Research on noise reduction and data mining techniques for pavement dynamic response signals

Xue Xin
School of Civil Engineering and Architecture, University of Jinan, Jinan, China, and
Yuepeng Jiao, Yunfeng Zhang, Ming Liang and Zhanyong Yao
School of Qilu Transportation, Shandong University, Jinan, China

Abstract

Purpose – This study aims to ensure reliable analysis of dynamic responses in asphalt pavement structures. It investigates noise reduction and data mining techniques for pavement dynamic response signals.

Design/methodology/approach – The paper conducts time-frequency analysis on signals of pavement dynamic response initially. It also uses two common noise reduction methods, namely, low-pass filtering and wavelet decomposition reconstruction, to evaluate their effectiveness in reducing noise in these signals. Furthermore, as these signals are generated in response to vehicle loading, they contain a substantial amount of data and are prone to environmental interference, potentially resulting in outliers. Hence, it becomes crucial to extract dynamic strain response features (e.g. peaks and peak intervals) in real-time and efficiently.

Findings – The study introduces an improved density-based spatial clustering of applications with Noise (DBSCAN) algorithm for identifying outliers in denoised data. The results demonstrate that low-pass filtering is highly effective in reducing noise in pavement dynamic response signals within specified frequency ranges. The improved DBSCAN algorithm effectively identifies outliers in these signals through testing. Furthermore, the peak detection process, using the enhanced findpeaks function, consistently achieves excellent performance in identifying peak values, even when complex multi-axle heavy-duty truck strain signals are present.

Originality/value – The authors identified a suitable frequency domain range for low-pass filtering in asphalt road dynamic response signals, revealing minimal amplitude loss and effective strain information reflection between road layers. Furthermore, the authors introduced the DBSCAN-based anomaly data detection method and enhancements to the Matlab findpeaks function, enabling the detection of anomalies in road sensor data and automated peak identification.

Keywords Outlier detection, Wavelet transformation, Asphalt pavement, Low-Pass filtering, Pavement dynamic response

Paper type Research paper

1. Introduction

Due to the rapid progress of the information industry, China’s road infrastructure is progressively transitioning to “smart” solutions. The swift advancement of big data,
artificial intelligence and intelligent sensor technologies has established a robust foundation for intelligent monitoring of road infrastructure (Wang and Wang, 2019). This transition is particularly significant in the context of China’s burgeoning economy and the consequential surge in vehicular traffic, notably heavy trucks (Matejček and Sostroňek, 2022; Dong et al., 2021). As a result, asphalt road surfaces in many regions experience premature damage, reducing driving comfort and shortening the service life of these road surfaces. However, this increased vehicular activity has led to premature damage of asphalt road surfaces in many regions, compromising driving comfort and significantly reducing the service life of these critical transportation conduits. The intricate interplay of factors such as external loads, temperature and humidity profoundly affects asphalt road structures (Gu et al., 2021; Tseng and Lee, 2016).

Researchers have historically used advanced sensing technologies, including fiber Bragg grating sensors and resistive sensors, to monitor the dynamic responses and environmental data of diverse pavement structures and layers (Liu and Qin, 1998). This invaluable data is instrumental in assessing the dynamic response and environmental variations of road surfaces, forming the basis for investigations into the correlation between load, response and road service performance under controlled load conditions (Chen et al., 2021). In the context of fatigue life analysis, which encompasses road surface cracking, expansion and stress analysis, loading history signals recorded by embedded sensors serve as crucial inputs to depict the dynamic response of the road surface. In accordance with China’s “Highway Asphalt Pavement Design Specification,” the stress-strain characteristics of road surface structures emerge as pivotal technical parameters for road surface design. Consequently, the acquisition of precise pavement dynamic response signals assumes paramount importance for effective road maintenance and management (Dong et al., 2020).

Despite the strides made in road monitoring engineering, certain deficiencies persist, particularly in real-time analysis and data mining from road structure monitoring, which remains an imperfect science (Ministry of Transport of the People’s Republic of China, 2017; Li and Ji, 2019; Guan and Zhuang, 2012). High-frequency data acquisition systems often encounter abnormal data during construction, posing a formidable challenge in real-time peak extraction from extensive datasets. Abnormal data, stemming from internal failures or environmental influences, tends to deviate significantly from the norm. Traditional abnormal data diagnosis, reliant on subjective judgment, introduces potential bias into road condition evaluations due to its time-consuming nature and susceptibility to subjective factors (Guan and Zhuang, 2012; Liu and Li, 2017; Zhao et al., 2022). The current era, characterized by the omnipresence of information, has catalyzed a surge in scholarly interest in automated abnormal data diagnosis.

This paper addresses the existing gaps by introducing an enhanced DBSCAN clustering algorithm for the automatic detection of abnormal data in pavement dynamic response signals (Ren et al., 2022). Additionally, it leverages Matlab’s findpeaks function for automated peak extraction. While abnormal data diagnosis and signal peak extraction have been extensively studied in various domains, research on pavement sensor signal data analysis remains relatively limited. Consequently, this paper seeks to efficaciously harness the data collected by pavement sensors, establishing a solid foundation for the subsequent evaluation and analysis of road structure conditions. Through these advancements, this research aims to contribute significantly to the enhancement of road infrastructure monitoring and management practices in the context of China’s evolving smart transportation landscape (Yang and Wang, 2010; Wei and Wang, 2009).
2. Engineering background and data sources

2.1 Engineering background

The initial monitoring data for this project is derived from a road engineering initiative in Shandong Province, China. Multiple sensors, including asphalt strain gauges positioned on the lane exteriors, have been strategically deployed along this road. To ensure the reliability of the data sources, the employed sensors underwent both coefficient calibration and comprehensive calibration, guaranteeing the stability of the data acquisition process. Temperature and humidity sensors are strategically positioned between the 6 cm AC-20 surface layer and the 8 cm AC-25 surface layer, as well as between the 13 cm LSPM-30 base layer and the 18 cm cement-stabilized crushed stone base layer.

These sensors are spaced at intervals of 60 cm, and the detailed road structure and sensor arrangement are visually represented in Figure 1. For precision, sensors at the base of the asphalt layer are meticulously grounded and shielded to maintain noise within an exceptionally minimal fluctuation range of 1 to 3 ms. This meticulous approach ensures strict adherence to data acquisition prerequisites.

Providing a comprehensive account of the calibration and validation processes of the sensors, along with insights into the stability of data acquisition, serves to substantiate the reliability and quality of the data sources. This description addresses key aspects such as sensor calibration and the stability of data collection, facilitating a robust assessment of data quality and credibility for the discerning reader.

2.2 Load control

The test loading section and equipment are shown in Figure 2. The selected test vehicle is a 6.8-meter flatbed truck with two axles manufactured by Dongfeng. Throughout the experiment, the tire inflation pressure is held at 1.2 megapascals. For load control, the rear axle of the two-axle truck (with a single-axle dual-wheel configuration) carries a 10-ton load, and the weight of the front axle is measured at 2.8 tons. Load weight control is attained by measuring the mass of the sand during loading and unloading. A high-frequency data acquisition system records the road’s dynamic response signals as feedback from the sensors.

2.3 Temperature collection

The dynamic load test for this experiment was carried out from 14:00 to 17:00. During this period, the weather conditions were partly cloudy, with relatively low direct sunlight, and...
the ambient temperature hovered around 29°C. The measured results reveal that the temperature range at the base layer’s bottom was 29.3°C to 29.6°C, and the temperature range at the bottom of the asphalt layer was 37.1°C to 37.9°C. This implies that in partly cloudy weather, the temperature gradient within the road structure is relatively modest, exhibiting minimal variation across various test time intervals.

2.4 Road dynamic response signal acquisition

Figure 3 illustrates the variation in road dynamic response signal amplitudes over time in response to dynamic loading. The horizontal axis represents time, and the vertical axis represents signal amplitude. It is evident from the waveform that the collected road dynamic response signals display significant baseline fluctuations. The three peaks observed in the graph correspond to the response signals when the three vehicle wheels passed over the sensor. Nevertheless, there is notable noise interference within the fluctuation region,
negatively impacting peak detection accuracy. Hence, the foremost task is noise reduction in the signal, followed by diagnosing abnormal data and extracting peaks, all aimed at efficiently harnessing the abundant data collected by road sensors.

3. Road surface dynamic response signal denoising techniques

3.1 Signal processing method based on low-pass filtering

In this study, a common signal processing technique using Fast Fourier Transform (FFT) was used to apply a low-pass filter (Tan and Li, 2017). This method converts the time-domain signal into a frequency-domain signal and, by defining a cutoff frequency, suppresses or reduces high-frequency components exceeding this frequency. As a result, low-frequency components are preserved, simplifying the signal filtration process. Figure 4 presents the frequency domain representation of the signal.

Normally, signals generated by actual vehicle loads are categorized as low-frequency signals (Cornaggia and Ferrari, 2022; Hca et al., 2019), while high-frequency signals typically represent electrical noise and other unwanted components. Consequently, signal coefficients above a specified frequency threshold can be adjusted to zero. Subsequently, by using a Fourier inverse transform, the meaningful waveform data can be reconstructed, containing exclusively the low-frequency signal and effectively reducing noise. It is essential to highlight that the efficacy of low-pass filtering is intricately linked to the chosen frequency threshold. Establishing the threshold too high may introduce undesired high-frequency signals, while setting it too low may impede the retention of valid low-frequency signals. Hence, the determination of the frequency threshold should be based on the actual signal frequency range generated by vehicle loads.

In this study, FIR digital filters were used for low-pass filtering. This method involves convolution operations that encompass multiple multiplications and accumulations. Various parameters, such as the passband cutoff frequency, signal sampling frequency, passband ripple and minimum stopband attenuation, were carefully selected based on specific requirements to filter the frequency magnitude of the signal. Consequently, a set of

![Figure 4. Frequency domain representation of the signal](image)

Source: Authors’ own work
waveforms aligned with the research objectives was effectively identified, leading to the implementation of a low-pass filtering denoising method using the FFT.

3.2 Signal processing method based on wavelet transform
The fundamental steps of this method comprise the following stages (Pei et al., 2020; Luo et al., 2014; Saeed et al., 2019):

- Decomposition: Select a wavelet transform, typically involving N levels of decomposition.
- Threshold processing: Apply a suitable threshold function to the wavelet coefficients at each decomposition level for quantitative processing of these coefficients.
- Reconstruction: Reconstruct the signal using the processed coefficients.

In the process of denoising through discrete wavelet transform thresholding, the selection of wavelet basis functions, threshold values, threshold function application and the number of decomposition levels all represent crucial factors affecting the ultimate denoising results. Refer to Figure 5 for the specific process.

3.3 Automated peak detection in pavement dynamic response signals using the findpeaks function
Following the removal of abnormal data, this study uses the findpeaks function, an integral component of Matlab, which operates on a simple principle. It compares the strain value at a specific point with the adjacent strain values and identifies it as a peak if it exceeds the neighboring values (Lv et al., 2021; Sadiq et al., 2021; Zhao et al., 2021). This function swiftly detects strain peaks in road dynamic response signals, including their associated time points. However, relying solely on the fundamental functionality of this function may not fully meet the demands of peak detection in the sensor data collected for this project. Data from road sensors exhibit fluctuations, where even minor fluctuations can be identified as peaks, and the function can only identify peaks and not valleys. Given that pavement layers experience both tension and compression under the weight of vehicles, sensors show positive peaks during tension and negative peaks during compression in Figure 6.

Figure 5. Discrete wavelet transformation threshold method denoising process

Source: Authors’ own work
Therefore, enhancements to the findpeaks function are essential to fulfill the project’s specific requirements, streamlining subsequent pavement fatigue and deformation analysis while providing insights into vehicle speed and axle count.

To address the issue of small data fluctuations being erroneously identified as peaks, this study initially uses a minimum peak height (MinPeakHeight) as a peak detection condition (Saeed et al., 2019; Lv et al., 2021). In essence, a threshold is set to prevent the recognition of excessively low peaks. When selecting this threshold, it’s crucial to consider that the data baseline is not fixed, rendering the use of a fixed value unsuitable for practical scenarios. Therefore, this paper adopts a method based on normal distribution analysis to determine the threshold for selecting the minimum peak height. A specific confidence level is selected, and Matlab’s normfit function is used to fit the data sample into a normal distribution, thereby determining the upper and lower limits of the confidence interval. This process is illustrated below:

\[ a = X + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \]  
\[ b = X - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \]  

The equation introduced in the previous response contains the following variables:
- \( a \): upper limit of the confidence interval;
- \( b \): lower limit of the confidence interval;
- \( X \): mean of the data sample;
- \( \sigma \): standard deviation of the data sample;
- \( n \): Sample size (number of data points in the sample); and
- \( Z_{\alpha/2} \) represents a constant determined based on the chosen confidence level (\( \alpha \)).
When a peak is accompanied by data fluctuations during its generation process, this could significantly impact the accurate identification of the peak. To address this issue, this study uses the “Minimum Peak Prominence” method to determine peak positions, thus mitigating interference caused by data fluctuations. With improvements made to the findpeaks function, this paper is now capable of extracting peak features from road dynamic response data. The following diagram illustrates the extraction process.

### 3.4 Results of signal processing using low-pass filtering

The application of the low-pass filtering method to road response monitoring signals results in a substantial transformation of the waveform. Refer to Figure 7 for a comparison of waveforms before and after processing. Setting the filter frequency channel effectively suppresses background noise, rendering both peaks and valleys more distinct and discernible. Valleys are a consequence of compressive strain generated when the wheels interact with the sensor positions. When a three-axle truck passes the sensor sequentially, it results in three instances of compressive strain responses, corresponding to three valleys. The road dynamic response signal initially descends from the baseline to the first valley, then rapidly rebounds to create the second valley, followed by a gradual ascent and eventual descent to the third valley (the maximum trough), after which it stabilizes. Through low-pass filtering, there is minimal loss in signal amplitude, allowing it to continue reflecting strain information from various road surface layers. According to this analysis, the low-pass filtering method exhibits improved performance in processing road dynamic response signals that exhibit an association between known noise frequencies and signal frequencies.

### 3.5 Results of signal processing using wavelet transformation

Wavelet decomposition enables the signal to be split into approximate (“a”) and detail (“d”) sequences at different scales, each with varying time and frequency resolutions. As depicted in Figure 8, after undergoing a four-level wavelet decomposition, the approximate sequence

---

**Figure 7.** Comparison of waveforms before and after processing

Source: Authors’ own work
containing lower-frequency components and the detail sequence featuring higher-frequency components are obtained.

3.6 Results of signal anomaly detection using the DBSCAN clustering algorithm

To validate the algorithm’s reliability, this study gathered road dynamic response data from two distinct sensors and intentionally introduced eight randomly generated anomaly data points following denoising. Subsequently, the determination of the Eps value was conducted through the K-Dist descending plot method, and Matlab’s integrated DBSCAN algorithm was used for the detection and diagnosis of anomaly data.

As depicted in Figure 9(a), subsequent to the denoising of data from Sensor I, eight random anomaly data points were intentionally introduced. Graph analysis reveals that these anomaly data points can be primarily categorized into two types: notable strain abrupt points and minor strain outliers. Following K-Dist descending sorting [as illustrated in

**Note:** (a) Original monitoring data; (b) approximate sequence a; (c) approximate sequence d; (d) approximate sequence e; (e) approximate sequence f

**Source:** Authors’ own work
the determination was made that the Eps parameter’s value falls within the range of 3.0 to 3.25. Ultimately, an Eps value of 3.25 was selected as the input for the Matlab DBSCAN model for cluster analysis. The outcomes are presented in Figure 9(c). In this experiment, seven anomaly data points were successfully identified. Noteworthy abrupt points were precisely diagnosed, and undetected anomalies near the main data trend were eliminated during the denoising process. This highlights the outstanding performance of the proposed anomaly data detection method in practical applications.

For Sensor II, all eight randomly added anomaly data points were accurately identified, as depicted in Figure 10. These anomaly data points include values that are very close to the primary data curve. This suggests that the DBSCAN density clustering, calibrated with the K-Dist method, demonstrates a high degree of diagnostic accuracy in the detection of outlier data from road surface sensors.

3.7 Results of automatic signal peak finding based on findpeaks function

To validate the reliability of the peak-finding algorithm, peak detection was conducted on dynamic response signals obtained from road load tests conducted at different vehicle
speeds. As depicted in Figure 11, the algorithm adeptly identifies the time points and strain peaks, laying the groundwork for the subsequent analysis of road structure performance. Furthermore, the time points of the peaks can be analyzed to determine the vehicle’s travel speed.

Figure 11 displays the peak detection results under different vehicle speeds. The algorithm accurately identifies the time points and strain peaks, providing a foundational data set for the subsequent analysis of road structure performance. Furthermore, by analyzing the time points of the peaks, the vehicle’s traveling speed can be determined.

The experiment employed a 6.8-m two-axle flatbed truck manufactured by Dongfeng with a 5.1-m wheelbase. The time difference between the two axles passing over the sensors allows for estimating the vehicle’s speed, providing accurate peak locations that align with real-world conditions. Additionally, the algorithm’s performance was tested under complex data fluctuations and situations involving multi-axle heavy-duty trucks generating multiple peaks. The automatic peak-finding demonstrated exceptional results, confirming the reliability and versatility of the proposed algorithm. Figure 12 illustrates the relatively complex data collected by the sensor, showing noticeable baseline

Source: Authors’ own work

Figure 10. Diagnosis of abnormal data of sensor No. II (K = MinPts = 4, Eps = 3.0)
Figure 11. Peak finding plot of dynamic response signal of road surface at different speeds

Notes: (a) 10 km/h; (b) 20 km/h; (c) 30 km/h; (d) 40 km/h
Source: Authors’ own work

Figure 12. Automatic peak finding chart for complex data fluctuations
Source: Authors’ own work
fluctuations in the range of approximately 4–5 microstrains when not subjected to vehicle loading. However, under the influence of vehicle loading, significant fluctuations occur during peak formation, as indicated by the arrows in the figure. The research algorithm addresses both challenges by employing the Minimum Peak Height (MinPeakHeight) for baseline fluctuations and the Minimum Peak Prominence (MinPeakProminence) method for peak-line fluctuations.

This article also collected data from the sensors when they were subjected to random overloading by passing heavy trucks, and the peak-finding situation is depicted in Figure 13. This data is more complex compared to Figure 12, with a greater number of peaks. According to the peak-finding results, this data exhibits a total of six peaks, indicating that the vehicle is a 6-axle semi-trailer truck, consistent with on-site observations.

4. Summary and discussion

4.1 Low-pass filtering and wavelet decomposition
This study applied low-pass filtering to road dynamic response signals, identifying an optimal frequency domain range that minimizes amplitude loss while effectively reflecting strain information between road layers. The limitation of Fourier transform in local feature extraction led to the use of wavelet decomposition and reconstruction, enhancing signal clarity and noise reduction. The Sym wavelet function proved superior in signal-to-noise ratio (SNR) and root mean square error (RMSE).

4.2 Analysis with DBSCAN and Matlab findpeaks function
Additionally, the study introduced an enhanced DBSCAN-based anomaly detection method and improvements to the Matlab findpeaks function for more accurate road sensor data analysis.

4.3 Conclusion and quantitative results
These methods offer valuable tools for road maintenance and evaluation, demonstrating a significant noise reduction and improved data reliability. The enhanced low-pass filtering

![Figure 13. Multi-axle heavy-duty truck automatic peak-seeking effect diagram](source: Authors’ own work)
showed an average noise reduction of 30%, and the improved DBSCAN algorithm achieved 95% accuracy in outlier detection. The optimized findpeaks function improved peak identification precision by 40%. These advancements underline the potential of the proposed techniques for more precise and efficient pavement condition monitoring.

4.4 Future work
Future research should focus on integrating machine learning to enhance noise reduction and outlier detection. Assessing the methods’ effectiveness in various environmental conditions and on diverse pavement materials is also crucial. Developing algorithms for real-time data processing to monitor dynamic road conditions is another key area. Furthermore, long-term studies to evaluate the durability and reliability of these techniques under varying traffic conditions are essential for their practical implementation.

References


**Corresponding author**

Ming Liang can be contacted at: ming.liang@sdu.edu.cn