extensively to investigate errors in pavement distress (Bianchini et al., 2010; Descornet et al., 2001; Flintsch & McGhee, 2009), and surface roughness (Hong & Prozzi, 2010; Ruotoistenmäki et al., 2006; Schmidt, 2001). For example, (Hong & Prozzi, 2010) collected data from in-service pavement sections as part of the Minnesota Road Test project to capture the real-world pavement deterioration process. Several variables involved in the system were thoroughly investigated, including pavement design, materials, traffic, environment, and maintenance factors. Performance uncertainty due to unobserved heterogeneity was incorporated into the proposed model. The unobserved heterogeneity was primarily caused by the variability in the materials and the construction process. The model was estimated through maximum simulated likelihood estimation, and it was demonstrated that the results were consistent with observations and engineering judgment in the context of pavement design and performance. In addition to population-level parameters representing the general deterioration mechanism, section-specific or individual-level parameters were obtained through the Bayesian approach. This is considering that two levels of parameters can be used to accommodate network- and project-level analysis in pavement management.

Additionally, Humplick (1992) developed a hybrid methodology, that integrates the above-mentioned methods and a latent Markov decision process (LMDP) to create an error model that is considered fundamental for later models (Kobayashi et al., 2012). Madanat, (1993) and Park et al. (2008) suggested some approaches to deal with sampling size and small sampling populations to reduce measurement errors. Others have focused on the use of Kalman filters and hidden Markov models (Chu & Durango-Cohen, 2007, 2008; Kobayashi et al., 2012; Lethanh & Adey, 2013).

Measurement System Analysis (MSA) is another statistical method used in accuracy and sensitivity estimations for measurement devices. Although not very common in the analysis of the road condition indicator variabilities, this method is very well-established in other fields including transportation (Joubert & Meintjes, 2015), spring measurement (Shi et al., 2014), wood industry (Li & Al-Refaie, 2008), automotive industry (Senol, 2004; Yeh & Sun, 2013), and steel industry (Pereira et al., 2016). Measurement system analysis has a pivotal role in assessing the measuring error when conducting empirical studies. Gage Repeatability and Reproducibility (GR&R) studies are classic methods used to evaluate measurement system adequacy for a particular application. For example, (Araújo et al., 2019) proposed new indicators for measurement error detection due to both repeatability and reproducibility variation in GR&R studies. These indicators were calculated using standardized scores from the analysis of variance. This study advanced the state of the art by being able to detect both reproducibility and repeatability sources of measuring error rather than just the reproducibility error. (Soares et al., 2022) provided a systematic literature review to assess the state-of-the-art in GR&R studies. In this study, articles were classified according to the manufacturing process, measuring instruments, quantities and units, design setup, statistical methods, and acceptability criteria. Finally, the main objectives and contributions of each study were highlighted. There are multiple examples of the application of MSA in industry. For example, in a study for the National Aeronautics and Space Administration (NASA) (Cmar et al., 2012), integrated system tests were conducted to demonstrate the control systems' functionality and capability. Modern measurement systems analysis (MSA) tools were developed to help verify system health and measurement integrity. (Wang et al., 2020) focused on the variance method to analyse the repeatability and the reproducibility of electrical measurement data. They studied the error of the measurement system and calculated the proportion of various error sources that affect the measurement result, the interaction, synthetically determination of the measurement system status and measuring data quality and adopted effective measures to reduce or eliminate the error.

2.2 Visual inspections

Pavement ratings were studied by Montgomery et al. 2019. The results indicate that pavement in good condition was rated more accurately, while pavement on the poor/fair condition boundary was rated less accurately. Participants were more accurate in assigning surface evaluation ratings after receiving specific training. Additionally, ratters that were more accurate were also more consistent in performing surface evaluation ratings. Participants with engineering roles, such as engineer, engineer technician, or engineer assistant, were more accurate. In contrast, participants with leadership roles, such as supervisor, manager/foreman, or team leader/elected official, or less than 1 year of surface evaluation rating experience were less accurate.

Jia et al. 2016 studied distress data underweening Monte Carlo simulation to quantify errors in distress extent, and transition matrices were applied to gauge errors in distress severity classification. Results show that the accuracy of assessing low-severity extent had minimal influence on the overall distress index. However, accurate assessment at moderate and high severity levels had a substantial impact on the overall distress index, with moderate-level assessments being the most influential. (Bogus et al. 2010) studied assessments of surface distress in pavements, including issues like cracking, bleeding, and ravelling, that are commonly incorporated into overall pavement condition evaluations. Both manual and automated survey methods are available for conducting pavement distress assessments. However, it is important to acknowledge that all distress evaluations inevitably exhibit a degree of variation in their outcomes. In particular, a high variability in the ratings for specific distress types, such as bleeding were reported.

2.3 Profilometers - roughness

(Freitas et al. 2014) studied data collected from five profilometers operating across six extended road segments (18 km). The parameters under investigation included the surface type, operator, system, and the number of repetitions during the runs. The findings revealed varying levels of repeatability for each parameter measured (mean profile depth, texture depth and IRI). Furthermore, the analysis of these factors underscores

the need for a more comprehensive investigation into how the surface type affects the variability of the data obtain from profilometers.
Transverse profile measurement data was analysed by (Descornet et al. 2001). The primary objective of this analysis was to assess the repeatability, reproducibility, and accuracy. Applying the International Organization for Standardization 5725 standard revealed the presence of a notable number of outliers in the results obtained from almost all the devices tested. The variation in (in)accuracy, ranging from 0.8 to 3.2 mm among different test sections, can be attributed to the influence of longitudinal unevenness. Importantly, the operating speed within the experiment's speed range did not significantly impact most of the measurements obtained. Additionally, speed had no discernible influence on the repeatability, reproducibility, or accuracy of the indices or profile measurements.

Road network managers employ inertial profilers to gather pavement profile data, which is then used to calculate the International Roughness Index (IRI) for each segment of the highway. An ongoing concern with the use of inertial profilers revolves around the accuracy of the IRI values computed from the profile data. To address this concern, typically, several test sections are established. Profile data is collected on these sections using both a reference device and the profiler. Subsequently, a comparison is made between the IRI values derived from these two sources. In many instances, a particular profiler may exhibit a high level of agreement in IRI with the reference device for some sections but noticeable discrepancies for others. An analysis of data gathered during Long-Term Pavement Performance profiler comparison studies was conducted to pinpoint factors that might contribute to variations between the IRI values obtained from the profiler and the reference device. The analysis revealed that the following factors can be responsible for differences in IRI values: the sampling quantity of the reference device, variations in the path followed by the profiler, the averaging effects on profiler data, errors in the data collected by the reference device, the computational and the operational procedures employed during the collection of profiler data (Perera et al. 1974).

(Jia et al., 2016) gauged data variance in pavement roughness and rut depth, by utilizing a ratio calculation based on the difference between two-wheel paths relative to their combined value. The findings revealed that, in terms of roughness, state routes exhibited more significant variations compared to Interstates. Additionally, it was observed that the left wheel path generally had a smoother pavement surface than the right, irrespective of route type.

(Jia et al. 2018) found that IRI is susceptible to measurement variability, especially for network evaluations where the IRI value of a road section is determined only based upon a single test run. The uncertainty of pavement evaluation due to the periodic measurement IRI errors was defined and quantified utilizing raw data from long-term pavement performance (LTPP). Three factors contributing to the uncertainty of pavement evaluation were investigated. They are (1) the performance thresholds, (2) variability of IRI measurement, and (3) distribution of IRI for a road network. Probabilistic relationship was constructed to consider the influence of periodical measurements variability of IRI for network-level pavement evaluation. Results indicated that the uncertainty of road sections rated as *good* were the lowest, whereas those rated as *fair* were generally high. The variability of IRI measurement significantly affected the pavement evaluation for state routes.

(Shalaby and Reggin 2007) used a nonparametric statistical test to assess the transverse variability of pavement condition data. The test compared the ratings for one lane with those of all lanes of each segment. The test concluded that the medians of the two groups are equal at a 92% confidence interval and that there are observed biases in the data. The bias can be eliminated if the surveyed lane is selected randomly.

Yin et al. 2006, offers recommendations for quantifying and managing the variability in longitudinal profile data and the resulting International Roughness Index (IRI) values. These recommendations are based on an analysis of profile data collected through the Long-Term Pavement Performance Program (LTPP). The primary focus was on devising methods to quantify variability that would be practical and cost-effective for network-level pavement management. In the initial phase of the variability analysis, the impact of geographical regions and seasonal variations was assessed using carefully screened profile data. Additionally, the relationship between the standard deviation and the mean of repeated profile measurements was modelled using raw profiles, which are more relevant to the typical conditions of single-pass network-level data collection. The second phase of the variability analysis focused on quantifying periodical measurement variability, which can be valuable for regular checks on control sections, device acceptance, or project-level data collection. It was determined that periodic measurement variability ranges between 6% and 8% for asphalt concrete (AC) pavements and between 6% and 9% for Portland cement concrete (PCC) pavements.

2.4 Friction measurements

A comprehensive review of the measurement and modelling of skid resistance of asphalt pavement was provided by (Yu et al., 2020). (Plati et al., 2020) quantified the skid resistance seasonal variation in asphalt pavements. In this study, skid resistance data were collected twice per year (after "wet" and "dry" seasons) along four trial sections of an urban highway with a GripTester system (during a six-year field experiment). For the estimation of a percentage reference change (%) between wet and dry skid resistance data, a methodology was developed based on distribution fitting of the percentage change of skid resistance index due to seasonal variation impact. The outcome of the analysis was then validated through a set of skid resistance data collected from an interurban highway pavement section with different traffic and environmental conditions. This developed methodology can be considered a promising tool for accommodating the seasonal variation of skid resistance. Furthermore, the resulting percentage reference change (%) enables the estimation of skid resistance level for the dry period based on the wet period measurements.

In another notable study (Kouchaki et al., 2018) performed a series of statistical analyses of field-measured friction and texture data to find the texture–friction correlation. Three test sections with different pavement types were selected within the state of Texas. Data were collected at three locations in the right wheel path and three locations in the centre of the lane for each test section. To measure the texture data, the researchers used the circular track meter (CTM) and a prototype measurement device developed in-house and consisting of a line laser scanner (LLS). Friction measurements were obtained with the dynamic friction tester (DFT) and Grip-Tester. The mean profile depth (MPD) was calculated by using the measured texture data. The relationship between the MPD values and the friction numbers obtained from the Grip-Tester and DFT was investigated at speeds of 50 and 70 km/h (31.1 and 43.5 mph). The repeatability and reliability of both the developed LLS prototype and the Grip-Tester were also evaluated, as well as the effect of test speed on friction measurement. The results indicated a strong positive correlation between the texture and friction data. In addition, the developed LLS prototype was able to scan the pavement surface texture more reliably and precisely than the CTM in terms of vertical and horizontal resolution. The Grip-Tester showed promising results compared with the DFT with regards to the friction measurement.

In two consecutive and more recent studies, (Edmondson et al., 2021, 2022) provided a unique analysis of long-term texture obtained using traffic speed condition survey (TRACS) data from 14 sites, located along a north to south transect spanning the longest highway in the UK. A total of 19 years of sensor measured texture depth (SMTD) data were analysed using spatial filtering techniques and compared with meteorological and traffic datasets. The results for hot rolled asphalt (HRA) surfaces revealed that changes to SMTD follow a linearly increasing trend with time. The “rate of change” is influenced by the order of magnitude of annual average daily traffic (AADT), when factored for the percentage of heavy goods vehicles. This linear trend is disrupted by environmental parameters, such as rainfall events and seasonal conditioning. In the summer, this signal is evident as a transient peak in the “rate of change” of texture greater than 0.04 mm, and in the winter as a reduction. The transient changes in texture corresponded to above average rainfall occurring in the week prior to SMTD measurement. The observed signal demonstrated an inverse pattern to the classically understood seasonal variation of skid resistance in the UK, where values are low in the summer and high in the winter. The findings demonstrated for the first time that texture measurements experience a seasonal signal and provided compelling evidence pointing toward surface processes (such as polishing and the wetting and drying of surface contaminants) causing changes to the texture that are affecting seasonal variation in skid resistance.

(Zhan et al., 2021) used a random forest analysis to determine the effect of aggregate properties on asphalt pavement friction. Aggregate properties were characterized in the laboratory using the Aggregate Imaging System (AIMS), and pavement texture and friction data were collected in parallel in the field using a three-dimensional laser imaging system and Grip Tester. The traditional multivariate friction models were developed while several regression coefficients did not follow engineering judgment. Subsequently, random forest analysis (RFA) was performed to evaluate the pavement surface friction subject to various aggregate properties. Loss of texture (%) was identified as the most significant aggregate characteristic for friction evaluation, accounting for a 26.54% contribution to pavement surface friction. Traffic volume and temperature also exhibited significant impacts, with 31%, and 14.7% contributions to pavement surface friction, respectively.

3 Norms

3.1 Status of the roads

Each part of the status of the road monitoring process is detailed in a specific norm but sometimes overlap between norms occur. In addition, it happens that several norms apply to the same part of the process depending on the circumstances. Figure 1 details the norms that are relevant in the road condition situation:

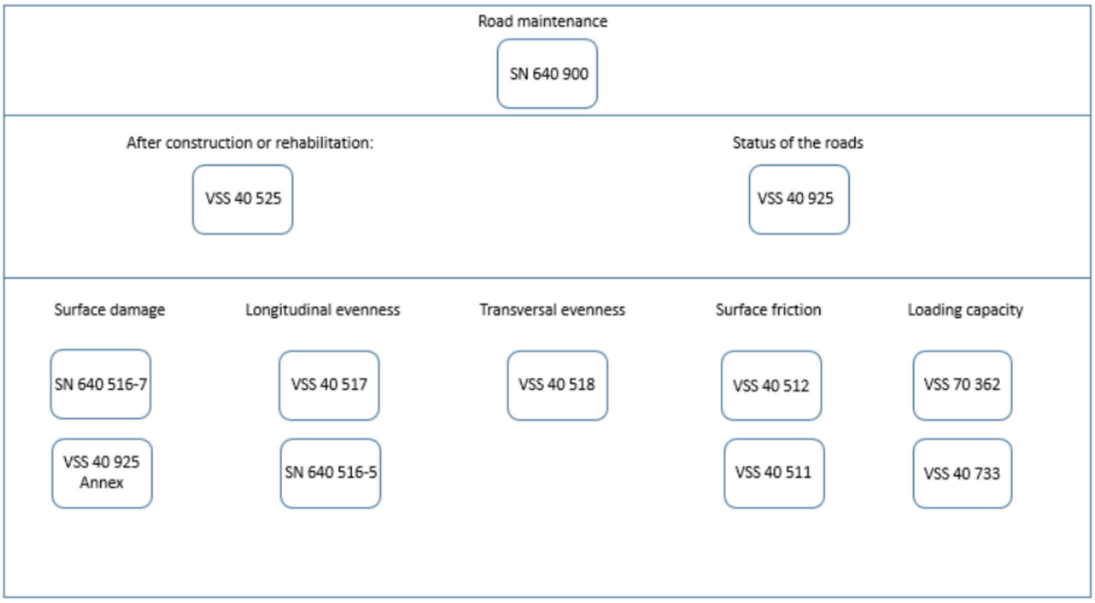

Figure 1: Relevant norms related to the assessment of road condition

The norm VSS 40 525 set the reception requisites for the road surface characteristics after construction or rehabilitation. Four characteristics are evaluated: local irregularities (surface damage), longitudinal evenness, transversal evenness and surface friction.

Local irregularities according to (SN 640 516-7) are used only at the reception of the work. Longitudinal evenness estimation is based on the longitudinal profile and two different methods of analysis described in the Norm VSS 40 517 are allowed. The first method consists on using the angular value W of the change in slope (considering the standard deviation Sw of the angular values and the W_{max}. The second method consist in using a waveband analysis (considering the NBO quality scores (*Notation par Bandes d'Ondes*) for the three wavelength bands: short (PO), medium (MO) and long waves (GO). For the waveband method, specific reception values are stipulated according to the wave's lengths, the maximal speed signalled, and the type and number of layers renewed. The norm mentions that the first method is less adapted because it only allows the appreciation of the short waves.

Transversal evenness is estimated using the rut depth and the theoretical water depth in the rut. Methods allowed are described in the Norm VSS 40 518. Because the theoretical water depth in the rut influences the security of the driving, threshold values applied. The rut depth is not subject to reception or threshold values and is only evaluated using the index I3 of the norm VSS 40 925 concerning the management and maintenance of roads. Reception values are stipulated according to the maximal speed.

Surface friction reception requisites are detailed in the norm VSS 40512. Two dynamic measurement systems applied: Skiddometer BV 8 (measuring the coefficient of friction on a locked wheel) according to the SNR 640 513-10 and SKM (measuring lateral friction coefficient) according to the SNR 640 513-8. And one static method described in SN 640 512-4 (pendulum SRT) combined with SN 640 511-3 (measurement of horizontal drainage of surface courses. Seasonality is considered (meteorological conditions) as well as the precision of the measuring systems used (technical specifications mentioned in SNR 640 513-8 and SNR 640 513-10. The measure of the receptions values is recommended 3-6 months after traffic is happening. Threshold value must be guaranteed during this period. Reception and threshold values are defined according to different speeds for both the dynamics and statics systems.

In order to evaluate the status of the roads, a series of characteristics need to be evaluated according to the norm VSS 40 925b. Accordingly, pavement conditions on a certain road section are estimated using the following four indices:

I0: Surface damage

I2: Longitudinal unevenness

I3: Transversal unevenness

I4: Surface friction

Each of these indicators are grade in a scale between 0 and 5 (where 1 means good, 5 means poor) and depends on one parameter (I2, I3 and I4) or on several parameters (I0). Measurements involve different equipment, people and methodologies that are described in specific guidelines. Each indicator value describes the status of a specific characteristic of the road.

The application of the road surveillance at a network level, using these indicators is in accordance to the road maintenance norm (SN 640 900) and specifically according to the surface of roads maintenance. The results of these evaluations could be uploaded into the data banks of road management according to VSS 40 944. For characterizing the status of network of roads composed indexes are used according to VSS 40 904.

At a project level, the status of the road can also be evaluated as stipulated in VSS 40 730 and according to VSS 40 925b. The status of the roads can be evaluated under all weather conditions if the surfaces are dry. For indices concerning surface friction and load carrying capacity, temperatures stipulation must be respected according to VSS 640510 and VSS 70 362.

The selection of characteristics to be measured depends on road characteristics (function, traffic load, etc. according to VSS 40 904) and on the methodology of road management used. The frequency of measurement will depend on the traffic conditions and the expected degradation rate but not longer than once every 5 years.

3.2 Surface damage

The surface damage (*i.e.* distresses) is evaluated differently for roads with bituminous surface and for roads with a concrete surface. For the former, surface smoothness, surface damage, surface deformation, structural degradation and repairing are the parameters evaluated. For the latter, surface smoothness, loss of material, border and joint degradation, vertical displacement, cracks and repairing are the parameters evaluated.

The norm recommends having units of survey that can be characterized by a unique value. The Indices I0 and I1 that measure the surface damage are monitored by visual inspection (details in the annex to the VSS 40 925b). Transversal evenness (I3) is measured manually or with specific measurement devices.

High-speed data collection of digital images connected to an automated image processing facilities is an alternative to the visual inspection (Ruotoistenmäki et al., 2006a) and are mentioned as an alternative in the annex of the norm. A summary status or a detailed status can be done. The first is normally performed for a complete network assessment or as a systematic procedure in order to establish maintenance needs, and the second one is performed more at a project level.

Surface distresses are described by two dimensions: extent (A) and severity (S). Extend is measure on a scale from 0 to 3 and severity on a scale from 1 to 3. Then an evaluation matrix allows combining both dimensions in a score M that goes from 0 to 9. For the case of the summary status, the parameters are grouped according to the characteristics of the road they are measuring, and M is calculated by group. Then each group has a specific weighing that is multiplied with the M values from the evaluation matrix. Finally, the weighted M values for all groups are summed and that value divided by ten will result on the index (values over 5 are scored as 5). It is recommended to use fixed units of analysis (50m long for the whole lane meaning when there is circulation in both senses, and 100m long for individual lanes) or the scale must be adapted consistently to the selected length of road analysed.

The extent of the damage is evaluated for each of the parameters, however for the severity, only the worst in each category is reported (this is the general case however exception to this rule may apply). For the severity, a representative extent for the whole group of parameters should be determined. For the case of detailed assessment for a project, both the extent and the severity in each parameter are reported.

The summary status is expected to be done by passing through the surveyed stretch on foot or by car and filling a form detailed in the annex to the VSS 40 925b. Specific descriptive sheets for each parameter of the visual inspection are detailed in the annex of the VSS 40 925b (see Figure 2 and Figure 3).

In practice, companies use more technological means to measure the indicator. The most common method consists in taking photos of the pavement thanks to cameras mounted in specially adapted vehicles. People with specific training analyse later the images and perform the evaluation. This process is very likely to be perform by artificial intelligence softwares in a near future.

	Relevé détaillé										Relevé sommaire									
Segment de: Segment à:	PR PR					km km					PR PR					km km				
Types de dégradations **Groupes principaux**	A	S	M	G	M·G	A	S	M	G	M·G	A	S	M	G	M·G	A	S	M	G	M·G
Polissage																				
Ressuage																				
Surface glissante				2					2					2					2	
Usure																				
Désenrobage, sablage																				
Perte de gravillons																				
Pelades																				
Nids de poule																				
Fissures de joint																				
Fissures transversales																				
Fissures diverses																				
Dégradations du revêtement				2					2					2					2	
Ornières																				
Bourrelets																				
Tôle ondulée																				
Déformations de poussée																				
Déformations du revêtement				2					2					2					2	
Fissures d'affaissement																				
Affaissements, flaches																				
Affaissements des bords																				
Soulèvements dus au gel																				
Fissures longitudinales																				
Faïençage																				
Fissures d'épaulement																				
Dégradations structurelles				3					3					3					3	
Réparations				1					1					1					1	
Somme $M_i \cdot G_i$																				
Indice I_1 ou **I_0**[1)]																				

1) I_1 et I_0 sont à calculer de la manière suivante: $1/10\ (\Sigma\ M_i \cdot G_i) \leq 5$

PR Point de repérage

Laisser vide

Matrice		Etendue A 0	1	2	3
Gravité S	1	0	1	2	3
	2	0	2	4	6
	3	0	3	6	9

Figure 2: Descriptive sheet for the visual inspection for the case of bituminous surfaces

	Relevé détaillé										Relevé sommaire									
Segment de: Segment à:	PR PR					km km					PR PR					km km				
Types de dégradations **Groupes principaux**	A	S	M	G	M·G	A	S	M	G	M·G	A	S	M	G	M·G	A	S	M	G	M·G
Polissage																				
Surface glissante				1					1					1					1	
Usure																				
Pelade																				
Ecaillage																				
Perte de matériaux				2					2					2					2	
Dégradations des bords, épaufrures																				
Jointoyage absent ou friable																				
Dégradations aux bords et aux joints				1					1					1					1	
Affaissements, soulèvements dus au gel																				
Formation de marches d'escalier																				
Pompage																				
Blow-up																				
Décalage vertical				3					3					3					3	
Fissures																				
Dalles cassées																				
Fissures, cassures				2					2					2					2	
Réparations				1					1					1					1	
Somme $M_i \cdot G_i$																				
Indice I_1 ou I_0 [1]																				

[1] I_1 et I_0 sont à calculer de la manière suivante: $1/10\ (\Sigma\ M_i \cdot G_i) \leq 5$

PR Point de repérage

Laisser vide

Matrice		Etendue A			
		0	1	2	3
Gravité S	1	0	1	2	3
	2	0	2	4	6
	3	0	3	6	9

Figure 3: Descriptive sheet for the visual inspection for the case of concrete surfaces

3.3 Longitudinal evenness

Longitudinal evenness (Indicator I2) is estimated according to the standard deviation (Sw) of the angles (W), as stipulated in VSS 40517 (previously on SN 640 520 and SN 640 521 now abolished). The norm EN 13036-5 (SN 640516-5) also provide details on the methods. The longitudinal profile is the intersection between the road surface and a reference plane that is perpendicular to surface and parallel to the circulation lane. Figure 4 details the measurement localization:

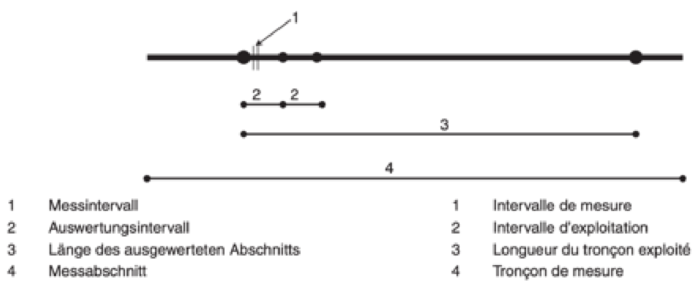

Figure 4: Schematics of the measurement localization and exploitation intervals

The measurement interval is the distance between two consecutive measures. The exploitation interval is the distance where a value is assigned. The length of an exploited part is where significant values are obtained, and the length of the part measured is normally longer for technical reasons. The measurements are expected to be done in a road distance of 250m.

The following measurements can be used to evaluate the longitudinal evenness:

- Values of angles W and standard deviation Sw
- International roughness Index (IRI). It is the most widely used measure of longitudinal unevenness (Sayers et al., 1986).
- Waveband analysis (considering the NBO quality scores (*Notation par Bandes d'Ondes*) for the three wavelength bands: short (PO), medium(MO) and long waves (GO).
- Weighted longitudinal Profile (BLP): standard deviation of amplitudes (SBW) and maximal amplitude (DMB).

Values of angles W and standard deviation Sw: The value W represent a slope change in ‰ in the longitudinal profile. W correspond to the angle between two neighbouring chords of 1 meter long (see Figure 5). The value Sw is the standard deviation of the W values for the analysed distance length.

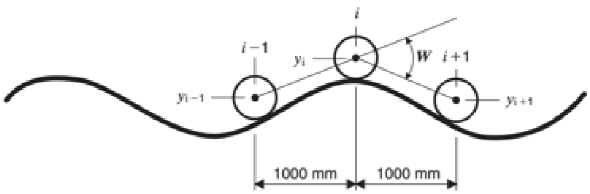

Figure 5: Detail for the measurement of the angle W in the longitudinal evenness evaluation

International roughness index: This index allows quantifying the comfort on the road for the user. It is calculated from simulations of the vertical movement of a modelled suspension of a quarter of a vehicle (one wheel) moving through the distance measured.

NBO analysis: This analysis allows to locate in which length of waves domain the irregularities are located. For the analysis, the longitudinal profile is decomposed in three, each corresponding to the wave length (PO, MO and GO) defined in EN 13036-5 (SN 640 516-5). The energy of the filtrated signal is calculated for each of the three profiles and then transformed in grades using a quality scale.

Weighted longitudinal profile: This method consists in defining a new profile, weighing the amplitudes measured in different wave lengths in a way so the take an equivalent weight in the analysis. The standard deviation of the amplitudes (SBW), and the maximal amplitudes (DBW) are calculated. They describe periodic or isolated defects respectively.

More details on each of the measurement's methods are described in VSS 40517. Then the Index is calculated according to the following graph (Figure 6):

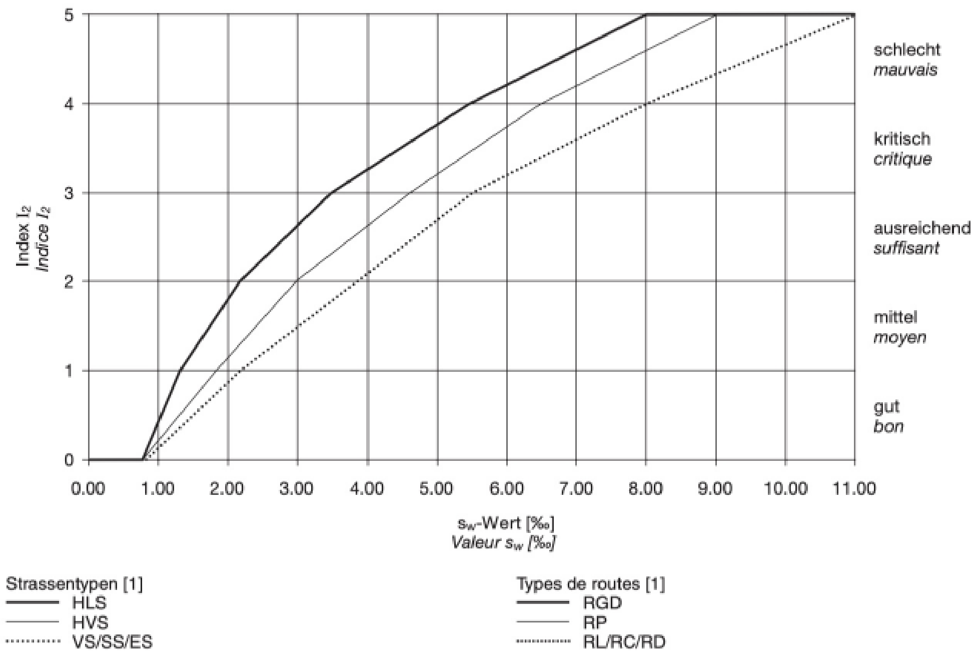

Figure 6: Graph used to transform the measured Sw into I2 values

There is no information on the norm on how to calculate the index I2 in relation to other measurement besides the one from the angle W. Only reception values after construction are listed in (VSS 40525).

In practice, companies use sophisticated vehicles with incorporated profilometers based on laser technology.

3.4 Transversal evenness

Transversal evenness (Indicator I3) is calculated based on the parameters: rut depth (T) and rut depth related to the horizontal or theoretical water depth (t) (VSS 40518).

Two methods exist to measure the rut depth: using a string or using a ruler. It must be indicated which method was used. Values must be measured for both ruts. Several measurements along the road are necessary to provide an estimate for a determined distance (see Figure 7 and Figure 8).

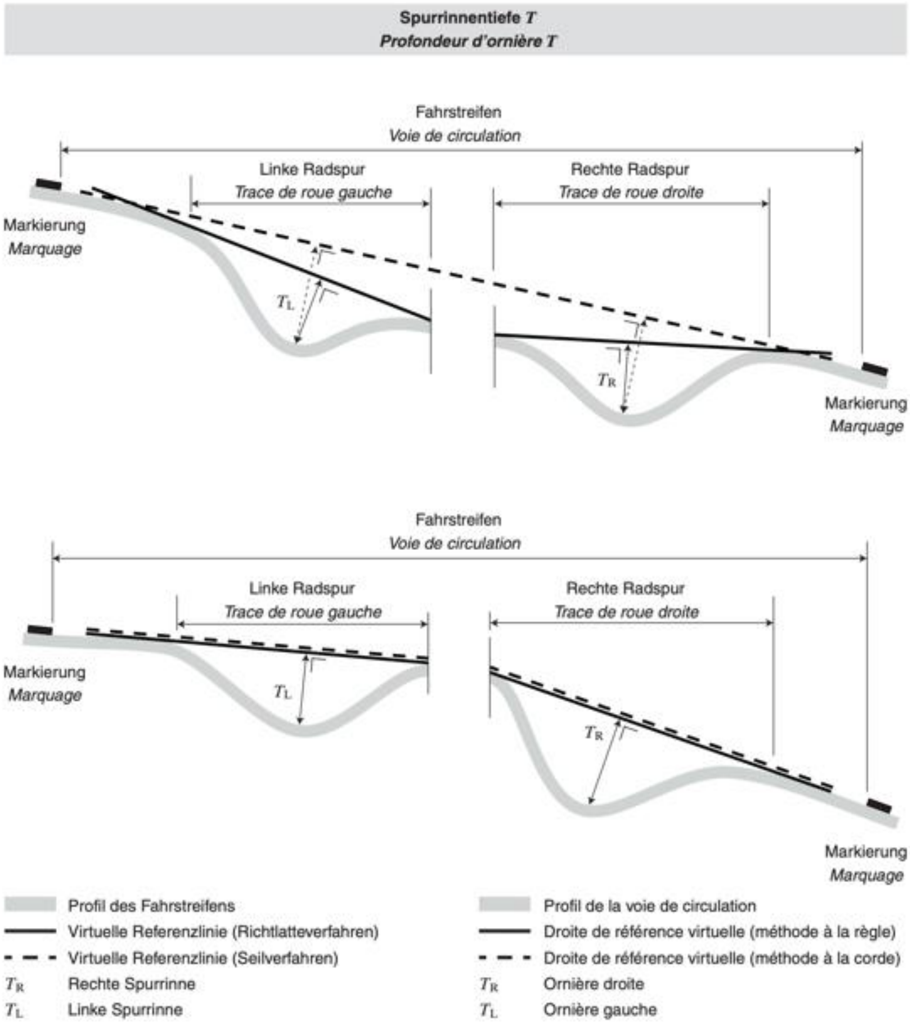

Figure 7: Graphical description of the methods used to measure the transversal evenness

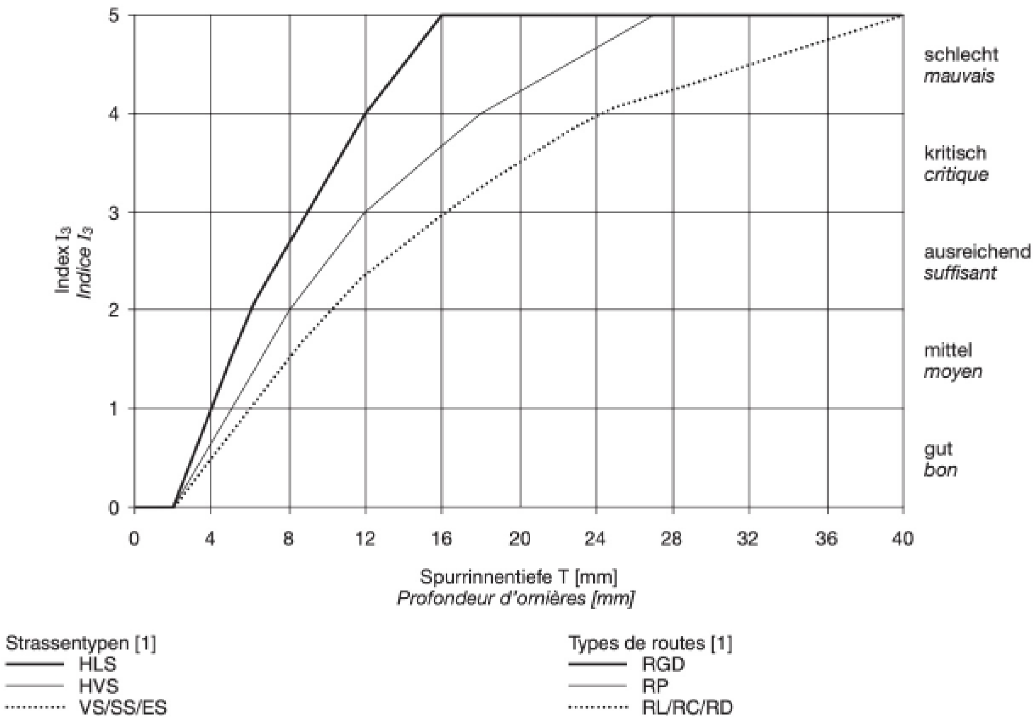

Figure 8: Graph used for the transformation of the measured values into the Indicator I3

Theoretical water depth (t) does not influence the index I3 but are evaluated from the security point of view and must respect the following threshold values: for highways and roads >80 km/h t_{max}= 4.0 mm, other roads, t_{max} = 8.0 mm.

Details on the measurement are shown in Figure 9 (the angle of the road is required. On the figure the road is completely horizontal):

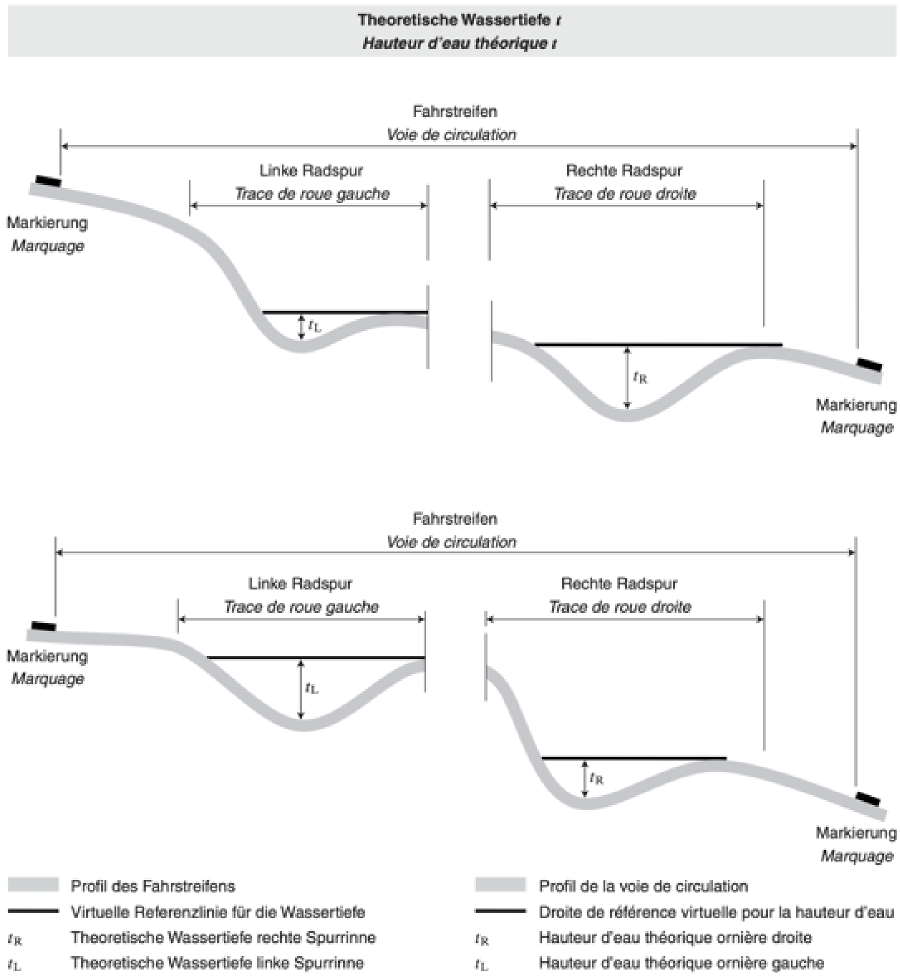

Figure 9: Graphical description of the influence of the water depth in the I3

The indices I2 and I3: Longitudinal and transversal evenness are monitored primary using vehicle-based longitudinal profilers. The norm VSS 40518 mention that for the case of measurement in a road network, only one rut can be measured. The transversal space between the measurement point of the profile must be maximum 10 cm. The precision of the points must be minimum 0.5 mm for dynamic measurements and 1.0mm for static measurements. When both ruts are measured, the maximal values are the ones that counts for the index. The norm mentions that statistic methods are normally employed. It is recommended to use a minimum of 10 measurements for a measured interval of road. The average and standard deviation of the values plus the maximal values must be recorded.

In practice, companies use sophisticated vehicles with incorporated profilometers based on laser technology.

3.5 Surface friction

The surface friction (Index I4) is estimated using the method of the "blocked wheel on wet surface" according to SN 640 510 and VSS 40 511. The parameter measured is the friction coefficient (μ) and is translated into the index according to the following graph (Figure 10):

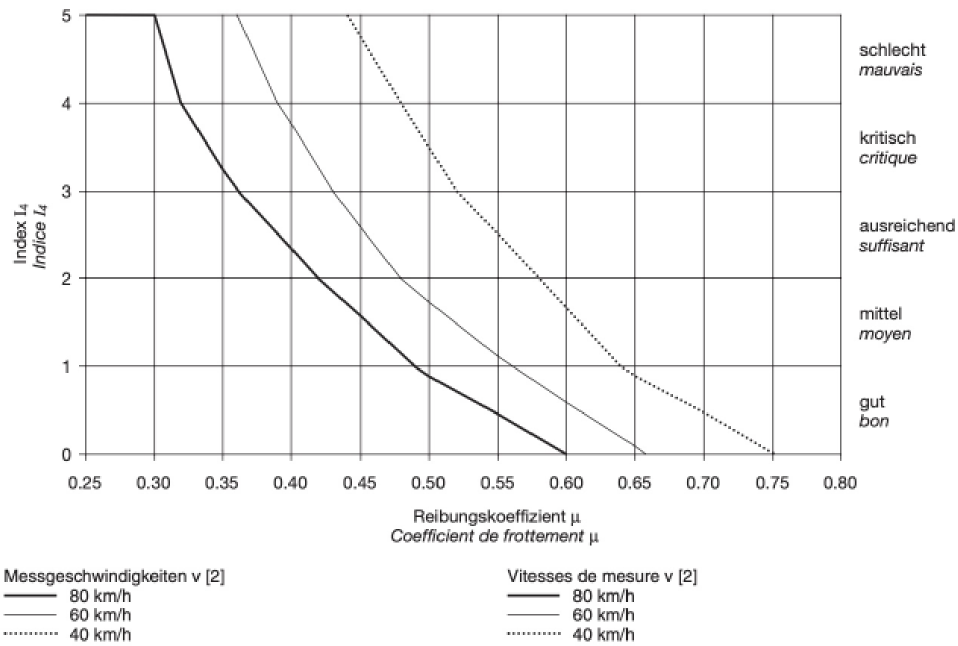

Figure 10: Graph illustrating the conversion from the friction coefficient into the Indicator I4

In practice, companies use devices mounted on vehicles that continuously measure friction while driving at a constant speed normally fixed at 80 km/h.

4 Methods

4.1 MISTRA Trassee database

The Federal Road Office (FEDRO) has two separated units considering road infrastructure named division West and division East. The division West include the *filialen* Estavayer-le-Lac (1) and Thun (2) and the division East include the *filialen* Zofingen (3), Winterthur (4) and Bellinzona (5). Each *filiale* has the responsibility of ensuring the measurement of the road condition indicators in his respective allocated area as shown in Figure 11.

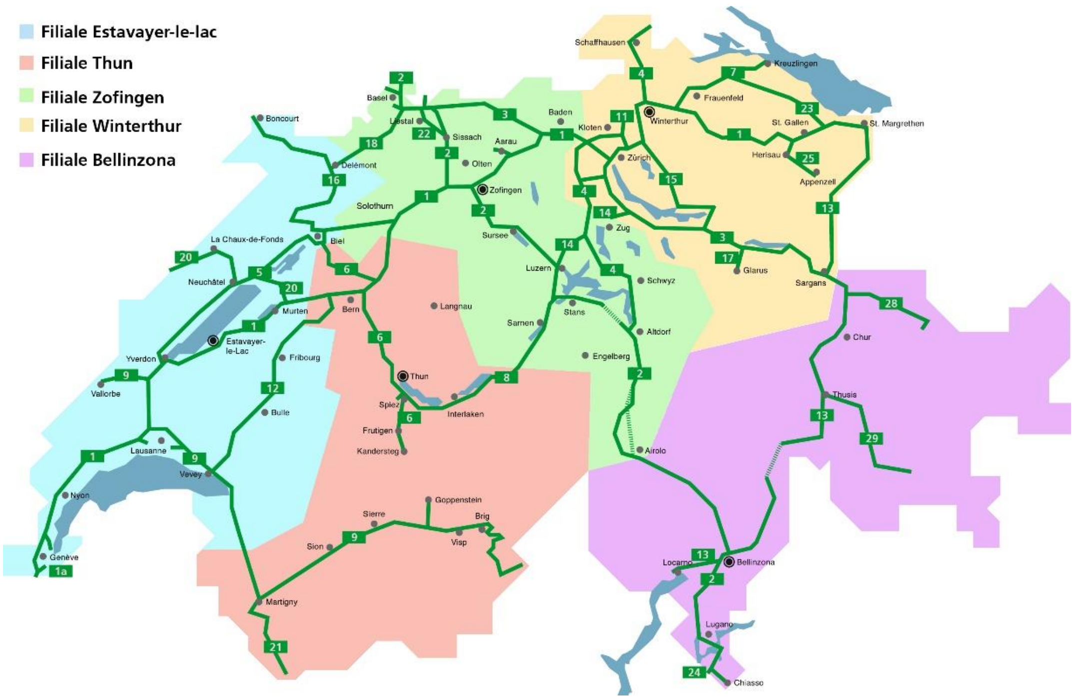

Figure 11: Separation of the national roads into the 5 *filialen*

Once the data is produced, meaning measured by the mandated companies, it is stored into the application TRA-Trassee. On FEDRO web site the following information about the application is mentioned: the business application Trassee manages data related to the structure, geometry, use, and condition of the roadway for the entire national road network. This information can be visualized on the map, on the longitudinal axis, and on the cross-sectional profile and can be the subject of statistical analyses.

In the context of asset management, this information is used to create maintenance objects (locations where interventions are needed), define maintenance measures (types of work to be carried out), and plan the implementation schedule (dates of work to be performed).

The data pass through a verification stage lead by Mr. Marc Delaby from the company Nibuxs and Mr. Jürg Bodemann from the company VICO, before entering the system. According to the interviews performed with them, there is no direct and systematic use of the data, so the information on the website was considered as a declaration of intentions more than a current reality. It was also mentioned that the time and human resources destined to verify the quality of the data is insufficient considering the huge amount of information.

The measurements on the national roads are generally performed by two companies: Schniering GmbH, now acquired by TÜV Rheinland and Infralab SA. Both companies rely mostly on advance continuous high speed measurement techniques. An intermediate step before reaching the Trassee platform is required where a Spatial adjustment is done. The company Geologix AG is involved. Figure 12 summarizes the data acquisition process.

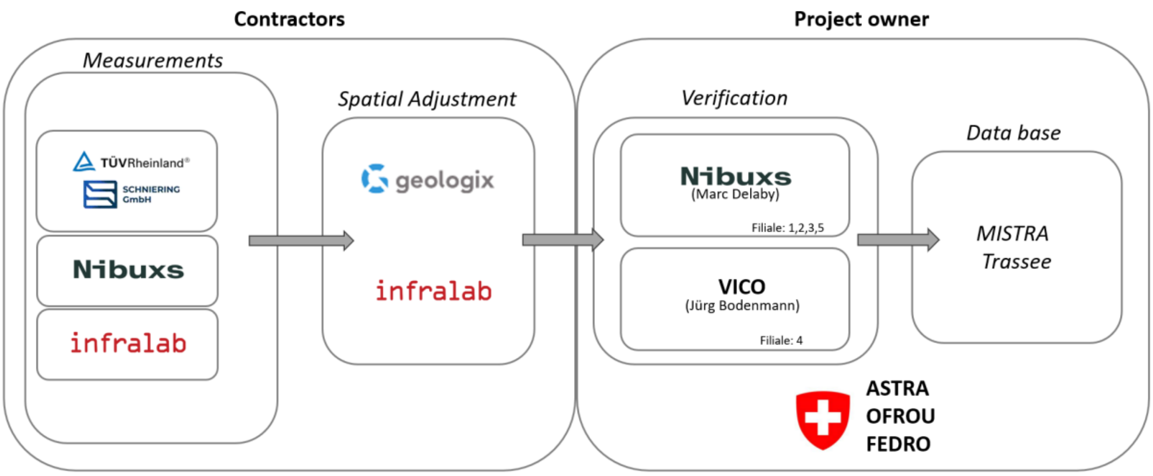

Figure 12: Data acquisition process including companies involved. (note that Nibuxs do measurements in cantonal streets but not in highways)

4.2 Visualisation process

When analysing large datasets, the initial visualization is a crucial first step in understanding the structure, distribution, and potential patterns within the data. Before applying statistical models or machine learning algorithms, visual exploration allows analysts to detect inconsistencies, outliers, and missing values that could impact the results. By leveraging the right visualization techniques, data scientists can uncover hidden insights, detect potential issues early on, and ensure that the dataset is well-prepared for more advanced analytical methods. Effective visualization not only enhances understanding but also provides a clear and interpretable starting point for stakeholders who may not have a technical background.

In this study, it was decided to create visualizations of the indicators data showing one of the lanes of the highway in the complete extension and showing the four indicators in parallel. It was also included the data measured in the different time steps to observe clearly how the indicators are evolving in time. Finally, it was also indicated when a major intervention has been performed. This way of presenting the data allows for a rapid understanding of the condition of the asset, and the evolution in time as well as the consistency among the four indicators.

4.3 Factor importance

The following figures presents the research framework that outlines the methodological approach employed in this study. The framework is structured into four distinct stages, each contributing to the analysis of pavement condition data variability using advanced modelling techniques. The first stage (Figure 13) , data collection and preparation, involves gathering information from two primary sources. Measurement data is obtained from FEDRO, which includes critical parameters such as grade value, measuring speed, intervention details, and organizational information. Complementing this, climate data is sourced from Meteoschweiz, providing temperature, wind speed, and relative humidity measurements. This raw data undergoes a series of preprocessing steps to ensure its quality and relevance. These steps include data cleaning to remove inconsistencies, temporal pairing where each pair consists of measurements from two consecutive survey years for the same road section, feature engineering to create meaningful variables, outlier removal to enhance data reliability, normalization to standardize scales, and data partitioning for subsequent analysis. The outcome of this stage is a set of final variables, including grade value difference, measuring speed differences, organization difference, intervention status, comparison period, and various climate variable differences.

For weather data, several key variables were consider: air temperature, dew point temperature, and relative humidity. Because the obtained road pavement data is categorized into five different *filialen*, specifically: *filiale* Estavayer-le-Lac, *filiale* Thun, *filiale* Zofingen, *filiale* Winterthur and *filiale* Bellinzona. The data for each *filiale* individually wased first process, as it is needed to account for the influence of weather conditions on the measurements at the time of data collection.

A total of 20 METAR weather datasets from 20 different stations were obtained, from which the weather stations with the most comprehensive historical data for each *filiale* was selected. The corresponding weather stations for each *filiale* are shown below:

Filiale	Weather Station
F1	LSGG (Geneva)
F2	LSGS (Sion)
F3	LSZG (Grenchen)
F4	LSZH (Zurich)
F5	LSZA (Lugano)

It is important to note, however, that the obtained weather datasets do not include any records earlier than 2002. Since the road pavement measures began in 2000, this means all weather data for the years 2000 and 2001 is missed. This gap in data may affect the analysis of the early years, as the influence of weather conditions on the pavement measurements during this period cannot be assessed, so there is no option than to give up these two years.

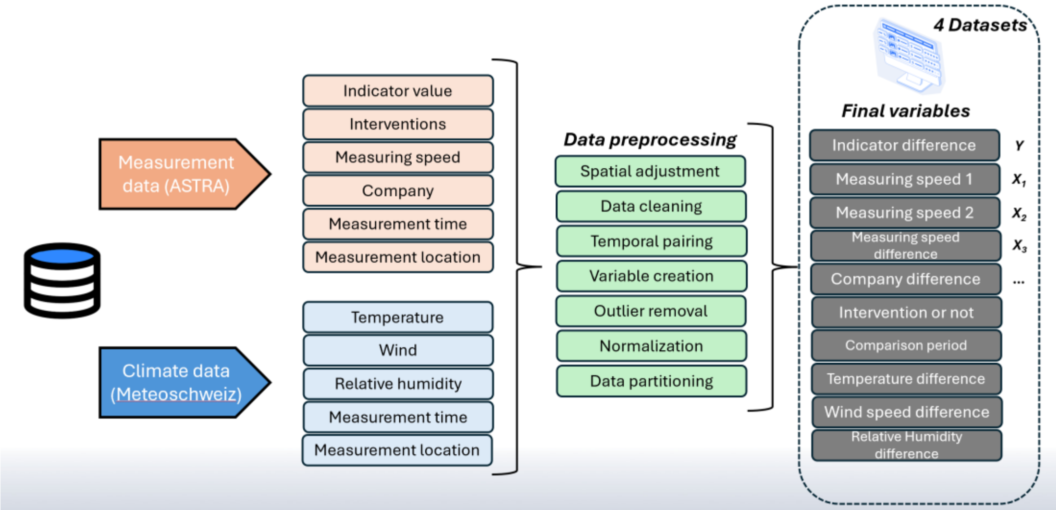

Figure 13: Data collection and preparation

At this stage it was used simple yet powerful charts such as histograms, bar graphs, and scatter plots that provide an immediate sense of the data's shape and spread. For example, a histogram can reveal whether a variable follows a normal distribution or is skewed, while a box plot helps detect extreme values that may need further investigation. Another essential aspect is exploring relationships between variables. Scatter plots can reveal correlations between two continuous variables, while pair plots or correlation matrices provide a broader view of dependencies across multiple features. Understanding these relationships is vital for feature selection and engineering, as it helps in identifying redundant or highly correlated variables that may influence model performance.

A spatial adjustment needs to be done in order to have consistency between data obtained in different years. As it can be seen in Figure 14, the stretches measured during for example the year 2000 (in yellow) will rarely corresponds exactly to the stretches measured during the following years (2009 in red). There are some physical marks on the roads that can be used but the measured stretches need to be adjusted. To overcome this problem, several approaches were used like using an average of 1 km for the analysis, using a prioritization of the longer stretch, averaging when 2 similar stretches or averaging with a distance weighted to length. This process will introduce some variability that need to be considered when analysing the evolution of the indicator in time.

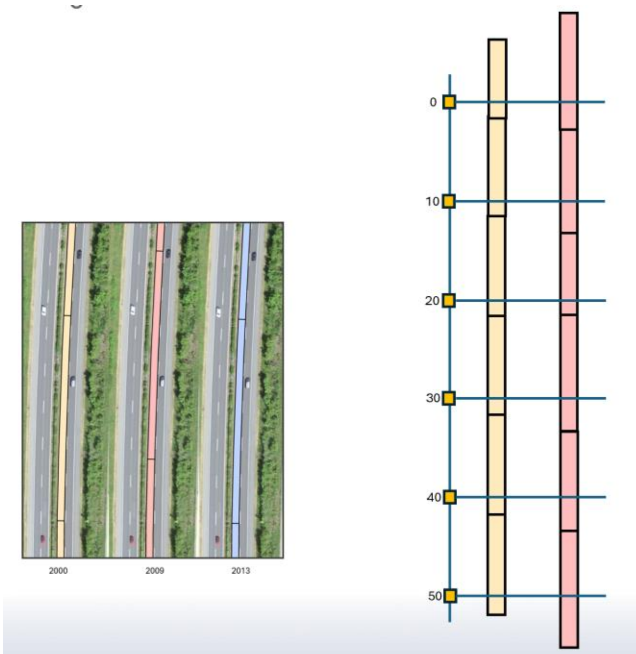

Figure 14: Spatial adjustment

The second stage focuses on the optimization of ensemble learning algorithms using Bayesian Optimization (BO). Five sophisticated ensemble algorithms are employed: XGBoost, AdaBoost, Random Forest, LightGBM, and CatBoost. Each of these algorithms brings unique strengths to the analysis, combining multiple models in different ways to capture complex patterns and relationships within the data. The BO process begins with the initialization of hyperparameters for each algorithm. A surrogate model is then constructed to approximate the relationship between hyperparameters and model performance. The optimization process iteratively selects new hyperparameter configurations based on an acquisition function, evaluates model performance on a validation dataset, and updates the surrogate model. This process continues until a stop criterion is met, resulting in a set of optimal hyperparameters for each algorithm. The BO approach aims to efficiently explore the hyperparameter space, balancing exploration and exploitation to find the best possible configuration for each ensemble model.

In the following the models are briefly presented, more details can be found in the references.

4.4 Predictive Models

To analyse the factors contributing to the variability in road surface conditions, a range of advanced machine learning algorithms was utilized. The following models were developed and evaluated (Winkler et al. 2018; Taiwo et al. 2024):

- **AdaBoost**: An ensemble learning method that combines multiple weak learners to create a strong predictor. It's particularly effective at reducing bias and variance in the model. For regression, the model can be expressed by Equation 1 (Aslam et al. 2022):

$$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \quad (1)$$

where $F(x)$ represents the final prediction, $h_t(x)$ are the weak learners, and α_t denotes their corresponding weights.

- **CatBoost**: A gradient boosting framework that incorporates ordered boosting and an innovative approach to handle categorical features. Its objective function is defined by Equation 2 (Prokhorenkova et al. 2018):

$$L(y, F) = \sum_{i=1}^{n} l\left(y_{i,} F(x_i)\right) + \Omega(\mathrm{F}) \quad (2)$$

where l is the loss function, F is the model, and Ω represents the regularization term.

- **LightGBM**: This gradient boosting framework employs histogram-based algorithms for efficient training (Chen et al. 2019; Ma et al. 2023). The tree learning process can be summarized by Equation 3:

$$Leaf_{split} = \max(\sum_{i \in I_L} g_i, \;\; \sum_{i \in I_R} g_i) \; + \; \lambda \quad (3)$$

where g_i are the gradients, I_L and I_R are the instance sets of the left and right nodes, respectively, and λ is the regularization parameter.

- **Random Forest (RF)**: RF is another ensemble of decision trees that aggregates the predictions from multiple trees. The regression prediction is given by Equation 4:

$$f(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x) \quad (4)$$

where B is the total number of trees, and $f_b(x)$ represents the prediction of the $b-th$ tree.

- **XGBoost**: This is an optimized gradient boosting algorithm designed for distributed computing. The objective function for XGBoost is expressed by Equation 5:

$$Obj\ (\theta) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \ l(\mathbf{y}_\mathrm{i}, \hat{\boldsymbol{y}}_\iota) + \sum_{k=1}^{K} \Omega(\mathrm{f}_k) \quad (5)$$

where l is the loss function, $\hat{\boldsymbol{y}}_i$ is the predicted value, and $\Omega(f_k)$ measures the complexity of the $k-th$ tree.

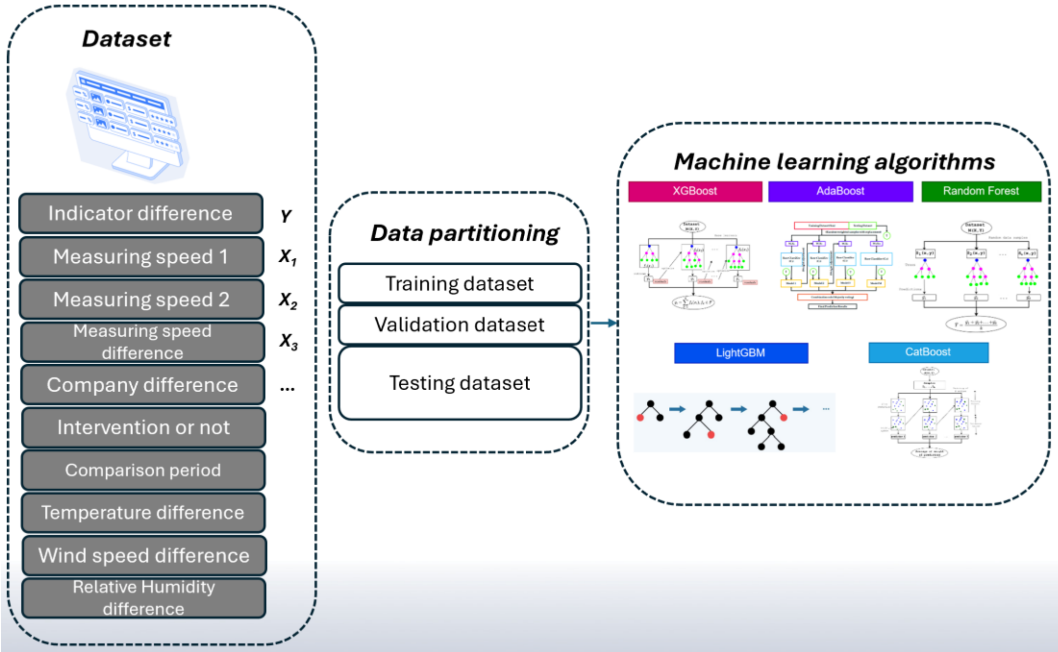

Figure 15: Ensemble learning algorithms

In the third stage, the performance of each ensemble model is evaluated using two key metrics: the coefficient of determination (R^2) and the root mean square error (RMSE). R^2 provides insight into the proportion of variance in the dependent variable that is predictable from the independent variables, while RMSE offers a measure of the differences between predicted and observed values. The model selection process involves choosing the algorithm that demonstrates the highest R^2 value alongside the lowest RMSE, indicating the best balance between explanatory power and prediction accuracy.

The final stage of the framework addresses model and variable explainability, utilizing Shapley additive explanations (SHAP) or permutation techniques. These advanced techniques allow for a nuanced interpretation of the selected model's decision-making process. The analysis of marginal contribution examines how each input variable influences the model's predictions, providing insights into the relative importance of different factors in determining pavement condition variability. Furthermore, the impact analysis offers a detailed visualization that simultaneously captures the distribution, density, and range of SHAP values for each feature. This comprehensive representation enables the identification of features with high overall impact, those with significant variability in their effects, and the detection of potential outliers or asymmetries in feature contributions. Thus, the analysis provides a rich, multidimensional understanding of how different factors contribute to measurement variability across various conditions and data points.

The SHAP method is grounded in cooperative game theory, provides a comprehensive approach to explaining the output of any machine learning model by calculating Shapley values (Fryer et al. 2021; Lundberg et al. 2017). These values quantify the importance of each feature in contributing to a specific prediction. The Shapley value for a given feature represents its average marginal contribution across all possible predictions. The SHAP method follows an additive feature attribution model, where the explanation of a prediction for an instance x is represented by Equation 6:

$$g(z') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i z_i' \qquad (6)$$

Here, $g(z')$ is the explanation model, $z' \in \{0,1\}^{\mathrm{M}}$, M is the total number of input features, and ϕ_0 represents the base value, or the average model output over the entire training dataset. The value ϕ_i denotes the Shapley value for the feature i, indicating its contribution to the prediction for instance x.

Several critical properties of SHAP make it a robust tool for model interpretation. It ensures local accuracy, meaning the sum of all feature attributions equals the model's output for any given instance. It also handles missing data effectively by assigning zero SHAP values to features that do not influence the model's prediction. Additionally, SHAP maintains consistency, meaning that if a model is altered so that a feature has a larger impact on the prediction, the feature's SHAP value will not decrease (Lundberg et al. 2017). For tree-based models, like the ensemble models in this study, SHAP uses an efficient polynomial-time algorithm that accounts for feature interactions. The Shapley value for a feature i is computed using Equation 7:

$$\emptyset_i = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{i\}) - f_x(S)] \qquad (7)$$

where N is the set of all features, S is a subset of features, and f_x is the model's prediction function. This formulation allows for the precise determination of each feature's contribution by considering all possible feature combinations. In this study, SHAP values were calculated for each feature across all instances in the dataset, using the selected model. This analysis provides a detailed understanding of how each feature influences the model's predictions of the explained variable.

5 Visualisation

The first step in order to understand the sources of variability was an exploration phase where the four indicators were observed together trying to extract meaningful information to explain the condition of the pavement. Also, this allowed checking consistency among indicators. For that it was decided to study two highways in each *filiale* since it was known that each *filiale* organize themselves separately. The selection of the particular highways in each *filiale* were random. Figure 16 shows the locations of the selected highways.

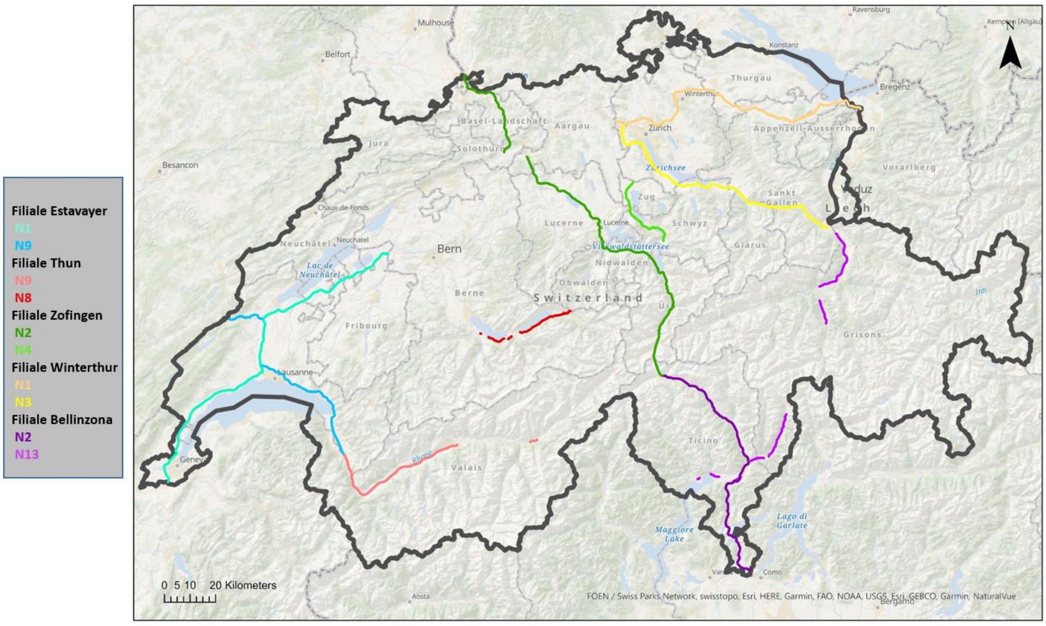

Figure 16: Selected national roads for the initial visualisation phase (all-indicators assessment).

For each highway it was used the information about the construction date (*Einbaudatum*) which indicate the last time the road experienced a major intervention. It was then split the selected highway lane into different sections according to the last time there was a major intervention (see Figure 17). The larger sections available for each road-age were then found and plotted together the four pavement condition indicators.

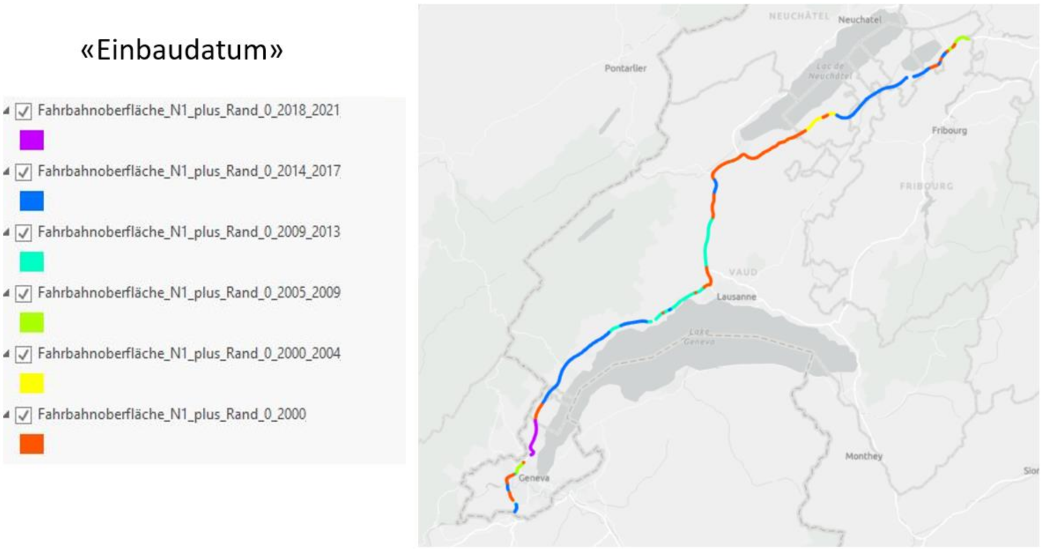

Figure 17: Different section according to their age (*Einbaudatum*) in the N1 highway (*Filiale* Estavayer).

For each highway only one lane was selected to perform a visualisation. A comparison check-up was done and it was found that the data is in general fairly similar in between lanes and traffic direction. This depends of course on interventions done on one or all the lanes at the same time but from the data it appears that in general all lanes are intervened at the same time or across shorter periods of time. Information about the construction date (*Einbaudatum*) was used which indicates the last time a road experienced a mayor intervention. The selected highway lane was then split into different sections according to the last time there was a major intervention (see Figure 18 for the N1 national road on the *Filiale* 1 stretch). It was then found the larger sections available for each road-age and plotted together the four pavement condition indicators.

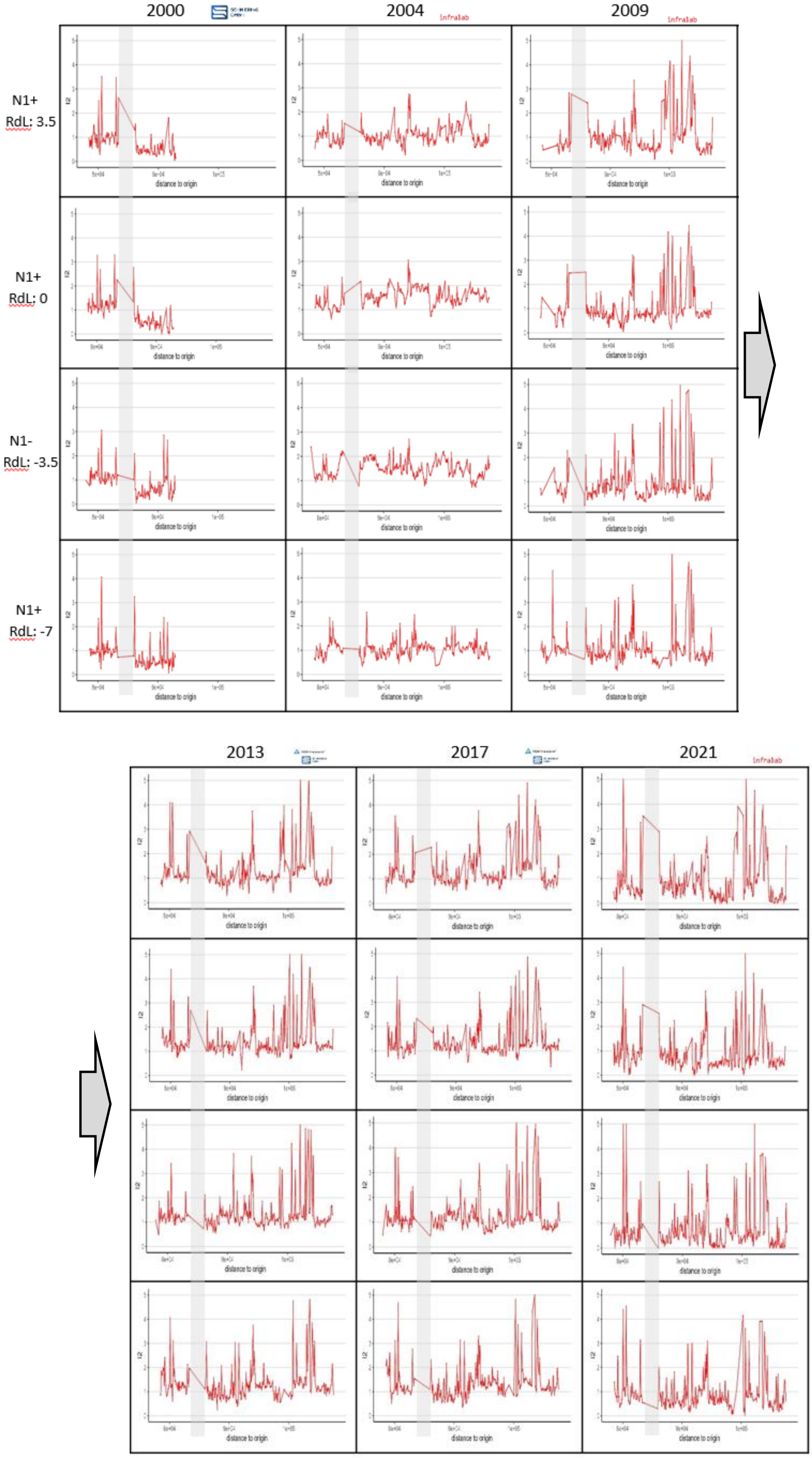

Figure 18: Comparison between different lanes on the Indicator I2 for the National Road N1 on the *filiale* 1. (on the small graphs, horizontal axis shows road distance to origin and vertical axis show the indicator grade, from 0 on the bottom to 5 on the top)

The indicator dataset is built from data recorded since the year 2000. It is expected that measurements will take place every 4 years but the spread in time can variate form one road to another, from indicator to another, and from one *Filiale* to another. It was found data from the following years: 2000, 2001, 2004, 2009-2010, 2013 2017, 2018, 2021, and 2022. New measurements are organized to take place between 2023 and 2027. Important is to mention that for each measurement campaign different companies can function as contractors. The two companies, Schniering/TÜV Rheinland and Infralab are the ones that have been selected so far. Table 1 provides details about in which year each company was in charge of the measurements, also split by *filiale*.

Measurement Companies

	2000-2001*	2004*	2009-2010*	2013	2017-2018-2022	2021	2023-2027
Filiale 1	SchT	Inf	Inf	SchT	SchT	Inf	-
Filiale 2	SchT	Inf	Inf	SchT	SchT	-	SchT
Filiale 3	SchT	Inf	Inf	SchT	SchT	-	SchT
Filiale 4	SchT	Inf	Inf	SchT	SchT	-	SchT
Filiale 5	SchT	Inf	Inf	SchT	SchT	Inf	-

Table 1: Companies in charge of the measurements. Inf is Infralab and SchT is Schniering/TUVRheinland. *Organized by FEDRO central office.

In the next sub-sections, the visualisations of the Indicators I1, I2, I3 and I4 are discussed. These values were measured during years 2000, 2004, 2009, 2013, 2017 and 2021 for different stretches corresponding to each *Filiale*. Figures show the indicator grades (from 0 to 5) along the road distance, as well as the major interventions conducted during the analysed period (in blue colour).

5.1 Filiale 1: Estavayer-le-lac

The *filiale* Estavayer-le-lac is located in Western Switzerland as seen in Figure 16. Highways N1 and N9 were chosen for visualization.

5.1.1 Highway N1

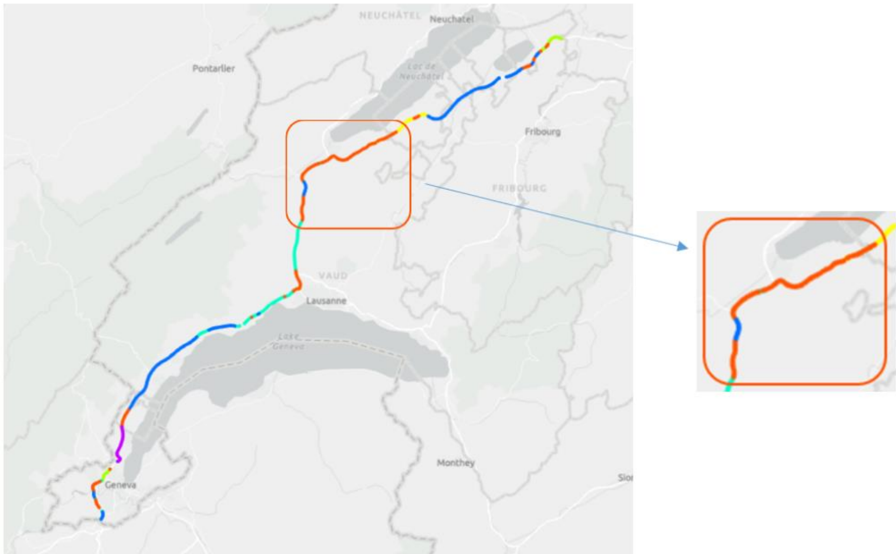

Figure 19: Detail of the N1 stretch with no major interventions after year 2000

Results in Figure 20 show that for the indicator I0, from 2004 to 2009 expected deterioration is observed, however from 2009 to 2013 a clear amelioration on the pavement condition is observed which is not expected. From 2013 to 2017, again expected deterioration we observe, but from 2017 to 2021 we observe a deterioration that is more than expected and more consistent with the values measured in 2004 and 2009.

For the indicator I2, first, we observe unreliable measurements during the year 2000 with big difference in between the same segment. Then on the year 2004, we observe more consistent measurement with values oscillating between 1 and 2. The measurements on the year 2009 are not expected since we observe an amelioration of the condition in the general tendency of the data. However, we also observe a market deterioration in a specific part of the road marked by a high quantity of peak values many of them reaching values of three to four. However, those peaks are somehow suspicious since we should expect a consistent deterioration through the road so they could constitute measurement problems. On the data obtained during the year 2013 we observe consistency in the measurements when compared to the ones made in 2009 and also an expected increase in the overall value indicating the expected deterioration in time. On the data measured during 2017 we observe consistency on the patterns as compared with 2013 but the normal deterioration is not observable in the data which is surprisingly stable when compared to the previous measurements during 2013. Finally, during the year 2021, we observe an amelioration of the condition which does not make sense at all.

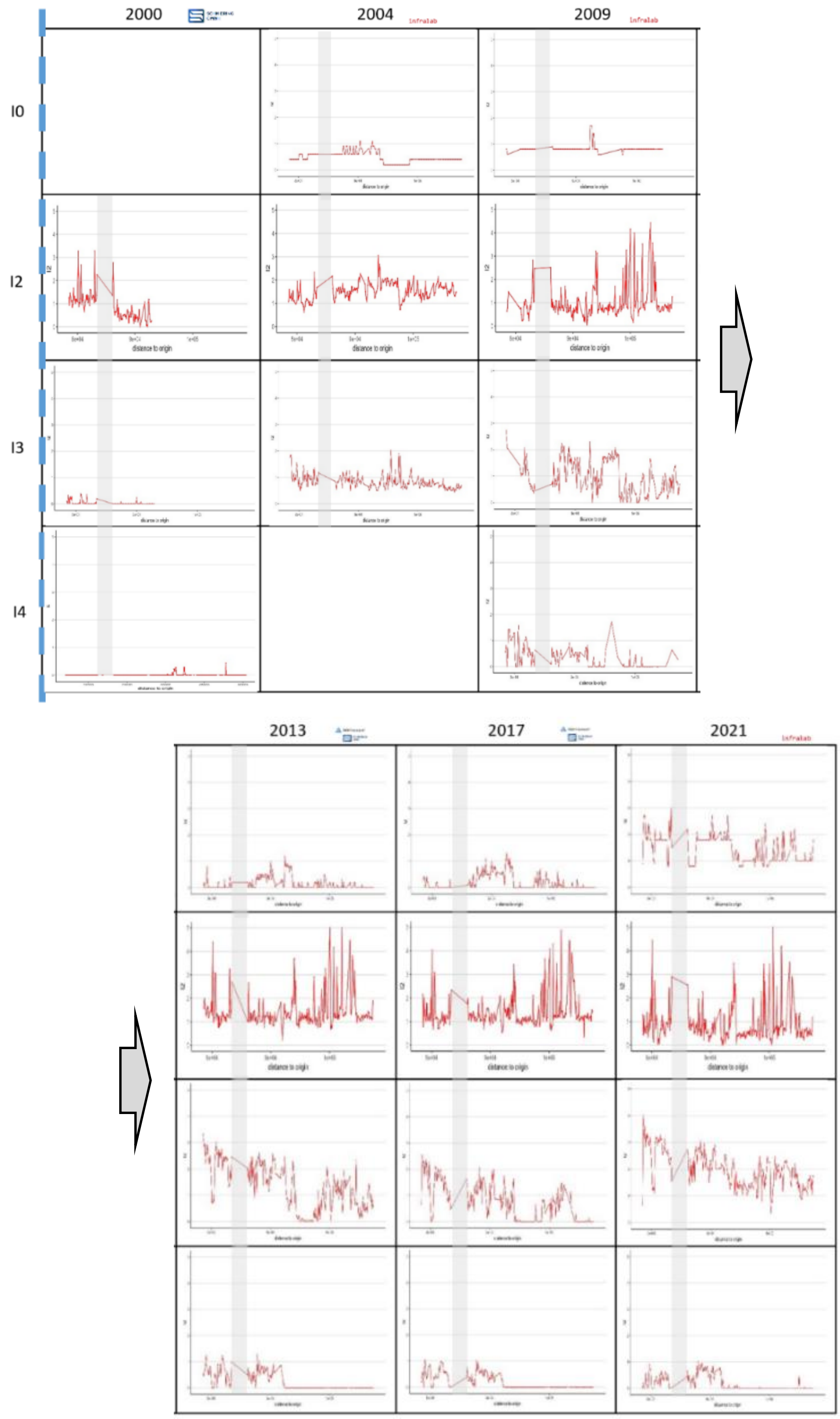

Figure 20: Indicators measured in the N1 stretch with no major interventions after the year 2000

For the indicator I3, we observe expected low values close to zero measured during the year 2000, despite that the measurement is incomplete for the whole segment. On the measurement done during 2004, we observe a deterioration into values slightly under 1 that is expected. Also, we observe a consistency on the data with values more or less similar with the exception of a few peaks that are not very high. However, on the data measured during year 2009, we observe a bigger variability on the data, where some parts of stretch show an improvement and other parts show a decrease in the condition. On the data of the year 2013, we observe an expected deterioration of the general measurements, except in a small part in the middle of the stretch where the condition seems to ameliorate bluntly. Measurements done during 2017 show a consistent pattern, but in general the condition is improving which make no sense without any intervention. Finally, the data obtained during 2021 show as expected high values that show the expected deterioration.

The indicator I4 measured during the year 2000 is as expected low then there is a gap in year 2004. On the year 2009, we have the following measurement, and we observe a deterioration that seems normal with values under 1 most of the time. We observe however that values on the second half of the stretch are somehow lower except a few pics. This tendency of having higher values in the first part of the stretch and lower values after will be maintained for the following measurements on year 2013, 2017 and 2021. On those years, it is also interesting to notice that we observe no deterioration at all which is not what is expected.

Another interesting aspect to notice is that we can identify the companies that made the measurements thanks to the label project on the dataset. So, we know that the company Schniering / TÜV Rheinland did the measurements on the year 2000, 2013 and 2017 and that Infralab did the measurements on the years 2004, 2009 and 2021. In consideration of that, we can observe that for the indicator I0 there is a clear difference on how the two companies assess the indicator since we observe important differences in between, with Schniering / TÜV Rheinland measuring lower values and with more spatial variability (values show more change for neighbouring units of measure) compared to the measurements done by Infralab where more consistency in neighbouring unit of measure are observed). This comparison between companies only reflects what can be seen in the data, but we have no way of knowing which of the two companies is more accurate. Then it is interesting to notice also that even when the company do the measurement in consecutive years, one indicator can show an amelioration of the condition and another can show normal deterioration in the condition. We observe this for example in between the years 2004 and 2009 where Infralab did the measurements and we see that I0 and I3 show deterioration but I2 shows amelioration. In between 2013 and 2017, we observe a similar phenomenon for measurements done Schniering / TÜV Rheinland where Indicator I2 and I4 show an unexpected status quo in the measurements, but I0 show a slight deterioration and I3 show a clear amelioration.

All these results are important to consider, since despite that each Indicator varies in his own way, they should all show a general tendency to a deterioration since no major intervention occurred.

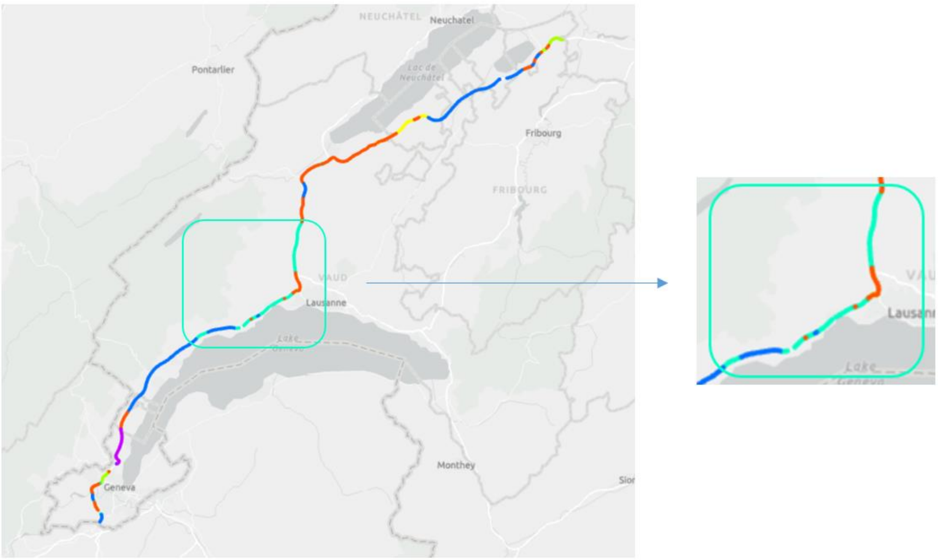

Figure 21: Detail of the N1 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 22 show that for the indicator I0, an expected deterioration can be seen between the years 2004 and 2009, after the intervention we observe that the values measured during 2013 are as expected very close to the value 0. Measurements done during 2017 show a slight deterioration which is expected and finally during the year 2021 we observe a bigger deterioration despite that in some parts of the stretch there are some unexpected improvements.

When checking the indicator I2, we observe that the beginning of the measurements in 2000 are already at values over 1 most of the time. Surprisingly we observe that during the year 2004 measurements, values seem to be maintained and even a very slight decrease can be observed at least in the first two parts of the stretch. On the third part of the stretch, we see what is expected so a deterioration in the indicator. Now checking at the data recorded during 2009, we observe that the first part of the stretch is maintained or slightly deteriorate but the rest of the stretch seems to stay stable or even ameliorate which is not expected of course. Checking at values measured in 2013, after the intervention, it was observed that the measurements are still at similar values than before the intervention which makes no sense at all. After that values measured during 2017 are maintained which could be possible if the road was recently renovated. However, values measured during 2021 show a marked amelioration of the condition which is not expected.

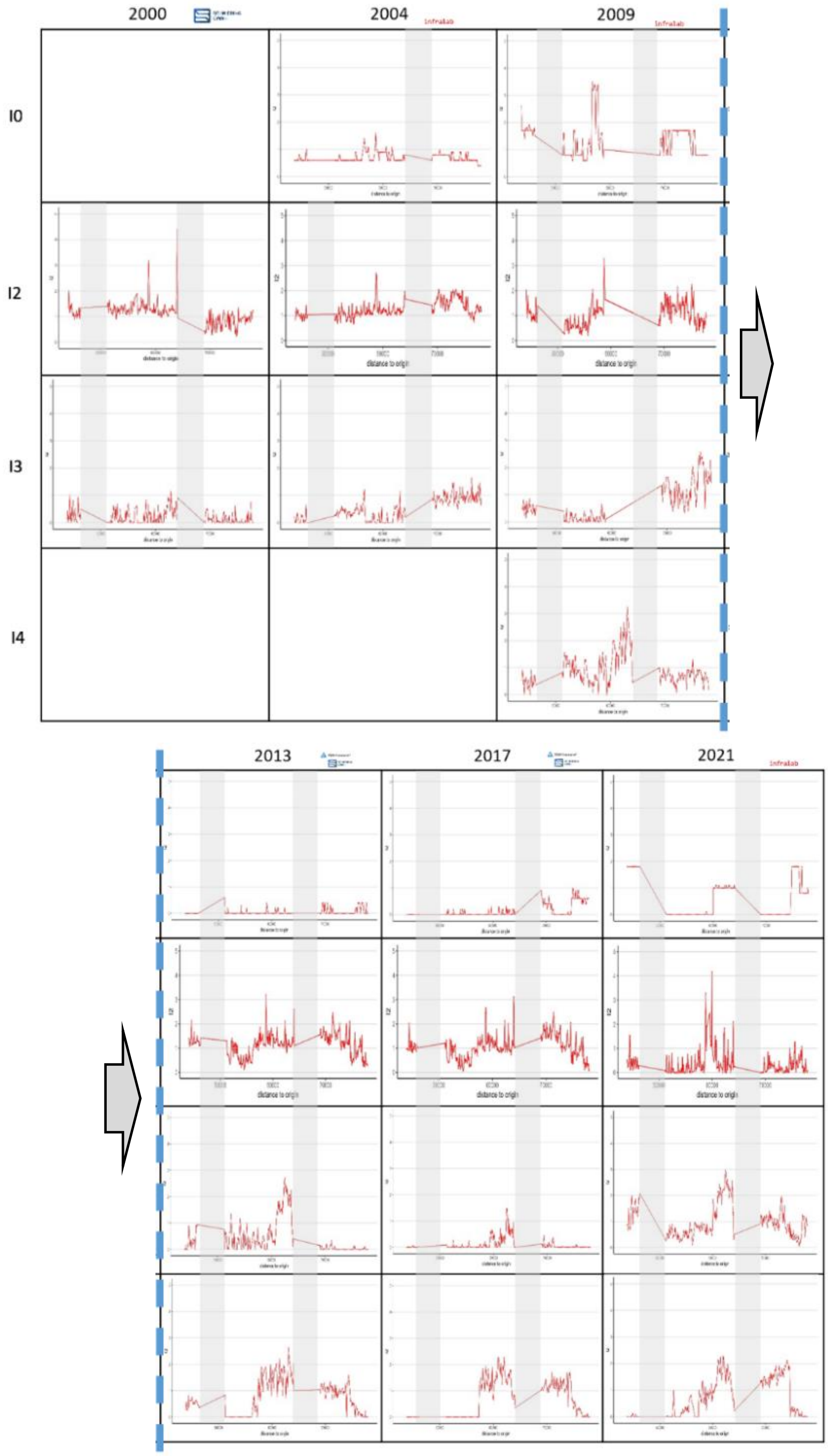

Figure 22: Indicators measured in the N1 stretch with a major intervention in between the years 2009 and 2013

For the case of the Indicator I3, we observe also different behaviour in between different parts of the stretch during the years 2000, 2004 and 2009. The third part of the stretch is the one that make more sense showing a continuous deterioration as expected. The other parts sometimes improve and sometime deteriorate which does not make real sense. It is interesting to notice that after the intervention, the mentioned third part of the stretch show expected results, ameliorating to values close to zero in year 2013 and increasing as expected until the year 2021. For the first and second part of the stretch it is more difficult to find logic in the data cause on the year 2013 we observe a deterioration despite the intervention and only for the measurements of the year 2017 we can see the improvement in the condition. From the year 2017 to 2021 the deterioration measured is consistent to what is expected.

For the case of the Indicator I4, where the measurement only started on 2009, we cannot really observe the effect of the intervention except for a small part of the stretch. The values measured after the intervention only deteriorate for that same small part of the stretch but are maintained more or less constant for the rest which is not really expected. When analysing all Indicators together, we can see that there is no consistency on what we would expect to see. I0 and some small parts of the stretch for indicator I3 and I4 are consistent with an intervention in between the years 2009 and 2013 but the indicator I2 and most of the stretch for indicators I3 and I4 are not showing the effects of the intervention. Data measured between the different companies show some consistency this time, except perhaps for the case of the I2 where an unexpected amelioration is measured between 2017 and 2021.

Figure 23: Detail of the N1 stretch with a major intervention in between the years 2014 and 2017

Results on Figure 24 show that for the indicator I0, stable measurement are perceived which is not expected. From 2004 to 2013 even a slight amelioration in some parts of the stretch are observed which is unexpected. More interesting is the fact that after the intervention in between the years 2013 and 2017, the indicator remains stable which is not expected. Only a slightly increase in the year 2021 is somehow expected.

Considering the indicator I2, measurements begin with fairly high values around 3 in the year 2000. Then during measurement done on the year 2004, we observe a slight amelioration which is not expected. On the year 2009, measurements stand stable instead of deteriorating. In the measurements of 2013, we can see an expected deterioration. After the intervention values ameliorate substantially as expected reach values around 0.5 or less indicating that the road is in excellent condition. On the measurements done during 2021, we observe a further improvement which is not expected. A slight deterioration should be perceivable at this point.

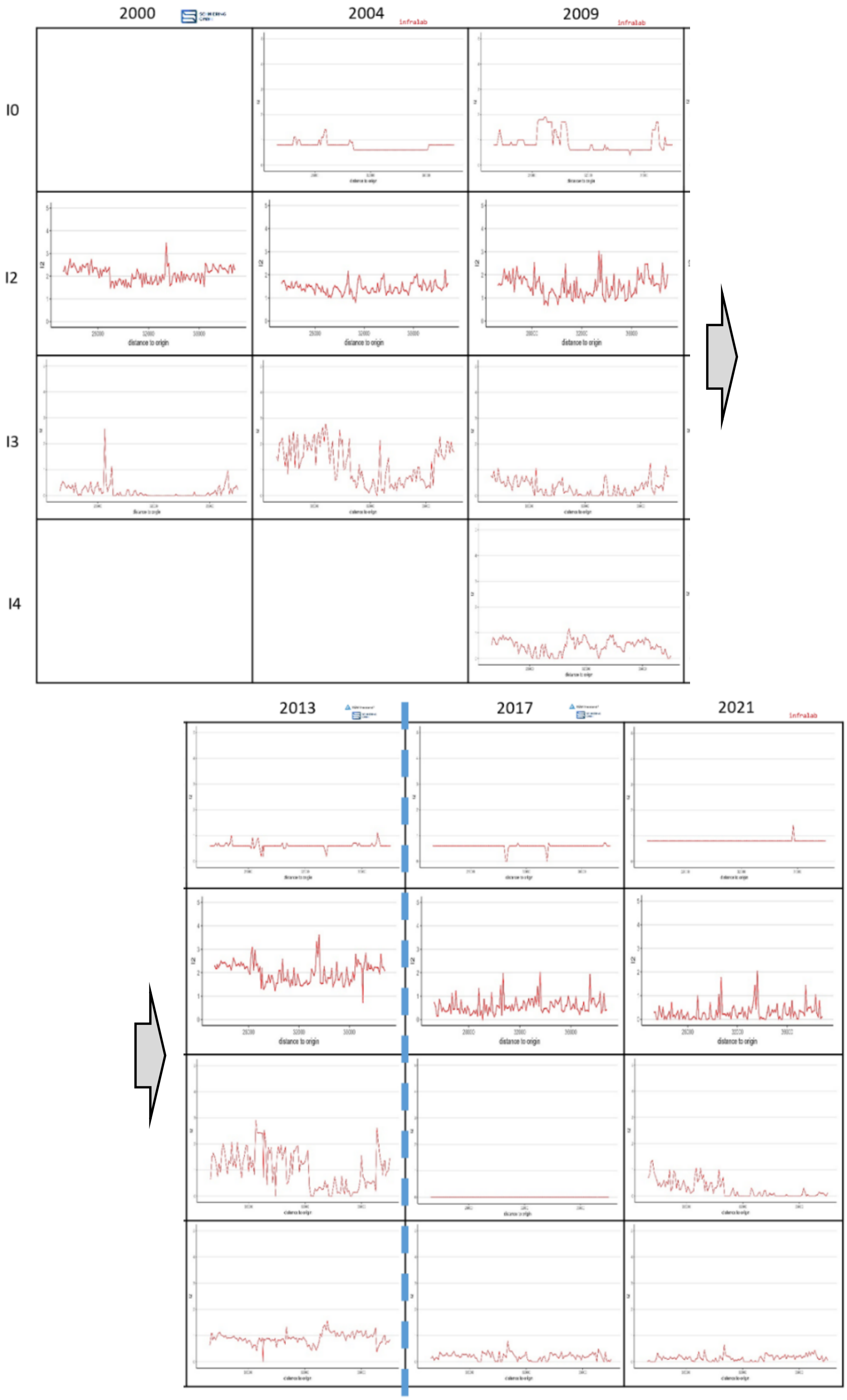

Figure 24: Indicators measured in the N1 stretch with a major intervention in between the years 2014 and 2017

Indicator I3 measurements are low in the beginning of the series in year 2000 mostly under 0.5. On the measurements of 2004 the indicator shows an important deterioration, reaching values around 3 in around half of the stretch. It is not expected to observe such a high increase in the values. Then on the year 2009, we observe an amelioration of the values, with most values under 1 which is unexpected without any intervention. On the year 2013, again we observe higher values, slightly under the ones measured in 2004 but very high for a normal increase since 2009. Then on the year 2017, we can see a clear influence of the intervention and the measured values are flat, which indicate a perfect condition of the road, which is somehow expected. On the measurement of the year 2021, we observe expected deterioration with values under 1.

For the case of the indicator I4, despite that we have no data for the years 2000 and 2004, we observe consistent measurements for the next years. Values are indicating a very good condition (under 1) on the year 2009, and a slight expected deterioration until year 2013 to values close to 1. After the intervention, measurements improve again to values close to 0 as expected. On the year 2021, we observe similar measurements, indicating that the excellent condition is maintained. When considering all indicators together, we observe that I2, I3 and I4 show clearly the influence of the intervention which is what we expect. For the case of the Indicator I0, it is not the case and results do not reflect the intervention which is strange.

5.1.2 Highway N9

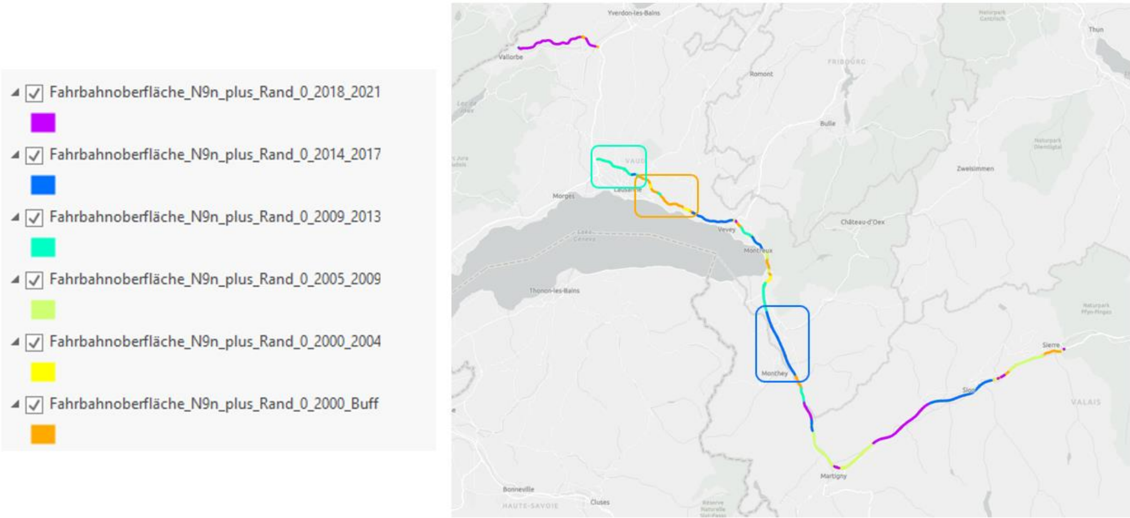

Figure 25: Different section according to their age (*Einbaudatum*) in the N9 highway (*Filiale* Estavayer)

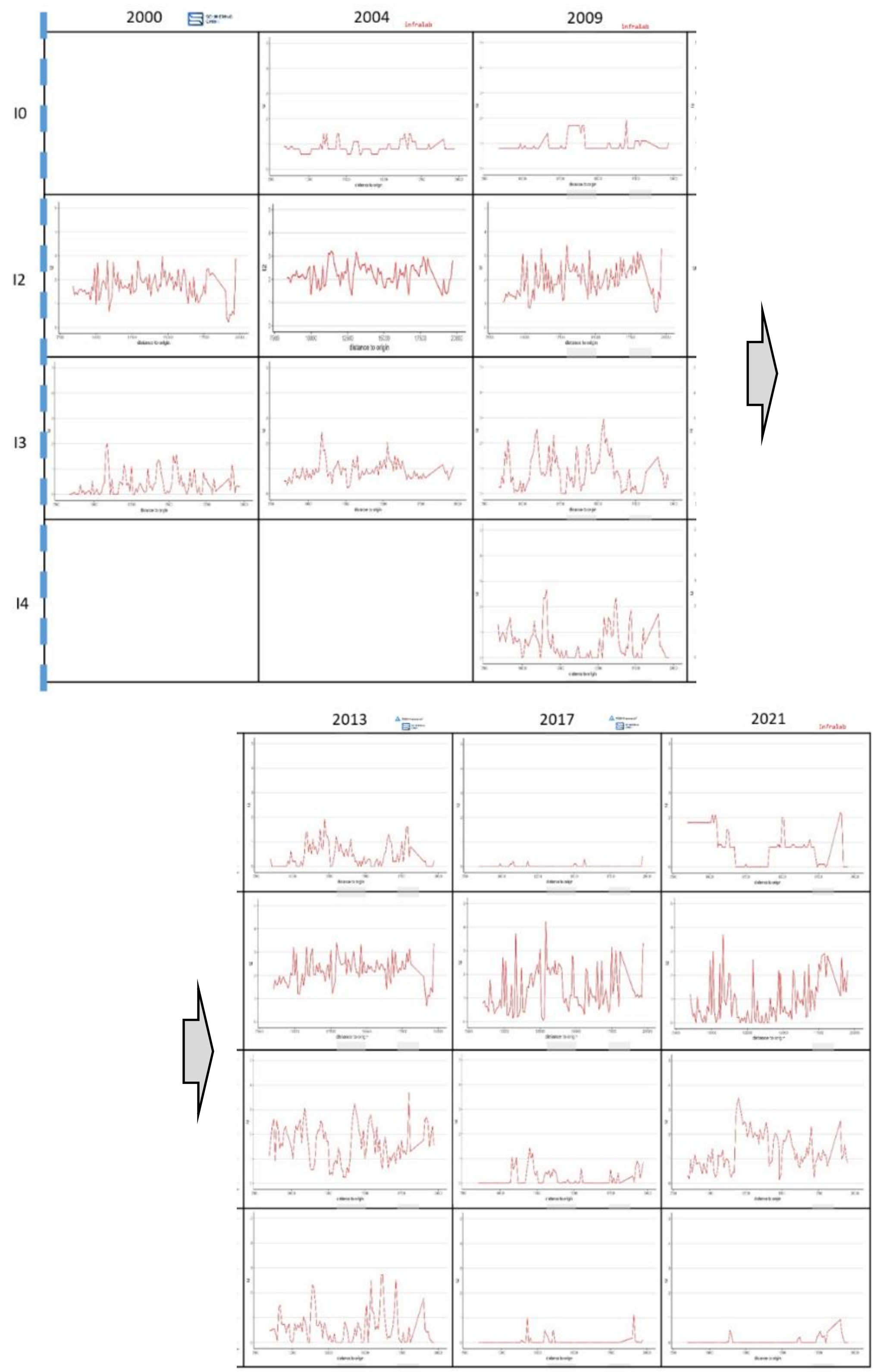

Figure 26: Indicators measured in the N9 stretch with no major interventions after the year 2000

Results on Figure 26 show that for indicator I0, measurement are stable during the years 2004 and 2009, but then an improvement in the condition is observed on the year 2013 which is not expected since there is no recorded intervention. Even more surprising is that we can see amelioration to values close to zero on the measurement of year 2017. In the measurements of the year 2021, we observe different results depending on the parts of the stretch: in some parts values are around 2, but in others around 1 and also some parts are close to value 0. This is not expected since more or less regularity should appear.

Considering the Indicator I2, we see that measurements from the year 2000 are around values of 2 but with variability ranging from 1 to 3. On the year 2004 we observe an expected decrease in the condition and thus a slight increase in the indicator. Values are more or less stable for the measurements done on 2009 and 2013 which is not expected as some deterioration should be observable. On the year 2017, despite big variability ranging from values over 3 to values close to 0,5, we observe an amelioration of the indicator which is not expected sine no intervention is reported. On the values measured during 2021, the road seems to have further improved which is not expected.

Checking Indicator I3, we observe lower values in the beginning of the series in 2000 indicating a very good condition. Afterwards we can see an expected slight deterioration on the year 2004, reaching average values close to 1. On the year 2009, further deterioration is noticed as expected but only in some parts of the stretch. Some parts seem to ameliorate which is not expected. Then, on year 2013, we observe a general deterioration which is consistent with the previous values despite that some areas seems to be in better conditions than others. On the year 2017, values show a clear amelioration of the indicator that can only be achieved by an intervention that is not reported. In the measurements form the year 2021, we observe a clear deterioration, somehow stronger than expected particularly in some parts of the stretch.

Considering indicator I4, measurements only start in the year 2009 where we observe lots of variability in the data, but a general good condition with values under 2. Those values stay stable in the year 2013 and only decrease substantially on the year 2017 were a clear amelioration to values close to 0 is observed. Values measured in 2021 are maintained very low, showing no deterioration. When observing all indicators together, we suspect that an intervention occurred in between 2013 and 2017, but that intervention was not reported on the road age data.

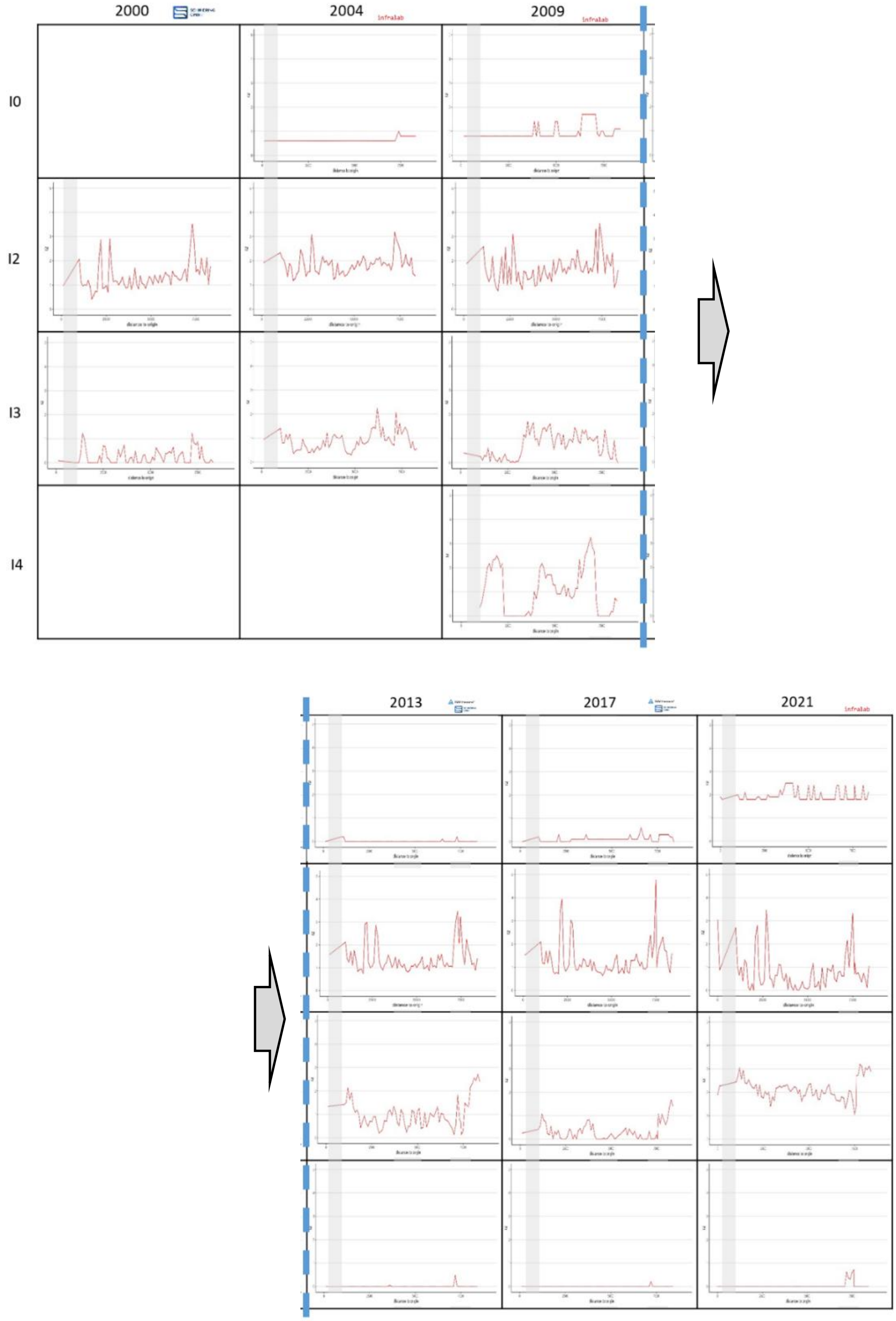

Figure 27: Indicators measured in the N9 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 27 show that for the indicator I0 measurements starts on the year 2004, we observe stable measurements indicating a good condition (values under 1). Regarding the measurements of 2009, we can see a slight increase in the values which is consistent with an expected slight deterioration. Then, in the measurements of 2013, we observe a clear effect of the intervention that is reported so values are close to 0 indicating excellent road conditions. It was observed that a slight deterioration until the year 2017, and a more important deterioration on the year 2021 reaching values over 2.

Considering the indicator I2, we observe that values begin in 2000 in values close to 1 with some pics and a tendency to higher values at the end of the stretch. In 2004, we observe an expected slight deterioration. It is interesting to notice that the values appear more similar along the stretch this time. Regarding the measurements of 2009, values are maintained and an even a very slight improvement is observed which is not expected. On the year 2013, measurements after the intervention show an amelioration which is expected; however, it is interesting to notice that it is only a slight decrease on the values, and in case of a major intervention, a more important effect is expected. In the measurements of 2017 values are maintained and a further amelioration is observable in the year 2021 which is not expected.

Considering indicator I3, we can see that in 2000 lower values close to 0.5 are visible indicating a very good road condition. Regarding the measurements of 2004, we observe a slight decrease which is expected indicating normal deterioration. However, on the year 2009, we observe improvement in parts of the stretch and slight deterioration in others. After the intervention, measurements on the year 2013 do not show the expected effect, and values are maintained or even show some deterioration in some parts of the stretch. Measurements of the year 2017 do show a clear improvement of the condition reaching values under 1 most of the time thus reacting to the intervention. On the year 2021, measurements show a big deterioration when compared to 2017, reaching values over 2 which can be expected considering the more than 14 years after the intervention but not the values measured in 2017.

When inspecting indicator I4, we can see that data is only available after 2009. On that year we observe variating values with some parts of the stretch showing excellent values close to 0 and other parts showing values of 1 and others 2 and even some peak reaching a value of 3. After the intervention we observe a clear improvement and measurements on the year 2013, 2017 and 2021 show values near 0, despite that some normal deterioration should be observable. When looking at all indicators together, we observe that Indicator I0 and I4 clearly show the intervention, but the effect is less perceivable on the other indicators which is unexpected.

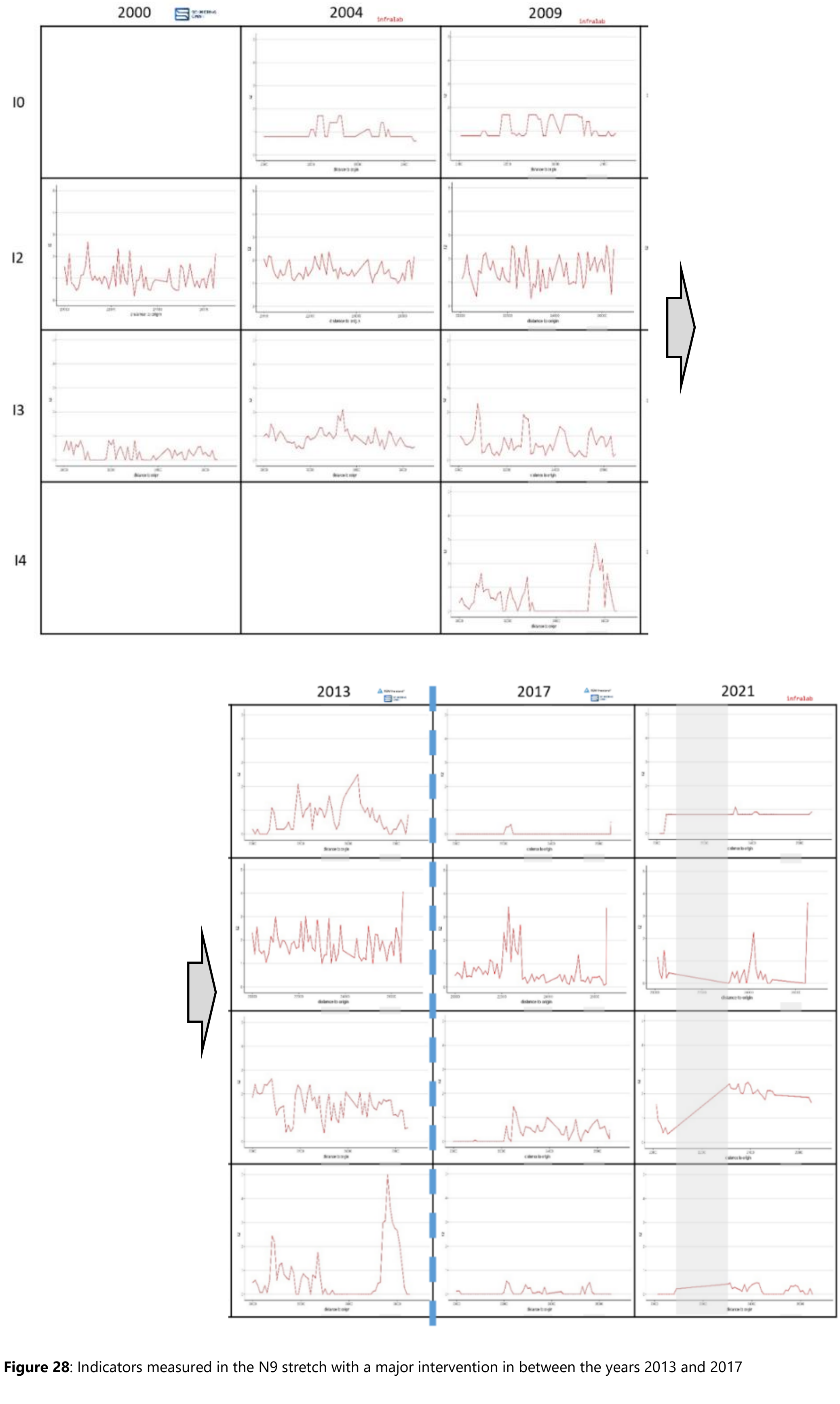

Figure 28: Indicators measured in the N9 stretch with a major intervention in between the years 2013 and 2017

Results on Figure 28 show that the indicator I0 begin in the year 2004, and we can observe good values close to 1 in average. In 2009, and expected slight deterioration can be observed so values increased but not reaching values of 2. On the year 2013, the measurements ameliorate which is not expected, also it seems to be more spatial variability. In the measurements of the year 2017, we can clearly see the effect of the intervention and values measured are close to 0 indicting almost perfect conditions. Measurements of the year 2021 still show good conditions but with an expected slight deterioration.

When considering the indicator I2, we observe that values begin in 2000 around 1 indicating good condition of the road. Then an expected slight deterioration occurs until 2004 where values have increase but not much. In the measurements of the year 2009, we observe that values are more or less stable, but we see an increase in the variability of the measurements. Then on 2013, we observe a slight deterioration which is expected. After the intervention in between 2013 and 2017, we can see that the values decrease showing clearly the effect of the work on the road. However, we can see that a part of the road does not improve which is strange. After, that values on 2021 are not complete, but from the data we have, it can be seen that the good condition is maintained with normal deterioration.

Considering the indicator I3, we can see that in the year 2000 the road condition is excellent with values close to 0 for most of the stretch. After that in 2004, we can see a slight deterioration of the road condition, which is expected from time and normal use. On 2009, however, we observe a slight improvement of the road condition which is not expected. Several peaks in the measurements are also visible which can be difficult to explain. In the measurements of 2013, we observe a deterioration that is expected. Road condition is still acceptable but values are in between 1 and 2 and reaching 3 in some small parts. After the intervention, we can see a clear improvement on the indicator showing values close to 0 again which is expected after work done on the road. During the year 2021, despite the missing data we observe a drastic decrease in the condition, which is not expected from the previous data.

Indicator I4 data starts only in 2009. We can see that values are not evenly distributed across the stretch. In general values around 1 or lower can be seen, which indicates a good condition of the road, but in one part we can see values over 2. In the measurements done in the year 2013, we can observe an expected slight deterioration of the values which is expected. However, in the last part of the stretch where values were already higher in 2009, we can see clear deterioration with a peak value reaching 5 which indicates that road is in very bad condition. When looking at all indicators together, we can observe that most of them are showing a consistent slight deterioration pattern until the intervention. After the intervention it can be seen expected values indicating a clear amelioration of the condition.

5.2 Filiale 2: Thun

The *filiale* Thun is located in southern Western Switzerland as seen in figure 11. Highways N8 and N9 were chosen for visualization.

5.2.1 Highway N8

Results on Figure 29 show that for the indicator I0, the measurements from 2004 to 2009 show a relatively stable trend with only minor variations. This is not entirely expected, as we would anticipate a gradual deterioration in pavement condition over time. From 2009 to 2013, an increase in peak values is observed, suggesting some level of deterioration, although the base line change is not drastic. However, due to the absence of data for 2018 and 2021, it is impossible to determine whether the expected deterioration continued or if an unexpected amelioration occurred.

For the indicator I2, the measurements taken in 2000 show large variations within the same segment. This suggests potential issues with measurement consistency. In 2004, we observe a more stable trend, with values oscillating between 1 and 2. Between 2004 and 2009, we see a general increase in peak values, but a similar trend of the base line which could aligns with the expected deterioration over time. This trend continues from 2009 to 2013, with increasing values and more pronounced peaks, indicating worsening conditions. Again, because of the absence of data for 2018 and 2021, it remains unclear whether the deterioration trend persisted or if unexpected anomalies, such as an improvement in condition, were recorded.

For the indicator I3, in the year 2000, the measured values are generally low, with some localized peaks indicating isolated issues. From 2004 to 2009, the values remain relatively stable but are higher in one part of the stretch and lower in another part, showing inconsistency in the measurements. During the period from 2009 to 2013, we observe an unexpected amelioration of the condition when checking the base line but with lots of peaks and thus an increase in variability across different sections of the road. This is not in line with what is expected, since road conditions should naturally degrade over time. As with other indicators, the missing data from 2018 and 2021 prevents a full analysis of long-term trends.

For the indicator I4, there is no available data for the years 2000 and 2004, making it difficult to analyse long-term trends. In 2009, we observe low values, suggesting that at this point, the pavement condition was still relatively good. The data from 2013 follows a similar pattern, with no major increase in values. This could indicate stable conditions. The missing data for 2018 and 2021 further complicates the analysis, as we are unable to determine whether the condition deteriorated as expected.

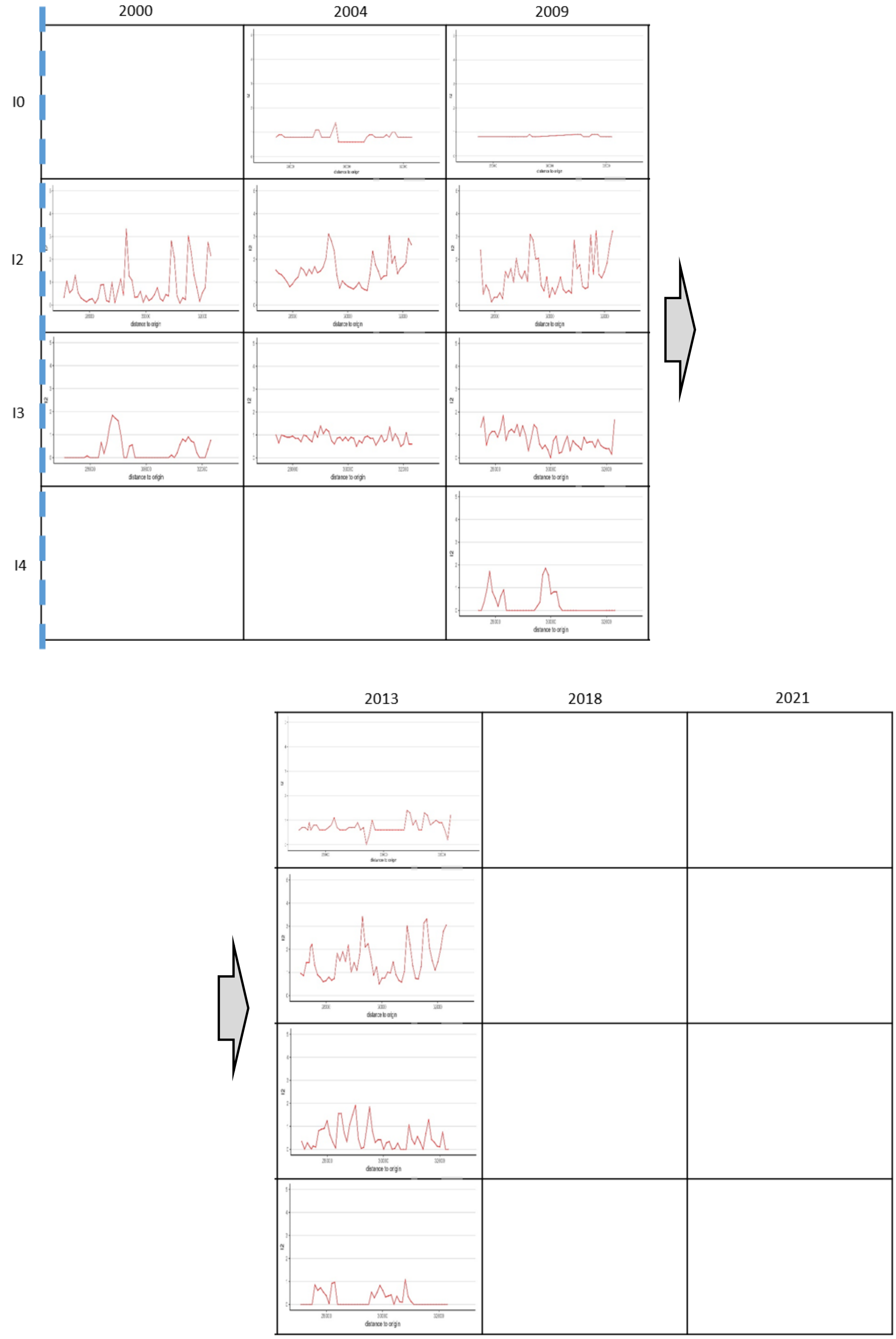

Figure 29: Indicators measured in the N8 stretch with no major intervention after year 2000

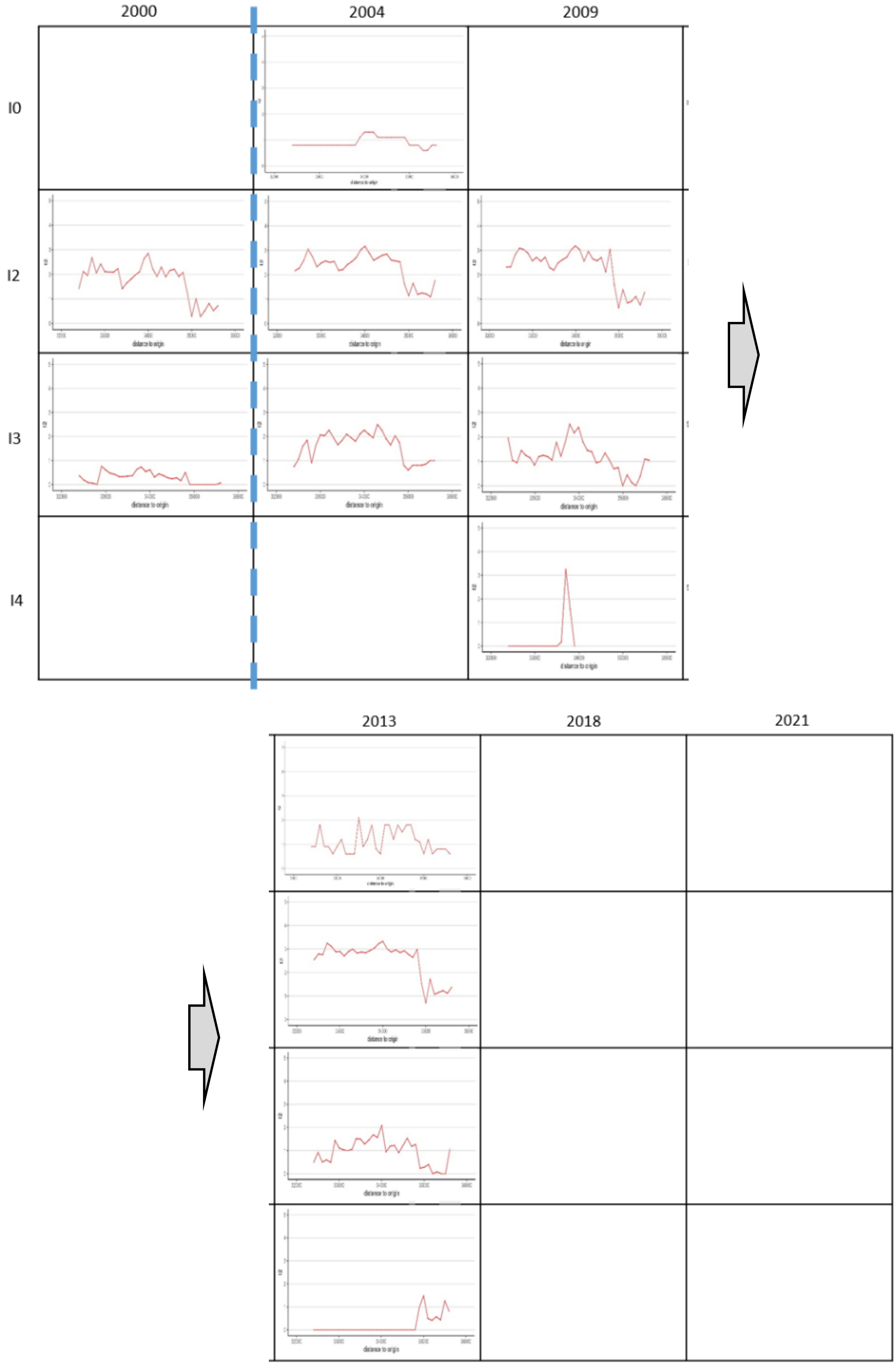

Figure 30: Indicators measured in the N8 stretch with a major intervention in between the years 2000 and 2004

Results on Figure 30 show that for the indicator I0, the measurements start in 2004, showing relatively good conditions with values close to 1 on average. In 2009, no data is available, which makes it difficult to assess the natural deterioration of the road. By 2013, a slight increase in values can be observed, indicating expected deterioration, though some fluctuations suggest possible inconsistencies in the measurements. The missing data for 2018 and 2021 prevents a full understanding of whether the deterioration continued as expected.

For the indicator I2, the first available measurements in 2000 show values oscillating during most of the stretch around 2 but also a part lower than 1, indicating a regular initial condition. A slight increase is observed in 2004, which aligns with expected gradual deterioration. The measurements in 2009 continue to show a consistent trend of degradation, with values increasing slightly compared to previous years. However, we see a peculiar pattern where part of the road segment shows a clear decrease in values, while the rest maintains a gradual deterioration. This unexpected partial amelioration raises questions about possible inconsistencies in the data or localized interventions. In the year 2013, results are consistent showing a slight deterioration all along the stretch. The lack of data for 2018 and 2021 prevents further verification of this trend.

For the indicator I3, the year 2000 measurements show values generally low values under 1, indicating an excellent road condition with only minor fluctuations. By 2004, we observe a strong so not completely expected deterioration, with values increasing over 2. The 2009 data show an amelioration of the condition to values around 1 but variable along the stretch which is not as expected. The 2013 data present a mixed pattern, with some sections showing deterioration while others exhibit a slight improvement, which is unexpected. As with other indicators, missing data for 2018 and 2021 limits the ability to confirm whether the deterioration followed its expected trajectory.

For the indicator I4, data first appears in 2009, showing mostly low values around 0, indicating excellent conditions. However, one noticeable peak value reaches 3, suggesting a localized severe deterioration. In 2013, the general condition remains stable, but another peak is visible, indicating that specific segments of the road are degrading more rapidly than others. The peak measured in 2009 disappeared which could indicate that it was a measurement problem. The missing data from 2018 and 2021 prevents further trend analysis. The major intervention recorded in between the years 200 and 2004 is not visible in the results indicating an inconsistency in either the intervention records or in the measurements.

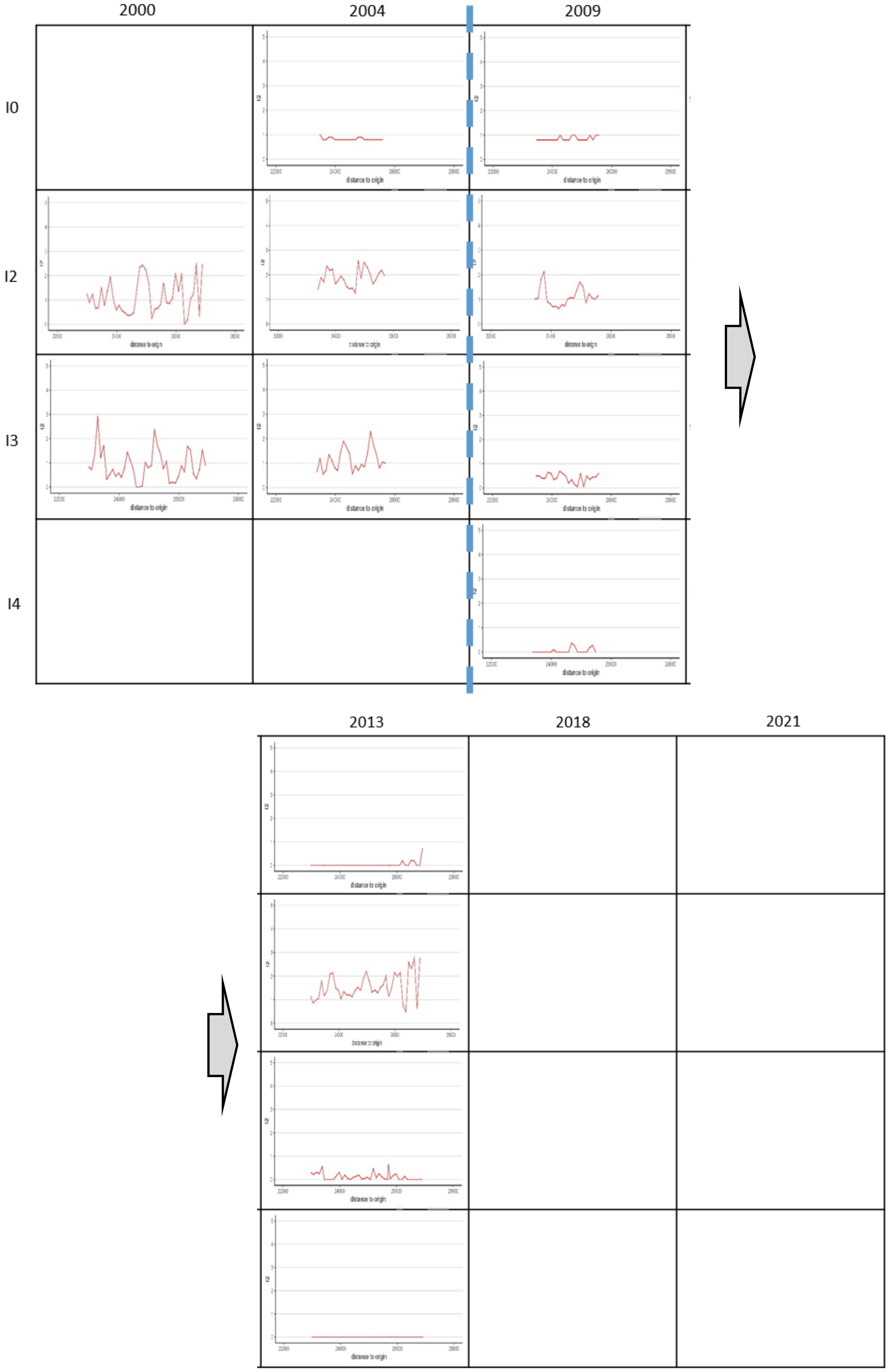

Figure 31: Indicators measured in the N8 stretch with a major intervention in between the years 2004 and 2009

Results of Figure 31 show that for the indicator I0, the first available measurements in 2004 show stable values close to 1, indicating good pavement conditions. In 2009, the values remain relatively consistent, with no significant deterioration or improvement observed, which is somewhat unexpected as normal wear over time should lead to a slight increase or in this case a strong improvement because an intervention is recorded in this stretch. By 2013, a clear amelioration of the condition can be observed, which is consistent with the intervention but not with the values of 2009. Also it can be observed that more data is available meaning a longer stretch was measured. Missing data for 2018 and 2021 prevents further analysis.

For the indicator I2, the measurements begin in 2000, showing substantial variability, with values oscillating between 0 and 2. This suggests inconsistent road conditions or possible measurement issues. In 2004, the values appear more stable, with a general trend close to 2. However, the stretch measured is smaller which is inconsistent. By 2009, the condition remains fairly stable with a slight increase in the condition which could be explained by the intervention reported but not fully since the amelioration should be stronger. In 2013, it was observed that a pattern similar to that of 2000, where values fluctuate significantly in some segments, raising questions about measurement consistency or potential external factors influencing the data. Also, the stretch is again longer as in the year 2000 which clearly indicate the discrepancies between companies, since the measurements form 2004 and 2009 are from one company and the ones from 200 and 2013 are from the other. Missing data from 2018 and 2021 does not allow further interpretation.

For the indicator I3, the data from 2000 shows significant variations, with several peaks, suggesting the presence of deteriorated sections. By 2004, the values are similar and appear also not very stable, indicating that either the conditions were maintained. The length of the stretch was reduced just like for the indicator I2. Data from the year 2009 show a clear increase in the condition which is consequent with the intervention that is recorded. By 2013, the values appear again lower, which is somehow unexpected but can be attributed to the change in company as well as the longer stretch. This anomaly could also be explained by smaller localized repairs or inconsistencies in the measurement methodology. Missing data for 2018 and 2021 prevents further trend verification.

For the indicator I4, no data is available before 2009. The available measurements from 2009 show very low values, close to 0, suggesting that the pavement condition was still very good at that point. By 2013, the trend remains similar, with minimal fluctuations, which could indicate that this indicator either does not register significant changes over time or that the road segment remained in a relatively good state. However, due to the absence of data for later years, it is unclear whether this stability continued or if deterioration eventually occurred.

5.2.2 Highway N9

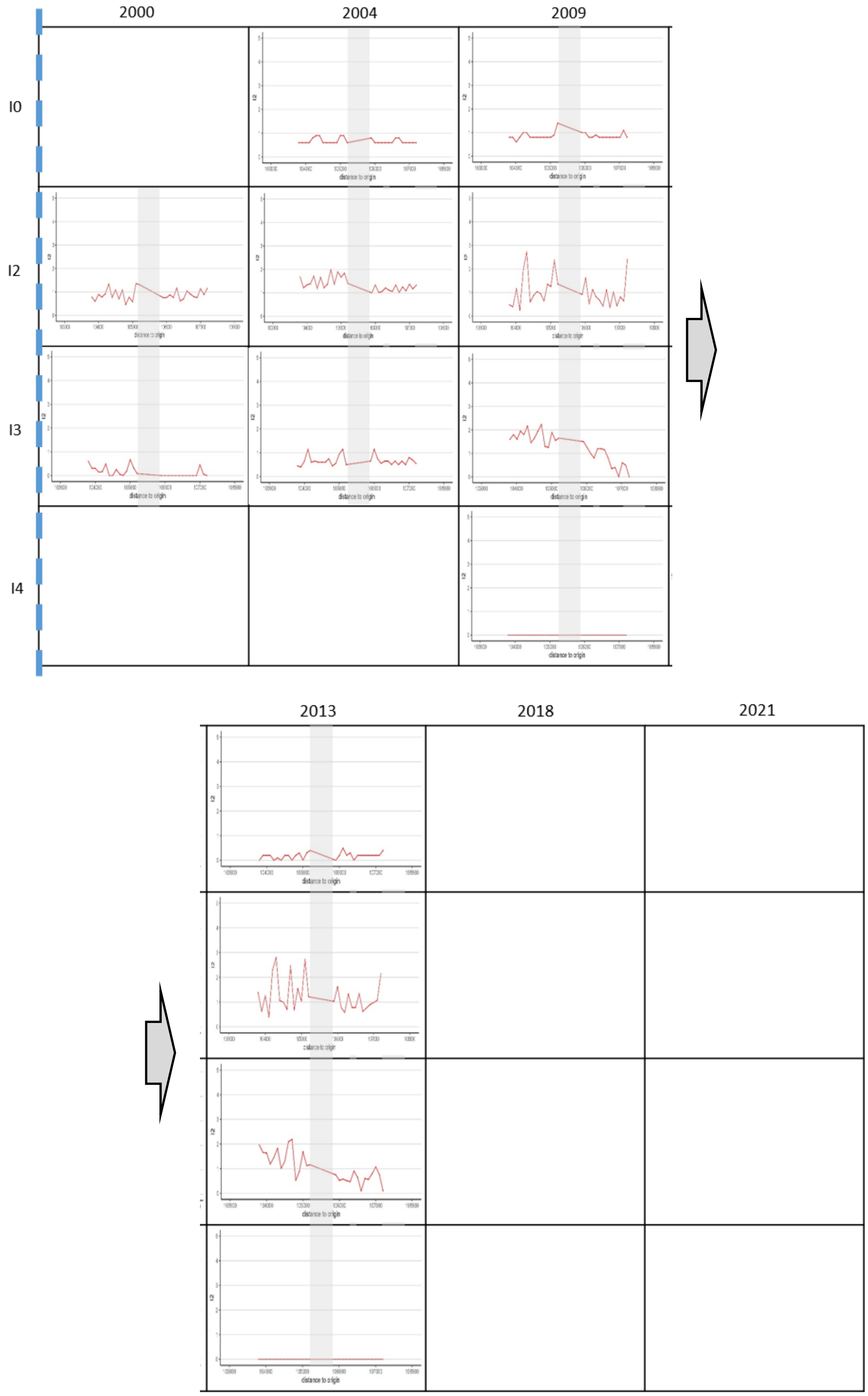

Figure 32: Indicators measured in the N9 stretch with no major intervention after the year 2000

Results on the Figure 32 show that for the indicator I0, the first measurements from 2004 show values slightly under 1, indicating a good pavement condition. By 2009, an expected slight deterioration is observed, with values increasing but still staying below 2. In 2013, the values plumbed indicating a strong amelioration in the condition, suggesting that there is a problem either with the measurement or with the interventions reporting.

For the indicator I2, the data from 2000 shows values generally around 1, indicating a road in good condition. By 2004, a slight increase is observed, which is expected due to gradual wear over time. The 2009 measurements however show a reduction in the baseline to values under 1 but with a more noticeable increase in variability, with several peaks reaching values of 3 or more. This suggests that certain segments of the road could have deteriorated more significantly. The general trend reduction in values does not make sense considering the lack of reported interventions. By 2013, the variability continues, with some peaks indicating poor condition, but the overall trend remains consistent with the previous measurement. As with other indicators, missing data for 2018 and 2021 prevents further analysis of whether this pattern persisted or worsened.

For the indicator I3, the values in 2000 start at a very low level, close to 0, which indicates excellent road conditions. By 2004, a slight increase in values is observed, indicating an expected deterioration. In 2009, the values remain within the expected range, though a slight downward trend can be seen in some sections. This is suspicious since it is very unlikely that an intervention was done only in one section of the stretch. In 2013, the values remain similar maybe a slight reduction in values in some sections, which is not expected unless some minor maintenance work was performed. The lack of data from 2018 and 2021 prevents further assessment.

For the indicator I4, the first available data appears in 2009, showing values generally below 1, which suggests that the pavement was still in good condition at that time. By 2013, the values remain mostly stable, with no significant signs of major deterioration. This stability is somewhat unexpected, as normal wear and tear should lead to a gradual increase in values over time. Without data from 2018 and 2021, it is unclear whether this trend continued or if further changes in road condition occurred.

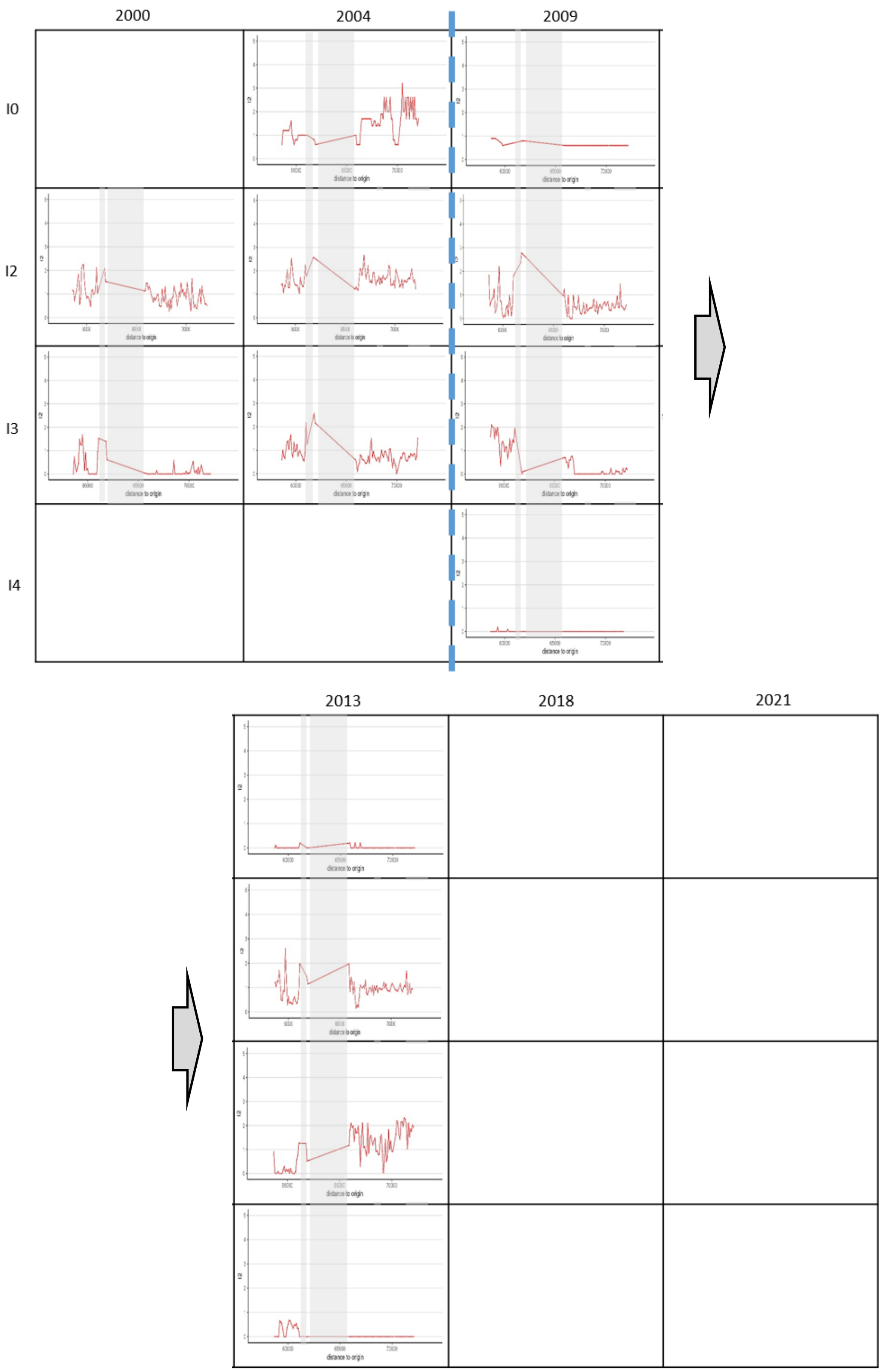

Figure 33: Indicators measured in the N9 stretch with a major intervention in between the years 2004 and 2009

Results on the Figure 33 show that for the indicator I0, the first available measurements in 2004 show values increasing up to approximately 2 and more in some sections, indicating early signs of deterioration. In 2009, however, a significant reduction in values is observed, with the measurements stabilizing under 1 or lower, which is expected since there is a recorded intervention. The data from 2013 shows extremely low values, close to 0, suggesting a further amelioration of the condition which is not expected.

For the indicator I2, the 2000 measurements show values fluctuating between under 1 and 2, indicating a relatively good initial condition with some localized deterioration. By 2004, a noticeable increase is observed, with some sections reaching values over 2, which is consistent with an expected deterioration pattern. In 2009, the values expectedly drop to lower levels, which is consistent with the intervention. The measurements from 2013 show similar trends to those in 2009, with values fluctuating and showing a slight worsening over time as expected with normal deterioration.

For the indicator I3, the 2000 data present mostly low values, close to 0, indicating excellent pavement condition with some section a bit worse. By 2004, a deterioration is observed, with values increasing to around 1 and up in some sections. In 2009, most of the stretch show an amelioration which is consistent with the intervention. However, there is a section that continue to deteriorate which does not make clear sense except if the intervention didn't happen in that part. In 2013, we see an increase in values again, showing the normal deterioration. Interestingly, the initial section previously mentioned that did not improve after the intervention, this time improve. It is possible then that the intervention happens around the time of the measurement in 2009 in that section this is why it was not visible before.

For the indicator I4, data first appears in 2009, showing very low values close to 0, indicating a well-maintained pavement. In 2013, the values remain almost the same, suggesting that there was little to no deterioration. Given the patterns observed in other indicators, this stability could be due to measurement differences rather than a lack of deterioration.

5.3 Filiale 3: Zofingen

The *filiale* Zofingen is located in northern Central Switzerland as seen in Figure 16. Highways N2 and N4 were chosen for visualization.

5.3.1 Highway N2

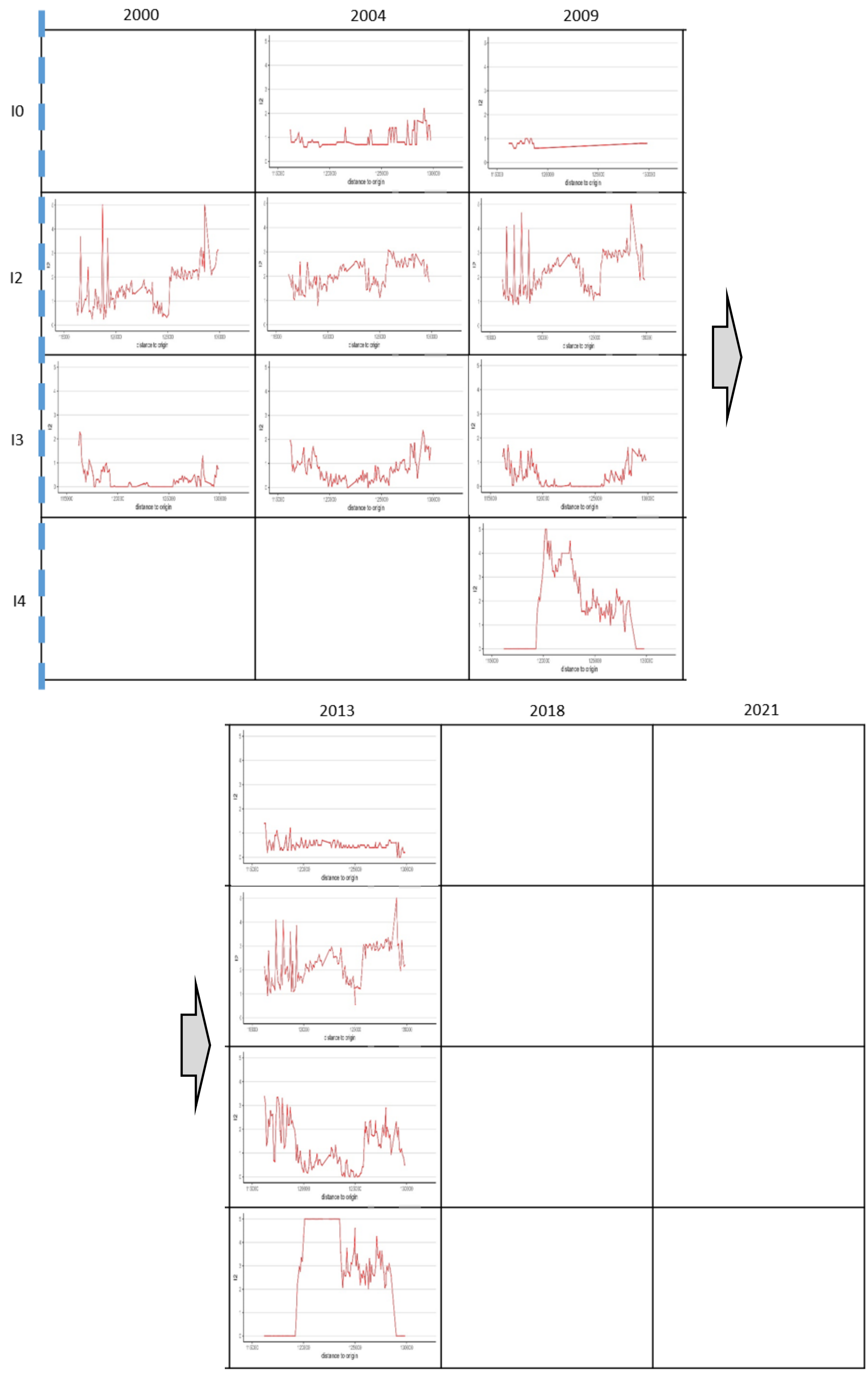

Figure 34: Indicators measured in the N2 stretch with no major intervention after the year 2000

Results on Figure 34 show that for the indicator I0, the first available measurements in 2004 show moderate variability, with values fluctuating around 1 and some peaks reaching to 2, indicating early signs of deterioration. By 2009, no slight deterioration is observed, and the values remain mostly stable compared to the previous measurement which could indicate problems with the measurement. In 2013, however, a reduction in values is observed, with measurements clustering under 1, which is not expected unless an intervention was performed or there were inconsistencies in measurement methods. The data is also very unstable showing lots of variability compared to previous measurements.

For the indicator I2, the 2000 measurements show a high degree of variability, with values frequently spiking up to 3 or 4, suggesting inconsistent pavement conditions or possible data collection issues. There are also difference in the sections, with values between 1 and 2 in the first section and values over 2 in the second section. A small section in between shows better values. By 2004, the values remain variable but less and a slight expected deterioration is observed. The three sections remain consistent. In 2009, the pattern remains mostly the same, with a slight deterioration happening. High variability can be observed in the beginning of the stretch. By 2013, values remain too similar to the previous measurement so no deterioration can be seen from the data which is not expected. The same variability is observed in the beginning of the stretch. This could suggest that certain sections of the road experienced rapid deterioration, possibly due to specific localized damage or environmental factors. Missing data for 2018 and 2021 prevents further assessment of whether this pattern persisted.

For the indicator I3, the initial 2000 data show relatively low values, indicating good pavement conditions, except in the beginning of the stretch. By 2004, a slight deterioration is observed, with values increasing slightly but remaining under 2. In 2009, the values remain relatively stable even ameliorating which is unexpected since there was no intervention recorded. By 2013, a strong an unexpected increase in variability is noted, with some sections reaching values around 3. This suggests that deterioration may have accelerated in certain areas. It can be seen also that there are differences in values in different parts of the stretch, having one deteriorating more than others.

For the indicator I4, data first appears in 2009, showing a clear distinction between sections in good condition and sections with significantly higher values above 3 and even reaching peaks over 4. This suggests uneven wear along the road. In 2013, the pattern remains the same, with some sections showing high deterioration while others appear relatively unchanged. Stable measurements of value 5 for a long part of the stretch suggested problems in the measurement.

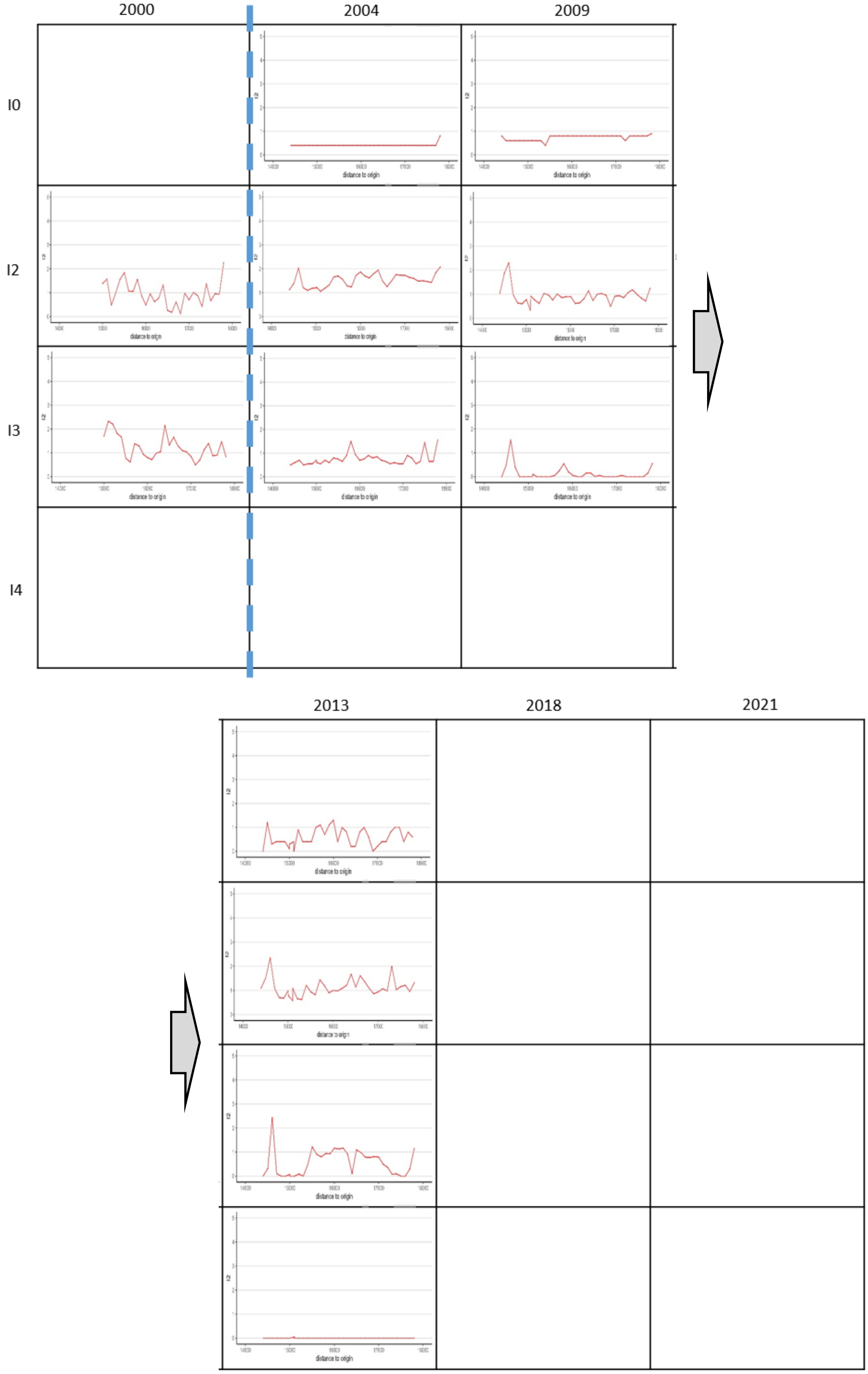

Figure 35: Indicators measured in the N2 stretch with a major intervention in between the years 2000 and 2004.

Results on the Figure 35 show that for the indicator I0, the first available measurements in 2004 show stable values under 1 1, indicating good pavement conditions. However, the data seems too stable to be reliable. In 2009, a slight but expected deterioration is observed, with values increasing but not exceeding 1. By 2013, the data shows higher variability in measurements, and a slight amelioration of the condition when not checking the pics which indicates that the different companies have used different methodologies.

For the indicator I2, the data from 2000 shows values oscillating around 1, indicating an overall good condition but with a lot of variability in the data. By 2004, a slight increase in values is observed, aligning with expected deterioration, however an intervention is recorded in between measurements so there is a discrepancy here. The 2009 measurements show an amelioration of the condition which is not expected but could make sense if the problem was with the measurement of 2004. By 2013, further deterioration is evident, though the increase is relatively moderate. The data this time is consistent with the previous measurement of 2009.

For the indicator I3, the 2000 data show relatively medium to low values, indicating a well-maintained road at that time. By 2004, we observe a slight amelioration of the condition which make sense considering the intervention that is recoded before the measurement. In 2009, the values show a further increase in the condition which is not expected and suggest some problem with the measurement. However, in 2013, the measurements show more pronounced variations, with an increase in higher values. This suggests that deterioration is accelerating in some parts of the segment.

For the indicator I4, the first available data appears in 2013, the trend is a very low value that is very stable which is not as expected when compared with the other Indicators.

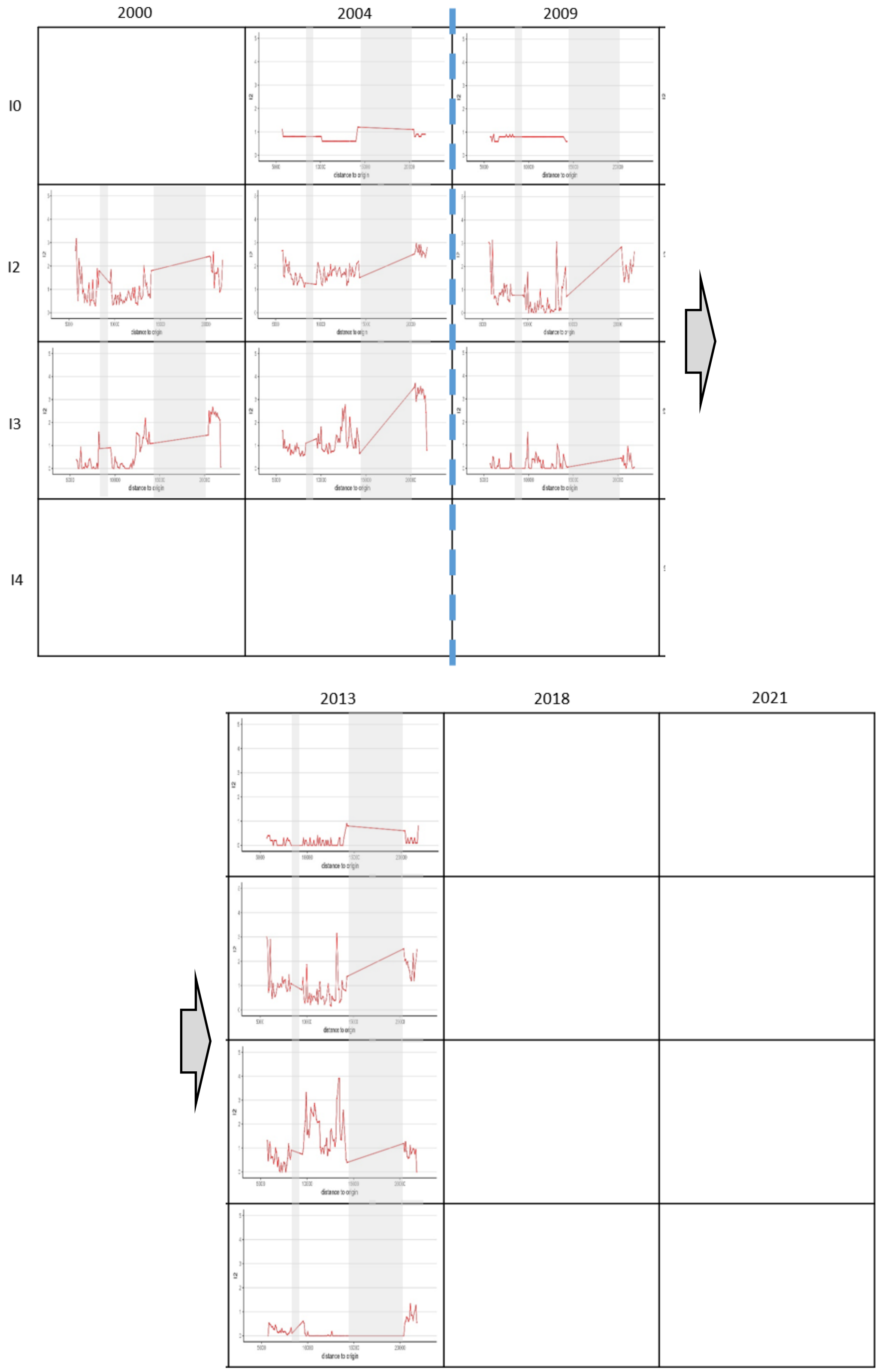

Figure 36: Indicators measured in the N2 stretch with a major intervention in between the years 2004 and 2009

Results on Figure 36 show that for the indicator I0, the first available measurements in 2004 show relatively stable values close to 1, suggesting a road in good condition.

By 2009, the values remain mostly stable with slight fluctuations, which is unexpected an intervention is recorded. In 2013, we observe a more significant variability in values, and a clear amelioration of the condition which this time could be consequent with the intervention previously mentioned.

For the indicator I2, the 2000 measurements show a high level of variability, with values fluctuating around 1 but and some sectors worst and some sector better. This suggests either inconsistencies in the road condition or possible measurement errors. In 2004, the values stabilize slightly but still show significant oscillations, which is unusual. A slight deterioration is observed which is as expected. By 2009, we see an expected amelioration consistent with the intervention, however there is high variability in the data. In 2013, the trend continues, showing higher variability, with some peaks reaching values of 3 or more. This suggests that certain segments have worsened more than others, possibly due to localized distress. No clear deterioration can be seen from the data which is not as expected.

For the indicator I3, the data from 2000 shows relatively low values with some peaks, indicating good road conditions with minor irregularities. By 2004, deterioration is observed, with values increasing to around 2 in some sections. In 2009, a clear amelioration of the condition can be observed which make sense considering the intervention. Variability, however, remains high which could indicate some measurement problems. In 2013, a noticeable rise in values is observed, with several peaks reaching 3, indicating a worsening condition particularly in one part of the stretch.

For the indicator I4, data first appears in 2013, a few localized peaks are observed, but the values are very low suggesting a good condition.

5.3.2 Highway N4

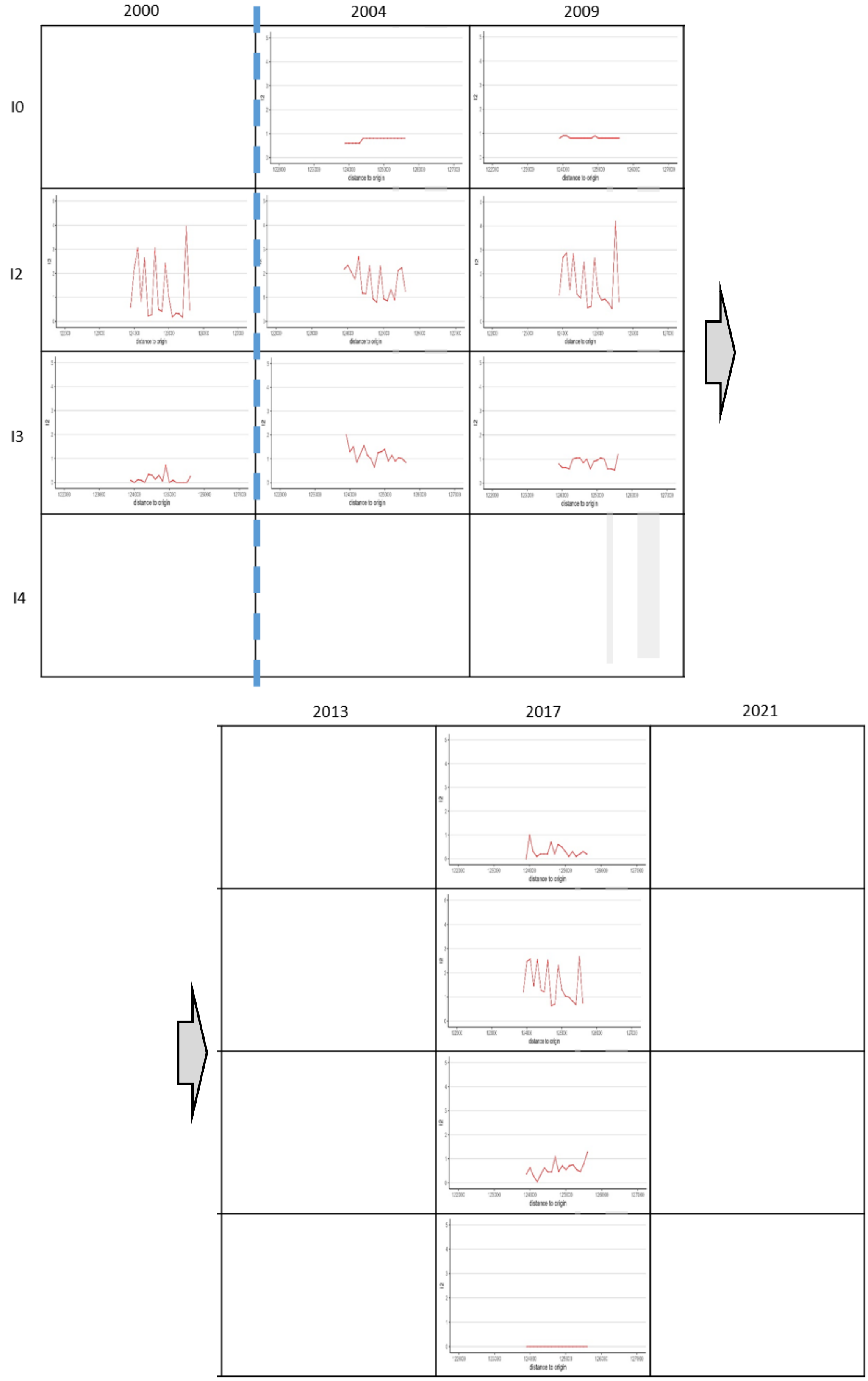

Figure 37: Indicators measured in the N4 stretch with a major intervention in between the years 2000 and 2004.

Results on Figure 37 show that the data available for this stretch is not very long. For the indicator I0, the first available measurements in 2004 show stable values close to 1, indicating a road in good condition. By 2009, there is no significant change, with values remaining relatively stable but this is unexpected since some normal deterioration should be visible. Data from 2017, show a drop in values, which is unexpected unless maintenance work was performed or there were differences in measurement methodology. Variability also increased in this last measurement year.

For the indicator I2, the 2000 measurements show a high level of variability, with values frequently oscillating between 1 and 3, indicating possible inconsistencies in the road surface or measurement anomalies. By 2004, the variability is still present, but the general trend remains within an expected deterioration range. However, an intervention is recorded here so the data is not consistent with that. In 2009, values show a very slight deterioration, and the variability is still present. In 2017, we observe a slight drop in values, which is unexpected without an explanation for improvement. The missing data for 2013 and 2021 prevents further analysis of whether this trend was maintained.

For the indicator I3, the data from 2000 shows relatively low values with some peaks, indicating mostly good road conditions with minor irregularities. By 2004, the values slightly increase, which is not expected since an intervention is recorded. In 2009, a slight improvement of the condition is observed, but overall values remain relatively stable. However, by 2017, we see a further improvement of the condition which does not make sense considering the time passed and the normal deterioration that should have happened.

For the indicator I4, there is not enough data to understand any trends. Values from 2017 are too stable and too low to be reliable.

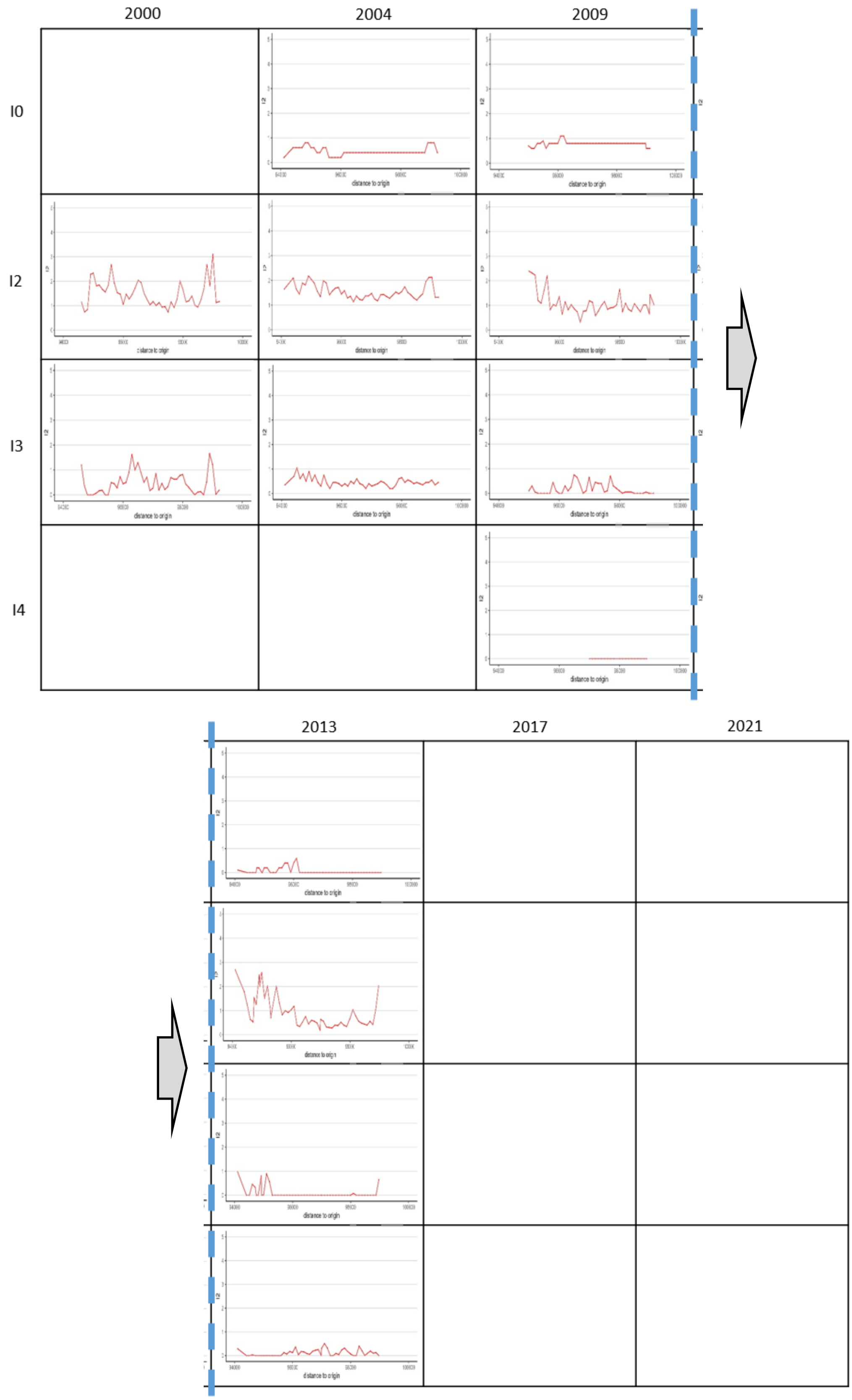

Figure 38: Indicators measured in the N4 stretch with a major intervention in between the years 2009 and 2013

Results on Figure 38 show that for the indicator I0, the first available measurements in 2004 show relatively stable values around 0.5, indicating that the pavement is in good condition. By 2009, a slight but expected deterioration is observed, though values remain below 1, suggesting that the pavement is aging at a normal rate. However, in 2013, a significant drop in values is noted, with most measurements close to 0, which is expected and consistent with the reported intervention.

For the indicator I2, the 2000 measurements show a high degree of variability, with values fluctuating between 1 and 2, suggesting inconsistent pavement conditions. In 2004, values stabilize slightly, but overall, the pattern remains similar, with the stretch showing normal deterioration. By 2009, a noticeable amelioration fo the condition is observed which is not expected since there are no interventions mentioned in between measurements. However, in 2013, an unexpected reduction in values is observed particularly in some areas, which is consistent with the reported intervention, but other parts of the stretch deteriorate. This anomaly raises questions about potential measurement differences or localized repairs.

For the indicator I3, the data from 2000 shows relatively low values but with high fluctuations, indicating a well-maintained pavement at that time but strong variability in the measurement. By 2004, a slight increase in values is observed, suggesting early signs of wear, but it is not clear since the variability is reduced and less peaks are observable. In 2009, values show an amelioration of the condition which is not expected. However, by 2013, a significant reduction in values is observed, which is expected because of the reported intervention. Some peaks are still present which could indicate inconsistencies in measurement or targeted maintenance work on specific sections of the pavement.

For the indicator I4, data first appears in 2009, showing values close to 0, indicating minimal pavement deterioration at that time. In 2013, after the reported intervention, the condition deteriorates which is not consistent with what is expected. The absence of data for later years makes it difficult to determine whether this pattern continued.

5.4 Filiale 4: Winterthur

The *filiale* Winterthur is located in northern easter Switzerland as seen in Figure 16. Highways N1 and N3 were chosen for visualization.

5.4.1 Highway N1

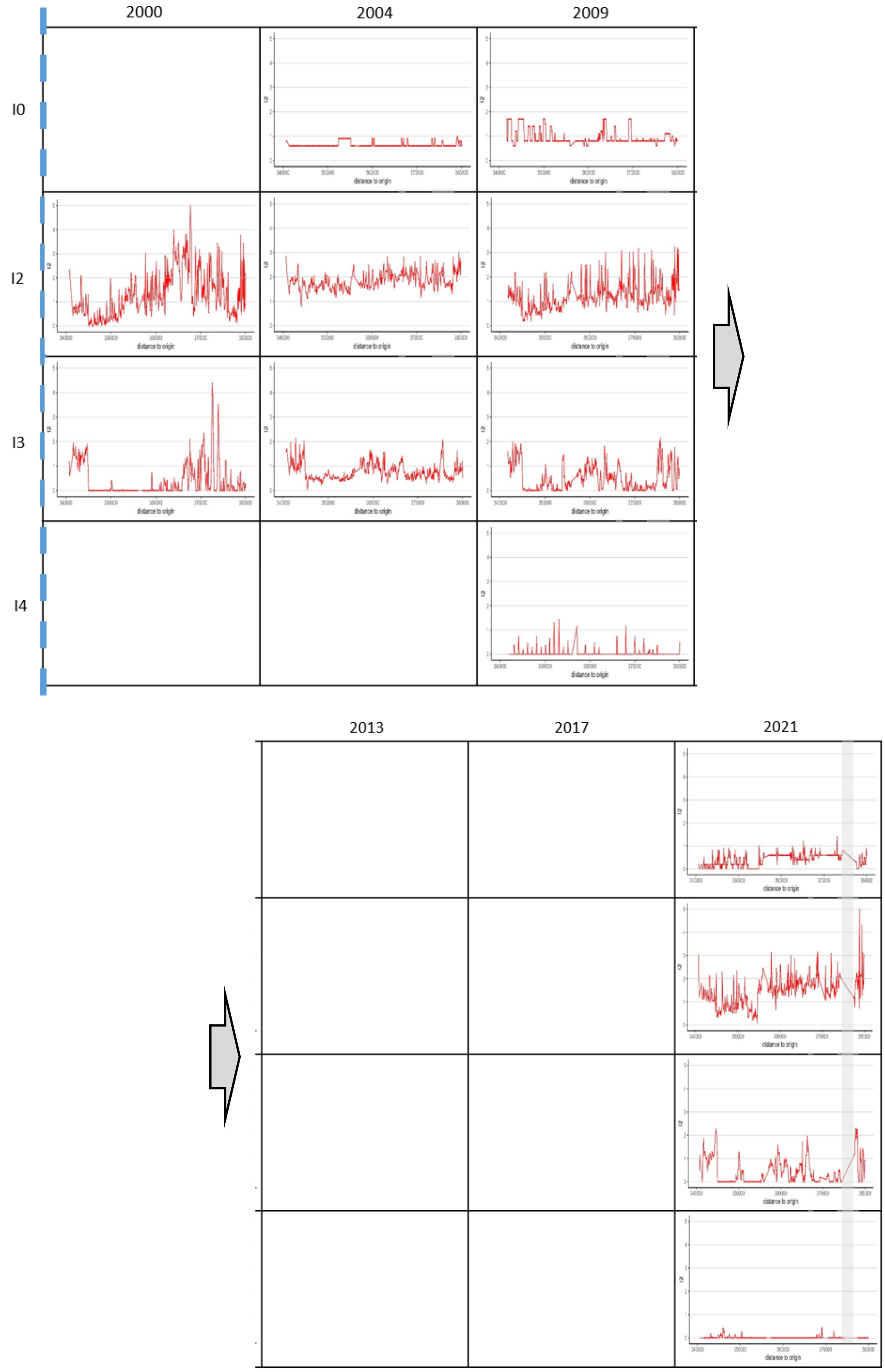

Figure 39: Indicators measured in the N1 stretch with no major intervention after the years 2000

Results in Figure 39 show that for the indicator I0, the first available measurements in 2004 show relatively stable values close to 1, indicating good pavement conditions. By 2009, we observe a slight but expected deterioration, with values becoming more variable but not exceeding 2. The increase in the variability is indicative a some kind of problem with the measurement or at least a market difference with the method used previously. The absence of data for 2013 and 2017 makes it difficult to determine how the road evolved during this period. In 2021, we observe that values have decrease indicating an amelioration of the condition which is not expected since there were no intervention reported. Interesting to see also is that, the spatial variability is high, suggesting that certain sections have deteriorated more significantly than others.

For the indicator I2, the 2000 measurements exhibit extremely high variability, with values frequently oscillating between 1 and 3, indicating possible inconsistencies in pavement conditions or measurement anomalies. A strong spatial variability is also observed with some parts of the stretch better than other. By 2004, the values stabilize but still show high variability. The spatial variability disappeared which is not expected and suggest some problems in the previous measurements. In 2009, the variability increases but the base line of the values show an amelioration of the condition which is not expected since there was no intervention. In 2021, the measurements still show a high variability and peaks reaching values above 3, indicating significant degradation in certain segments. The missing data for 2013 and 2017 prevents further verification of how this pattern developed over time.

For the indicator I3, the data from 2000 shows high fluctuations, with significant peaks suggesting localized pavement distress. In general, the condition is good for most of the stretch. By 2004, a more consistent spatial pattern appears, with values mostly below 2 but still showing significant variability. A general deterioration can be observed which is expected. In 2009, the condition ameliorates which is not expected without an intervention. By 2021, we see a similar trend, with even lower values in general despite some pics across the dataset, suggesting an overall poor condition in some sections.

For the indicator I4, data first appears in 2009, showing mostly low values below 1, indicating relatively good pavement conditions at that time but showing some variability. By 2021, the measurements remain mostly stable, with only a few peak values. This suggests that either the indicator is less sensitive to pavement deterioration or that this aspect of the road condition has not worsened as significantly as others or has even ameliorate.

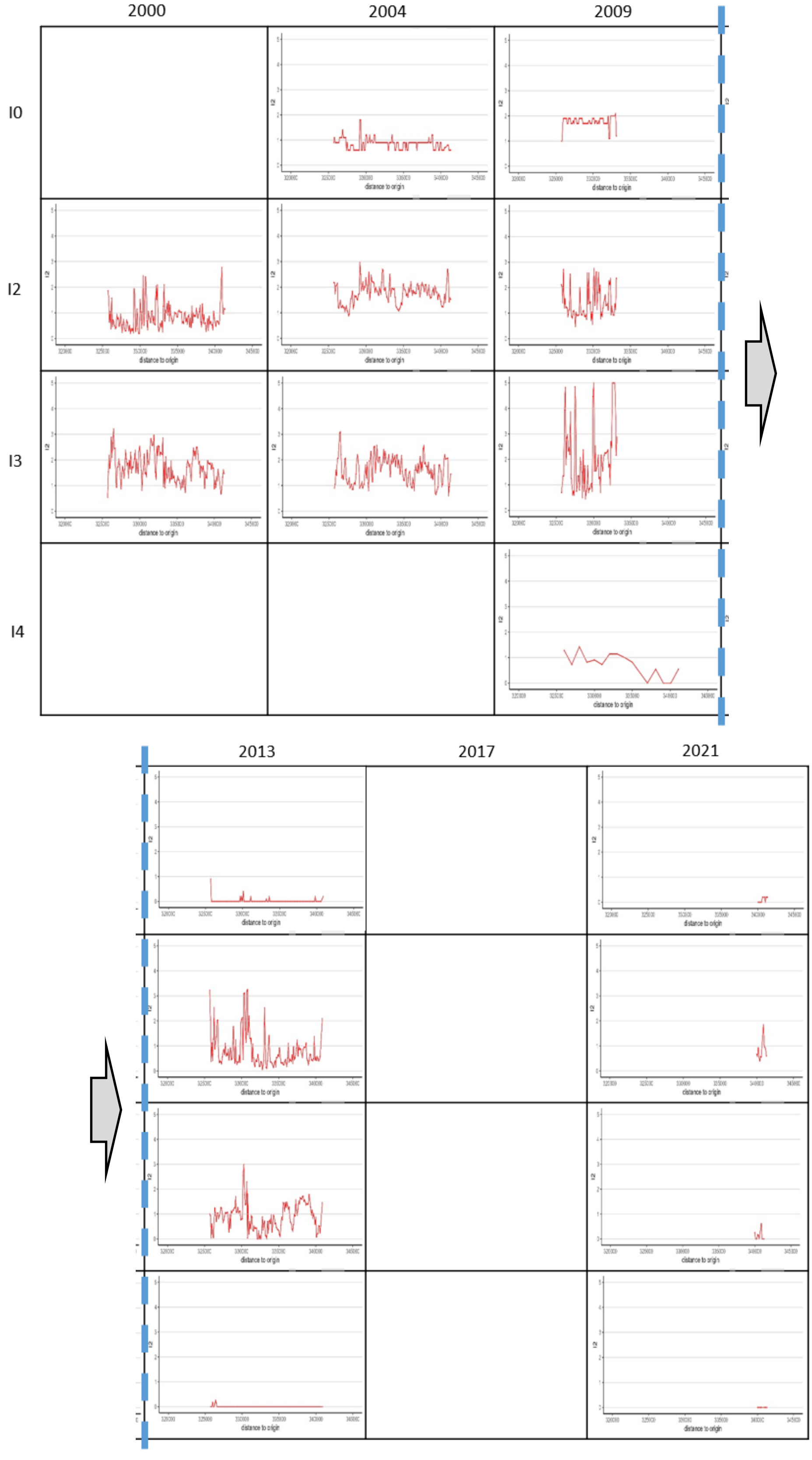

Figure 40: Indicators measured in the N1 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 40 show that for the indicator I0, the first available measurements in 2004 show relatively stable values close to 1, indicating a road in good condition. By 2009, the values show a slight increase, reflecting expected gradual deterioration. However, the data is smaller meaning a shorter part of the stetch was measured which is not expected. In 2013, a sharp decrease in values is observed, with most sections showing values close to 0, which is consistent with the reported intervention. By 2021, The data available only corresponds to a very small part of the stretch which does not allow further analysis.

For the indicator I2, the 2000 measurements show significant variability, with values mostly under 2. It can be observed inconsistent spatial pavement conditions, but still good in general terms. In 2004, the variability persists, but no deterioration of the condition is observed which is unexpected. By 2009, an big increase in the variability is observed, suggesting localized deterioration. The data is only available for a small stretch which is not optimal. The 2013 data show an amelioration of the condition which is expected since there is an intervention reported for this stretch. In 2021, again there is not enough data to make any conclusions.

For the indicator I3, very similar patter like for indicator I2 can be seen. The data from 2000 shows substantial fluctuations, with a few peaks reaching values of 3 or higher, indicating existing pavement distress. By 2004, values become slightly more stable, but fluctuations remain in this case no clear deterioration can be inferred from the data which is not expected. The 2009 data show great variability and the expected further deterioration is difficult to grasp because of the fluctuation in measurements. Just like for Indicator I2, the measured stretch is shorter which is strange. In 2013, an increase in the condition can be observed which is consistent with the reported intervention. Finally, by 2021, the measurements are too few and does not allow any interpretation and highlight inconsistencies in the data collection method.

For the indicator I4, the first available data in 2009 shows low values below 1, indicating minimal pavement deterioration but still measurable. In 2013, a strong downward trend is observed, as expected because of the intervention.

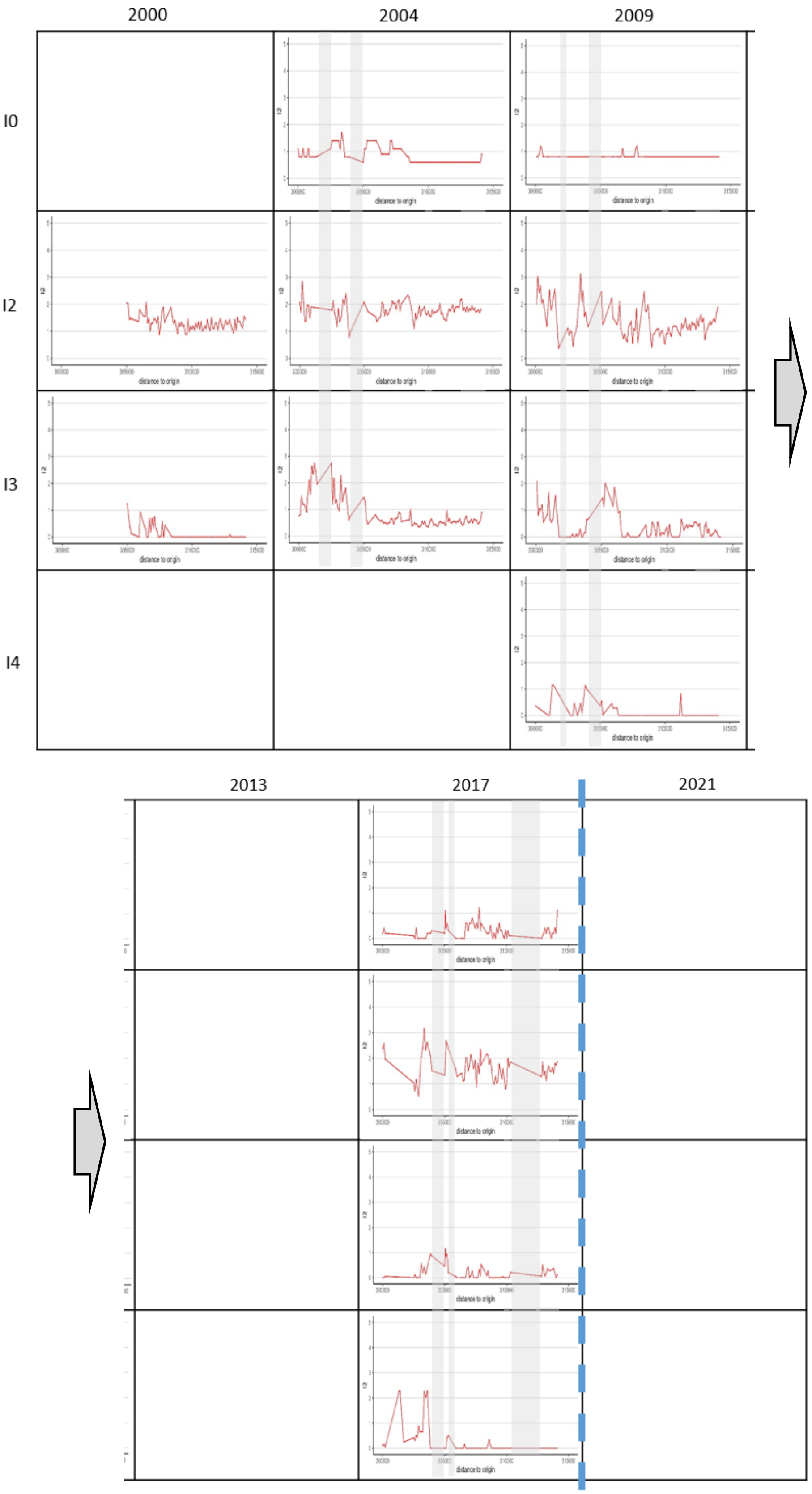

Figure 41: Indicators measured in the N1 stretch with a major intervention in between the years 2017 and 2021

Results on Figure 41 show that for the indicator I0, the first available measurements in 2004 show relatively stable values around 1, indicating good pavement conditions. By 2009, a slight but expected deterioration is observed, with values fluctuating but remaining close to 1. In 2013, data is missing, making it difficult to determine how the trend evolved. However, by 2017, a clear increase in variability is observed, as well as an amelioration of the pavement condition which cannot be explained without an intervention.

For the indicator I2, the data from 2000 shows values slightly over 1 indicating a good condition. By 2004, A slight deterioration is observed as expected with values reaching close to 2. However, the 2009 data does not shows further deterioration, instead it can be observed an amelioration which is unexpected without an intervention. The variability is also stronger . In 2017, the general trend show a slight deterioration but not as expected after 8 years of use.

For the indicator I3, the 2000 data shows relatively low values, indicating good pavement conditions. By 2004, an expected slight deterioration is observed, with values increasing more than others in certain areas. The 2009 data show an amelioration of the condition despite some areas that deteriorate which is not consistent since no intervention took place. By 2017, the measurements show further improvement in the condition which cannot be explained since no intervention took place.

For the indicator I4, data first appears in 2009, showing relatively low values, suggesting minimal pavement distress. However, by 2017, significant peaks appear in certain sections, indicating localized deterioration which is as expected.

5.4.2 Highway N3

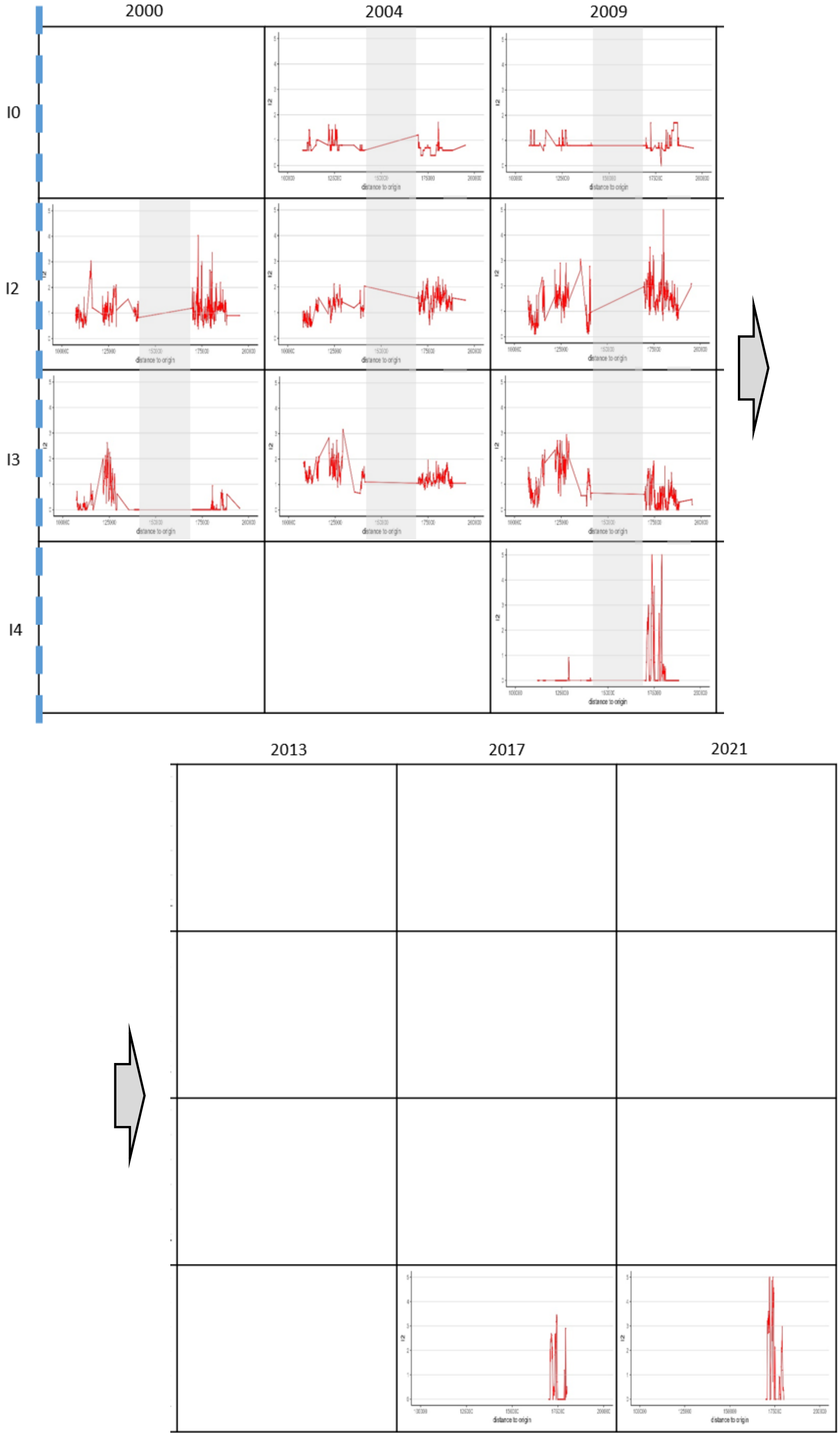

Figure 42: Indicators measured in the N3 stretch with no major intervention after the year 2000

Results on Figure 42 show that for the indicator I0, the first available measurements in 2004 show values around 1 indicating a good condition. In 2009, the values show a slight deterioration which is expected. The absence of data for 2013, 2017, and 2021 show inconsistencies in the data acquisition process between *filialen*.

For the indicator I2, the 2000 measurements indicated good condition with values slightly higher than 1 but with high variability. By 2004, a slight deterioration is observed in most of the stretches as expected but one part show an amelioration in the condition which does not make clear sense. In 2009, the variability increases importantly, with more frequent peaks suggesting worsening conditions and a general trend towards deterioration which in expected.

For the indicator I3, the 2000 data shows excellent conditions, but with significant spatial variability and peaks indicating localized pavement distress. By 2004, a clear deterioration is observed as expected . The 2009 data however, show a clear improvement of the condition which cannot be explained without an intervention. Spatial variability in the measurement remain high so it is difficult to analyse a general trend since small zones seems to behave differently.

For the indicator I4, data first appears in 2009, showing localized peaks that signal either a problem in the measurements or zones with severe deterioration in some sections while other parts remain relatively stable. By 2017 and 2021, the high peaks persist, suggesting continued pavement degradation. However, the missing data from other years makes it difficult to confirm whether the deterioration was gradual or abrupt Data from the rest of the stretch outside the peaks in not present anymore which make impossible to understand well what is going on.

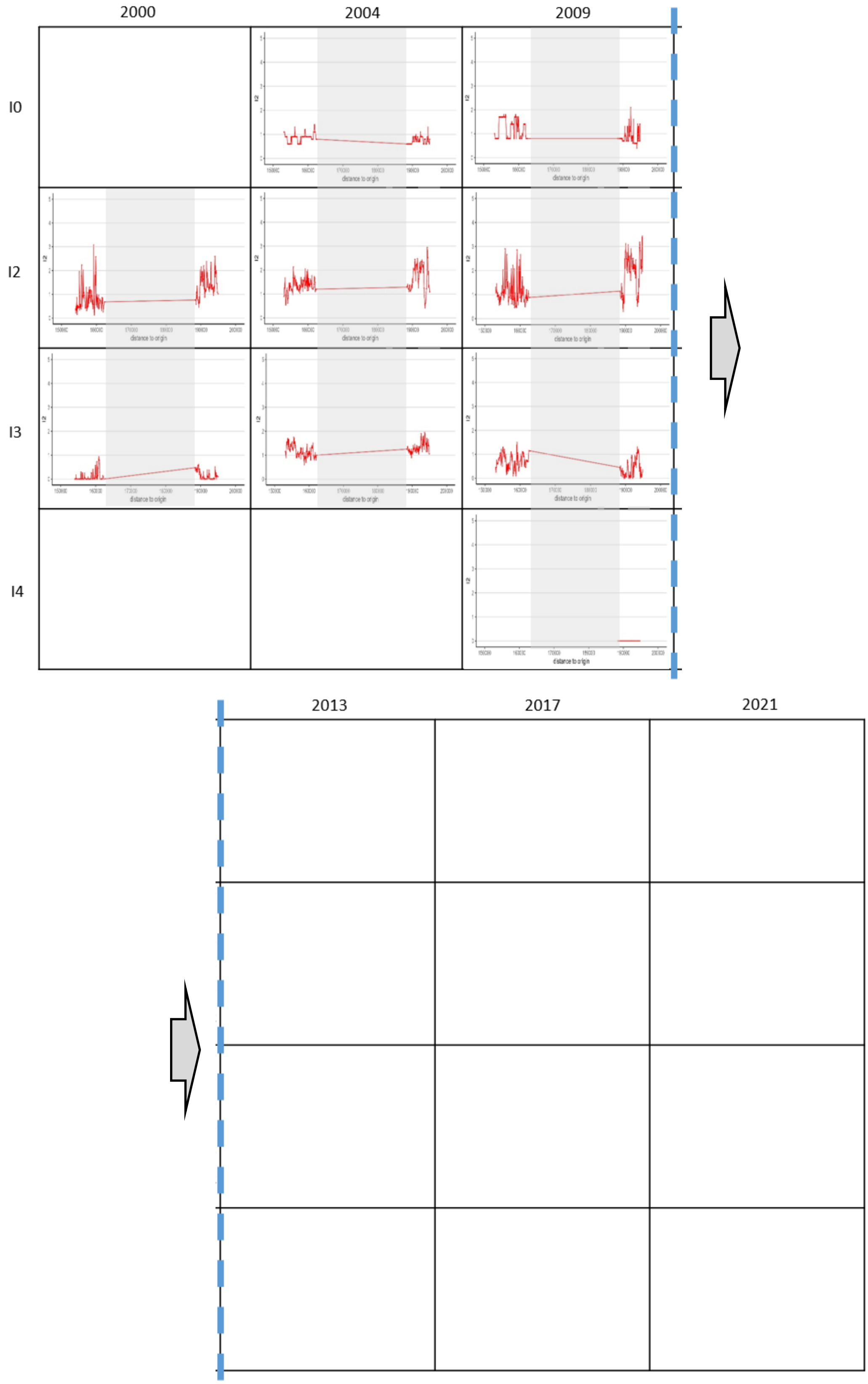

Figure 43: Indicators in the N3 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 43 show that for the indicator I0, the first available measurements in 2004 show only scarce measurements but in general they show good condition with values around 1. The 2009 measurements show a slight increase in values, with localized peaks, suggesting more significant deterioration in some segments. The values are consistent with a normal deterioration.

For the indicator I2, the 2000 data display high variability, with values oscillating around 1 or even a little more in another zone. In 2004, this variability continues, but a slight upward trend is visible, suggesting a gradual deterioration of the road condition. The 2009 data confirm this trend, with peaks reaching values above 3, indicating worsening conditions in certain areas.

For the indicator I3, similar results as for indicator I2 can be seen. The main difference is that measurements of the year 2009 show an improvement of the condition which is not expected since there was no intervention. .

For the indicator I4, only one point is available and with very few data signalling the inconsistencies in the measurements but nothing about the road condition.

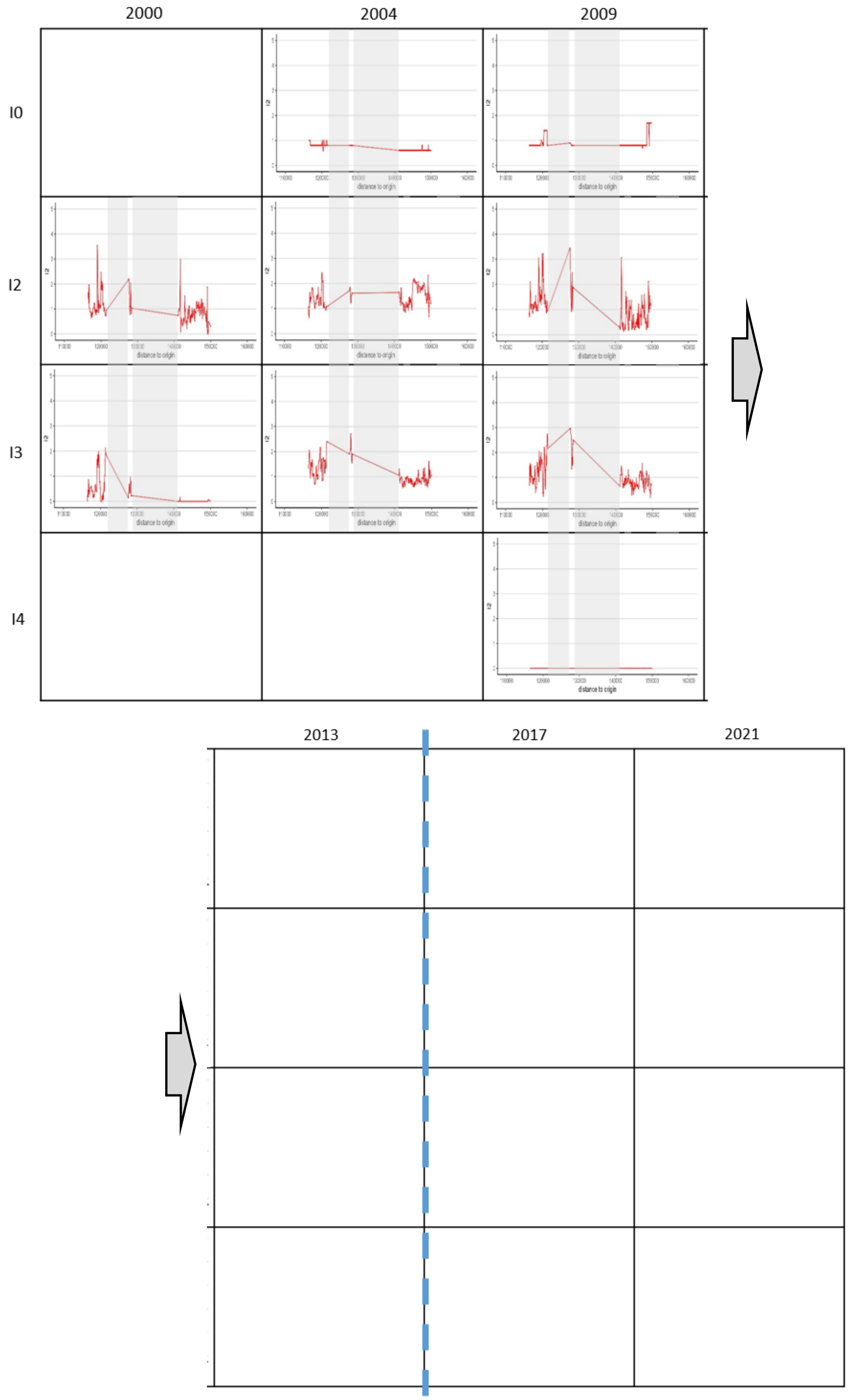

Figure 44: Indicators measured in the N3 stretch with a major intervention in between the years 2013 and 2017

Results on Figure 44 show that for the indicator I0, the first available measurements in 2004 show values close to 1, indicating good pavement conditions. However, like there is not a lot of data for this stretch. By 2009, a slight but expected deterioration is observed, with increased variability and some peaks.

For the indicator I2, the 2000 measurements show significant variability, with values fluctuating over 1 in one segment and under 1 in the other. By 2004, the trend shows a deterioration in both segments which is expected. In 2009, the variability intensifies, with frequent peaks reaching values above 2, indicating worsening pavement conditions. One of the segments show an amelioration of the condition which cannot really be explained.

For the indicator I3, the data show a similar trend like the data from the Indicator I2. Again, the step between 2004 and 2009 does not show a deterioration expected which is suspicious.

For the indicator I4, the data is insufficient to understand trend since there is only one measurement and furthermore the values are 0 indicating a too good condition that is not consistent with the data from the other indicators.

5.5 Filiale 5: Bellinzona

The *filiale* Bellinzona is located in Eastern southern Switzerland as seen in Figure 16. Highways N13 and N2 were chosen for visualization.

5.5.1 Highway N13

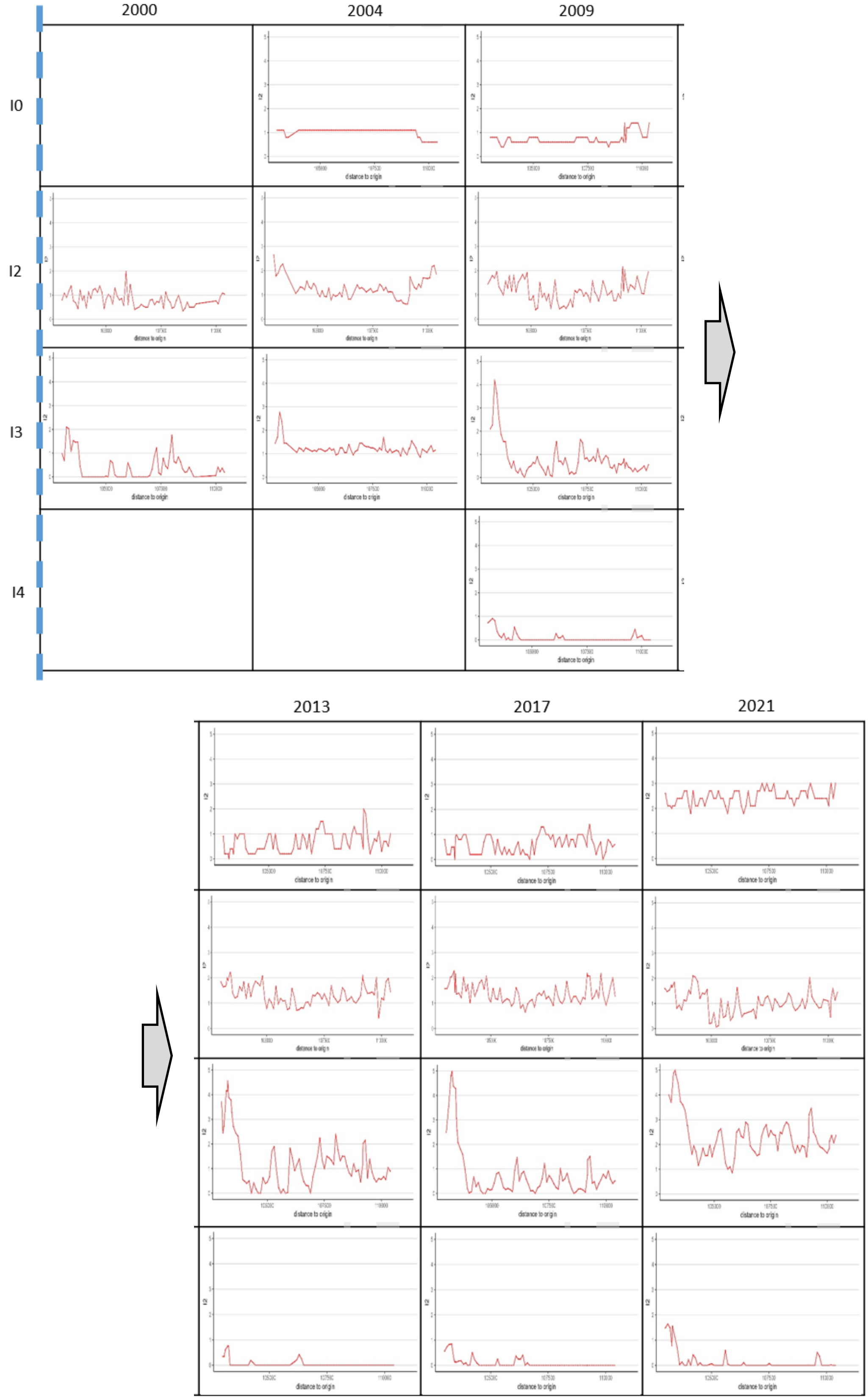

Figure 45: Indicators measured in the N13 stretch with no major intervention after the years 2000

Results on Figure 45 show that for the indicator I0, the first available measurements in 2004 show relatively stable values around 1, indicating good pavement conditions. In 2009, the variability of the measurement increased and in general term a slight amelioration of the condition can be seen. By 2013, we see a further increase in values variability, but the condition is not worse, perhaps even a bit better. In 2017, this trend continues, with no clear deterioration, except in one part of the stretch. By 2021, the measurements confirm a significant increase in pavement deterioration, with most values reaching between 2 and 3, suggesting extensive wear and tear over time.

For the indicator I2, the data from 2000 shows values fluctuating around 1, indicating a relatively good condition. In 2004, a slight increase is observed, aligning with expected deterioration however spatial variability appear with some parts of the segment showing more deterioration. The 2009 measurements show a continuation of this trend, but the general evaluation show a slight amelioration of the condition which is not expected. By 2013, the values remain consistent with previous years, but with a slight deterioration that make sense. In 2017 the trend is similar, but deterioration is barely perceivable. By 2021, however, A clear improvement of the condition can be seen which cannot be explained since there is no intervention recorded.

For the indicator I3, the 2000 data present mostly low values, indicating excellent pavement conditions at that time but with some peaks in some areas. By 2004, an expected increase in values is observed, but the trend remains relatively stable which is a strong contrast with the previous measurement. In 2009, variability increases, and the data show an amelioration of the condition which is not expected (except in the beginning of the stretch where a huge peak is recorded. By 2013, this trend continues, with more fluctuations and some peaks reaching values above 2. This time a slight deterioration is visible which make sense considering the normal use. In 2017, however, the condition seems to improve slightly which is not expected and point at a problem in the measurement since there was no intervention recorded. The 2021 measurements are opposed to this trend, and show the expected deterioration also maintaining the spatial variability and the big peak at the beginning of the stretch.

For the indicator I4, data of the four available measurements show lower values and does not allow to infer any normal deterioration which is not as expected. However, some variability is present newer measurements that could indicate deterioration.

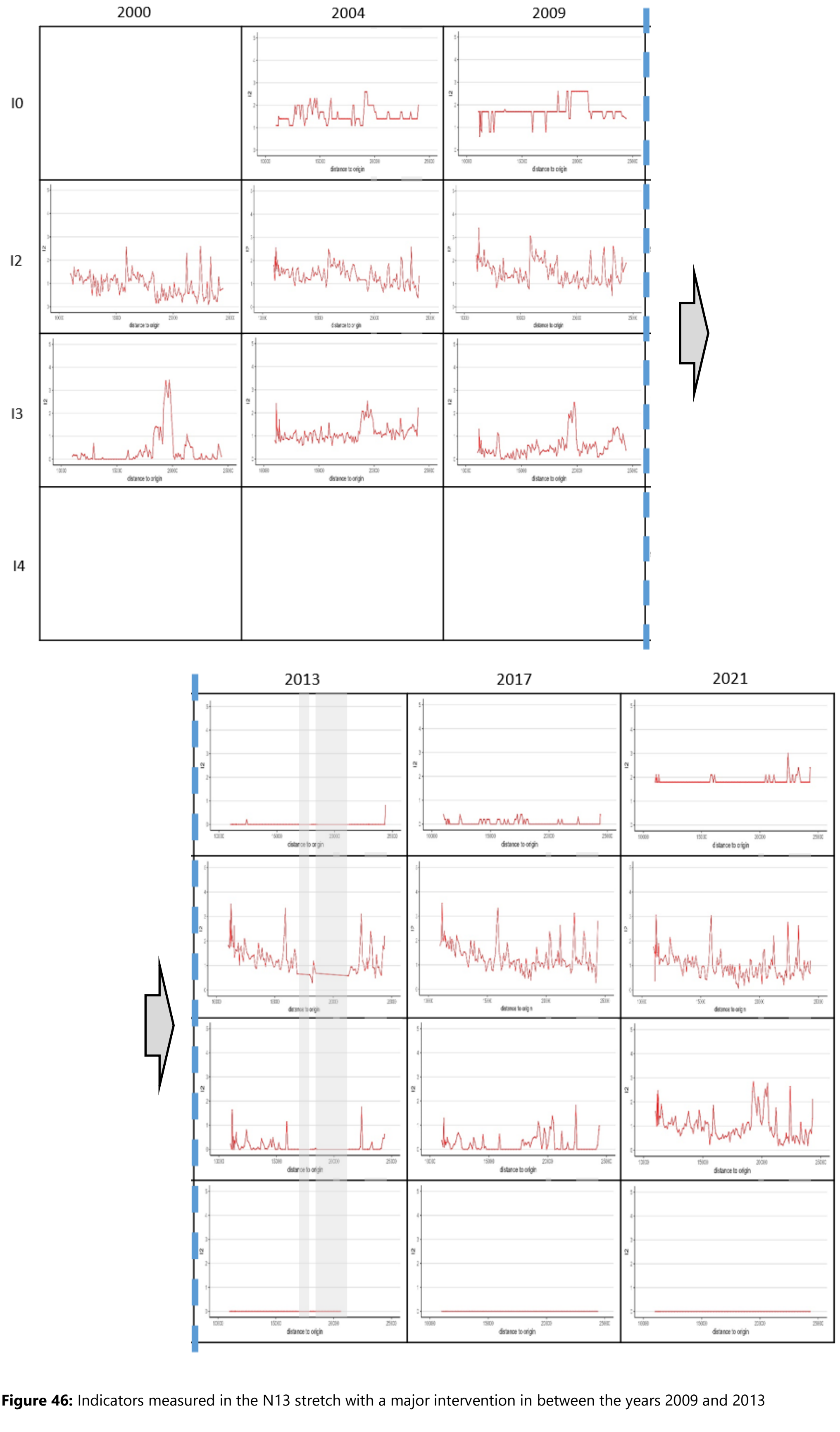

Figure 46: Indicators measured in the N13 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 46 show that for the indicator I0, the first available measurements in 2004 show values fluctuating around 1 to 2, indicating early signs of deterioration but still within a reasonable range. A fair amount of variability is present. By 2009, an slight to no increase in values is observed, except in one section, indicating localized distress. However, in 2013, a noticeable drop in values occurs, with measurements stabilizing near 0, which is expected since a maintenance intervention was performed. By 2017, low values persist, suggesting continued good conditions but variability in the measurements appear. In 2021, a big increase is observed, jumping to values around 2 which indicate a big deterioration in a short period of time. The variability remains low in these measurements which is also a change from previous measurements.

For the indicator I2, the 2000 measurements exhibit a high degree of variability in the measurements and spatially, with values oscillating around 1, suggesting inconsistent pavement conditions. In 2004, a similar pattern is observed, with a general trend of increasing values which is consistent with a normal deterioration. By 2009, variability continues to increase, with peak values reaching close to 2.5, suggesting worsening pavement conditions in some areas and in general an expected deterioration trend. By 2013, this pattern remains similar, which does not make any sense since an intervention took place. Values should have decrease showing the amelioration in the condition. Also, a lack of data in a part of the stretch is not normal. In 2017, the values are quite similar, so there is no sign of the expected normal deterioration. By 2021, the values show an amelioration of the condition which make no sense at all without an intervention and keeping the general trend of the measurement.

For the indicator I3, the data from 2000 shows a mix of low values and sudden peaks, indicating early signs of pavement degradation in certain sections. By 2004, a slight increase in variability is observed, with an expected level of degradation. However, the values are stable now and not spatially distributed differently like the previous measurement. The 2009 data show an amelioration of the condition with lower values which is not expected since there were no interventions. By 2013, values drop consistently with the intervention and the condition of the road is thus very good again. In 2017, a expected worsening occurs, but very slight with values becoming more erratic, and in 2021, this trend continues, suggesting slower but constant pavement deterioration.

For the indicator I4, data first appears in 2013, showing relatively low values, suggesting minimal pavement distress at that time which make sense after an intervention. In 2017, the values remain stable, with no major signs of degradation. By 2021, the pattern remains unchanged which is not expected since some deterioration should be visible.

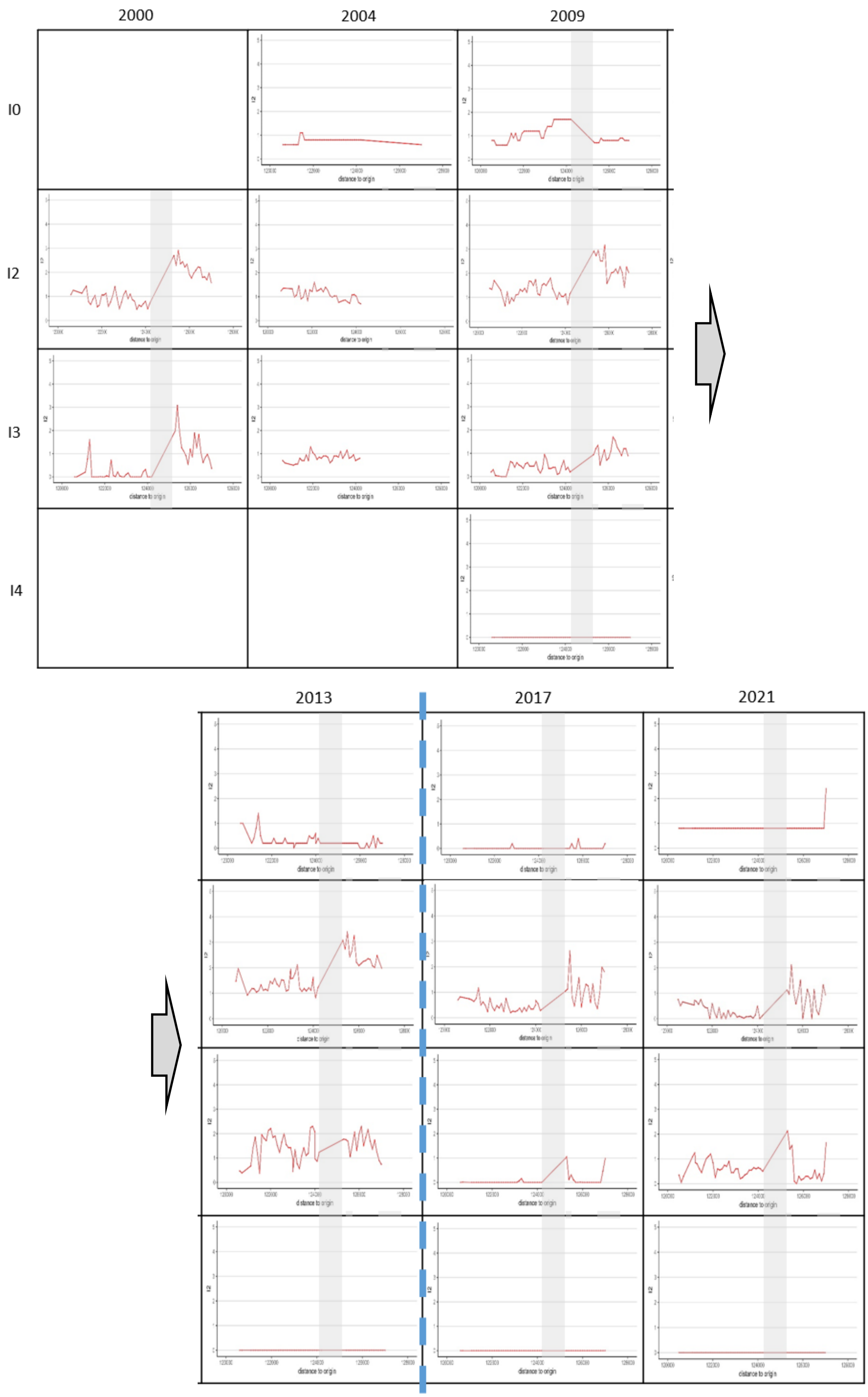

Figure 47: Indicators measured in the N13 stretch with a major intervention in between the years 2013 and 2017

Results on Figure 47 show that for the indicator I0, the first available measurements in 2004 show values around 1, suggesting a road in relatively good condition. Data is however extremely stable which is suspicious. By 2009, a slight but expected deterioration is observed, with values increasing and fluctuations becoming more evident. In 2013, values drop significantly, showing an unexpected amelioration in pavement condition, which is unusual unless maintenance was performed. In 2017, values drop completely which is consistent with the reported intervention. By 2021, an increase is observed which is consistent with regular deterioration.

For the indicator I2, the 2000 measurements show values under 1 for one sector of the stretch but values around 2 for another sector. By 2004, values are stable, and no deterioration can be inferred. Also, data is only available for the first section of the stretch which is not ideal. The 2009 data follow an expected trend, showing progressive slight deterioration. The values for the second section of the stretch are available again and are consistent with the ones measured during the year 2000. In 2013, The trend to deterioration continues very slightly. By 2017 and 2021, after the intervention, the condition increases as expected. It is however strange that instead of a regular slight deterioration, a slight amelioration is observed between 2017 and 1021 which reflect some problems in the measurements.

For the indicator I3, the data from 2000 to 2013 is quite consistent and show the expected slight deterioration in each measurement as well as a general consistency in the spatial variability of the measurements. After the intervention, the data is also consistent with expectations showing an important amelioration of the condition followed by a slight deterioration.

For the indicator I4, data first appears in 2009 and show flat 0 values which is suspicious considering the deterioration that could be observed in the other indicators. All measurements of this indicator show the same trend which does not allow to understand the development of the road condition.

5.5.2 Highway N2

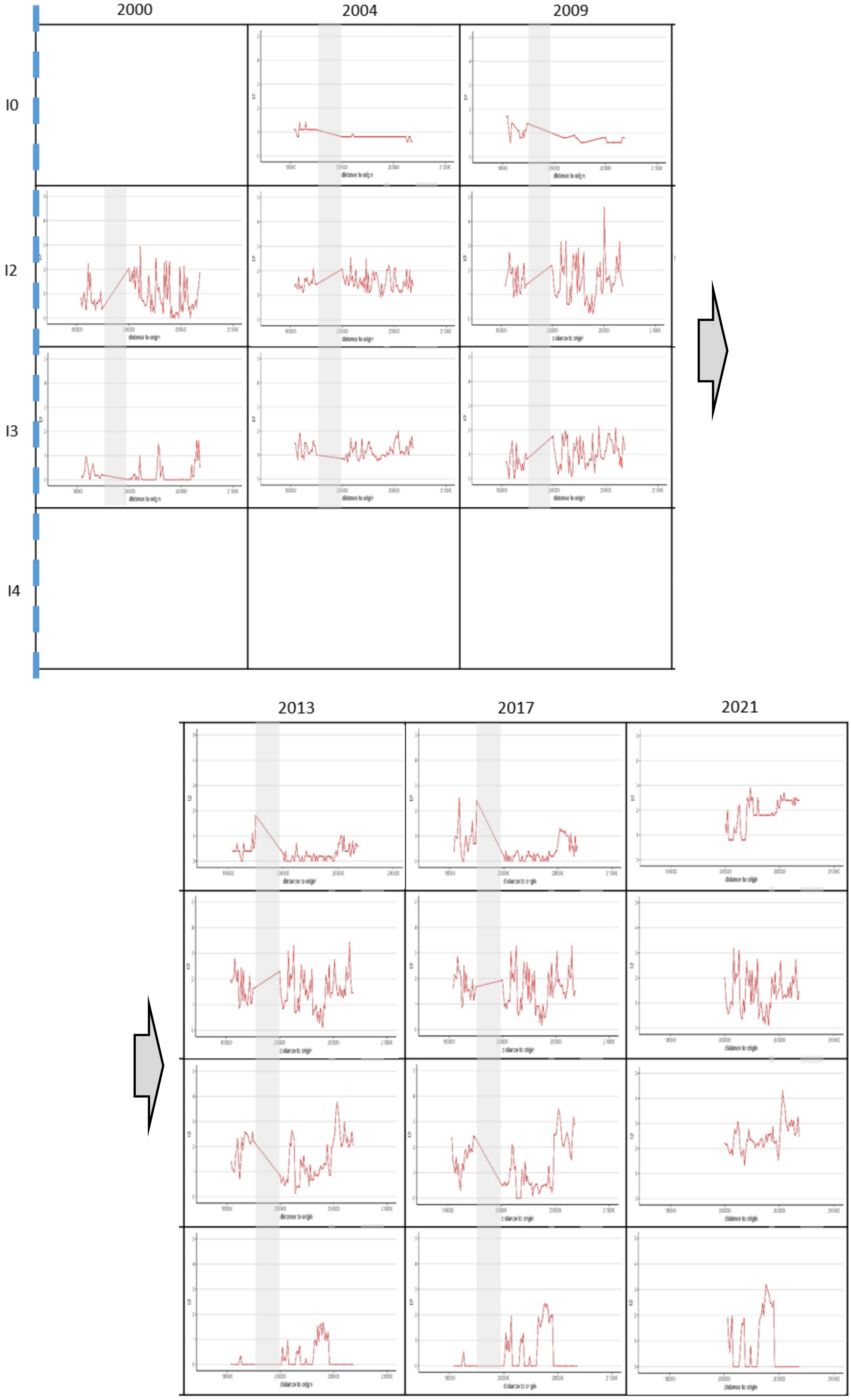

Figure 48: Indicators measured in the N2 stretch with no major intervention after the year 2000

Results in Figure 48 show that for the indicator I0, the first available measurements in 2004 show relatively stable values close to 1, indicating a road in good condition. By 2009, the values remain good but some variability appear and no deterioration or very slight. In 2013, a noticeable decrease in values is observed, suggesting ameliorating pavement conditions. Also, the data show more variability which could signal some problems or difference in the measurements. By 2017, a slight deterioration is observed with more fluctuations and increased peaks. In 2021, the data is only available for one part of the stretch and the deterioration observed is big considering the same time step as before in previous years.

For the indicator I2, the 2000 measurements exhibit high variability, with values fluctuating but most of the time under 2, and lots of spatial variability suggesting inconsistent pavement conditions. By 2004, the values stabilize slightly but still show notable fluctuations. A slight increase in values show the normal deterioration that is expected. In 2009, deterioration continues, with more frequent peaks reaching values above 3, indicating worsening pavement conditions. The variability of the data also increases significantly. By 2013, the trend persists, but no real deterioration can be observed, and the data looks very similar to the one from the previous measurement. The same can be said about the measurement of 2017 and 2021, where in addition a part of the measurement is not available and even a slight amelioration of the condition can be seen.

For the indicator I3, the data from 2000 shows good conditions with a few peaks over values of 1 that indicate some localize problems. By 2004, more variability appears, and a deterioration is clear reaching most values around 1 or above. By the year 2009, the variability increases substantially but the deterioration is not clear, even it seems like the general trend is to an amelioration of the condition which is not expected. By the year 2013, a clear deterioration can be seen from the data but there is a lot of spatial variability. By the year 2017, the patter is quite similar, but a slight amelioration of the condition can be interpreted form the data which is not expected as some deterioration should have happened. By 2021, the deterioration is evident so the values are as expected even if a part of the stretch was not measured.

For the indicator I4, the first available data appears in 2013, showing relatively low values, suggesting minimal pavement distress at that time. But with the presence of peak areas thus showing spatial variability In 2017 a slight deterioration can be observed which is expected. The trend remains similar in the year 2021, with no major changes observed except a slight and expected deterioration.

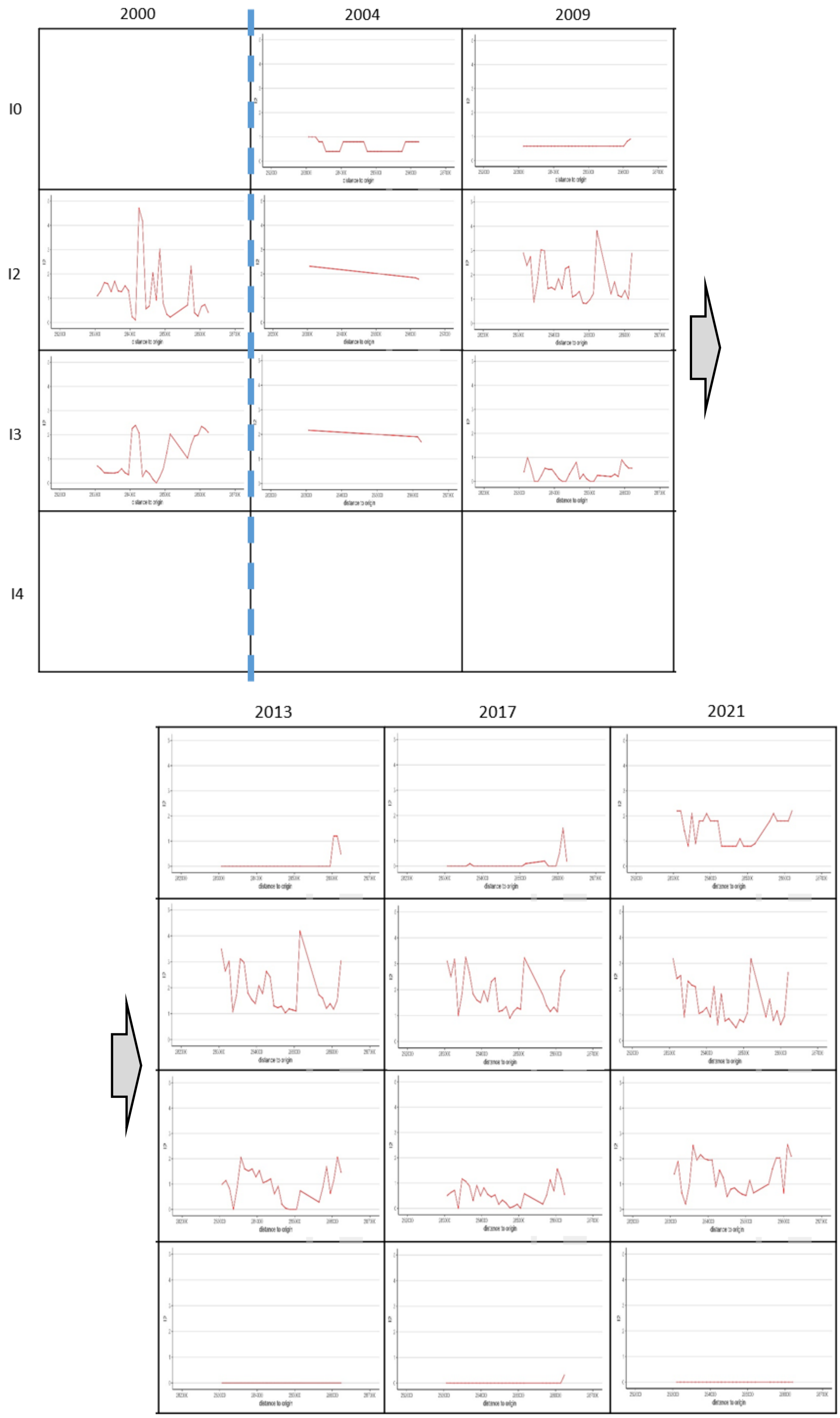

Figure 49: Indicators measured in the N2 stretch with a major intervention in between the years 2000 and 2004

Results on Figure 49 show that for indicator I0, the first available measurements in 2004 show relatively stable values around 1, indicating good pavement conditions. However, in some sections, lower values are observed, suggesting either localized maintenance or inconsistencies in measurement. Considering that an intervention took place a few years before these measurements, the values are high. By 2009, the values remain stable, but in 2013, a drop can be observed, so an increase in the pavement condition which in not expected. In 2017, a few peaks start to emerge, suggesting localized pavement distress. But in general, the condition is excellent, and the deterioration is as expected. By 2021, the values increase significantly, with a higher degree of variability, confirming widespread pavement degradation.

For the indicator I2, the 2000 data shows high fluctuations, with peak values reaching close to 3, indicating inconsistent pavement conditions or measurement anomalies but in general the condition is not optimal. By 2004, after the intervention, measurements show a high value over 2, with no variability at all which suggest that the data is not reliable. By 2009, more reliable data is available, but the condition of the pavement is not good or according to the intervention that happened a few years before. The data also show a strong spatial variability. In 2013 the values are similar, and as expected a small degradation of the condition can be observed. By 2017 and 2021 however, the measurements confirm an amelioration for the condition which is not expected and suggested problems in the measurements.

For the indicator I3, the data from 2000 shows noticeable fluctuations, with signs of pavement degradation in some sections reaching values of 2. By 2004, after the intervention, the data just like for Indicator I2 is not reliable being to stable and not accounting for the improvement of the pavement condition. By 2009, measurements are consistent and show a good condition with values under 1 but some spatial variability is present. In 2013, The data show expected deterioration, and some areas appear to be degrading faster than others. The 2017 data show an amelioration of the condition which does not make sense, so measurement errors or different methods are to be suggested to explain the problem. In 2021, the measurements are most consistent with what should be expected and show a normal deterioration according to the time passed.

For the indicator I4, the first available data appears in 2013 showing low values, indicating minimal pavement distress at that time. By 2017, the trend remains similar, suggesting little deterioration and data shows the same in 2021 which is not as expected since some deterioration should take place.

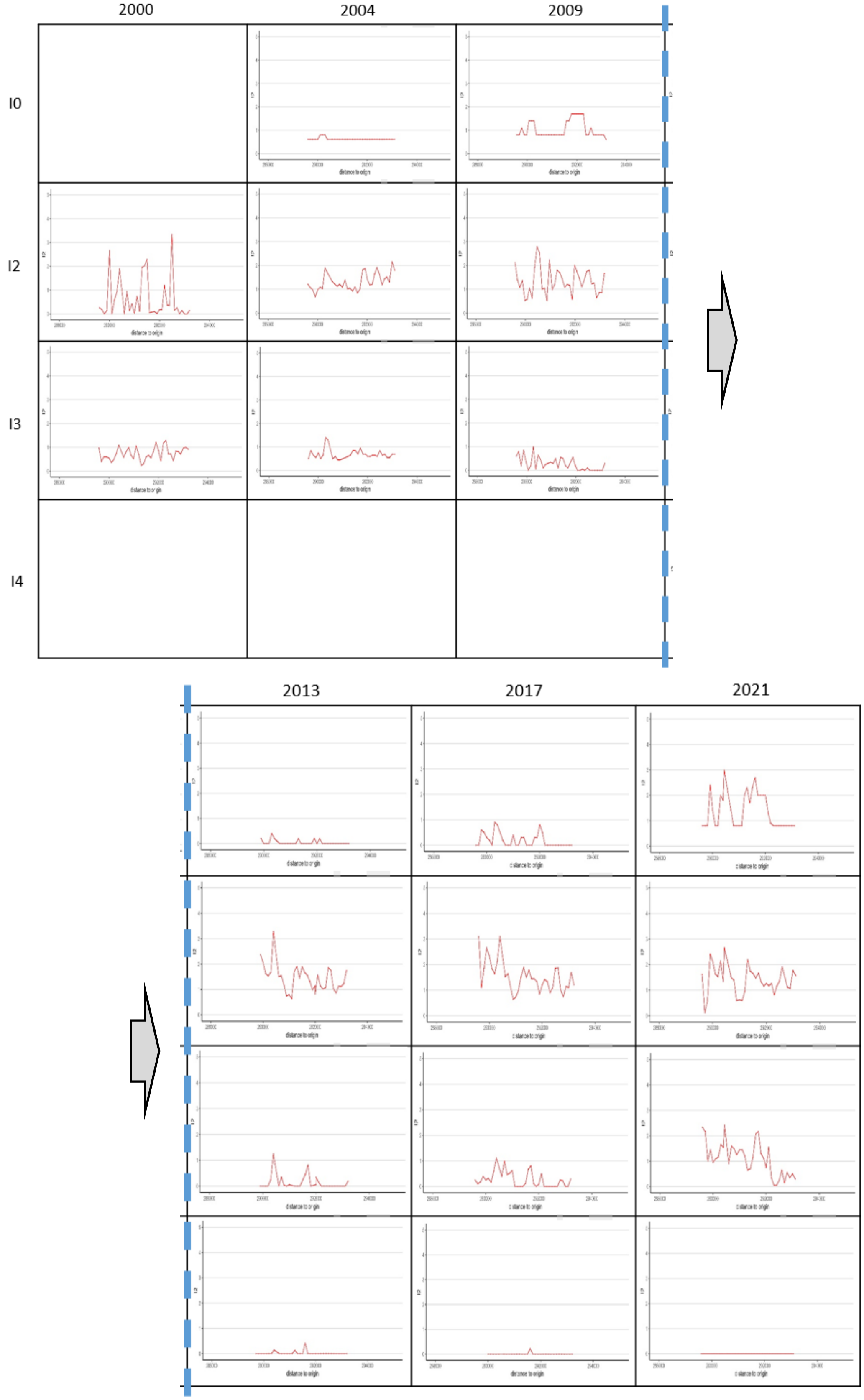

Figure 50: Indicators measured in the N2 stretch with a major intervention in between the years 2009 and 2013

Results in Figure 50 show that for the indicator I0, the first available measurements in 2004 show stable values close to 1, indicating good pavement conditions. By 2009, a slight increase in values is observed, with some segments reaching close to 2, suggesting mild expected deterioration. In 2013, values drop significantly, stabilizing near 0, which is expected cause a maintenance intervention occurred. In 2017, small fluctuations appear, indicating early signs of pavement wear. By 2021, a noticeable increase in values is observed, with more fluctuations and higher peaks, indicating significant pavement deterioration. The deterioration appears to abrupt but is realistic that this could have occur.

For the indicator I2, the 2000 measurements exhibit high variability, with values fluctuating between 1 and 3, indicating inconsistent pavement conditions. In 2004, the trend stabilizes slightly, with values mostly between 1 and 2. By 2009, an increase in peaks is observed, so the variability is high but it is difficult to assess if ta general degradation happened as expected. In 2013, values increase slightly, which is not expected since an intervention occurred. By 2017, the fluctuations increase again, showing signs of progressive pavement degradation. The 2021 data confirm a worsening trend, or perhaps a stabilisation with values remaining high and fluctuating more frequently. The data does not reflect the intervention at all which is a problem.

For the indicator I3, the data from 2000 shows relatively low values with some localized peaks, indicating good pavement conditions with minor distress in certain areas. By 2004, a stable trend is observed, with values remaining in the same average so there is not clear sign of normal deterioration. In 2009, values are even lower suggesting an improvement in the condition that cannot be explained. By 2013, after the intervention, low values are measured but there are a couple of peaks that show some localized deterioration already. By 2017, the trend continues, with an increase in values and more fluctuations. In 2021, the worsening pattern continues, indicating a significant decline in pavement conditions.

For the indicator I4, data first appears in 2013, showing very low values but with some variability and peaks that is not expected just after an intervention. By 2017, and 2021 the values are lower, most of the time 0 which indicate an amelioration in the condition that is not expected.

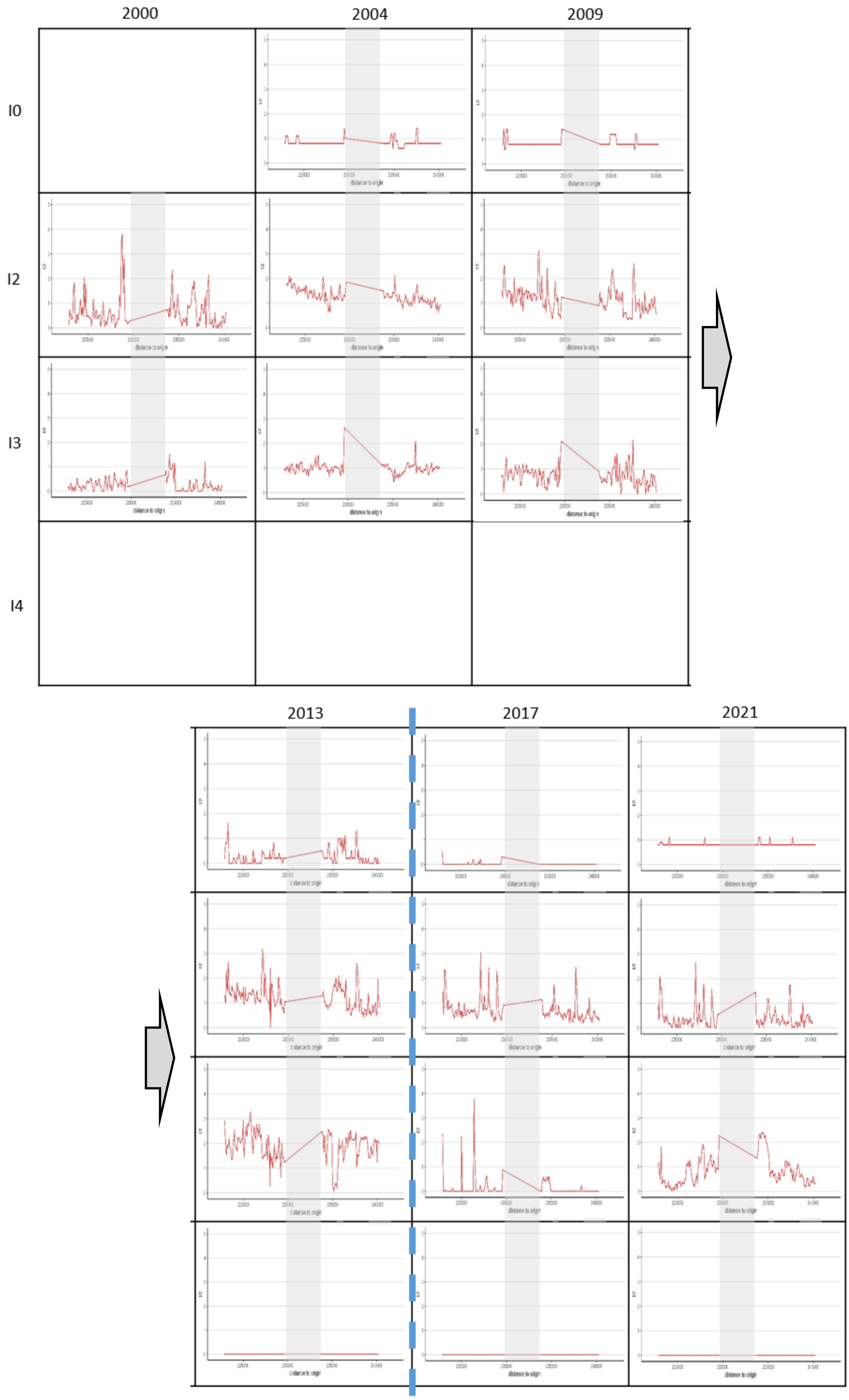

Figure 51: Indicators measured in the N2 stretch with a major intervention in between the years 2013 and 2017

Results in Figure 51 show that for the indicator I0, the first available measurements in 2004 show relatively stable values around 1, indicating good pavement conditions. By 2009, a slight expected deterioration is almost not observable. In 2013, however, the condition improves even if there is a big variability in the measurements. It is not clear why the condition improves with no reason, so measurement problems are possible. By 2017, after the intervention a sharp drop in values is observed, which is expected after maintenance was performed. In 2021, values remain stable, but with a slight increase indicating a normal deterioration trend.

For the indicator I2, the 2000 data show high variability, with values frequently oscillating between 0 and 2, indicating good pavement conditions but with potential measurement noise. By 2004, the trend stabilizes slightly, but some fluctuations remain. A slight deterioration can be observed which is as expected. In 2009, the values increase in variability, with more peaks reaching values above 2, suggesting worsening pavement conditions, but the general trend does not indicate clearly a deterioration. By 2013, this pattern persists, and there is no clear sign of the expected normal deterioration which suggest some problems in the measurements. By 2017, after the intervention the condition improve as expected but there is still notable variability that should not be there after an intervention. The same applies to some extreme spikes that are present. The 2021 data confirm a continued trend of improvement despite maintaining the same peaks and fluctuations, suggesting problems in the measurements.

For the indicator I3, the data from 2000 shows a mix of low values with some variability, indicating early signs of pavement degradation in some sections. By 2004, the trend remains relatively stable, though an expected slight increase in values is observed. In 2009, However, the values show a slight decrease meaning an improvement in the condition which cannot be explained. The increase in variability is also difficult to explain so is possible that measurements methods play a role. By 2013, more fluctuations are present, but this time a normal deterioration is clearly observable and as expected the pavement has deteriorated reaching values in between 2 and 3. In 2017, after the intervention and as expected a sharp drop is observed, similar to I0. Data from 2021 show an expected increase in values suggesting a slight deterioration.

For the indicator I4, data available only from 2013 show values completely flat close to 0 and does not allow further interpretation except that the normal deterioration that should be measurable is not present so some kind of measurement problem is to be considered.

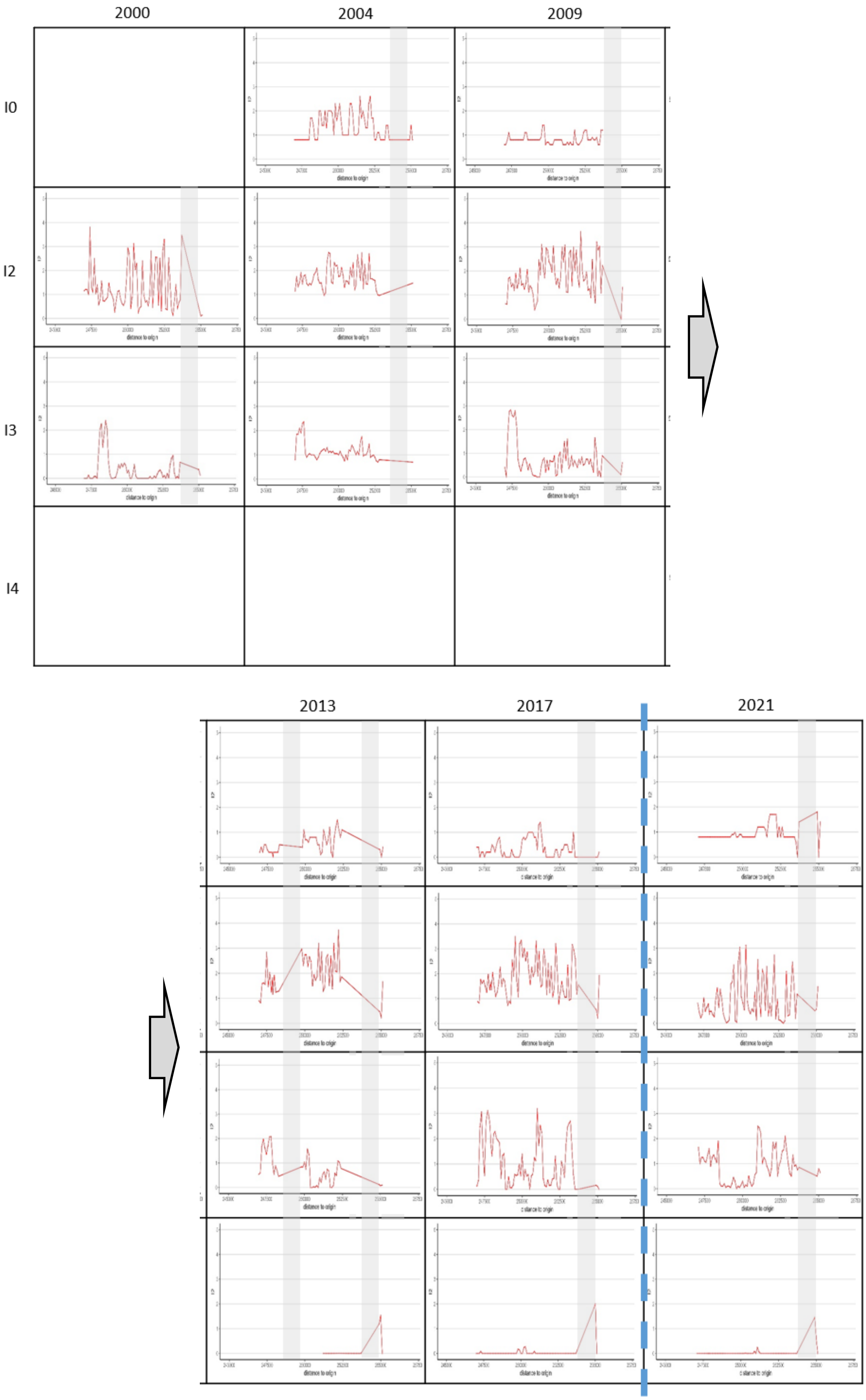

Figure 52: Indicators measured in the N2 stretch with a major intervention in between the years 2017 and 2021

Results in Figure 52 show that for the indicator I0, the first available measurements in 2004 show significant fluctuations, with most values oscillating between 1 and 2, indicating lots of variability but still good pavement conditions across different sections. By 2009, the values show that most of the variability disappeared and not the average is close to 1 suggesting an improvement in the condition which is unexpected. In 2013, there is a slight decline in values, suggesting an amelioration of the condition that is unexpected. In 2017, the values are even lower suggesting further improvement which is not expected. By 2021, after the intervention the values show a deterioration in the pavement condition which is not expected at all and suggest some problems either with the measurements or with the recording of the intervention.

For the indicator I2, the 2000 measurements exhibit high variability, with values fluctuating significantly. By 2004, this pattern persists, but the variability is lower, and some normal deterioration could be inferred. In 2009, the overall trend remains the same, but the high variability is back, with frequent peaks suggesting worsening conditions in some areas. By 2013, the values continue to increase, confirming expected deterioration. In 2017, however, a slight decrease in values is observed, that is not expected. By 2021, after the intervention, values decrease as expected but the variability and peak values remain very high which suggest some problems pointing at wrong measurements or distress and potential localized failures.

For the indicator I3, the data from 2000 shows some signs of pavement distress, with some noticeable peaks but is in general very low indicating good conditions. By 2004, a slight increase in values show the expected deterioration, maintaining an expected amount of variability in the measurements. In 2009, a decrease in values and is recorded, suggesting better conditions which is not expected, however the variability increases suggesting measuring problems. By 2013, this trend continues, and the condition appears even better. In 2017, however, there is a strong deterioration of the condition and big peaks reaching values between 2 and 3 are visible which is expected but there is lots of spatial variability also, so some parts of the stretch improve, and others are worst. By 2021, after the recorded intervention, values do not seem to improve enough and there is still strong spatial variability which suggest that there is a problem with the data.

For the indicator I4, there is not enough data, and the values are constantly around 0 which suggested that the data is not reliable to be able to evaluate the pavement condition.

6 Factors importance

6.1 Indicator I0 surface damage

6.1.1 Variables description

Dependent variable:
Grade difference: Indicator measurement difference between two consecutive survey years for the same road section of 100m. It is calculated by subtracting the value of the second survey year to the value of the first survey year.

Independent variables:
Measuring speed survey year 1 (MS1)

Measuring speed survey year 2 (MS2)

Measuring speed difference (MSD): The difference in measuring speed between the survey year 2 and the survey year 1.

Intervention or not: A binary indicator of whether there was an intervention between the two surveyed years.

Comparison Period: The time elapsed between the two measurements in each pair of surveyed years.

Organ difference: A binary indicator of whether the inspection organization changed between the two survey years (0 if the same, 1 if different).

Climate Factor Differences: Differences in temperature, wind speed, and relative humidity between measurement in survey year 2 and measurement in survey year 1.

Table 2 presents the summary statistics of the data, including their mean, standard deviation, minimum and maximum values. To better understand the dataset, Figure 53 provides a visual representation of the distribution for each variable. The *Grade_difference* distribution shows a roughly symmetrical pattern with a slight positive skew, indicating a tendency for minor road deterioration over time. *MS1* and *MS2* display bimodal distributions, suggesting two common speed ranges for surveys, while MSD exhibits a trimodal distribution, indicating significant variations in survey speeds between years. The binary variable "*Organ_difference*" shows that nearly half of the road sections had a change in the inspection organization between the two survey years, which could introduce variability in the grade assessments when different methods or equipment are used. The distribution for "*Intervention_or_Not*" indicates that major interventions occurred in approximately 15% of the road sections, suggesting a relatively low frequency of significant maintenance activities. The *Comparison_period* distribution is primarily bimodal, reflecting a more or less stable survey schedule. *Temperature_difference, Wind_speed_difference*, and *Relative_humidity_difference* all demonstrate approximately normal distributions centred near zero. However, it can be seen that many pairs of data show important climate differences like for example differences over 10 °C in temperature even reaching 20 °C many times. These distributions offer valuable insights into the nature of the data and potential factors influencing road condition variability. These tendencies will be put to test by more sophisticated modeling techniques to unravel the complex relationships between these variables and changes in road surface conditions.

Data Statistics

Variable	Unit	Mean	Standard Deviation	Minimum	Maximum
Grade_difference	-	0.090	0.759	-2.000	0.600
MS1	km/h	41.887	33.775	0.000	88.000
MS2	km/h	59.224	33.429	0.000	101.000
MSD	km/h	17.337	58.291	-88.000	82.000
Organ_difference	-	0.480	0.499	0.000	1.000
Intervention_or_Not	-	0.132	0.339	0.000	1.000
Comparison_period	Years	4.268	0.442	4.000	5.000
Temperature_difference	ºC	1.287	10.654	-27.180	28.440
Wind_speed_difference	km/h	-0.988	5.258	-14.000	13.000
Relative_humidity_difference	%	1.022	13.393	-30.980	33.890

Table 2: Summary statistics of the processed data

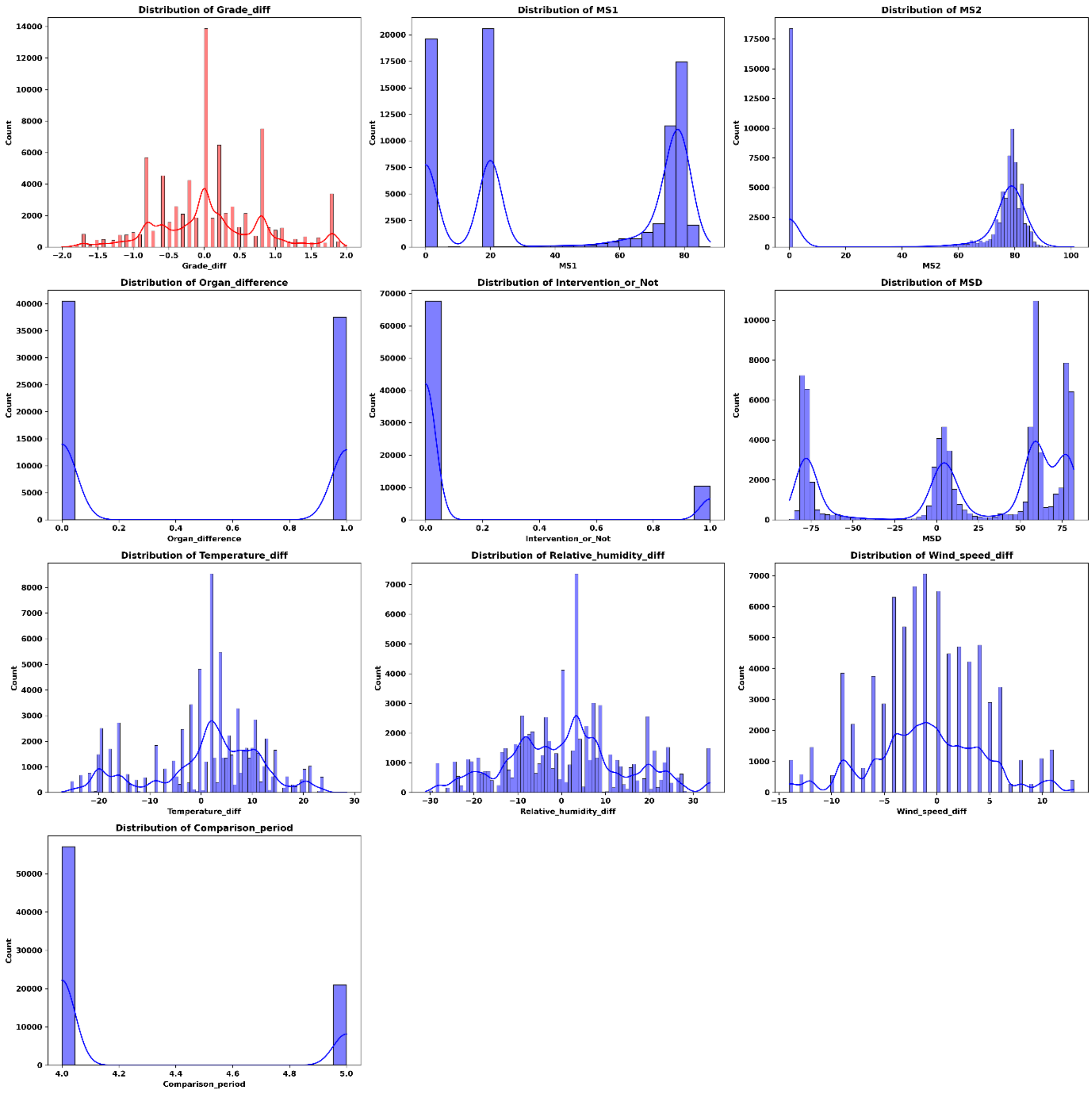

Figure 53: Visualization of the variable distribution

Figure 54 presents a correlation matrix heatmap that reveals the complex interrelationships among variables affecting road surface conditions. Notable correlations include a moderate positive relationship (0.59) between *Grade_difference* and *MS2*, suggesting higher measuring speeds in the second measurement of comparison may be associated with larger grade differences. The negative correlation (-0.32) between *Grade_difference* and *Intervention_or_Not* indicates that when road interventions occur, they are generally associated with improvements in road conditions. Specifically, this means that when an intervention takes place (*Intervention_or_Not* = 1), the *Grade_difference* tends to be lower or negative, signifying an enhancement in road surface quality compared to the previous survey. High correlations exist between related variables such as *MS1* and *MS2*, and between *MSD* and its component speeds which is expected. *Organ_difference* shows moderate correlations with measuring speeds, hinting at potential procedural changes when inspection organizations change. Climate variables and comparison period demonstrate weak correlations with most other factors. Again, these relationships contribute to understand the complex relational puzzle between variables that will be further explored by advanced modeling techniques to fully understand the factors influencing grade variability.

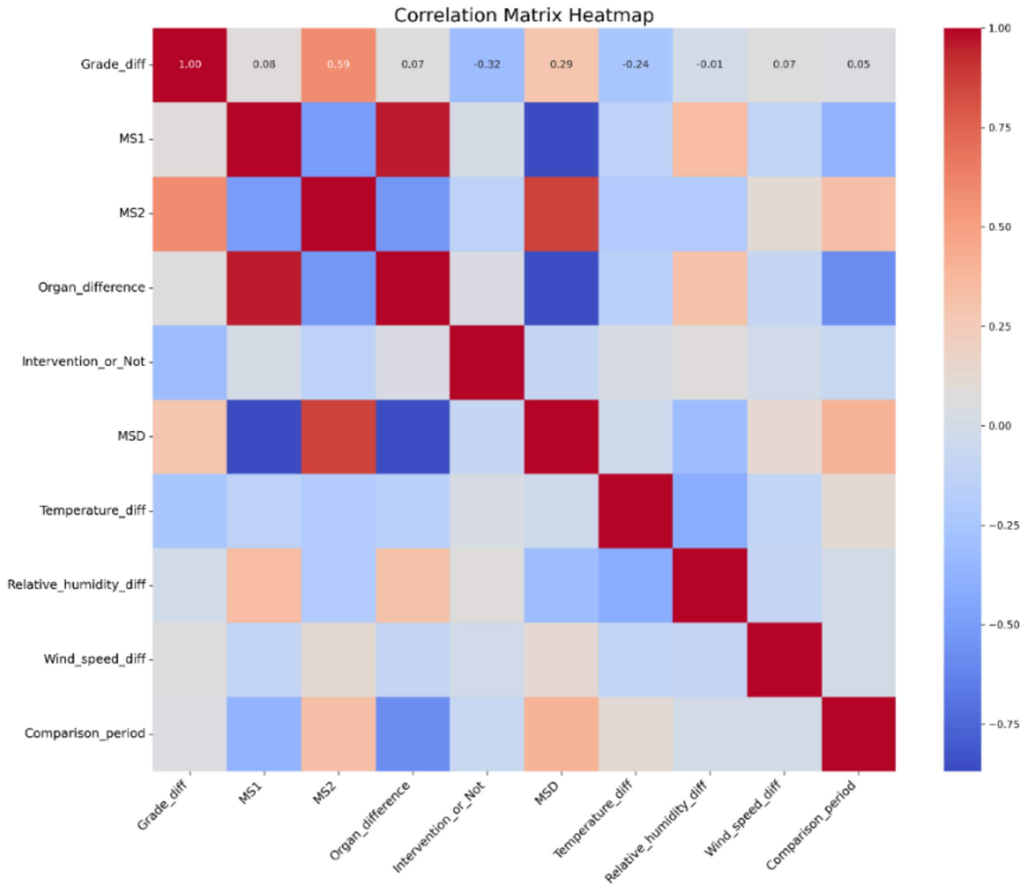

Figure 54: Correlation map of the variables

6.1.2 Data fitting

Tables 3 and 4 present the performance metrics of five ensemble learning algorithms applied to predict road surface condition variability, before and after optimization using BO respectively. The metrics used for evaluation are the R^2 and the RMSE. R^2 indicates the proportion of variance in the dependent variable (*Grade_difference*) that is predictable from the independent variables, with higher values suggesting a better fit. RMSE measures the standard deviation of prediction errors, where lower values indicate better model performance.

Before optimization (Table 3), CatBoost and XGBoost demonstrated the highest predictive power, both achieving an R^2 of 0.689. Random Forest followed closely with an R^2 of 0.681, while LightGBM and AdaBoost showed slightly lower performance with R^2 values of 0.680 and 0.603, respectively. In terms of RMSE, Random Forest performed marginally better than the other algorithms with the lowest value of 0.422, closely followed by CatBoost (0.424) and XGBoost (0.425). After applying BO (Table 4), a notable improvement in the performance of all models was observed. CatBoost remains the top performer, with its R^2 increasing to 0.693 and RMSE slightly improving to 0.422. XGBoost, LightGBM, and Random Forest now show identical performance with R^2 of 0.691 and RMSE of 0.423 (0.422 for Random Forest). AdaBoost demonstrates the most substantial relative improvement among all models, with its R^2 increasing from 0.603 to 0.681 and RMSE decreasing from 0.479 to 0.429. This represents a 12.9% increase in R^2 and a 10.4% decrease in RMSE. Although AdaBoost's final performance still falls slightly below that of the other models, this marked improvement underscores the significant impact of BO on its predictive capabilities. The optimization process has effectively narrowed the performance gap between the different algorithms, with all models now explaining approximately 69% of the variance in the measurement data. This suggests that BO has successfully fine-tuned the hyperparameters of each model, leading to more consistent and improved performance across the board.

Given these results, the optimized CatBoost model, with its highest R^2 (0.693) and lowest RMSE (0.422), is selected for further analysis using the SHAP algorithm to interpret feature importance and impacts on road surface condition variability. This choice is primarily based on CatBoost's superior performance metrics, indicating its strong predictive power for this specific dataset.

Base models – I0

Model	R^2	RMSE
AdaBoost	0.603	0.479
CatBoost	**0.689**	0.424
LightGBM	0.680	0.430
Random Forest	0.681	**0.422**
XGBoost	**0.689**	0.425

Table 3: Result of fitting the data on the base models

Optimized models – I0

Model	R^2	RMSE
AdaBoost	0.681	0.429
CatBoost	**0.693**	**0.422**
LightGBM	0.691	0.423
Random Forest	0.691	**0.422**
XGBoost	0.691	0.423

Table 4: Result of fitting the data on the optimized models

6.1.3 Variable importance

Figure 55 and Figure 56 present the SHAP analysis results for the optimized CatBoost model, providing insights into the importance and impact of each feature on road surface condition variability. Figure 55 displays the SHAP feature importance bar plot, ranking features by their mean absolute SHAP values, while Figure 56 shows the SHAP feature importance violin plot, illustrating the distribution and direction of each feature's impact.

The SHAP analysis reveals that *MS2* is the most influential feature in predicting *grade differences*, with a significantly higher mean absolute SHAP value compared to other features. This suggests that the speed at which the second measurement is taken has a substantial impact on *grade difference* predictions (Adlingen et al. 2013). The violin plot further shows that *MS2* generally has a positive impact on *grade difference*, especially at higher values, indicating that higher speeds during the second measurement tend to be associated with larger grade differences. *Temperature difference* emerges as the second most important feature. Its bimodal distribution in the violin plot suggests that both high positive and high negative *temperature differences* can influence *grade differences*, but in opposite directions (Qin et al. 2022). This highlights the complex relationship between temperature changes and road surface condition variability. "*Organ_difference*" ranks third in importance, with the violin plot displaying a clear separation. This indicates that a change in the surveying organization (value 1) tends to result in higher grade differences compared to no change (value 0). This finding underscores the potential impact of methodological differences between different surveying organizations on road condition assessments. *MS1* and *MSD* also play significant roles in the model's predictions (Adlingen et al. 2013). *MS1* shows a more complex impact pattern. While there are some negative SHAP values, the distribution is predominantly positive, especially for higher feature values. This suggests that higher speeds during the first measurement tend to be associated with larger grade differences. The impact of *MS1* is not uniform across its range, indicating a non-linear relationship with grade differences. *MSD* displays an interesting symmetric distribution around zero, but with distinct impacts at different values. For negative *MSD* values (indicating the second measurement was slower than the first), the SHAP values tend to be negative, suggesting a decrease in grade differences. Conversely, for positive *MSD* values (indicating the second measurement was faster than the first), the SHAP values tend to be positive, suggesting an increase in grade differences. This symmetry implies that the relative change in speed between measurements, rather than just the absolute speeds, plays a crucial role in predicting grade differences. The magnitude of the impact appears to increase as the absolute value of *MSD* increases, highlighting the importance of consistent measurement speeds in road condition assessments.

Figure 55 reveals that both low and high values of the comparison period are associated with negative SHAP values. This suggests a non-linear relationship between the *time between measurements* and *grade differences*. Shorter and longer comparison periods tend to be associated with smaller *grade differences*. This complex relationship may reflect various factors such as the timing of natural degradation processes, maintenance cycles, or seasonal effects on road conditions. The finding underscores the importance of carefully considering the timing of road condition assessments and suggests that the relationship between measurement frequency and observed changes in road conditions is not straightforward. Environmental factors such as *wind speed difference* and *relative humidity* difference show relatively symmetric distributions around zero, indicating varied but generally smaller impacts on grade differences. These factors appear to have more nuanced effects on road surface condition variability.

Lastly, the "*Intervention_or_Not*" primarily shows the impact when interventions are absent (value = 0), which is associated with positive SHAP values. This indicates that the absence of interventions tends to increase grade differences. The plot does not clearly show the impact when interventions are present (value = 1) due to the limited number of data points for this case, as evidenced in Figure 53 (the variable distribution graph). The scarcity of intervention cases makes it difficult to draw definitive conclusions about their impact from the SHAP plot alone. However, the positive SHAP values associated with no intervention suggest that interventions, when they do occur, might have a mitigating effect on *grade differences*. This interpretation aligns with the expectation that road maintenance or repair activities would improve road conditions, but the limited data for intervention cases in the dataset means this conclusion should be treated with caution. Further investigation with a more balanced dataset would be necessary to fully understand the impact of interventions on road surface condition variability.

These findings highlight the complex interplay of factors affecting road surface condition variability. The significant impact of measurement speeds suggests that measurement methodology plays a crucial role in assessing road conditions. Environmental factors, organizational changes, and interventions also show influences on grade differences.

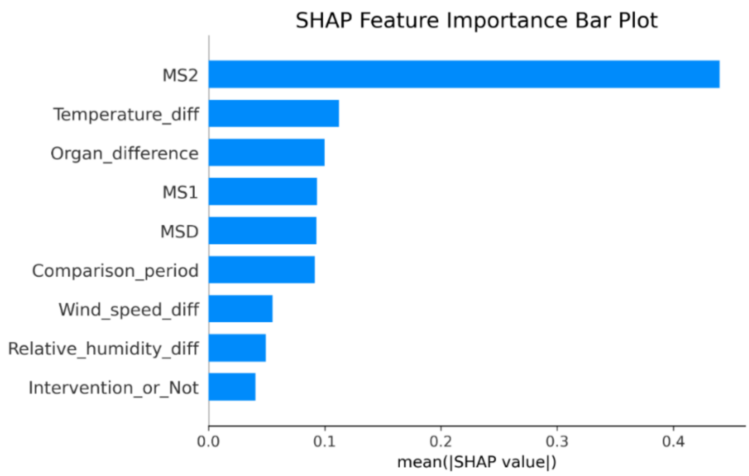

Figure 55: SHAP feature importance for the optimized CatBoost model

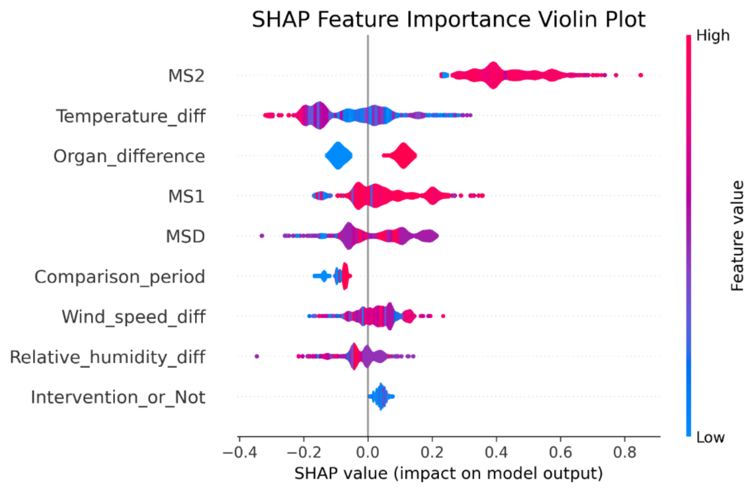

Figure 56: Directional impact of the features on the model output

6.2 Indicator I2 Longitudinal evenness

6.2.1 Variables description

Six input variables were selected for the models, with the measurement interval (100 m) defined as the unit of the road segment.

<u>Dependent variable</u>:
Grade difference (mean_note_final_diff): Indicator measurement difference between two consecutive survey years for the same road section of 100m. It is calculated by subtracting the value of the second survey year to the value of the first survey year.

<u>Independent variables</u>:
Intervention: This variable indicates whether an intervention was performed on the same road segment between two consecutive survey years. If an intervention occurred, the value is set to 1; otherwise, it is set to 0.

Test year length: This represents the time interval (in years) between two consecutive surveys measurements on the same road segment.

Project difference: This variable capture whether the same company conducted the measurements in two consecutive surveys. If the measurement company is the same, the value is 0; if different, the value is 1.

Temperature difference: The difference in air temperature (*tmpf*) between two consecutive measurement surveys on the same road segment.
Dew point difference: The difference in dew point temperature (*dwpf*) between two consecutive measurement surveys on the same road segment.

Relative humidity difference: The difference in relative humidity (*relh*) between two consecutive measurement surveys on the same road segment.

After filtering out measurement years without weather data and those where the specific measurement company was not identified, a dataset containing 95,439 rows was obtained. It was then visualized the distribution of this data to gain further insights (Figure 57).

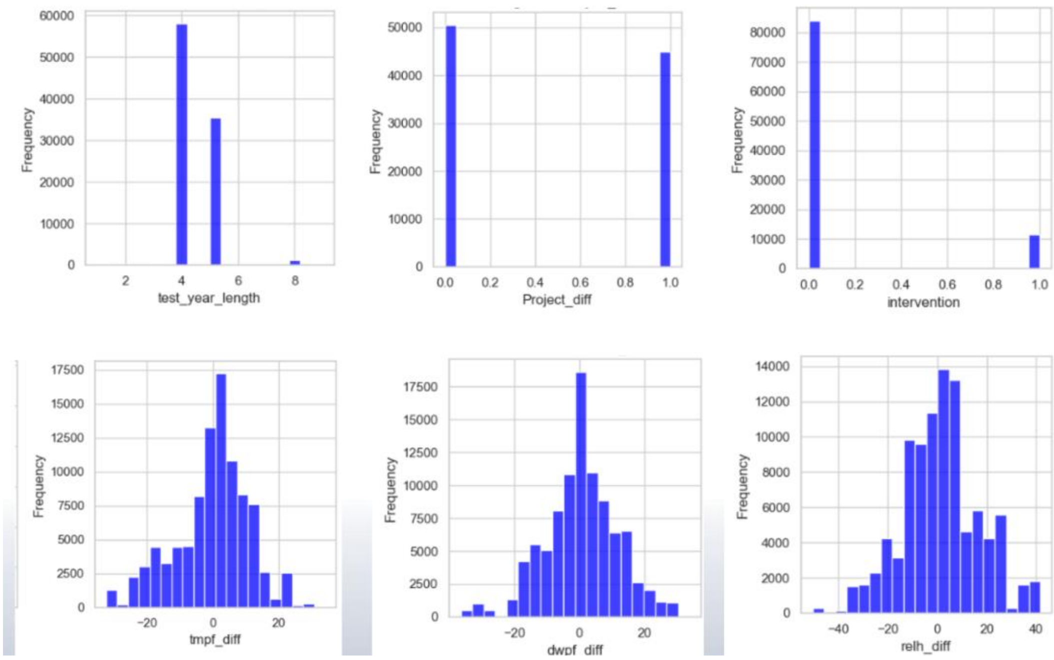

Figure 57: Distribution of input variables for indicator I2

From the visualization, it is evident that most road segments under consecutive measurement years did not undergo any interventions. The typical interval between test years is 4 to 5 years. Additionally, the temperature and relative humidity differences between consecutive measurements are generally significant, indicating that, in measurements taken in different years, the weather conditions can differ substantially. As for the project or company differences, the proportions are relatively close, but there are slightly more road sections where the same company performed the measurements in consecutive years.

It was also aimed to investigate the correlation between independent factors and dependent variables (Figure 58). To achieve this, heatmaps for each *Filiale* were generated (not shown here), as well as conducted a correlation analysis by combining the data from all *Filialen*. The results show low correlations between the variables. The correlations between the *mean_note_final_diff* (the I2 difference) and the values of independent variables are low, suggesting that the index value difference is influenced by multiple factors, making it difficult to infer a direct influence of one variable. However, special attention should be paid to the variable intervention. Regardless of whether the analysis was done for an individual *Filiale* or for the combined data from all *Filialen*. Intervention consistently shows the highest correlation with the I2 difference compared to other factors. Although the correlation value hovers only around -0.3, it is negative, indicating a trend where interventions are associated with a reduction in the index value difference, which aligns with the expectations. This negative correlation suggests that interventions tend to reduce the pavement index difference, as a smaller index value indicates less deterioration. In some cases, the intervention can even lead to a significant improvement, where the index value in the next measurement year is lower than the previous one, potentially resulting in a negative difference.

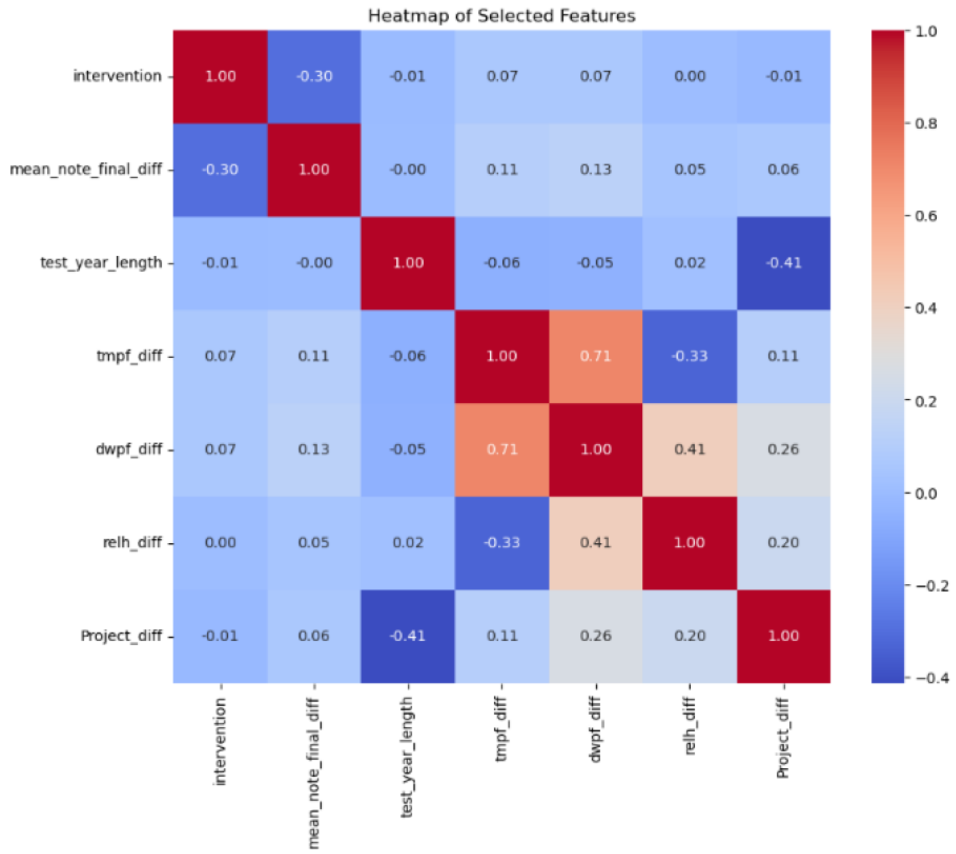

Figure 58: Heat map of correlations between variables

6.2.2 Data fitting

The overall fitting performance of all models was relatively weak (Table 5). The highest performance was observed with the CatBoost model, achieving a R^2 value of 0.3508. parameter tuning was undertaken, including increasing the complexity of the tree-based models, such as adjusting the depth and the number of estimators, but these adjustments did not lead to a significant improvement in the model's performance. In addition to tree-based models, it was also experimented with other advanced methods, such as neural networks and ensemble learning techniques. However, these approaches also failed to yield substantially better results.

These findings suggest that the target variable is likely influenced by a considerable amount of randomness and data noise or simply by other variables not measured. Such factors limit the ability of any model, even complex ones, to achieve a high R^2 value. This observation aligns with the conclusions drawn during the data exploration stage, where patterns indicating that the target variable is affected by multiple unpredictable factors were noticed. Consequently, it becomes challenging for machine learning models to fully capture the underlying relationships in the data. However it is still interesting to explore more in detail the 35% variability that the recorded variables are able to explain.

Tree-based ensemble models – I2			
Model	**Parameters**	**R^2**	**RMSE**
Random Forest	n estimators = 50 max depth = None min samples split = 2	0.3506	0.513
LightGBM	boosting type = "gbdt" n estimators = 200 max depth = None	0.3473	0.516
XGBoost	n estimators = 50 max depth = None	0.3459	0.517
CatBoost	Iterations = 200 depth = 15	**0.3508**	**0.512**

Table 5: Parameters and results of tree-based ensemble models

6.2.3 Variable importance

Feature importance rankings were conducted for all models and results can be seen in Figure 59. It is evident that the feature '*intervention*' consistently ranks as the top important. This finding is consistent with the conclusions drawn from the heatmap analysis. Weather variables are also important. *Relative humidity* and *temperature difference* playing an important role in explaining the variability of the data.

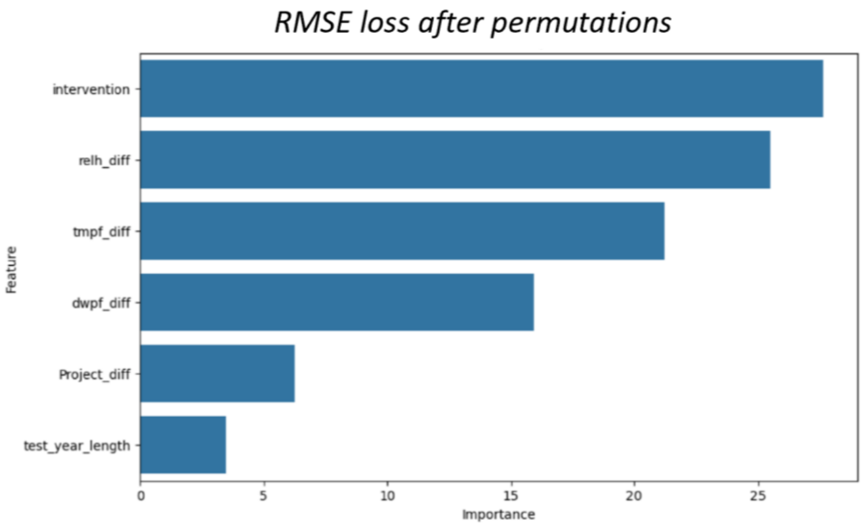

Figure 59: RMSE loss after permutations

6.3 Indicator I3 Transversal evenness

6.3.1 Variable description

Seven input variables were selected for the modeling efforts, with a measurement interval of 100 m defined as the unit of the road segment.

<u>Dependent variable</u>:

Grade difference (note_diff): Indicator measurement difference between two consecutive survey years for the same road section of 100m. It is calculated by subtracting the value of the second survey year to the value of the first survey year.

<u>Independent variables</u>:

Intervention: This variable indicates whether an intervention was performed on the same road segment between two consecutive measurement years. If an intervention occurred, the value is set to 1; otherwise, it is set to 0.

Company: These variables captures whether the same company conducted the measurements in two consecutive survey years. If the measurement company is the same, the value is 0; if different, the value is 1.

Temperature difference: The difference in air temperature between two consecutive measurement surveys on the same road segment.

Wind direction difference: The difference in wind direction between two consecutive measurement surveys on the same road section.

Wind speed difference: The difference in wind speed between two consecutive measurements surveys on the same road segment.

Visibility difference: The difference in visibility between two consecutive measurement survey on the same road segment.

Figure 60 provides a visual representation of the distribution for each variable. The *Grade difference* distribution shows a roughly symmetrical pattern which could indicate that the slight increases in values determined by the normal deterioration could be counteracted by the more important decrease in values due to maintenance works (i.e. interventions). The distribution for "*Intervention_or_Not*" indicates that major interventions occurred in a small percentage compared to the stretches with no intervention suggesting a relatively low frequency of significant maintenance activities. The binary variable "*company*" shows that nearly half of the road sections had a change in the inspection organization between the two survey years, which could introduce variability in the grade assessments. Weather variables difference show approximately normal distributions centred near zero. However, it can be seen that many pairs of data show important climate differences like for example differences over 10 °C in temperature even reaching 20 °C many times.

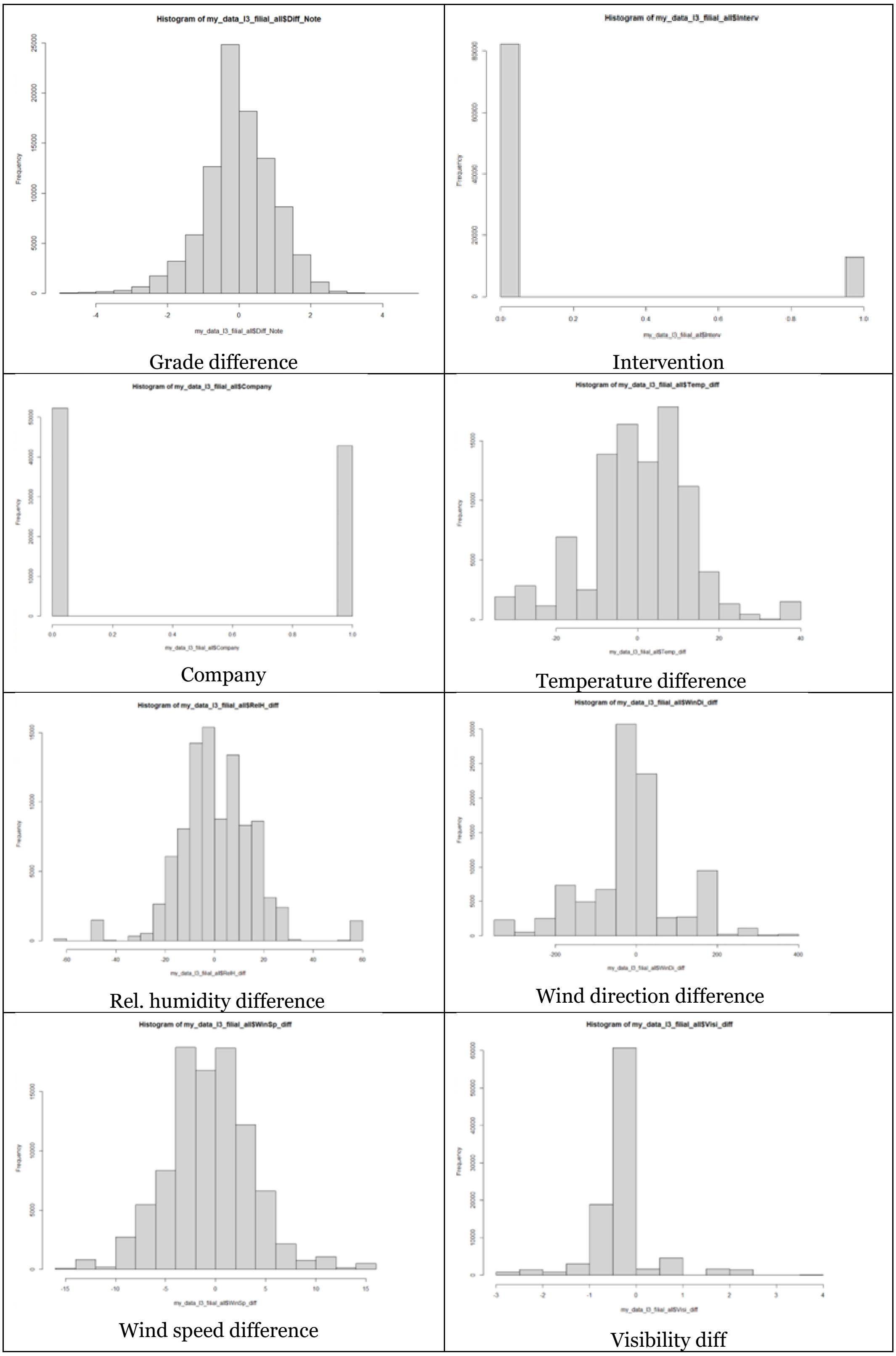

Figure 60: Distribution of variables used in the modelling

The correlation exploration (Figure 61) shows low correlations between the explanatory variable, and also between the explanatory variables and the explained variable.

As expected, weather variables show some correlation, but not very high so it still makes sense to use them in the modelling. The different companies contracted for consecutive surveys show the higher correlation with the different in the grade so it is possible that it plays an important role. Further analysis will help validate or discard this hypothesis.

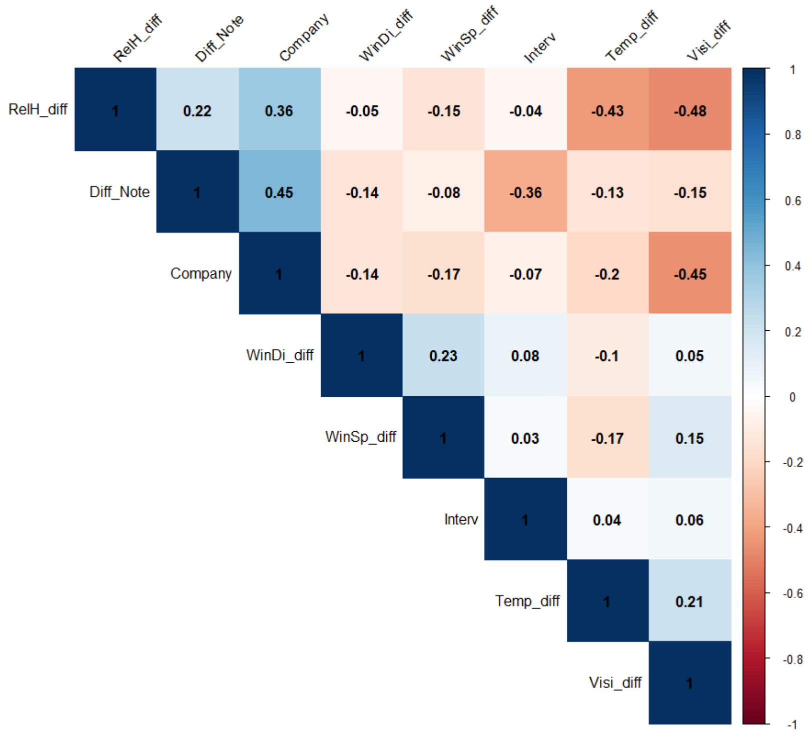

Figure 61: Correlation heat map between variables

6.3.2 Data fitting

The fitting performance of all models accounted for around 50% of the variability which is a moderate value (Table 6). The highest performance was observed with the LightGBM model, achieving a R^2 value of 0.4958 and a RMSE value of 0.441. Parameter tuning, including increasing the complexity of the tree-based models, such as adjusting the depth and the number of estimators, did not lead to a significant improvement in the model's performance or even deteriorating the fittings.

These findings suggest that the target variable is likely to be influenced by other non-measured variables and some inherent randomness and data noise. This observation aligns with the conclusions drawn during the data exploration stage through correlations, where patterns indicating that the explained variable is affected by multiple explanatory factors but not very strongly by any of them were noticed. Consequently, it is challenging for machine learning models to fully capture the underlying relationships in the data. However, it is still interesting to explore more in detail the 50% variability that the recorded variables are able to explain.

Tree-based ensemble models – I3			
Model	**Parameters**	**R^2**	**RMSE**
Random Forest	n estimators = 50 max depth = None min samples split = 2	0.4956	0.457
LightGBM	boosting type = "gbdt" n estimators = 200 max depth = None	**0.4958**	**0.441**
XGBoost	n estimators = 50 max depth = None	0.4803	0.441
CatBoost	Iterations = 200 depth = 15	0.4808	0.474

Table 6: Parameters and results of tree-based ensemble models

6.3.3 Variable importance

Feature importance rankings were conducted for all models and results can be seen in Figure 63. From the results it can be seen that having different *companies* on consecutive survey years play the most important role in the variability of the dataset. It can be seen that '*intervention*' is also important which makes sense considering that a big amelioration in the condition and thus a strong decline in the grade values is expected after an intervention. These findings are consistent with the conclusions drawn from the heatmap analysis. Weather variables play also a role, but their combined effects must be considered to be able to compete with the previously mentioned variables, in particular *Temperature* and *Relative humidity*.

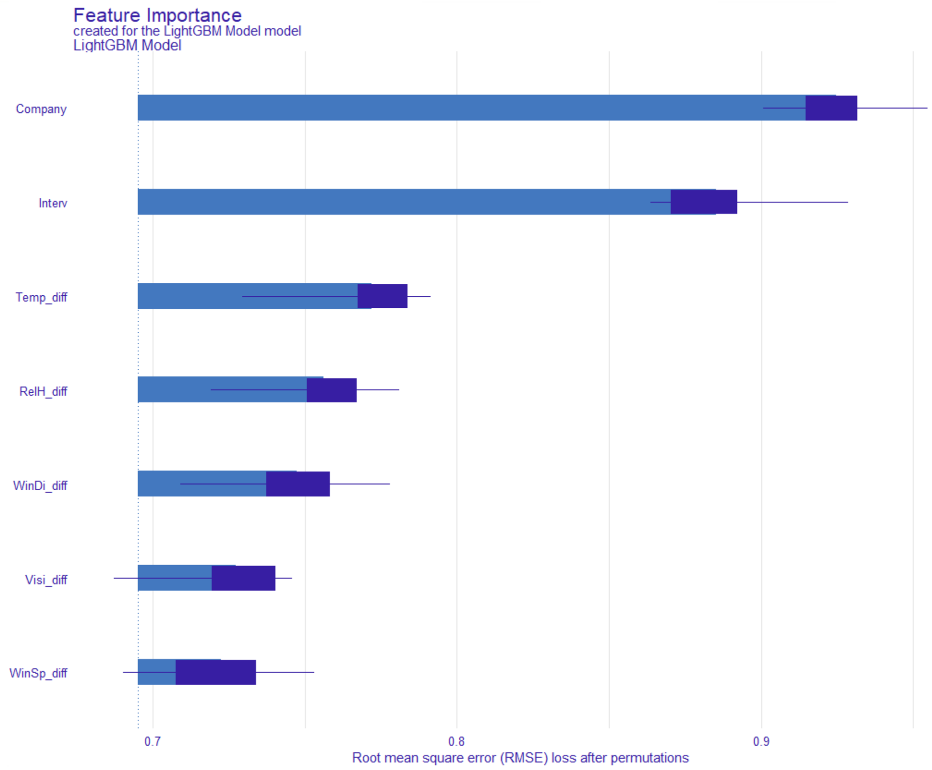

Figure 62: Results of the variable importance study performed with the root mean squared error loss after permutation method

6.4 Indicator I4 Surface friction

6.4.1 Data description and data processing

A more comprehensive analysis was conducted for the indicator I4. For this parameter, the road friction coefficient is measured by the RK-SKM 80 test vehicle, whose standard testing speed is 80 km/h. In addition to the target value, the database includes corresponding information such as lane, distance markers, distances from the markers, project details (including inspection companies), measurement equipment, dates, and speeds. The SKM-80 records data every one hundred meters. In practical testing, it is challenging to guarantee the consistency of the starting point location, which fluctuates within a range of 5 meters. It was assumed that the road coefficient friction remains the same within this range of starting point fluctuations. To standardize the data, for this indicator the nearest-neighbour principle was used to ensure that the data collected in different inspection years share consistent location sections (every 100 meters from the start location to the end location). This standardization process commenced at the 0-meter markers and aligned the data at 100-meter intervals. Analysis was conducted on the centre lane of each road and will consider stretches of 1 km as explained later with the created average difference variable.

Data of the highway N13 was chosen, spanning twenty years, as an example, displayed in Figure 63. It can be seen that the entire road section was not inspected within a year, which was often the case for all highways. However, there were some exceptions, like for the N13 in 2001. In general, the inspection was completed over several years. Since the friction measurement data were expected to vary annually, the data covered by the green boxes in Figure 63 was specifically selected, as they correspond to the same road section across four or five inspection years. Consistent criteria for region and year selection across all road sections was applied.

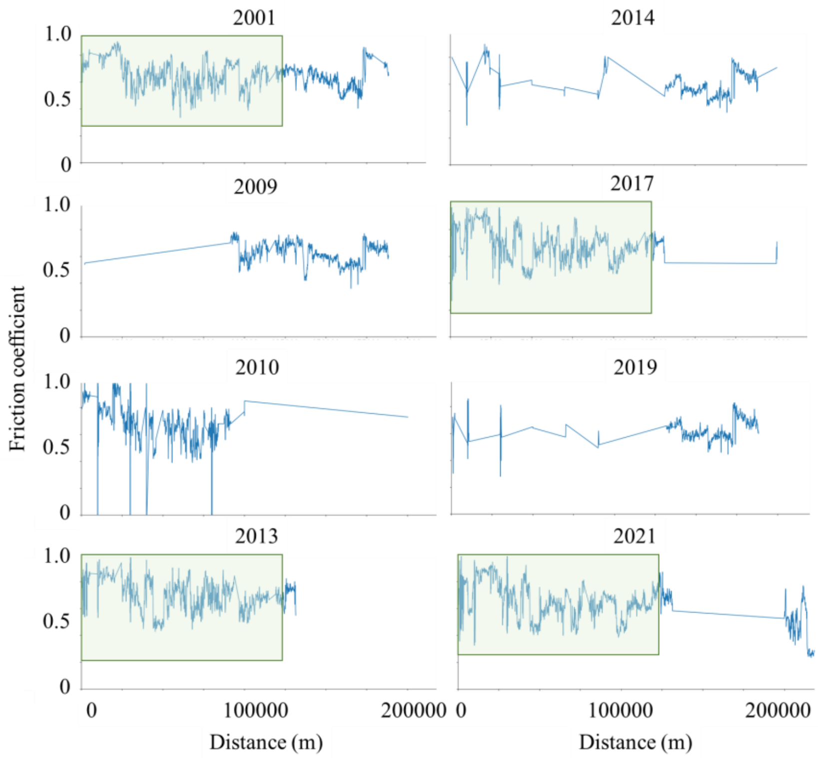

Figure 63: Example of the data on the highway N13 from years 2001, 2009, 2010, 2013, 2014, 2017, 2019 and 2021. The green box indicates when consistent data was found to be used in the analysis

The objective was to ensure completeness and comparability of data from different years for the same road section. Table 7 presents a comprehensive list of all highway sections used in the study of the indicator I4, encompassing 12 highways, denoted as N1 through N16 (excluding N8, N10, N13, and N15). Specific details are also provided on the starting and ending points, road section lengths, and the selected inspection years. The cumulative length of the selected roads totals 474 kilometres.

Highways (analysed sections)				
Road	Start (km)	End (km)	Length (km)	years
N1	30	150	120	2001, 2009, 2013,2017,2021
N2	200	270	70	2001, 2010, 2013, 2017,2021
N3	165	180	15	2001, 2009, 2015, 2019
N4	95	105	10	2001, 2010, 2013, 2019,2020
N5	8	54	46	2009, 2013, 2017,2021
N6	1	38	37	2001, 2009, 2014,2022
N7	2	30	28	2001, 2009, 2013, 2018
N9	0	50	50	2001, 2009, 2013, 2017, 2021
N11	1	3	2	2009, 2017, 2022
N12	0	70	70	2001, 2009, 2013, 2017, 2021
N14	22	28	6	2001, 2009, 2013, 2019
N16	18	38	20	2001, 2009, 2013, 2017, 2021
Total			474	

Table 7: Highway section used in the analysis, including information about the number of kms.

Data from two consecutive inspection years for the same road section was compared, as depicted in Figure 64. The blue line represents data from the former year, while the red line represents data from the subsequent year. The blue area indicates a decrease in measured values during this period, while the red area indicates an increase in measured values. In addition to the inspection data, records of road interventions carried out in Switzerland over the past two decades were also obtained, including specific maintenance locations and years. The intervention information with inspection data using distance and time were matched. In Figure 64, the road sections that underwent maintenance between two inspection years are masked. It can be observed that variability within the same road sections, particularly the unexplained variability involving value increases in the latter year without intervention and decreases with intervention.

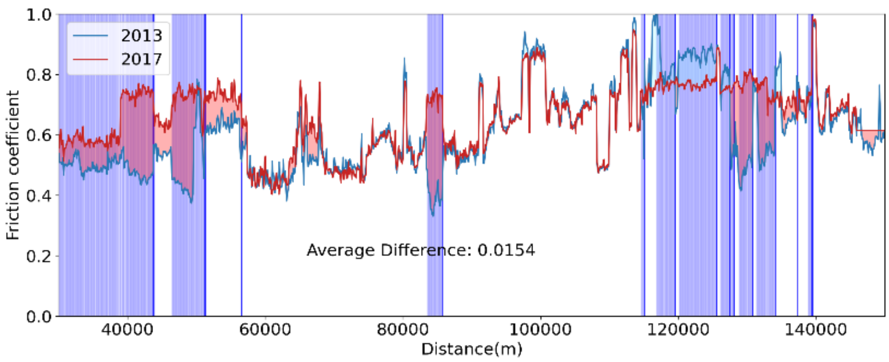

Figure 64: Example of the matching of two years (2013 and 2017) including intervention information in purple.

6.4.2 Variables description

Even though road managers can observe such variability, there is currently no unified quantitative assessment metric to quantify data variability between two inspection years. Therefore, this study introduced the average difference (AD) as a quantitative evaluation metric to describe the extent of data variability. The AD is calculated as follows:

$$AD = \sum_{i=0}^{n} \frac{(V_{2i} - V_{1i})}{n} \tag{8}$$

n denotes the quantity of data (number of measurements in one interval), while V_1 and V_2 represent the measurement friction values (friction coefficient in the paper) in the former year and the latter year. If two consecutive missing values (NA) of AD appeared within a data sample, the entire sample was removed to ensure data quality. Therefore, there was a total of 1'322 samples for further analysis.

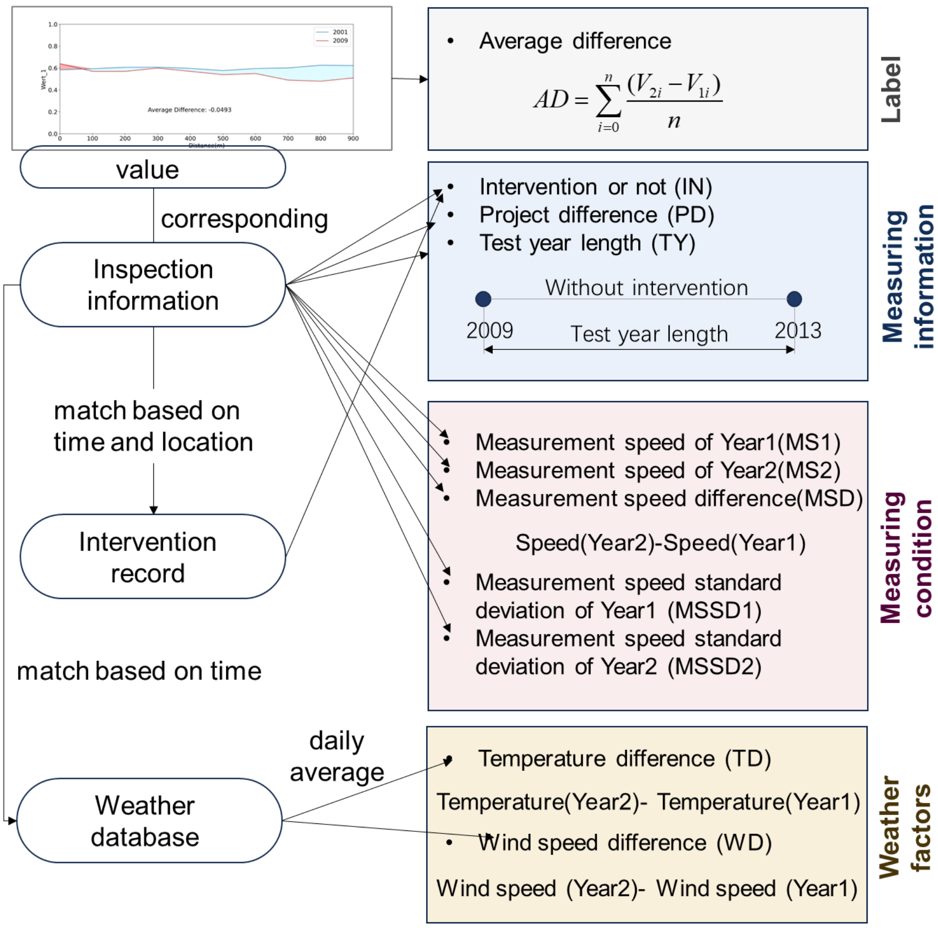

Figure 65: Description of quantitative evaluation metric (AD average difference) and various factors

Additionally, the data was partitioned into 1-kilometer intervals, with each interval sample comprising ten values from two consecutive testing years. Simultaneously, additional information on the potential influence factors from the samples was extracted. These factors were categorized into three main groups: measuring information, measuring conditions, and weather factors. Figure 66 illustrates the categorization, calculation, and sources of the quantitative evaluation metric AD and various factors. Figure 65 displays the value distribution of the factors. The specific characteristics of each factor are described in detail below. The value distribution of AD is essentially a slightly skewed Gaussian distribution centred at a value smaller than 0 (Figure 66(a)).

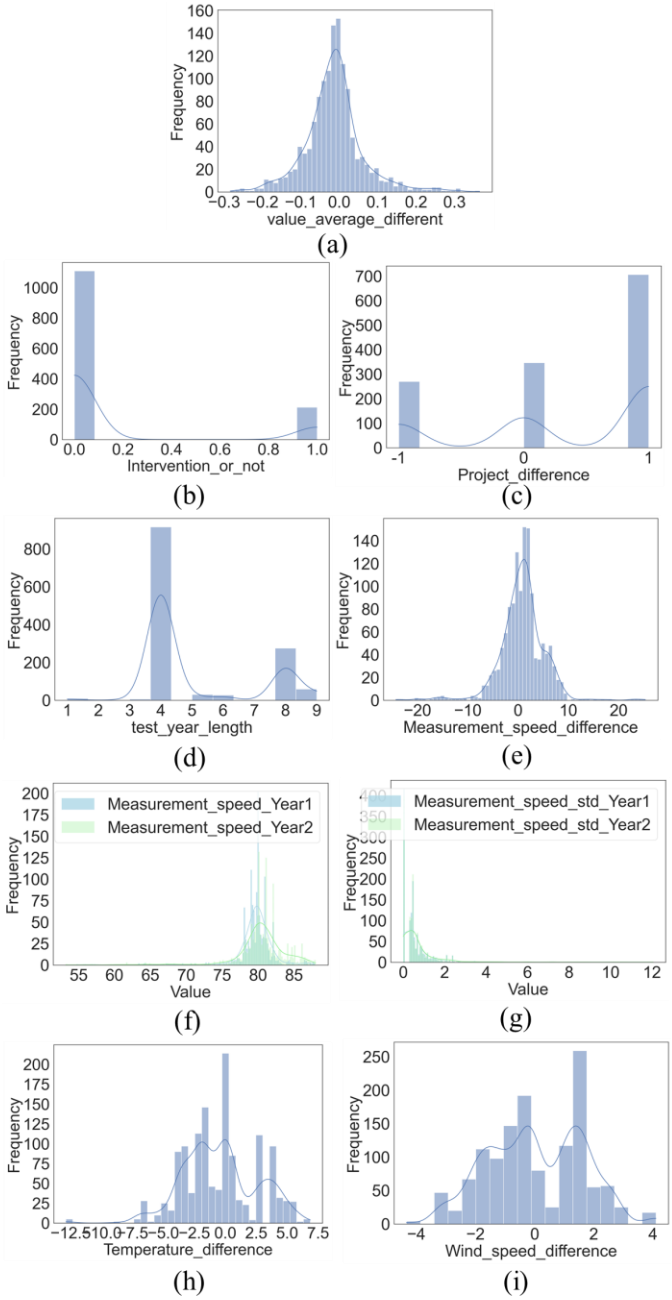

Figure 66: Distribution of the quantitative evaluation metric (AD average difference) and various factors.

Measuring information includes Intervention or not (IN), project difference (PD), and test year length (TY). IN is a binary variable, where 0 signifies the absence of intervention between two inspection years, while 1 indicates the presence of intervention, accounting for 16% of the dataset. PD denotes whether the inspection company remains consistent over two consecutive years. There were two inspection companies involved. Company A conducted the inspections in 2000, 2001, 2013, 2017, 2018, and 2022, while Company B conducted the inspections in 2004, 2009, 2010, and 2021. A PD value of 0 indicates the same companies were involved, while a value of 1 or -1 signifies different companies. If Company A is in the former year and Company B in the latter year, the value of PD is 1. Conversely, if the value of PD is -1, Company B is in the former year and Company A in the latter year. TY represents the time between two inspections. The values for the TY are primarily concentrated at four years, followed by eight years. IN is derived from intervention records, while PD and TY are derived from the inspection information.

Measuring conditions include Measurement speed of year 1 and year 2 (MS1 and MS2), measurement speed difference (MSD), and measurement speed standard deviation of year 1 and year 2 (MSSD1 and MSSD2). As shown in Figure 66, MS1 and MS2 fluctuate around 80 km/h, with approximately 5% being less than 75 km/h. MSD refers to the subtraction of MS1 and MS2. MSSD1 and MSSD2 exhibit similarity in the distribution.

The weather factors include temperature difference (TD) and wind speed difference (WD). The data are collected from the meteorological stations in Switzerland. Firstly, the daily average value of temperature and wind speed from 2000 to 2022 in Switzerland were computed. Subsequently, the TD and WD were computed by subtracting the daily average value of temperature and wind speed in the latter year from that in the former year, respectively.

The correlation heatmap between the AD and the values of the factors is shown in Figure 67. There are low correlations between AD and the values of individual factors, indicating that AD is influenced by multiple factors, making it difficult to observe its correlations directly. In the following section, it was used a powerful machine learning model and causal inference framework to address this challenge. In this section, a dataset for training the upcoming model was constructed. AD was considered as a label, the ground truth for the model to learn, representing the level of variability of the data. The values of the different factors were regarded as features and served as input variables for the model. In this case, features represented the potential factors affecting the label AD.

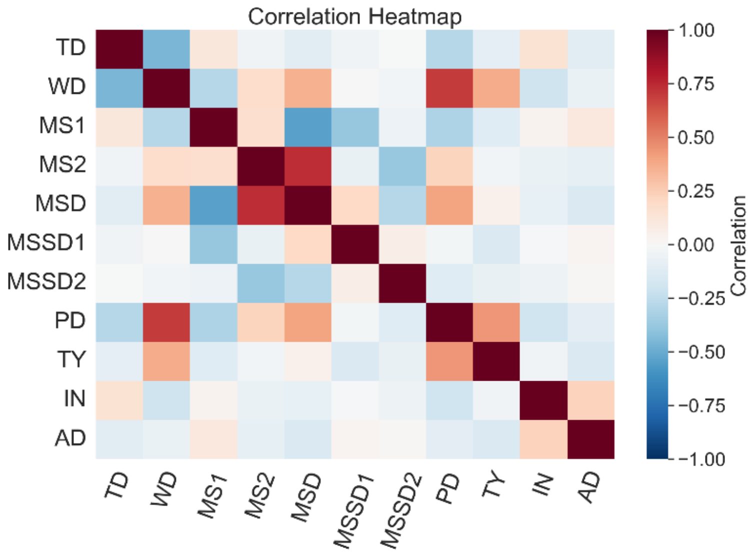

Figure 67: Correlation heatmap among factors and the explained variable.

6.4.3 Data fitting

The fitting performance of all models accounted for around 92% of the variability which is a high value (Table 8). The highest performance was observed with the Cat-Boost model, achieving a R^2 value of 0.9186 and a RMSE value of 0.362. Parameter tuning, including increasing the complexity of the tree-based models, such as adjusting the depth and the number of estimators, did not lead to a significant improvement in the model's performance. These findings suggest that the target variable can be almost completely explained by the available explanatory variables.

Tree-based ensemble models – I4			
Model	**Parameters**	**R^2**	**RMSE**
Random Forest	n estimators = 50 max depth = None min samples split = 2	0.8738	0.365
LightGBM	boosting type = "gbdt" n estimators = 200 max depth = None	0.8594	0.361
XGBoost	n estimators = 50 max depth = None	0.8408	0.385
CatBoost	Iterations = 200 depth = 15	**0.9186**	**0.362**

Table 8: Parameters and results of the tree-based ensemble model tested to fit the data

6.4.4 Causal inference modelling

In this case because the fitting of the models is so high, it was decided to also test a causal inference modelling technique that is another option available to study relation among variables.

Traditional machine learning methods can effectively identify correlations between the features and a label. However, it is essential to note that pseudo-correlations can sometimes arise, and the results do not inherently imply a causal relationship between the observations and the outcome. The fundamental principle of establishing causality hinges on comprehending how changes in a treatment under a series of conditions affect the outcome. This level of inquiry is challenging to address using traditional machine learning methods. In this study, it was used a structural causal framework (DoWhy (Sharma and Kiciman, 2020)) to specify assumptions about the mechanisms underlying observed data and test whether they are valid and to what extent.

This structural causal approach incorporates prior knowledge as input and ensures that the model reflects real-world causal relationships. For example, according to the results of the SHAP beeswarm summary plot and our domain knowledge, it can be known that the temperature difference, measurement speed difference, etc., might be causes of the data variability. Therefore, a causal graph as a directed acyclic graph was constructed that is used to specify and test causal assumptions. Each node represents an input variable (features), and an arrow indicates a causal link with a direction. Ten features were considered to establish a causal graph and structural assumptions between the features and the label, as shown in Figure 68. U represents an unobserved confounder.

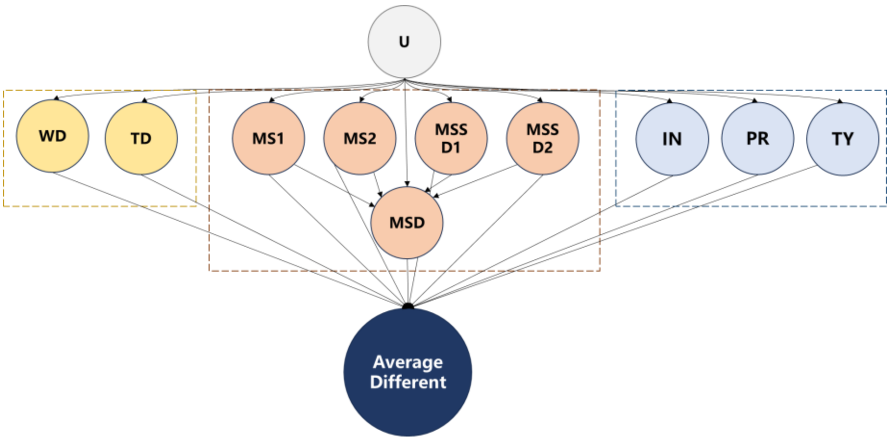

Figure 68: Causal graph

The causal effect refers to the effect of a treatment variable on an outcome variable, indicating the causal relationship between them. Specifically, it quantifies the change in the outcome variable caused by the treatment. The causal effects are estimated as the following equation:

$$P(Y|do(X = x)) = \sum_S P(Y = y \mid X = x, S = s)P(S = s) \tag{9}$$

where $do(X = x)$ means the specific action of treatment that is the focus of the causal analysis, while X indicates the treatment. Y and S indicate the outcome and the sufficient set variables, respectively.

For example, if a study is about the effect of a new drug on patient recovery, the use of the new drug would be the treatment, and the patient recovery would be the outcome.

$P(Y \mid do(x))$ represents the probability distribution; *x*, *y*, and *s* indicate the value in the corresponding variables (P. Judea, 2010). If a study on a new drug on patient recovery is conducted, using the new drug would be considered the treatment. In a causal inference setting, you might want to investigate how this treatment influences the outcome, such as the recovery rate of patients. The causal effects of each treatment can be quantitatively evaluated by the average treatment effect (ATE), which can be calculated as follows:

$$ATE = \frac{\sum_{i=1}^{N}(y_1(i) - y_0(i))}{N} \tag{10}$$

where $y_0(i)$ is the outcome value when feature *i* is not treated, and $y_1(i)$ is the outcome when feature *i* is treated. *N* represents the total number of the test data. When the variable's value is continuous rather than binary, y_1 signifies that the individual is treated with one additional unit compared to y_0. The estimated causal effects of selected features are shown in Figure 9, where the value represents the estimated effect. The higher its absolute value, the greater the degree of effect. For example, in a study that focuses on the effect of a new drug on patient recovery. A positive value of 0.2 means that patients who received the new drug recovered, on average, 20% faster than those who did not.

The obtained estimate should be validated process through robustness checks and sensitivity analyses. The validation of the causal estimate serves as a foundation for the conclusions and ensures their robustness under varying conditions. There are three checks for varied conditions to ensure that the conclusions hold not only in specific scenarios but also maintain their reliability in broader contexts: Add Random Common Cause: A synthetic independent random variable as a common cause is added to determine whether the estimation method alters its estimate in response to this change, where, ideally, the effect should not change. Placebo Treatment: when the true treatment variable was replaced with an independent random variable, the estimated causal effect should go to zero. Data Subsets Validation: the new effects should not change significantly when the given dataset is replaced with a randomly selected subset.

The results of the refutation are presented in Table 9. It is evident from the outcomes that all test results align with the robustness check expectations and affirm the validity of the estimated effects, thereby supporting further analysis.

Robustness checks

Feature	Estimated effect	Refute: Add a random common cause	Refute: Use a Placebo Treatment	Refute: Use a subset of data
IN	0.7193	0.7193	0.0020	0.7739
PD	-0.3608	-0.3607	0.0000	-0.3576
TY	-0.0169	-0.0170	0.0000	-0.0163
TD	-0.1856	-0.1857	0.0000	-0.1863
WD	0.0292	0.0293	0.0000	0.0277
MS1	0.1390	0.1392	0.0000	0.1379
MS2	-0.3062	-0.3063	0.0000	-0.3058
MSD	-0.2270	-0.2271	0.0000	-0.2243
MSSD1	0.0327	0.0327	0.0000	0.0292
MSSD2	0.0405	0.0406	0.0000	0.0384

Table 9: Results of the robustness checks

6.4.5 Variable importance

In Figure 69, temperature difference and intervention are the top two most important features. The SHAP value distributions for temperature difference, intervention, and test year length are notably distinct between positive and negative values. In contrast, the measurement speed standard deviation exhibits a bidirectional influence from the outlier values. Despite wind speed difference showing relatively high feature importance, its specific influence mechanism remains somewhat ambiguous.

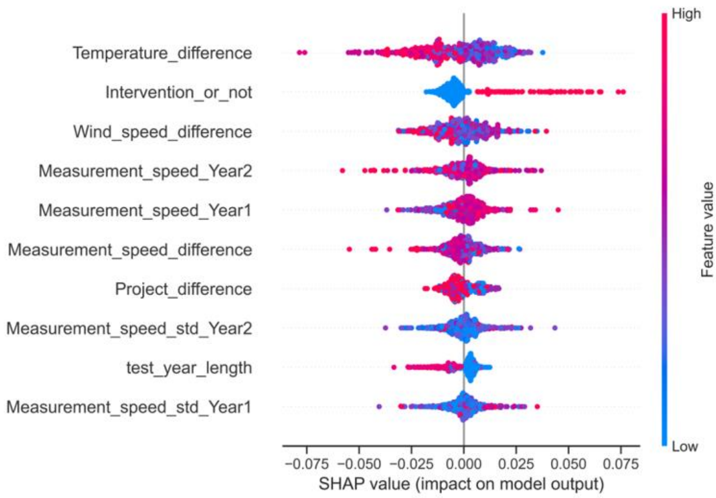

Figure 69: Beeswarm summary plot for feature interpretability of the CatBoost mode

Figure 70 contains the results of the average treatment effects analysis. It can be seen that intervention (IN) has a significant positive effect. This suggests that friction measurement values in the latter year tend to be higher when an intervention has occurred.

The causal effect of IN has the highest value of 0.7193, which means that with interventions, the AD would be 71.93% higher than in scenarios without interventions. In other words, interventions are highly likely to enhance friction measurement values. In contrast, PD, MS2, MSD, and TD have a noticeable negative effect on the Average difference. The importance of the influence of the factors is the information that was kept at this point since because these variables are not being Boolean, their interpretation is more difficult concerning the positive or negative effect on the AD.

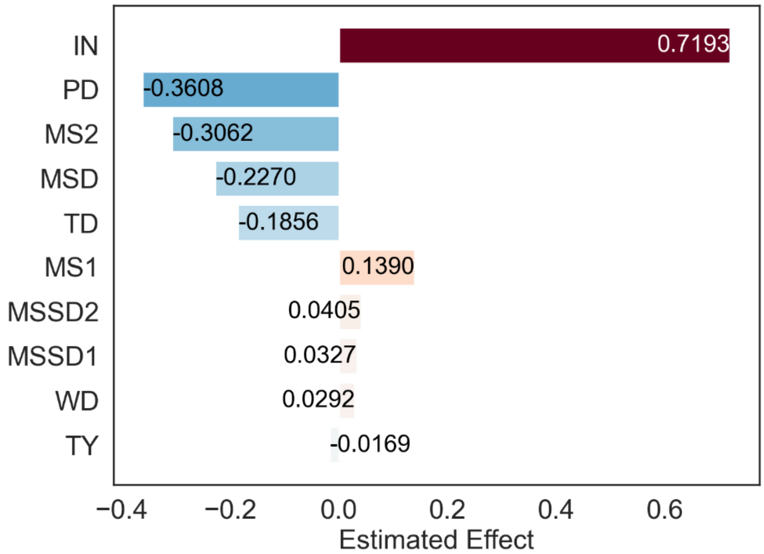

Figure 70: Estimated casual effect of the factors

6.5 Differences between *Filialen*

In order to understand possible differences between *filialen*, separate analyses for the case of the indicators I3 and I0 were performed. Results are resumed in the following Figure 71 and Figure 72:

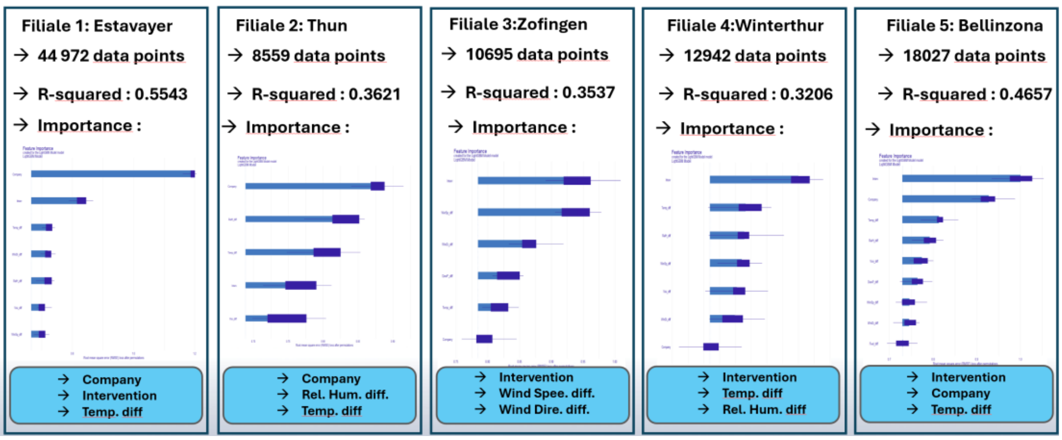

Figure 71: Results separated by *filiale* for the indicator I3

It can be seen that important differences are present according to which *filiale* is analysed. First of all the amount of data points show important differences. Some *filiale* like the *filiale* 1 have complete datasets with consistent data during most of the survey periods and for most of the indicators. It is not the case for example for *filiale* 2 or 3 that show less data, with some survey years completely absent. It would be required to clarify if the data was not yet uploaded to the system or if it was never measured. These inconsistencies however make difficult to have reliable knowledge of the roads at a country level. It can also be seen that in some *fililale* the most important factor affecting the variability in the data can be different. This can have several explanations like for example different geographical conditions, and different level of climatic and traffic stress.

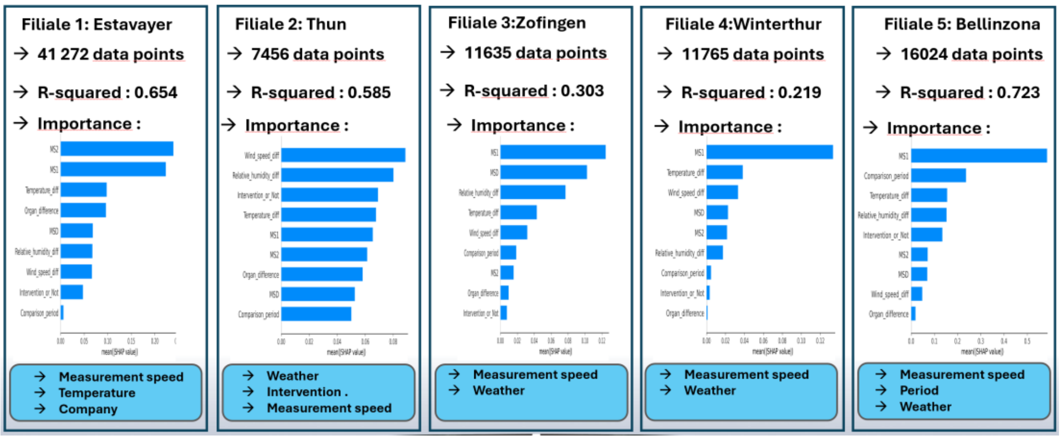

Figure 72: Results separated by filialen for the indicator I0

For the case of the Indicator I0, similar conclusions can be made. In addition to the amount of data per *filiale*, it was observed that results can also be very different. For example for the case of the *filiale* 2, it was observed that weather factors play a more important role than measurement speed for example that is the main important factor in the other *filiale*. This can be explained by geographical differences for example. Interesting is also to notice that the amount of variability that can be explained changes by *filiale*, which indicates that some *filiale* have more unexplained variability than others. All this information is important to evaluate if a general knowledge of the whole road network wants to be evaluated and at the moment of setting general rules for measurements or modernized norms.

7 Conclusions and recommendations

At the end of the project, it seems important to remark that although the general objectives outlined in the initial project proposal may not have been fully achieved, it can be confirmed that the specific goals were successfully met. In line with the aim of improving the *accuracy, understandability, repeatability, and reproducibility of road condition measurements*, the development of a comprehensive simulation model encompassing the entire road surface measurement process, as described in project proposal, could not be pursued as originally envisioned. This limitation stemmed from the nature of the data available.

Consequently, the project was redirected to focus on the creation of in-depth visual representations of key indicator data. Additionally, road condition metrics with relevant meteorological data was integrated to investigate the contribution of accessible factors to variability. These analyses were performed using advanced machine learning techniques—specifically CatBoost and interpretability tools such as SHAP—to gain deeper insights into the model's decision-making processes.

Results from the **visualisation** exploration showed that in each indicator there are variations in time that are not expected. Instead of observing a consistent slight deterioration from survey to survey, it can be seen that increase or deterioration in the road condition are almost random. This is particularly the case for Indicators I0, I2 and I3. Indicator I4 show a little more consistency in time.

It is quite interesting to observe that the indicators appear to be disconnected from each other, which was unexpected. Although they measure different aspects, they should ideally align in reflecting either a continuous deterioration trend or a significant improvement following a recorded intervention. Another important aspect to mention is that from one measurement to another, some increase in the variability was observed, meaning values are way more disperse, even when the same company is doing the measurements.

It was also observed that for the same highway lane, where an intervention took place recently, despite that some variability is expected due to use for example, some parts have much better condition than others parts which is not expected since a certain level of consistency should be present for a highway with the same history of interventions and similar use.

When an intervention occurs between two measurements, its effect is visible in some indicators but not in others. In case of minor interventions this could be the case but not in case of major interventions like the ones considered. This inconsistency is difficult to explain, and measurement issues may only partially account for this phenomenon. Additionally, the influence of different companies on the measurements appears to be significant, though its impact varies depending on the specific indicator being assessed.

The visualization also revealed discrepancies in the amount of data available across different Filialen. Some exhibit high consistency, with data available for all survey years and minimal missing information. In contrast, others have significant gaps in their datasets, making it difficult to accurately assess the development of pavement conditions.

Table 10 summarizes the main findings of the analyses on the different indicators. It highlights the percentage of variability that can be explained by the analysed variables and which ones present more influence. As it could have expected, the maintenance works (interventions) conducted on the roads between survey campaigns show a significant effect on the evolution of certain indicators. Likewise, the fact that different companies carry out the measurement adds a grade of uncertainty.

Indicators variabilities and variable importance

Indicators (VSS 40 925b)	Measured Parameter	Norm	Fitting Model	Variable Importance
I0 – Surface Damage	*1/10* $(\sum M_i{\cdot}G_i)$ M_i: Extent (A), Severity (S) G_i: Group Weight	VSS 40 925b (Annex) SN 640 516-7	$R^2 = 0.693$ (CatBoost)	Measuring speed Temperature Company
I2 – Longitudinal Evenness	Standard Deviation (Sw) of angles (W)	VSS 40517	$R^2 = 0.351$ (CatBoost)	Intervention Relative Humidity Temperature
I3 – Transversal Evenness	Rut Depth (T)	VSS 40518	$R^2 = 0.496$ (LightGBM)	Company Intervention Temperature
I4 – Surface Friction	Friction Coefficient (μ)	VSS 40 511 SN 640510	$R^2 = 0.919$ (CatBoost)	Temperature Wind speed Intervention

Table 10: Summary of the indicators and variable importance.

The important influence that other variables such as weather conditions or measuring speeds seem to have a less direct understanding from a practical perspective. Nowadays, the use of new technologic methodology allows the companies to conduct more precise measurements that would be independent of the speed and, therefore, this should have no impact on data quality. For example, while it is generally assumed that the indicator I0 should remain unaffected by measurement speed, data analyses suggest otherwise, revealing a significant influence of speed on the results. However, the speed of measurements has varied over the years due to changes in condition survey methods. Consequently, the survey method itself may be a confounding variable underlying the observed relationship with speed.

Results from the variable importance analysis on the **indicator I0** show that the measuring speed is the most important factor driving the variability between successive measurements. Nonetheless, this phenomenon may be indirectly associated to the changes in visual survey methodologies over the past two decades. In general the model is able to explain around a 70% of the variability. This mean that 30% of the variability is still to be determined and could not be assessed in this study. After the measuring speed, the temperature plays also a considerable role and having different companies making the measurements comes in third place.

Results from the variable importance analysis on the **indicator I2** show that only a small amount of the variability could be explained by the available data. Best models could only explain 35% of the variability. Further factors like traffic for example, type of road material or factors related to the measurement process could be included to increase the model fitting. As expected, the presence of an intervention is the most important variable, but it was found that weather parameters like relative humidity and temperature also play a role.

The results from the variable importance analysis for the **indicator I3** indicate that the company plays a significant role in the variability of measurement differences across surveys. Although only about 50% of the variability could be explained, this finding is crucial, as it highlights the need for greater consistency in the measurement methods used by the company for this specific indicator. The second most important factor is the intervention, which aligns with expectations. In third place are the weather parameters, which, while not as influential as for the previous indicators, still contribute to the observed variability.

Results from the influencing factors analysis on the **indicator I4** show a better fit of the data to the variables available accounting for up to 91% of the variability which is particularly high. The main factor influencing the measurements is the temperature, but also the wind speed plays an important role. After the weather factors it was found that the presence of an intervention which is expected but not far behind the measuring speed that also plays an important role.

Finally, taking into consideration the results from the analysis, the following recommendations are proposed:

1 – The visualisation of the data plays a key role in allowing a rapid understanding of the pavement condition and an assessment of the amount of data available and the consistency between the different indicators. Considering the possibility of systematizing the elaboration of the visualization graphs and only built in with the newer data, it is highly recommended to develop a visualization of every lane of every highway that should be easily accessible for managers but also to measuring companies so they can

rapidly estimate if there are some problems with their measurements. Furthermore, the visualisation allows to immediately spot missing data that constitute a problem when assessing the condition in the long term.

2- Weather parameters should be consistently measured at the same time as the pavement indicators. Nowadays this will not constitute an excessive added complexity or costs since portable weather stations are very reliable and not very expensive. The inclusion of weather parameter like air and surface temperatures as well as relative humidity could allow to really understand possible discrepancies in the data and also to eventually adjust the measurements. Furthermore, it is highly recommended to use the weather data to decide when to do the measurements and try when possible, to have the most similar condition possible than the ones from the previous measurements. Nevertheless, dedicated studies should be conducted in order to understand the effect of the temperature on the surface parameters and calculate correction factors that allow normalising the obtained measurement results to a pre-defined reference temperature. Recent examples of scientific publications and/or technical reports focused on this topic are shown in Table 11.

Influence of weather conditions on surface parameters

Surface parameters	Reference Studies
Surface Damage	Zhao et al., 2025 Drumm and Meier, 2003 Abdulsalam and Omar, 2023 Kapela et al., 2015
Longitudinal Evenness	Homsi et al., 2017 Vámos and Szendefy, 2024
Transversal Evenness	Banerji et al., 2021 Guo et al., 2025
Surface Friction	Anupam et al., 2013 Luo, 2003

Table 11: Reference studies on the influence of weather conditions on surface parameters.

3- Considering the influence that having different companies plays in the variability of the data it is recommended that contracts are made for two consecutive survey years with one company so there is a certainty to avoid the company variability at least every two consecutive surveys. Moreover, it is recommended that companies validate their measurement equipment by calibration protocols specifically designed for each performance evaluation. In addition, companies should be advised to participate in regular testing campaigns to assess the reliability and accuracy of measurements in order to ensure that the measurements are equivalent and thus can provide precise estimations of the condition.

4- Considering the discrepancies observed with the interventions, it is recommended to develop a more systematic way of recording these procedures, including not only major interventions (rehabilitation) but also minor ones (maintenance) that can also play a role in the development of the pavement condition, particular for some indicators. This points also highlights the importance of geographically referencing the measurements in order to enhance their traceability, repeatability, and reproducibility as well as allow a better comparison between survey campaigns.

5- It is recommended that the analysis performed during this project should be expanded to include more data that could explain the variability that the current variables could not address. In particular it is recommended including the traffic data, especially heavy traffic data to further explain the pavement deterioration. Also, it is recommended including characteristics of the type of road material, topographic data and soil parameters to understand each influence on the development of the pavement condition.

6- Finally, it is recommended investing in the further development of the Trassee application to consolidate all relevant data linked to pavement conditions. This should include enhanced visualization features and an intuitive, user-friendly interface to ensure accessibility. By doing so, managers, measurement companies, scientists, and decision-makers will have an efficient and straightforward way to access and utilize this critical information.

7- It should be ensured that updates to transformation functions or classification thresholds (e.g. in VSS 40925b) are accompanied by correlation analysis with prior versions to preserve continuity in index interpretation.

Table 12 presents the recommendations organized by priority level and designated stakeholder responsibility.

Actions recommended after the project

Priority Level	Recommendations	Responsible Stakeholder
Short-Term	Improve data visualization to support consistency checks, early issue detection, and usability.	FEDRO
	Standardize weather data collection during measurements using portable weather stations.	Measurement companies, FEDRO, Cantons
	Enforce equipment calibration and comparative testing across all measurement firms.	Measurement companies, FEDRO, Cantons
	Require minimum 2 consecutive survey years continuity for contractors to reduce company-related variability.	FEDRO, Cantons
	Clearly document and track interventions, including minor maintenance events	FEDRO, Cantons
	Check and complete missing datasets in TRA-Trassee, especially for gaps in specific *Filialen*	FEDRO
Mid-Term	Harmonize measurement protocols across firms to reduce organizational bias in indicators	FEDRO, Measurement companies, Standardization (VSS)
	Introduce mandatory training and calibration for visual inspection teams	FEDRO, Cantons, Measurement companies
	Improve TRA-Trassee functionality to support interactive analytics and better user access	FEDRO
	Systematically integrate and store weather data alongside condition data in TRA-Trassee	FEDRO
Long-Term	Include traffic load, material, and soil data in PMS datasets for more robust deterioration modeling	FEDRO, Cantons
	Develop and implement national-level standards for indicator synchronization and segment consistency	FEDRO, Standardization (VSS)
	Extend data collection to include topographic and environmental variables to improve deterioration models	FEDRO, Cantons
	Coordinate a national program for automated measurement consistency studies	FEDRO

Table 12: Prioritization of the actions recommended in the project.

Bibliography

Adamczak, S., Świderski, J., Dobrowolski, T., & Durakbasa, N. (2018). Application of MSA analysis for the evaluation of instruments for measuring the surface texture. *Mechanik*, *91*(5–6), 434–436.

Adlinge, S.S. & Gupta, P.K. (2013). Pavement Deterioration and its Causes. *International Journal of Innovative Research and Development*, 9–15. Retrieved from https://www.iosrjournals.org

Adulsalam, B. A., &, Omar H.(2022) Pavement Crack Detection and Width Estimation Using Digital Image Processing Technique. *Proceedings of the 9th National Conference on Construction Materials and Structural and Geotechnical Engineering*, Department of Civil Engineering, University of Tripoli, Libya.

Anupam, K., Srirangam, S. K., Scarpas, A., & Kasbergen, C. (2013) Influence of Temperature on Tire–Pavement Friction Analyses. *Transportation Research Record: Journal of the Transportation Research Board, 2369*, Transportation Research Board of the National Academies, Washington, D.C.

Araújo, L. M. M., Paiva, R. G. N., Peruchi, R. S., Junior, P. R., & de Freitas Gomes, J. H. (2019). New indicators for measurement error detection in GR&R studies. *Measurement*, *140*, 557–564. https://doi.org/10.1016/j.measurement.2019.04.010

Asplund, M., & Lin, J. (2016). Evaluating the measurement capability of a wheel profile measurement system by using GR&R. *Measurement*, *92*, 19–27. https://doi.org/10.1016/j.measurement.2016.05.033

Aslam, F. , Alyousef, R., Hassan Awan, H., & Javed, M.F. (2023). Forecasting the self-healing capacity of engineered cementitious composites using bagging f and stacking regressor. *Structures*, 54, 1717–1728. https://doi.org/10.1016/j.istruc.2023.05.140

Baig, A. A., Ruzli, R., & Buang, A. B. (2013). Reliability analysis using fault tree analysis: A review. *International Journal of Chemical Engineering and Applications*, *4*(3), 169.

Banerji, A.K., Topdar, P., & Datta, A.K. (2023). Study on the Effect of Wheel Load and Temperature on Rutting Damage of Composite Flexible Pavement Using Finite Element Method. In: Fonseca de Oliveira Correia, J.A., Choudhury, S., Dutta, S. (Eds.) *Advances in Structural Mechanics and Applications (ASMA 2021) Structural Integrity,* Springer. https://doi.org/10.1007/978-3-031-10501-1_15

Beylihan, E., & Elevi, S. (2022). Fuzzy Measurement System Analysis Approach: A Case Study. *International Journal of Engineering Research and Development*, *14*(1), 176–185.

Bianchini, A., Bandini, P., & Smith, D. W. (2010). Interrater reliability of manual pavement distress evaluations. *Journal of Transportation Engineering*, *136*(2), 165–172. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000092

Blaško, P., & Petrík, J. (2015). FTA analysis of the fault in the horizontal fluidity test. *Environment (Temperature, Pressure, Geo-Climatic Conditions)*, *3*, 6-12.

Blumenfeld, T., Elvarsson, A., Schiffmann, F., Grund, C., Kempf, D., Balck, H., Bald, S., & Hajdin, R. (2018). *M*ulti-kriterielle probabilistische Prognose der Zustandsentwicklung ProZEnt. *Bundesant für Strassen (ASTRA)*

Bogus, S. M., Song, J., Waggerman, R., & Lenke, L. R. (2010). Rank correlation method for evaluating manual pavement distress data variability. *Journal of Infrastructure Systems*, *16*(1), 66–72. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000005

Breiman, L. (2001). Random forests. *Machine learning,* 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chen,C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemometrics and Intelligent Laobroatory Systems,* 191, 54–64. https://doi.org/10.1016/j.chemolab.2019.06.003

Chen, Y., Wang, K., Xu, C., Shi, Q., He, J., Li, P., & Shi, T. (2018). Evaluation of the safety performance of highway alignments based on fault tree analysis and safety boundaries. *Traffic Injury Prevention*, *19*(4), 409–416. https://doi.org/10.1080/15389588.2018.1441963

Chen, Y., Wang, K., Zhang, Y., & Shi, Q. (2021). Identification of black spots on highways using fault tree analysis and vehicle safety boundaries. *Journal of Transportation Safety & Security*, *13*(1), 46–68. https://doi.org/10.1080/19439962.2019.1601001

Chen T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chu, C.-Y., & Durango-Cohen, P. L. (2007). Estimation of infrastructure performance models using state-space specifications of time series models. *Transportation Research Part C: Emerging Technologies*, *15*(1), 17–32. https://doi.org/10.1016/j.trc.2006.05.002

Chu, C.-Y., & Durango-Cohen, P. L. (2008). Estimation of dynamic performance models for transportation infrastructure using panel data. *Transportation Research Part B: Methodological*, *42*(1), 57–81. https://doi.org/10.1016/j.trb.2007.09.002

Cmar, M. D., Maloney, C. T., & Butala, V. D. (2012). Modernization of B-2 Data, Video, and Control Systems Infrastructure. *58th Instrumentation Symposium*, (*IIS12-P051).*

Corley-Lay, J. (2014). Pavement performance measures: How states see good, fair, and poor. *Transportation Research Record*, *2431*(1), 1–5. https://doi.org/10.3141/2431-01

Davis-McDaniel, C., Chowdhury, M., Pang, W., & Dey, K. (2013). Fault-tree model for risk assessment of bridge failure: Case study for segmental box girder bridges. *Journal of Infrastructure Systems*, *19*(3), 326–334. .https://doi.org/10.1061/(ASCE)IS.1943-555X.0000127

Descornet, G., Berlemont, B., & Martin, J.-M. (2001). Study of precision of transverse evenness measurements in FILTER experiment: forum of european national highway research laboratories investigation on longitudinal and transverse evenness of roads experiment. *Transportation Research Record*, *1764*(1), 210–220. https://doi.org/10.3141/1764-25

Dillon, B. S. (2009). *Human Reliability, Error, and Human Factors in Engineering Maintenance*. CRC Press. https://doi.org/10.1201/9781439803844

Doshi, J. A., & Desai, D. A. (2019). Measurement system analysis for continuous quality improvement in automobile SMEs: multiple case study. *Total Quality Management & Business Excellence*, *30*(5–6), 626–640. https://doi.org/10.1080/14783363.2017.1328274

Drumm, E.C., & Meier, R. (2003) *LTPP Data Analysis: Daily and Seasonal Variations in Insitu Material Properties*. NCHRP Web Document 60.

Edmondson, V., Ardill, O., Martin, J., Lim, M., Kane, M., & Woodward, J. (2022). Quantifying long-term rates of texture change on road networks. *International Journal of Pavement Engineering*, *23*(6), 1957–1969. https://doi.org/10.1080/10298436.2020.1835732

Edmondson, V., Martin, J., Ardill, O., Lim, M., Kane, M., & Woodward, J. (2021). Seasonal Signals Observed in Non-Contact Long-Term Road Texture Measurements. *Coatings*, *11*(6), 735. https://doi.org/10.3390/coatings11060735

Evdorides, H., Burrow, M., & Sulaiman, S. (2013). A methodology to model the variability in pavement performance. *Proceedings of the Institution of Civil Engineers-Transport*, *166*(4), 233–240. https://doi.org/10.1680/tran.10.00086

Flintsch, G. W., & McGhee, K. K. (2009). *Quality management of pavement condition data collection* (Vol. 401). Transportation Research Board.

Freitas, E., Freitas, C., & Braga, A. C. (2014). The analysis of variability of pavement indicators: MPD, SMTD and IRI. A case study of Portugal roads. *International Journal of Pavement Engineering*, *15*(4), 361–371. https://doi.org/10.1080/10298436.2013.806864

Fryer,D, Strumke, I., & Nguyen, H. (2021). Shapley values for feature selection: The good, the bad, and the axioms. *IEEE Access*, 9, 3–10. https://doi.org/10.1109/ACCESS.2021.3119110

Giraud, L., & Galy, B. (2018). Fault tree analysis and risk mitigation strategies for mine hoists. *Safety Science*, *110*, 222–234. https://doi.org/10.1016/j.ssci.2018.08.007

Guo, M., Liu, P., & Feng, W. (2025) Analysis of tire contact stresses and asphalt pavement rutting under gradient temperature and typical driving conditions. *Scientific Reports, 15,* 2985. https://doi.org/10.1038/s41598-025-02985-2

He, X., Wang, Y., Shen, Z., & Huang, X. (2008). A simplified CREAM prospective quantification process and its application. *Reliability Engineering & System Safety*, *93*(2), 298–306. https://doi.org/10.1016/j.ress.2006.10.026

Homsi, F., Bodin, D., Chabot, A., & Hornych, P. (2017) Effect of temperature on asphalt pavement strain responses: APT data and thermo-viscoelastic pavement modelling. *Proceedings of the 17th AAPA International Flexible Pavements Conference*, Melbourne, Australia

Hong, F., & Prozzi, J. A. (2010). Roughness model accounting for heterogeneity based on in-service pavement performance data. *Journal of Transportation Engineering*, *136*(3), 205–213. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000091

Humplick, F. (1992). Highway pavement distress evaluation: Modeling measurement error. *Transportation Research Part B: Methodological*, *26*(2), 135–154. https://doi.org/10.1016/0191-2615(92)90021-4

Jain, A. (2017). A case study on measurement system analysis (MSA) at a pump company. *International Research Journal of Engineering Technology*, *4*(5), 1567–1571.
Jabeur, S.B., Mefteh-Wali, S., & Viviani, J.-L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research,* 1-21. https://doi.org/10.1007/s10479-021-04423-6
Jia, X., Huang, B., Dong, Q., Zhu, D., & Maxwell, J. (2016). Influence of pavement condition data variability on network-level maintenance decision. *Transportation Research Record*, *2589*(1), 20–31. https://doi.org/10.3141/2589-03
Jia, X., Huang, B., Zhu, D., Dong, Q., & Woods, M. (2018). Influence of measurement variability of international roughness index on uncertainty of network-level pavement evaluation. *Journal of Transportation Engineering, Part B: Pavements*, *144*(2), 4018007. https://doi.org/10.1061/JPEODX.0000052
Joubert, J. W., & Meintjes, S. (2015). Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transportation Research Part B: Methodological*, *76*, 81–92. https://doi.org/10.1016/j.trb.2015.03.004
Judea, P. (2010). An introduction to causal inference. *The International Journal of Biostatistics,* 6(2), 1-62.
Kalita, K., & Rajbongshi, P. (2015). Variability characterisation of input parameters in pavement performance evaluation. *Road Materials and Pavement Design*, *16*(1), 172–185. https://doi.org/10.1080/14680629.2014.939667
Kamil, S., & Pawel, B. (2017). Measurement system analysis for one-sided tolerance. *ITM Web of Conferences*, *15*, 5011. https://doi.org/10.1051/itmconf/20171505011
Kapela, R, Śniatała, P.,Turkot, A., Rybarczyk, A., Pożarycki, A., & Rydzewsk, P. (2015) Asphalt surfaced pavement cracks detection based on histograms of oriented gradients. P*roceedings of the 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES)*, Toruń, Poland.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems,* 30, 3146-3154.
Klaput, P., Vykydal, D., Tošenovský, F., Halfarová, P., & Plura, J. (2016). Problems of application of measurement system analysis (MSA) in metallurgical production. *Metalurgija*, *55*(3), 535–537.
Kobayashi, K., Kaito, K., & Lethanh, N. (2012). A statistical deterioration forecasting method using hidden Markov model for infrastructure management. *Transportation Research Part B: Methodological*, *46*(4), 544–561. https://doi.org/10.1016/j.trb.2011.11.001
Kouchaki, S., Roshani, H., Prozzi, J. A., Garcia, N. Z., & Hernandez, J. B. (2018). Field investigation of relationship between pavement surface texture and friction. *Transportation Research Record*, *2672*(40), 395–407. https://doi.org/10.1177/0361198118797002
Lapp, S. A. (2005). Applications of fault tree analysis to maintenance interval extension and vulnerability assessment. *Process Safety Progress*, *24*(2), 91–97. https://doi.org/10.1002/prs.10068

LeBeau, K. H., & Wadia-Fascetti, S. J. (2007). Fault tree analysis of Schoharie Creek Bridge collapse. *Journal of Performance of Constructed Facilities*, *21*(4), 320–326. https://doi.org/10.1061/(ASCE)0887-3828(2007)21:4(320)

Lethanh, N., & Adey, B. T. (2013). Use of exponential hidden Markov models for modelling pavement deterioration. *International Journal of Pavement Engineering*, *14*(7), 645–654. https://doi.org/10.1080/10298436.2012.731720

Li, M. C., & Al-Refaie, A. (2008). Improving wooden parts' quality by adopting DMAIC procedure. *Quality and Reliability Engineering International*, *24*(3), 351–360. https://doi.org/10.1002/qre.923

Lokuge, W., Wilson, M., Tran, H., & Setunge, S. (2019). Predicting the probability of failure of timber bridges using fault tree analysis. *Structure and Infrastructure Engineering*, *15*(6), 783–797. https://doi.org/10.1080/15732479.2018.1557163

Losa, M., & Leandri, P. (2011). The reliability of tests and data processing procedures for pavement macrotexture evaluation. *International Journal of Pavement Engineering*, *12*(1), 59–73. https://doi.org/10.1080/10298436.2010.493830

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems,* 30, 4765-4774.

Luo, Y. (2003). Effect of Pavement Temperature on Frictional Properties of Hot-Mix-Asphalt Pavement Surfaces at the Virginia Smart Road (Master's Thesis). *Vir*ginia Polytechnic Institute and State University.

Ma, J., Bai, Y., Shen, J., & Zhou, F. (2014). Examining the impact of adverse weather on urban rail transit facilities on the basis of fault tree analysis and fuzzy synthetic evaluation. *Journal of Transportation Engineering*, *140*(3), 4013011. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000658

Ma, H., Wang, M., Geng, Y., Ai, W., Zhang, W., & Zheng, W. (2023). A new hybrid approach model for predicting burst pressure of corroded pipelines of gas and oil. *Engineeing Failure Analysis*, 149, 107248. https://doi.org/10.1016/j.engfailanal.2023.107248

Madanat, S. (1993). Optimal infrastructure management decisions under uncertainty. *Transportation Research Part C: Emerging Technologies*, *1*(1), 77–88. https://doi.org/10.1016/0968-090X(93)90010-T

Magnusson, G., Dahlstedt, S., & Sjögren, L. (2002). Measurements of the road surface longitudinal unevenness – methods and necessary accuracy. *VTI Rapport*, *475*.

Mahmood, Y. A., Ahmadi, A., Verma, A. K., Srividya, A., & Kumar, U. (2013). Fuzzy fault tree analysis: A review of concept and application. *International Journal of System Assurance Engineering and Management*, *4*, 19–32. https://doi.org/10.1007/s13198-013-0122-8

Mia, M. N. U., Henning, T. F. P., & Costello, S. B. (2017). Application of fault tree analysis to identify moisture damage potential in flexible road pavements. *Journal of Infrastructure Systems*, *23*(4), 4017026. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000372

Montgomery, S. R., Gkritza, K., & Haddock, J. E. (2019). Factors Affecting the Accuracy and Variability of Pavement Surface Evaluations and Ratings. *Journal of Infrastructure Systems*, *25*(2), 4019008. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000485

Mottahedi, A., & Ataei, M. (2019). Fuzzy fault tree analysis for coal burst occurrence probability in underground coal mining. *Tunnelling and Underground Space Technology*, *83*, 165–174. https://doi.org/10.1016/j.tust.2018.09.005

Ozer, D.J. (1985). Correlation and the coefficient of determination. *Psychological bulletin,* 97(2), 307-315. https://doi.org/10.1037/0033-2909.97.2.307

Park, E. S., Smith, R. E., Freeman, T. J., & Spiegelman, C. H. (2008). A Bayesian approach for improved pavement performance prediction. *Journal of Applied Statistics*, *35*(11), 1219–1238. https://doi.org/10.1080/02664760802328346

Pereira, R. B. D., Peruchi, R. S., de Paiva, A. P., da Costa, S. C., & Ferreira, J. R. (2016). Combining Scott-Knott and GR&R methods to identify special causes of variation. *Measurement*, *82*, 135–144. https://doi.org/10.1016/j.measurement.2015.12.003

Perera, R. W., Kohn, S. D., & Wiser, L. J. (2006). Factors contributing to differences between profiler and the international roughness index. *Transportation Research Record*, *1974*(1), 80–88. https://doi.org/10.3141/1974-10

Plati, C., Pomoni, M., & Georgouli, K. (2020). Quantification of skid resistance seasonal variation in asphalt pavements. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(2), 237–248. https://doi.org/10.1016/j.jtte.2019.10.002

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems,* 31, 6638-6648.

Qin, X., Zhang, K., Tan, J., & Wang, J. (2022). A review on the influencing factors of pavement surface temperature. *Environmental Science Pollution Research*, 29(45), 67659–67674. https://doi.org/10.1007/s11356-022-22295-3

Ruijters, E., & Stoelinga, M. (2015). Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review*, *15*, 29–62. https://doi.org/10.1016/j.cosrev.2015.03.001

Ruotoistenmäki, A., Seppälä, T., & Kanto, A. (2006). Comparison of modeling and measurement accuracy of road condition data. *Journal of Transportation Engineering*, *132*(9), 715–721.

Schlotjes, M. R., Henning, T. F. P., Burrow, M. P. N., & St. George, J. D. (2013). Descriptive fault trees for structural pavement failure mechanisms. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, *22*(4), 3–14.

Schmidt, B. (2001). EVEN project: experiment to compare and harmonize methods for assessment of longitudinal and transverse evenness of pavement. *Transportation Research Record*, *1764*(1), 221–231. https://doi.org/10.3141/1764-26

Senol, S. (2004). Measurement system analysis using designed experiments with minimum α–β Risks and n. *Measurement*, *36*(2), 131–141. https://doi.org/10.1016/j.measurement.2004.05.003

Setunge, S., Zhu, W., Gravina, R., & Gamage, N. (2016). Fault-tree-based integrated approach of assessing the risk of failure of deteriorated reinforced-concrete bridges. *Journal of Performance of Constructed Facilities*, *30*(3), 4015058. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000810

Shalaby, A., & Reggin, A. (2007). Optimization of data collection needs for manual and automated network-level pavement condition ratings based on transverse variability and neural networks. *Canadian Journal of Civil Engineering*, *34*(2), 139–146. https://doi.org/10.1139/l06-116

Shapley, L.S. (1953) A value for n-person games. In: H. Kuhn & A. Tucker (Eds.), *Contributions to the Theory of Games II* (pp. 307-317). Princeton University Press.

Sharma A., & Kiciman, E. (2020). DoWhy: An end-to-end library for causal inference. *arXiv preprint* arXiv:2011.04216.

Shi, L., Chen, W., & Lu, L. F. (2014). An approach for simple linear profile gauge R&R studies. *Discrete Dynamics in Nature and Society*, *2014*, 1-6. https://doi.org/10.1155/2014/869586

Soares, W. de O. S., Peruchi, R. S., Silva, R. A. V., & Rotella Junior, P. (2022). Gage R&R studies in measurement system analysis: A systematic literature review. *Quality Engineering*, *34*(3), 382–403. https://doi.org/10.1080/08982112.2021.1906197

Taiwo, A. M., Yussif, M. E. A., Ben Seghier, A., & Zayed, T. (2022). Explainable ensemble models for predicting wall thickness loss of water pipes. *Artificial Intelligence in Civil Engineering, 5(2)*, 45–59.

Ung, S.-T. (2019). Evaluation of human error contribution to oil tanker collision using fault tree analysis and modified fuzzy Bayesian Network based CREAM. *Ocean Engineering*, *179*, 159–172. https://doi.org/10.1016/j.oceaneng.2019.03.065

Vámos M.J., & Szendefy, J. (2024). T*emperature Effects on Traffic Load-Induced Accumulating Strains in Flexible Pavement Structures*. *International Journal of Pavement Research Technolgy, 17(5)*, 553–564.

Vilaça, J. L., Fonseca, J. C., Pinho, A. C. M., & Freitas, E. (2010). 3D surface profile equipment for the characterization of the pavement texture–TexScan. *Mechatronics*, *20*(6), 674–685. https://doi.org/10.1016/j.mechatronics.2010.05.004

Volkanovski, A., Čepin, M., & Mavko, B. (2009). Application of the fault tree analysis for assessment of power system reliability. *Reliability Engineering & System Safety*, *94*(6), 1116–1127. https://doi.org/10.1016/j.ress.2008.12.006

Wang, Y., Zhang, D., Zhang, S., Jia, Q., & Zhang, H. (2020). Quality confirmation of electrical measurement data for space parts based on MSA Method. *In Proceedings of the 6th International Conference on Signal and Information Processing, Networking and Computers (ICSINC)* (pp. 344–351). https://doi.org/10.1007/978-981-15-7347-3_50

Winkler, M., Haltmeier, M., Kleidorfer, W., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*, *14(10)*, 1402–1411. https://doi.org/10.1080/15732479.2018.1443145

Xu, C., Wang, C., Ding, Y., & Wang, W. (2020). Investigation of extremely severe traffic crashes using fault tree analysis. *Transportation Letters*, *12*(3), 149–156. https://doi.org/10.1080/19427867.2018.1552144

Yang, X., Wang, R., Wang, H., & Yang, Y. (2020). A novel method for measuring pose of hydraulic supports relative to inspection robot using LiDAR. *Measurement*, *154*, 107452. https://doi.org/10.1016/j.measurement.2020.107452

Yanyan, Z., & Renzuo, X. (2008). A human factors fault tree analysis method for software engineering. *In Proceedings of the 2008 IEEE International*

Conference on Industrial Engineering and Engineering Management (pp. 1971–1975). https://doi.org/10.1109/IEEM.2008.4738068

Yao, L., Leng, Z., Jiang, J., & Ni, F. (2022). Modelling of pavement performance evolution considering uncertainty and interpretability: a machine learning based framework. *International Journal of Pavement Engineering*, *23*(14), 5211–5226. https://doi.org/10.1080/10298436.2021.1903671

Yeh, T.-M., & Sun, J.-J. (2013). Using the Monte Carlo simulation methods in gauge repeatability and reproducibility of measurement system analysis. *Journal of Applied Research and Technology*, *11*(5), 780–796.

Yin, H., Stoffels, S. M., & Antle, C. A. (2006). Profile data variability in pavement management: Findings and tools from LTPP. *In Airfield and Highway Pavement: Meeting Today's Challenges with Emerging Technologies* (pp. 984–995). https://doi.org/10.1061/40838(191)75

Yu, M., You, Z., Wu, G., Kong, L., Liu, C., & Gao, J. (2020). Measurement and modeling of skid resistance of asphalt pavement: A review. *Construction and Building Materials*, *260*, 119878. https://doi.org/10.1016/j.conbuildmat.2020.119878

Yuan, Q., Yin, X., Yan, W., Tan, Y., & Li, Z. (2016). A new approach for the capability analysis based on measurement system analysis: a case study for coal quality detection equipment. *International Journal of Applied Decision Sciences*, *9*(1), 17–38. https://doi.org/10.1504/IJADS.2016.073998

Zaib, A., Yin, J., & Khan, R. U. (2022). Determining role of human factors in maritime transportation accidents by fuzzy fault tree analysis (FFTA). *Journal of Marine Science and Engineering*, *10*(3), 381. https://doi.org/10.3390/jmse10030381

Zhan, Y., Li, J. Q., Liu, C., Wang, K. C. P., Pittenger, D. M., & Musharraf, Z. (2021). Effect of aggregate properties on asphalt pavement friction based on random forest analysis. *Construction and Building Materials*, *292*, 123467. https://doi.org/10.1016/j.conbuildmat.2021.123467

Zhao, Q.-M., Liu, Z.-H., Pan, C.-L., Meng, F.-Y., & Guo, G.-H. (2025). Accurate identification of pavement cracks based on an improved segmentation model. *International Journal of Pavement Engineering, 26*(1).

Project conclusion

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK

Version vom 09.10.2013

Formular Nr. 3: Projektabschluss

erstellt / geändert am: 10.06.2025 / 20.10.2025

Grunddaten

Projekt-Nr.:	TRU_20_02B_01
Projekttitel:	Removing the unexplained variability in road condition indicator values (COEUS)
Enddatum:	30.09.2025

Texte

Zusammenfassung der Projektresultate:

In this study we focus our analysis on two methods to explore the complex and large datasets. On one side, visualizations play a key role in data analysis, helping identify inconsistencies and trends over time in a simple and intuitive way. Various statistical methods, such as histograms and scatter plots, aid in assessing data distribution, detecting outliers, and ensuring consistency in road condition indicators. On the other hand, a more complex approach like factor analysis is also a tool that can be considered a critical part of pavement condition assessment. In this study, we integrate weather data from MeteoSwiss, matching relevant meteorological stations to each Filiale. The combination of weather data and measurement factors allowed us to explore more in detail the importance when explaining the variability in the datasets. To enhance pre-diction accuracy, advanced machine learning models—including AdaBoost, CatBoost, LightGBM, Random Forest, and XGBoost—are optimized using Bayesian Optimization (BO). These ensemble learning methods help capture complex relationships in pavement condition data. The model performance is assessed using R^2 and RMSE metrics to determine the most accurate predictions. SHAP (Shapley Additive Explanations) and permutations techniques provide interpretability for the predictive mod-els, allowing insight into the influence of different variables on road condition variability. The quantification of the importance of features such as climate, intervention history, and measurement methods ensure a data-driven approach to road infrastructure planning and maintenance.

In general, the results from the visualization exploration indicate that the variations observed in the different indicators over time are inconsistent with expected patterns. Rather than a gradual and predictable deterioration in road conditions from one survey to the next, the observed changes appear almost random. This issue is most pronounced in indicators I0 (Surface Damage), I2 (Longitudinal Evenness), and I3 (Transversal Evenness), whereas indicator I4 (Surface Friction) exhibits slightly more stability over time. Another unexpected finding is the disparity in road conditions within the same stretch of pavement. While it is reasonable to assume that some sections of a road may deteriorate at different rates due to varying traffic loads or environmental influences, a certain level of consistency is still expected. The fact that some segments appear to be in significantly better condition than others raise further concerns about the accuracy of the measurement process.

The variable importance analysis conducted for each indicator provided further in-sights into the key factors influencing data variability. Results on the Indicator I0 show the measuring speed as the most important factor driving the variability between successive measurements. In general the model is able to explain around a 70% of the variability. Results on the indicator I2 show that only a small amount of the variability could be explained by the analyzed variables (35%) with interventions being the most influential factor, followed by weather parameters. Also, only about 50% of the variability could be explained for indicator I3 where results indicate that the company plays a significant role. Finally, results from indicator I4 show a better fit of the data to the variables analyzed accounting for up to 91% obtaining temperature as the most significant influencing factor.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

Zielerreichung:

The main goal of COEUS was to improve the accuracy, understandability, repeatability, and reproducibility of road condition measurements in Switzerland. This results in improved estimates of asset deterioration, optimal intervention strategies, and timing of specific interventions.
Throughout different data visualizations we were able to illustrate the extent of the unexplained variability using collected historical data according to the existing practice of condition recording and assessment.
In addition, we have identified the main sources of the variability as well as the extent of their contribution for each indicator .
After the project, we can recommend practical measurements to reduce this variability in the future.

Folgerungen und Empfehlungen:

- The visualisation of the data plays a key role in allowing a rapid understanding of the pavement condition and an assessment of the amount of data available and the consistency between the different indicators. Considering the possibility of systematizing the elaboration of the visualization graphs and only built in with the newer data, it is highly recommended to develop a visualization of every lane of every highway that should be easily accessible for managers but also to measuring companies so they can rapidly estimate if there are some problems with their measurements. Furthermore, the visualisation allows to immediately spot missing data that constitute a problem when assessing the condition in the long term. In addition, along with the values of the different indicators, the storage of the recorded raw data associated to the measured parameters could significantly improve future complex analyses that are now eased by the use of more powerful analytics models.

- Weather parameters should be consistently measured at the same time as the pavement indicators. Nowadays this will not constitute an excessive added complexity or costs since portable weather stations are very reliable and not very expensive. The inclusion of weather parameter like air and surface temperatures as well as relative humidity could allow to really understand possible discrepancies in the data and also to eventually adjust the measurements. Furthermore, it is highly recommended to use the weather data to decide when to do the measurements and try when possible, to have the most similar condition possible than the ones from the previous measurements. Nevertheless, dedicated studies should be conducted in order to understand the effect of the temperature on the surface parameters and calculate correction factors that allow normalising the obtained measurement results to a pre-defined reference temperature.

- Considering the influence that having different companies plays in the variability of the data it is recommended that contracts are made for two consecutive survey years with one company so there is a certainty to avoid the company variability at least every two consecutive surveys. Moreover, it is recommended that companies validate their measurement equipment by calibration protocols specifically designed for each performance evaluation. In addition, companies should be advised to participate in regular testing campaigns to assess the reliability and accuracy of measurements perform a calibration of their equipment or at least a regular comparison in order to ensure that the measurements are equivalent and thus can provide precise estimations of the condition.

- Considering the discrepancies that we observed with the interventions, it is recommended to develop a more systematic way of recording these procedures, including not only major interventions (rehabilitation) but also minor ones (maintenance) that can also play a role in the development of the pavement condition, particular for some indicators. This points also highlights the importance of geographically referencing the measurements in order to enhance their traceability, repeatability, and reproducibility as well as allow a better comparison between survey campaigns.

- We recommend that the analysis performed during this project should be expanded to include more data that could explain the variability that the current variables could not address. In particular we recommend including the traffic data, especially heavy traffic data to further explain the pavement deterioration. Also, we recommend including characteristics of the type of road material, topographic data and soil parameters to understand their effect on the development of the pavement condition.

- Finally, we recommend investing in the further development of the Trassee application to consolidate all relevant data linked to pavement conditions. This should include enhanced visualization features and an intuitive, user-friendly interface to ensure accessibility. By doing so, managers, measurement companies, scientists, and decision-makers will have an efficient and straightforward way to access and utilize this critical information.

Publikationen:

"Exploring pavement friction variability factors using ensemble trees and causal inference." Zihang Weng, Marcelo Galleguillos-Torres, Bryan T. Adey, Saviz Moghtadernejad, Yuchuan Du. Infrastructure Asset Management (In Press)

Der Projektleiter/die Projektleiterin:

Name: Öngel Vorname: Aybike

Amt, Firma, Institut: Berner Fachhochschule - Architektur, Holz und Bau (AHB)

Unterschrift des Projektleiters/der Projektleiterin:

Forschung im Strassenwesen des UVEK: Formular 3 **Seite 2 / 3**

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK

Formular Nr. 3: Projektabschluss

Beurteilung der Begleitkommission:

Beurteilung:

Von den Hauptzielen, Verbesserung
- der Genauigkeit
- der Verständlichkeit
- der Wiederholbarkeit
- und der Reproduzierbarkeit

der maschinellen Zustandserfassung konnte aufgrund der Datenlage lediglich die Verständlichkeit erhöht werden.

Die Unterziele
- Veranschaulichung der Messschwankung
- Identifizierung von Ursprüngen der Messschwankungewn
- Vorschläge, Messchwankungen herauszufiltern
- Vorschläge, wie die Messchwankungen reduziert werden können

wurden im Wesentlichen erreicht.

Umsetzung:

Kantonale Daten waren im Projekt vorgesehen, wurden aber nicht berücksichtigt da teilweise die Georeferenzierungen fehlten und es im gegebenen Rahmen zu einem Zeit- und Ressourcenmangel kam. Trotz diversen personellen Änderungen wurde das Projekt mit einer 6-monatigen Verlängerung zeitgerecht zum Abschluss gebracht.

weitergehender Forschungsbedarf:

Der Forschungsbericht enthält diverse Verbesserungsvorschläge, die mit weiteren Untersuchungen in der Praxis valdiert werden sollten.

Einfluss auf Normenwerk:

Direkte Anpassungsvorschläge für das schweizerische Normenwerk waren im Projekt nicht vorgesehen.

Der Präsident/die Präsidentin der Begleitkommission:

Name: Bühlmann Vorname: Erik

Amt, Firma, Institut: Grolimund + Partner AG

Unterschrift des Präsidenten/der Präsidentin der Begleitkommission:

Forschung im Strassenwesen des UVEK: Formular 3 Seite 3 / 3