Journal of Signal Processing Systems
https://doi.org/10.1007/s11265-024-01939-2

Adaptable Accelerometer Signal Processing Pipelines for Smartphone based Evenness Estimation

Friedrich Rieken Münke[1,2] · **Manuel Schenk**[2] · **Sandra Murr**[2] · **Markus Reischl**[1]

Received: 10 August 2024 / Revised: 23 October 2024 / Accepted: 27 October 2024

Abstract
Evenness is an essential indicator of road quality. Accelerometer sensors in smartphones offer an accessible and cost-efficient solution for monitoring road evenness. However, the accelerometer signal from smartphones is influenced by various internal and external factors beyond the road's actual evenness. External factors, in particular, can introduce systematic bias due to differences in vehicle suspension or smartphone mounting methods. In this study, we investigate specific external factors affecting windshield-mounted smartphones and propose a method to automatically adapt accelerometer processing pipelines, improving robustness against such external influences.

Keywords Evenness · Road · Infrastructure · Accelerometer · Adaptable processing

1 Introduction

The quality of a road is closely linked to its evenness, which is a critical factor in ensuring safe and efficient transportation networks. Continuous monitoring of road evenness facilitates timely repairs and enables authorities to improve overall road safety. Accelerometer-equipped smartphones offer significant opportunities for assessing road infrastructure. They have high-quality sensors, are widely available, easily adaptable, and provide seamless data transfer.

However, the raw accelerometer signals recorded by smartphones do not correlate directly with the underlying road surface. The signals are influenced by a combination of internal and external factors. On the one hand, internal factors such as vehicle speed, gear ratio, braking and driving direction vary during the recording of the accelerometer signal. On the other hand, external factors, including suspension characteristics, vehicle type, tire pressure, sprung mass, sensor position, and sampling rate [1–5], remain constant during recording and uniformly affect the accelerometer signal. Therefore, a robust approach is needed to evaluate road evenness based on mobile phone sensors, taking into account both internal and external factors.

Previous studies [6–13] compute road quality measures based on accelerometer signals and vehicle speed, focusing primarily on compensating for internal factors while overlooking external factors. This leads to unpredictable and inconsistent results when the external and internal factors change (e.g., a different vehicle is used). Yu et al. [14] propose a comprehensive framework for comparing accelerometer-based evenness estimation methods. The framework considers the accuracy of the evenness estimation as well as the robustness to external factors, taking into account the vehicle type, vehicle speed, and mounting type. The article goes on to evaluate three different methods and selects the method in [2] as the best performing in terms of evenness prediction and robustness against external factors. The method in [2] is a calibration-based approach that creates a physical (half car) model with unknown parameters to compensate for external factors such as sprung mass. These parameters are estimated based on the response of

✉ Friedrich Rieken Münke
friedrich.muenke@me.com

Manuel Schenk
m.schenk@vialytics.de

Sandra Murr
s.murr@vialytics.de

Markus Reischl
markus.reischl@kit.edu

[1] Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz, 76344 Karlsruhe, Baden Württemberg, Germany

[2] Data Science, vialytics GmbH, Silberburgstraße 187, 70178 Stuttgart, Baden Württemberg, Germany

Published online: 29 November 2024

the physical model to a known flatness (5 cm high speed bump) at a known speed (10 km/h). This method is capable of estimating evenness consistently at the cost of dedicated calibration drives. However, the approach fails when the user changes vehicles and frequently reinstalls a smartphone. Another method with a self-calibrating algorithm, as proposed by Yamamoto et al. [15], calibrates the parameters of a physical model during signal acquisition, eliminating the need for a separate calibration run, but at the same time limiting the possible smartphone positions to a rigid mounting on the vehicle floor. This limitation reduces the number of parameters to be calibrated, but is particularly problematic if the goal is to support other types of smartphone mounting (e.g., windshield mounting), as it introduces additional parameters such as mount stiffness and smartphone weight. External factors are also a challenge for commercial applications such as RoadLab [16], RoadRoid [17], and RoadBump [18], which rely on the user to manually select the vehicle type and its suspension type (hard/medium/soft). This allows the application to select the appropriate set of parameters for its underlying algorithm and adapt the flatness estimation to the current vehicle. This method only supports a limited number of options and is not able to adapt to more complex scenarios involving complex attachments such as windshield mounts.

While there are methods that can adapt to external factors, they come with significant drawbacks such as dedicated calibration runs or do not support complex mounts (e.g., windshield mount). The windshield mount configuration is of particular interest because it allows users to conveniently access smartphone information while driving and enables data collection using the smartphone's camera.

In this paper, we propose a method that is able to automatically adapt to external factors and achieve accurate evenness estimation, while not restricting the type of mount or requiring calibration runs. To evaluate our method, we present a diverse real-world dataset consisting of different vehicles, smartphones, mounts, and speeds. The use case presented in this paper focuses on windshield-mounted smartphones.

2 Method

2.1 Overview

Our method is outlined in Fig. 1. In the first step, we define an algorithm that is able to provide an evenness prediction $\hat{y}$ based on an accelerometer signal x. This includes the general structure, such as the number and type of processing steps and the order of processing steps.

In the second step, we adapt the algorithm blueprint to our use case. When recording accelerometer signals x, we consider external factors as static and summarize them as the sensor setup. Each use case requires a different set of sensor setups, and it is important to adapt the algorithm blueprint accordingly. We accomplish this by introducing a scoring method that considers both internal factors and the sensor setup when evaluating the evenness prediction.

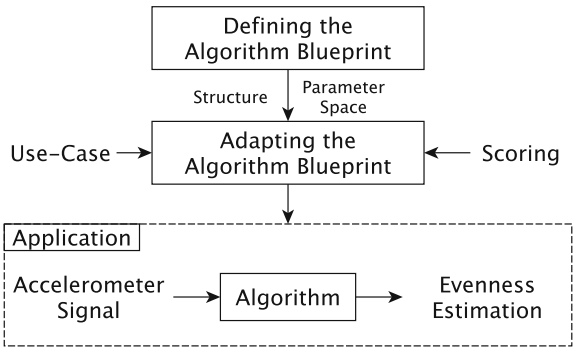

Figure 1 We aim to predict the evenness of the road based on an accelerometer signal while being robust against deviations from the sensor setup. First, we define the general algorithm. This general algorithm is then adapted to the given data set based on our scoring system.

The adapted algorithm can then be used to process accelerometer signals and estimate road surface evenness. This method allows the user to quickly adapt and optimize an algorithm for different sensor setups.

2.2 Definition Algorithm Blueprint

In our application, we require an algorithm that can transform an accelerometer signal x of arbitrary length to an evenness prediction $\hat{y}$. We call this algorithm an Accelerometer Signal Processing Pipeline (ASPP). An ASPP is a series of signal processing operations with an aggregation operation at the end. The signal processing operations transform the input signal x into a new signal x^*. This can remove noise or certain frequencies from a signal. Aggregation summarizes the processed signal into a scalar evenness prediction $\hat{y}$. The evenness estimation $\hat{y}$ is not calibrated and does only correlate with the evenness y. We define the algorithm blueprint as a general ASPP *GASPP-C* of complexity C, where C is the number of signal processing operations before aggregation. The *GASPP-C* is visualized in Fig. 2.

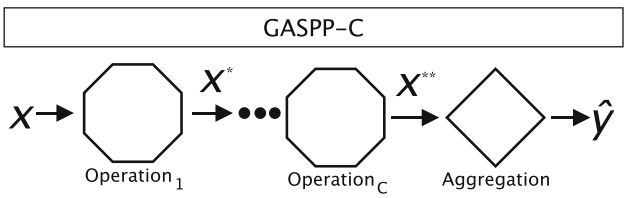

Figure 2 The *GASPP-C* is able to compute an evenness prediction $\hat{y}$ for an accelerometer signal x. The complexity C indicates the number of signal processing operations before the mandatory aggregation. The *GASPP-C* defines the parameter spaces P with N_P possible configurations, where a specific configuration is referred to as ASPP.

Besides the general structure and number of processing operations, we can further specify the type of operations for each step and the corresponding parameters. These define the parameter space P of the *GASPP-C*, where an ASPP is a parameterized version of the *GASPP-C*. For signal processing operations, we consider the Moving Average Filter (AVG) and Bandwidth Filtering (BND) operations as suggested in [4]. As a BND filter we have chosen the Butterworth bandwidth filter [19]. In addition, we test other filter kernels such as the Ramp Filter (RMP). For the average and ramp filters, we consider kernel sizes 3, 5, 7, 9, and 11. The bandwidth filters are parameterized to filter three wide frequency bands of 0-25Hz, 25-50Hz, and 10-40Hz, and five smaller frequency bands of 0-10Hz, 10-20Hz, 20-30Hz, 30-40Hz, and 40-50Hz.

Signal aggregations are commonly used in other publications, so we evaluate Root Mean Square (RMS) [3, 9, 11], Mean of the FFT magnitudes (MOM) [6], Maximum of the FFT (MFFT) [13], and Standard Deviation (STD) [13]. While STD and RMS are similar, STD removes the mean and may produce more stable results when going up and down a hill.

We evaluate three categories of *GASPP-C* with increasing complexity ($C = 0, 1, 2$) to assess the impact of adding more signal processing operations before aggregation.

Depending on the complexity C and the given operations and their parameters, we can specify the size of the parameter space P, where N_P is the total number of possible parameter combinations of *GASPP-C*. With the specified operations and corresponding parameters we get $N_P(C = 0) = 6$, $N_P(C = 1) = 108$ and $N_P(C = 2) = 1944$ possible parameter combinations.

2.3 Adaption Algorithm Blueprint

Our proposed method, as visualized in Fig. 3, automatically adapts the *GASPP-C* to a specific use case represented by the provided dataset. The dataset contains a set of N accelerometer signals, where each signal x_i with $i \in N$ is a time series of arbitrary length. Each signal x_i has a corresponding evenness grade y_i and supplementary information such as sensor setup s_i, location l_i of the recording, and velocity v_i during the recording.

Using this information, we score all possible ASPP of the *GASPP-C* in the predefined parameter space P based on two factors: the inter-setup stability and the correlation with the reference evenness y. This ensures that the selected configuration can be applied to all relevant sensor setups while providing a high quality evenness estimation as well. The scoring process takes velocity into account because previous research has shown that velocity affects the accelerometer signal. By considering velocity in the scoring process, we aim to account for its impact on the overall assessment of setup and evenness. Scores are computed using the evenness predictions $\hat{y}$ of each ASPP.

The **setup score**, denoted as S_S, quantifies the robustness against inter-setup deviation on a range from 0 to 1. To compute this score, we group accelerometer signals x_i by their location l_i. On the same location an ASPP needs to have the same evenness prediction $\hat{y}$ regardless of the sensor setup used to record the signal. In the following we compare all setups pairwise for all locations. We normalize all evenness predictions $\hat{y}$ by dividing them by their maximum and minimum values, resulting in a range from 0 to 1. Next, we calculate the absolute average difference between the response values of each pair of setups at each location and average these differences. The resulting average differences represent the dissimilarity between the setups. To convert this dissimilarity into a similarity score, we subtract the average differences from 1. In this way, higher values of the Setup Score indicate greater consistency between setups, while a score of 0 indicates no similarity.

The **evenness score**, denoted as S_E, assesses the correlation between the evenness y in the dataset and the

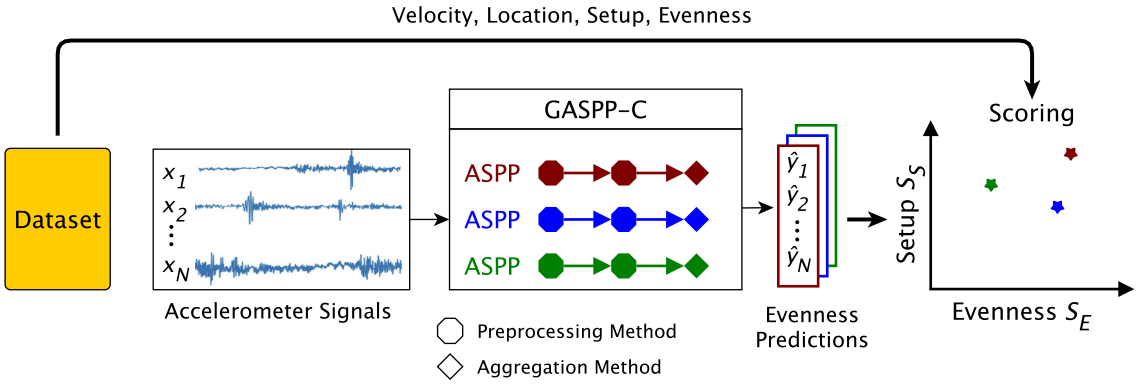

Figure 3 The method adapts the *GASPP-C* to the data set by evaluating all configurations in the predefined parameter space. Each ASPP is scored for their inter-setup deviation (S_S) and their evenness prediction capabilities (S_E). The best performing ASPP on the given data set is selected and can be further applied to the use case.

evenness predictions $\hat{y}$ of an ASPP. The velocity v_i affects the accelerometer signal $\bar{x}_i$. So we group the accelerometer signals $\bar{x}_i$ based on their velocity v_i and divide them into bins with 10 km/h intervals. This allows us to score the evenness predictions $\hat{y}$ without considering the velocity within the ASPP itself. Within each bin, we calculate the Pearson correlation coefficient between evenness y and the corresponding evenness prediction $\hat{y}$ as an individual score for each velocity bin. The final value is obtained by averaging the correlation coefficients across all velocity bins.

The overall score S is then calculated as:

$$S = \frac{1}{2}(S_S + S_E). \tag{1}$$

Finally, our method selects the ASPP with the highest score S. This evaluation process ensures that the ASPP that performs best on the given dataset is selected. It is performed automatically, making the method transferable to other datasets and use cases. By leveraging these findings, researchers and practitioners can apply the optimal ASPP to estimate road evenness in different contexts and improve the reliability of their results.

2.4 Application

The final and parameterized ASPP is static and can predict accelerometer signals x of arbitrary length. Its evenness prediction $\hat{y}$ is not calibrated and the user may set thresholds to match their desired evenness prediction standards.

3 Dataset

The data set used in this publication serves as a benchmark for the presented method and represents the use case. The evaluated *GASPP-C* are tested for evenness correlation and inter-setup deviation. Thus, we require data that covers many separate individual setups as well as different degrees of evenness. We have recorded the same area with 10 different setups. These setups include five different vehicles and three different iPhone models, as shown in Table 1. The accelerometer signal is recorded at a sample rate of 100Hz and we only consider the vertical axis in the smartphone accelerometer. Since we are focusing on windshield mounted smartphones, all data is recorded using a windshield mount as shown in Fig. 4.

For the ground truth evenness y, we rely on the German standardized system for grading road conditions known as ZEB (Zustandserfassung und -bewertung) [20]. The ZEB system uses special vehicles equipped with laser scanners and multiple cameras to assess road conditions. Within the ZEB system, road condition is graded based on various factors, including evenness along and perpendicular to the direction of travel. We focused solely on evenness in the direction of travel, which is assigned a grade ranging from 1 (good) to 5 (bad).

Table 1 Summary of all test runs and setups.

Id	Vehicle	iPhone version
SUV	Cupra formentor (SUV)	11 / 13
Van 1	Opel vivaro (Van)	11
Van 2	VW Caddy (Van)	11 / 13
Van 3	Mercedes benz vito 116L (Van)	11 / 13 / XR
Car	Nissan micra (Car)	11 / 13

In total, we collected 10 individual setups, with each setup covering a distance of 17,171m (324 segments)

The ZEB divides roads into segments and assigns a grade to each segment. These segments are typically either 20 meters long in cities or 100 meters long outside of cities. A segment corresponds to a location and has an evenness grade y_i attached to it. A data point in our dataset corresponds to a recording of a segment, adding the average speed in km/h, the setup identifier, and the accelerometer signal x_i. In total, our dataset contains 324 segments with a total length of 17.171 km. All segments were recorded 10 times (once with each setup).

In order to explore the malicious effects of different setup parameters, we conduct a dedicated test series focusing on different aspects of windshield-mounted smartphones, such as the position behind the windshield. By systemati-

Figure 4 Exemplary sensor setup used to record the dataset.

Table 2 Results of the *GASPP-0* focusing on different aggregation methods applied to raw accelerometer signals without pre-processing.

Aggregation	S_E	S_S	S
RAW-MAX	0.502	0.937	0.72
RAW-STD	0.58	0.909	0.744
RAW-RMS	**0.586**	0.909	**0.747**
RAW-MOM	0.532	0.937	0.735
RAW-MFFT	0.359	**0.946**	0.653

The best results are highlighted in **bold**

cally varying these parameters and evaluating the resulting accelerometer signals, we gain insight into the potential biases and limitations introduced by windshield mounting. This analysis provides valuable information on the challenges and constraints associated with utilizing smartphones as sensors for road evenness estimation in this particular setup.

- Windshield Position: Comparison of two equally configured smartphones, both positioned at different locations behind the windshield. We compare a position in the center and to the right of the windshield.
- Mount Strength: Mounts can be installed with tightened and loose screws.
- Mounting Type: Different windshield mounts can affect the accelerometer signal. We compare a one-joint and a two-joint mount.
- Different iPhone types: The type of smartphone can make a difference, especially considering the corresponding weight differences (iPhone 13 Pro / iPhone 13 XR with and without case).

4 Results

4.1 Overview

In this section, we present and discuss the results of our evaluation. We examine three versions of the *GASPP-C*, where

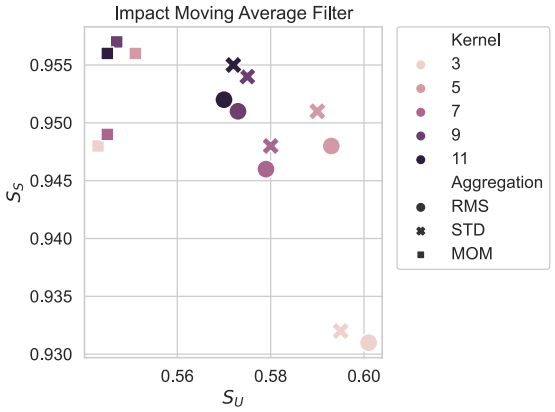

Figure 5 Effect of the average filter with different kernel sizes on the two scores S_S and S_E for the aggregation methods RMS, STD and MOM. We observe that larger kernel sizes improve the setup score S_S, while reducing the evenness score S_E.

$C = 0, 1, 2$ represents the number of signal processing operations before aggregation. By categorizing the *GASPP-C* based on their complexity C, we can systematically analyze the impact of different processing techniques on the estimation of road evenness using accelerometer signals. This approach allows us to understand the effectiveness of different combinations of preprocessing and aggregation methods in capturing the relevant information and achieving accurate results.

4.2 GASPP-0

The results in Table 2 provide insight into the performance of different aggregation methods for estimating road evenness using the raw accelerometer signals without pre-processing. Among the aggregation methods considered, those that take the full signal into account, such as STD and RMS, exhibit the highest evenness scores ($S_E \geq 0.58$), indicating a strong correlation between the prediction $\hat{y}$ and the evenness y. This suggests that these aggregations effectively capture the variation in road conditions.

Table 3 The top 5 results for the different configurations of *GASPP-1*.

ASPP Conf.	S_E	S_S	S
AVG5-STD	0.590	0.951	0.770
AVG5-RMS	0.593	0.948	0.770
RMP3-RMS	0.601	0.937	0.769
RMP3-STD	0.596	0.938	0.767
RMP5-STD	0.582	0.952	0.767

Sorted by highest overall score S

Table 4 The top 5 results for the different configurations of *GASPP-2*.

ASPP Conf.	S_E	S_S	S
AVG3AVG3-RMS	0.599	0.943	0.771
AVG3RMP3-RMS	0.597	0.944	0.771
RMP3AVG3-RMS	0.597	0.944	0.771
RMP3RMP3-RMS	0.595	0.945	0.770
RMP3RMP3-STD	0.592	0.948	0.770

Sorted by highest overall score S

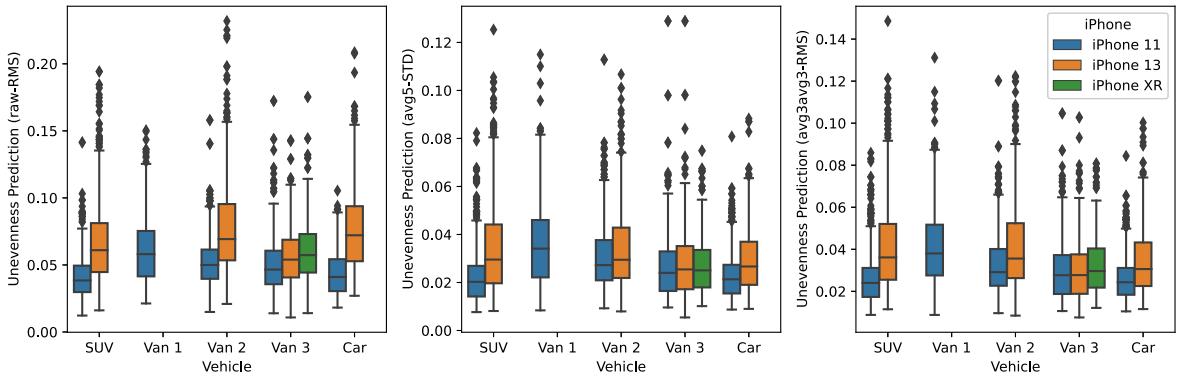

Figure 6 The evenness predictions $\hat{y}$ of the highest performing *GASPP-C* separated by individual setups. [Left: *GASPP-0* (Configuration: RAW-RMS), Middle: *GASPP-1* (Configuration: AVG5-STD), Right: *GASPP-C* (Configuration: AVG3AVG3-RMS)].

When it comes to evaluating the consistency across different setups, the MFFT, MOM, and MAX aggregations show the highest setup scores ($S_S = 0.946$), indicating a relatively low deviation of about $S_S \pm 0.054$ between different setups. The other aggregations show a higher setup deviation of about $S_S \pm 0.1$. This suggests that the former aggregations are more robust and less affected by variations in setup parameters.

Considering the overall performance, the RMS method stands out as the most effective approach, achieving the highest average score ($S = 0.747$). The RMS method takes into account the offset to zero, which contributes to its superior performance compared to the STD method ($S = 0.744$).

In the following evaluations we will focus only on the aggregation methods RMS, STD, since they have a higher score S compared to other aggregation methods.

4.3 GASPP-1

The next set of experiments evaluates a *GASPP-1* with a signal processing operation. As specified in Section 2.2, we consider three different filters (AVG, RMP, and BND). The results are shown in Table 3. The highest scores were achieved by the AVG5-STD configuration with $S_E = 0.590$, $S_S = 0.951$ and $S = 0.770$, and by the AVG5-RMS configuration with $S_E = 0.593$, $S_S = 0.948$ and $S = 0.770$. The results of the ramp filter are similar to those of the average filter. The BND filter, on the other hand, cannot compete with the best performing configuration, BND00/10-STD ($S = 0.760$).

In comparison to the previous results of *GASPP-0*, the introduction of one preprocessing operation improves the evenness score $S_E = +0.004$ and the setup score $S_S = +0.042$.

We have done a detailed study of the effect of the moving average filter with different kernel sizes (3, 5, 7, 9, 11). The results are visualized in Fig. 5. We observe that the filtering process with increasing kernel size reduces the evenness score S_E while increasing the setup score S_S, indicating a trade-off between the two.

4.4 GASPP-2

In this section we cover the category *GASPP-2*, which uses two preprocessing methods. The results are summarized in Table 4. The best performing *GASPP-2* is the AVG3AVG3-

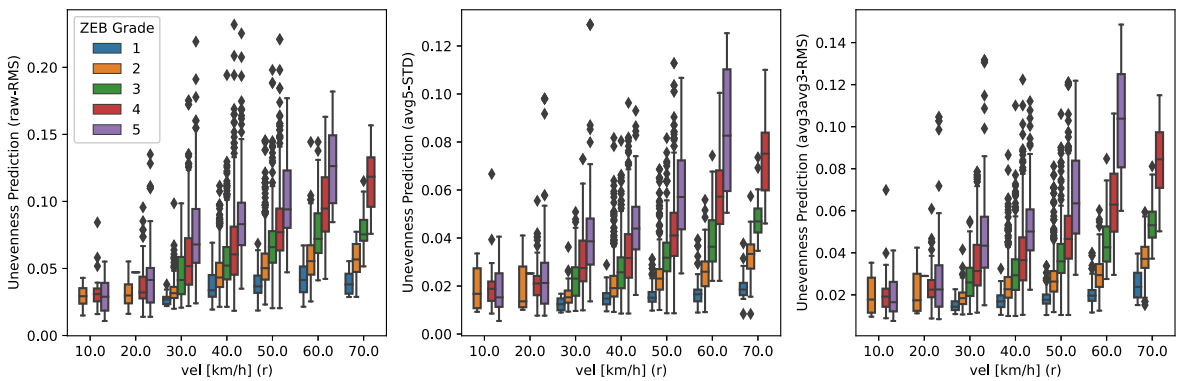

Figure 7 The evenness prediction $\hat{y}$ of the highest performing *GASPP-C* divided into velocity bins (10 km/h) and corresponding ZEB grades. [Left: *GASPP-0* (Configuration: RAW-RMS), Middle: *GASPP-1* (Configuration: AVG5-STD), Right: *GASPP-C* (Configuration: AVG3AVG3-RMS)].

Table 5 Comparison of the setup score S_S for the three top performing ASPP for each complexity category.

ASPP Conf.	*GASPP-0*	*GASPP-1*	*GASPP-2*
Position	0.952	0.978	0.975
Mount strength	0.895	0.853	0.841
Mount type	0.898	0.976	0.938
Phone type	0.895	0.970	0.942

RMS configuration ($S_E = 0.599$, $S_S = 0.943$ and $S = 0.771$). Increasing the number of preprocessing operations results in an overall improvement of $+0.001$ compared to *GASPP-1*. The evenness score $S_E = +0.009$ improves, compensating for the diminished setup score $S_S = -0.008$.

4.5 Visualization

We visualized the results in two plots, the first focusing on the interpretation of the setup score S_S (Fig. 6) and the second focusing on the evenness score S_E (Fig. 7). For both plots, we predict the accelerometer signal x of each segment from our dataset using the three best performing ASPPs for each complexity (RAW-RMS, AVG5-STD, AVG3AVG3-RMS).

In Fig. 6 we group all evenness predictions $\hat{y}$ by their setup, where each setup has a unique combination of smartphone and vehicle type. Since all setups cover the same locations, we expect the distribution of evenness predictions $\hat{y}$ of each setup to be the same.

We can observe a systematic impact of the sensor setup on the evenness prediction $\hat{y}$, even when the accelerometer signal x was recorded from within the same vehicle. Specifically, we observe that the iPhone XR has the highest evenness prediction $\hat{y}$ range, followed by the iPhone 13 Pro and then the iPhone 11. These differences show a clear trend, but they are of different magnitude, so we expect other external factors besides smartphone type to influence the signal. The differences between cars are less obvious and seem to be random.

The *GASPP-0* (RAW-RMS) is most affected by the influence of the sensor setup and shows the highest relative deviation related to the sensor setup. Our experiments in Sections 4.3 and 4.4 show the positive impact of preprocessing operations on the setup score S_S. This can also be seen in Fig. 6. We can see that the different setups are more homogeneous than before.

We compare the evenness prediction $\hat{y}$ and the evenness grade y in Fig. 7. We observe a clear influence of the vehicle velocity on the evenness prediction $\hat{y}$. There is also a noticeable correlation between the evenness y and the prediction $\hat{y}$, although there is some overlap, especially at lower velocities. The evenness prediction $\hat{y}$ is increasingly distinct for more preprocessing operations. The overlap between different grades is highest for *GASPP-0* (RAW-RMS) and lowest for *GASPP-2* (AVG3AVG3-RMS), which is particularly noticeable for the 60 km/h speed bin. Here the classes are clearly separated from *GASPP-1* (AVG5-STD). The general problem of speeds below 30 km/h is still present, where there are hardly any differences between the grades.

4.6 Windshield Related Parameters

In this section, we will discuss the impact of windshield-specific factors on the sensor setup. For each isolated factor, we collected a test series and evaluated the impact on different ASPP configurations. For each test series, we compare the behavior of the best performing ASPP for each complexity category (*GASPP-0*: RAW-RMS, *GASPP-1*: AVG5-STD, *GASPP-C*: AVG3AVG3-RMS) to provide detailed insight into the effect of using pre-processing operations.

The results are summarized in Table 5. We can see that the impact of the windshield position is small compared to the other test series, as the setup score is generally high $S_S \geq 0.952$ even when no preprocessing is used. Overall, the introduced preprocessing operations improve the setup score S_S (position: $+0.026$ / $+0.023$, mounting type: $+0.078$ / $+0.04$ and phone type: $+0.065$ / $+0.047$). The only setup parameter that cannot be improved by preprocessing is the mounting strength (-0.042 / -0.054). In this case, the mounting screws were loose and we suspect that the mounting was permanently displaced. All other cases were successfully compensated by our *GASPP-1* and *GASPP-2*.

5 Conclusion

In this paper we present a general algorithm (*GASPP-C*) that is able to predict the evenness $\hat{y}$ based on a provided accelerometer signal x. We provide a new method to adapt this *GASPP-C* to a data set. Our method automatically evaluates all configurations of the *GASPP-C* on the dataset considering two main aspects: robustness against sensor setup deviations and correlation with evenness. The method selects the configuration with the highest combined score. This process ensures that the ASPP is adapted to the intended use case presented by the dataset. This method allows the user to adapt the *GASPP-C* to a new use case as long as a dataset can be provided.

We benchmark our method using our dataset, which is focused on evenness detection with windshield-mounted smartphones. The dataset is diverse, with different sensor setups for recording the smartphone accelerometer signals, and corresponding high quality ground truth recorded with dedicated measurement vehicles.

We compare three different complexities of ASPP and conclude our tests with the best performing ASPP being AVG3AVG3-RMS, which uses two moving average filters with kernel size 3 and root mean square as aggregation. In additional experiments, we show that averaging filters increase robustness to variations in sensor setup. We identify loose mountings as the remaining setup parameter that cannot be compensated by our method.

In future work, we aim to record a continuous velocity signal and all accelerometer axes to prevent events such as breakage from affecting the signal processing. Additionally, we want to analyze the impact of individual shocks unrelated to road evenness, such as from driving over manholes, speed bumps, and sidewalks. These events can trigger a high evenness prediction while not necessarily reflecting the evenness of the road surface.

Appendix A Score Evaluation

We assessed the functionality of the evenness score S_E and the setup score S_S through three experiments. The results are presented in Table 6. We use the dataset presented in Section 3 as a foundation and create hypothetical ASPPs to generate evenness predictions for each accelerometer signal $\bar{x}_i$ in the dataset.

The first hypothetical ASPP (RNG), generates random evenness predictions. As expected, the evenness score (S_E) is low at 0.154 since there is no correlation possible with the random evenness predictions. On the other hand, the setup score (S_S) is high at 0.985 since there is no pattern corresponding to the setup a_i. The second hypothetical ASPP (GTR) outputs the existing reference evenness $\hat{y}$ for any accelerometer signal $\bar{x}_i$. As the reference evenness $\hat{y}$ represents the true evenness without any correlation with the setup, it achieved perfect scores ($S_E = 1$ and $S_S = 1$) for both evenness and setup. The last hypothetical ASPP (SET) outputs for each accelerometer signal $\bar{x}_i$ the unique ID for each setup as referenced in the data set. Since these fixed values had no correlation with evenness, both evenness ($S_E = 0.114$) and setup ($S_S = 0.0$) scores were low.

Table 6 These experiments showcase the general functionality of the introduced scoring system to evaluate different response values.

Experiment	S_E	S_S	S
RNG	0.154	0.985	0.569
GTR	1.0	1.0	1.0
SET	0.114	0.0	0.057

[RNG: Randomly generated response values, GTR: evenness ground truth as response value, SET: Encoded setup as response values]

Appendix B GASPP-1

In the evaluation of the top 20 *GASPP-1* configurations (Table 7), we can clearly observe the influence of the RMP filter and the BND filter on the scores. The RMP filter shows a remarkable performance that is close to the average filter. Overall, the RMP filter shows a higher evenness score (S_E), whereas the average filter outperforms in terms of the setup score (S_S). On the other hand, the BND filter performs worse, but still better than all ASPP-0 (without any pre-processing) with an overall score of $S = 0.760$. Note that only the low frequency of the BND filter performs well.

Table 7 The top 20 results for the different configurations of ASPP-0 and ASPP-1.

ASPP Conf.	S_E	S_S	S
AVG5-STD	0.590	0.951	0.770
AVG5-RMS	0.593	0.948	0.770
RMP3-RMS	0.601	0.937	0.769
RMP3-STD	0.596	0.938	0.767
RMP5-STD	0.582	0.952	0.767
AVG3-RMS	0.601	0.931	0.766
RMP5-RMS	0.582	0.950	0.766
AVG9-STD	0.575	0.954	0.764
AVG7-STD	0.580	0.948	0.764
AVG3-STD	0.595	0.932	0.764
AVG11-STD	0.572	0.955	0.763
AVG7-RMS	0.579	0.946	0.763
AVG9-RMS	0.573	0.951	0.762
RMP7-STD	0.570	0.954	0.762
AVG11-RMS	0.570	0.952	0.761
BND00/10-STD	0.576	0.945	0.760
BND00/10-RMS	0.576	0.945	0.760
RMP9-STD	0.566	0.954	0.760
RMP7-RMS	0.568	0.951	0.760
RMP11-STD	0.563	0.954	0.759

Sorted by highest overall score S

Appendix C GASPP-2

We show the top 20 *GASPP-2* considering all complexity categories (*GASPP-0*, *GASPP-1*, *GASPP-2*). We notice that the improvement from *GASPP-1* to *GASPP-2* (+0.1%) is not as significant as from *GASPP-0* to *GASPP-1* (+2.4%). In addition, we find that the bandwidth filter still does not perform as well as the average and ramp filters.

Table 8 The top 20 results for all ASPP complexity categories (ASPP-0, ASPP-1, ASPP-2).

ASPP Conf.	S_E	S_S	S
AVG3AVG3-RMS	0.599	0.943	0.771
AVG3RMP3-RMS	0.597	0.944	0.771
RMP3AVG3-RMS	0.597	0.944	0.771
AVG5-STD	0.590	0.951	0.770
AVG5-RMS	0.593	0.948	0.770
RMP3RMP3-RMS	0.595	0.945	0.770
RMP3RMP3-STD	0.592	0.948	0.770
RMP3AVG3-STD	0.594	0.947	0.770
AVG3RMP3-STD	0.594	0.947	0.770
AVG3AVG3-STD	0.595	0.945	0.770
RMP3-RMS	0.601	0.937	0.769
AVG3AVG5-STD	0.585	0.952	0.768
AVG5AVG3-STD	0.585	0.952	0.768
AVG5RMP3-STD	0.584	0.952	0.768
RMP3AVG5-STD	0.584	0.952	0.768
AVG3AVG5-RMS	0.586	0.949	0.768
AVG5AVG3-RMS	0.586	0.949	0.768
RMP3-STD	0.596	0.938	0.767
AVG5RMP3-RMS	0.585	0.949	0.767
RMP3AVG5-RMS	0.585	0.949	0.767

Sorted by highest overall score S

Author Contributions **Conceptualization:** Friedrich Münke
– **Methodology:** Friedrich Münke
– **Formal analysis and investigation:** Friedrich Münke, Manuel Schenk, Markus Reischl
– **Writing – original draft preparation:** Friedrich Münke, Markus Reischl
– **Writing – review and editing:** Friedrich Münke, Sandra Murr, Manuel Schenk, Markus Reischl
– **Funding acquisition:** Sandra Murr, Markus Reischl
– **Supervision:** Markus Reischl

Funding Open Access funding enabled and organized by Projekt DEAL. This publication results from the project "Sensorik- und KI-basierte Straßenzustands-Analyse - SEKISA," funded by the BMDV mFUND, Förderrichtlinie 1 "Mobilität der Zukunft."

Availability of Data and Materials The dataset is available upon request from the corresponding author.

Code Availability The code is available upon request from the corresponding author.

Declarations

Conflict of Interest/Competing Interests The authors declare that they have no financial, non-financial, or other competing interests.

References

1. Islam, S., Buttlar, W., Aldunate, R., & Vavrik, W. (2014). Use of cellphone application to measure pavement roughness. In: *Conference: Second Transportation Development Congress 2014,* (pp. 553–563). https://doi.org/10.1061/9780784413586.053
2. Zhao, B., & Nagayama, T. (2017). IRI Estimation by the frequency domain analysis of vehicle dynamic responses. *Procedia Engineering, 188*, 9–16.
3. Wang, G., Burrow, M., & Ghataora, G. (2020). Study of the factors affecting road roughness measurement using smartphones. *Journal of Infrastructure Systems, 26*(3), 04020020. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000558
4. Alatoom, Y. I., & Obaidat, T. I. (2022). Measurement of street pavement roughness in urban areas using smartphone. *International Journal of Pavement Research and Technology, 15*, 1021–122.
5. Múčka, P. (2023). Relation between seated person vibrations and the international roughness index. *Transportation Research Record, 2677*(2), 1–14. https://doi.org/10.1177/03611981221147210
6. Douangphachanh, V., & Oneyama, H. (2013). A study on the use of smartphones for road roughness condition estimation. *Journal of the Eastern Asia Society for Transportation Studies, 10*, 1551–1564.
7. Alessandroni, G., Klopfenstein, L., Delpriori, S., Dromedari, M., Luchetti, G., Paolini, B., Seraghiti, A., Lattanzi, E., Freschi, V., Carini, A., & Bogliolo, A. (2014). SmartRoadSense: Collaborative road surface condition monitoring. In: *In Proceedings of the UBICOMM 2014: The Eighth International Conference on MobileUbiquitous Computing, Systems, Services and Technologies*. https://doi.org/10.13140/RG.2.1.3124.2726
8. Alessandroni, G., Carini, A., Lattanzi, E., Freschi, V., & Bogliolo, A. (2017). A study on the influence of speed on road roughness sensing: The smartroadsense case. *Sensors, 17*(2)
9. Bisconsini, D., Nicoletti, R., Nuñez, J. Y., & Fernandes Jr, J. (2018). Pavement roughness evaluation with smartphones. *International Journal of Science and Engineering Investigations*
10. Masino, J., Thumm, J., Levasseur, G., Frey, M., Gauterin, F., Mikut, R., & Reischl, M. (2018). Characterization of road condition with data mining based on measured kinematic vehicle parameters.

Journal of Advanced Transportation, 2018, 1–10. https://doi.org/10.1155/2018/8647607
11. Yeganeh, S. F., Mahmoudzadeh, A., Azizpour, M. A., & Golroo, A. (2019). Validation of smartphone based pavement roughness measures. arXiv:1902.10699. [cs.HC]
12. Martinelli, A., Meocci, M., Dolfi, M., Branzi, V., Morosi, S., Argenti, F., Berzi, L., & Consumi, T. (2022). Road surface anomaly assessment using low-cost accelerometers: A machine learning approach. *Sensors, 22*(10)
13. Sabapathy, A., & Biswas, A. (2023). Road surface classification using accelerometer and speed data: evaluation of a convolutional neural network model. *Neural Computing and Applications,* 1–12
14. Yu, Q., Fang, Y., & Wix, R. (2023). Evaluation framework for smartphone-based road roughness index estimation systems. *International Journal of Pavement Engineering, 24*(1), 2183402. https://doi.org/10.1080/10298436.2023.2183402
15. Yamamoto, K., Shin, R., Sakuma, K., Ono, M., & Okada, Y. (2023). Practical application of drive-by monitoring technology to road roughness estimation using buses in service. *Sensors, 23*(4), 1–17. https://doi.org/10.3390/s23042004
16. Wang, W., & Guo, F. (2016). RoadLab: Revamping road condition and road safety monitoring by crowdsourcing with smartphone app. In: *Transportation Research Board 95th Annual Meeting*
17. Forslof, L., & Jones, H. (2015). Roadroid: Continuous road condition monitoring with smart phones. *Journal of Civil Engineering and Architecture, 9*, 485–496. https://doi.org/10.17265/1934-7359/2015.04.012
18. Grimmer, D. (2017). RoadBump [Mobile App]. Google Play. https://play.google.com/store/apps/details?id=com.grimmersoftware.roadbumpfree&hl=en_US
19. Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer, 7*, 536–541.
20. Löcherer, Anger, Bühler, Buslaps, Gast, Klinkhart, Krause, Oertelt, Ohmen, Schniering, Socina, Ueckermann. (2006). Zusaetzliche Technische Vertragsbedingungen und Richtlinien zur Zustandserfassung und -bewertung von Strassen. Forschungsgesellschaft für Strassen- und Verkehrswesen: Technical report.