Vehicle trajectory fractal theory for macro-level highway crash rate analysis

Yuhan Nie [a,d] , Min Zhang [a,*] , Bo Wang [b,c,d] , Chi Zhang [b,d] , Yijing Zhao [a,d]

[a] School of Transportation Engineering, Chang'an University, Xi'an, China
[b] School of Highway, Chang'an University, Xi'an, China
[c] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, Singapore
[d] Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an, China

ARTICLE INFO

ABSTRACT

Vehicle trajectory data can reveal actual driving behavior patterns reflected by different road geometric designs, providing important insights for road safety analysis and improvements. This study aims to is to explore the correlation between vehicle trajectory fractal dimension (FD) and highway crash rate (CR) using large-scale telematics trajectory data. Specifically, we propose three methods to measure the FD of vehicle trajectories, and developed fractal parameter estimation technology. The results show that FD differences between road segments have a statistically significant effect on CR. A comparison of FD with five common surrogates in identifying high-risk crash sections reveals that FD reduces the false alarm rate from 52% to 94% (other surrogates) to 46%, with a recall rate of 95%. The fractal method enhances the dimensionality of trajectory feature analysis, refining the granularity of road safety analysis. It fully considers the interaction between road geometry design and driving behavior, revealing the complex dynamic movement of vehicles within the road system. This study provides methodological support for improving road geometry design and enhancing road safety.

1. Introduction

The highway network density has rapidly increased to meet the growing demands of traffic and mobility. However, the increase in design complexity and diversity has also led to a rise in accidents and fatalities (Dai et al., 2022). As a result, road management department and policymakers have been striving to monitor unsafe areas on the highway. The first step in implementing road safety management plans is screening, which primarily involves reducing the list of hotspots to a manageable one (Wang et al., 2024a). While the analysis method based on crash data can identify more reliable high-risk spatial locations, it requires long-term accumulation and is costly due to the small sample size and unobservable heterogeneity (Khan et al., 2023). Therefore, road safety monitoring and management are shifting towards using larger scale trajectory data as safety surrogates (or crash surrogates) (Gedamu et al., 2024).

Safety surrogates describes the parameters of network and vehicle attributes on the road, which are easier to record or collect (Nikolaou et al., 2023). Among them, micro level surrogates focus on the impact of

the gap and speed difference between the target vehicle and surrounding vehicles on crash risk (Wang et al., 2021). However, collecting trajectory data for the entire traffic flow is difficult and costly. More importantly, in most cases, trajectories are limited in spatiotemporal coverage (Arman and Tampere, 2022). The only exception is the recent high-cost automatic trajectory data collection project, I-24 MOTION, which covers several kilometers and spans several days (Wang et al., 2024b). Furthermore, researchers cannot plan passive trajectory collection based solely on the observation of traffic phenomena. Many vehicles are now equipped with mobile sensors, which, through their real-time and extensive coverage, enable data-assisted traffic management and safety assessments. The data is then uploaded to storage servers via continuous cellular or delayed Wi-Fi connections (Alrassy et al., 2023). Compared to other data collection devices, mobile sensors are low-cost and can cover any required time and spatial range (Arman and Tampere, 2024). On the other hand, the accuracy of mobile sensors' positioning and the convenience of data collection have significantly improved in recent years (Zhang et al., 2021), providing data support for analyzing the patterns of single vehicle's kinematics at different spatial and temporal moments

with finer granularity and on a larger scale (Arman and Tampere, 2024).

Kinematic triggers surrogates extracted from large-scale telematics data have been shown to be highly correlated with crashes risks (Kim et al., 2024). When kinematics triggers exceed predefined thresholds, they are considered as near-crash or safety–critical events (Pinnow et al., 2021). However, these kinematic triggers also lead to some false positive rates. On one hand, the factors contributing to crashes may come from different risk sources and result from the interaction of multiple factors, such as road geometric design, which can potentially influence driving behavior (Khan et al., 2023). On the other hand, these kinematic triggers focus only on a single dimension of kinematics or a single pattern leading to high crash rates (CR) (Ye et al., 2024), hindering a comprehensive assessment of actual vehicle motion.

Fortunately, advancements in nonlinear analysis have made it possible to rapidly, accurately, and comprehensively extract information from recorded signals (Lee and Park, 2021). This type of nonlinear analysis can utilize fractal theory (Zhao et al., 2022) to extract finergrained information from trajectories. Fractal theory is a mathematical tool used to study complex and nonlinear phenomena, focusing on analyzing the self-similar characteristics of irregular shapes and patterns across different scales (Li et al., 2024a).If we compare vehicle trajectories through road space with natural phenomena such as tree branches, leaf veins, human lungs, and exoskeletons (Moriguchi, 2023), we find that shape features may be an appropriate form to reflect the driving characteristics of trajectories. The unevenness and complexity of shape evolution reflect the deviations of actual driving paths of numerous vehicles (Hamedi and Shad, 2022), which in turn mirror the differences in road geometric elements. Compared to vehicle kinematic triggers, fractal analysis integrates partial road geometry and vehicle motion, taking into account interactions and capturing changes in both local and global patterns of vehicle paths (H. Li et al., 2023).

In order to use fractal characteristics as safety surrogates, they should be relevant to the outcomes: in our context, crash frequency and CR (Pinnow et al., 2021). The main focus of this work is to understand whether there is spatial correlation between the proposed safety surrogates and crashes, rather than inferring a causal relationship between the two. Therefore, this study examined and compared the degree of spatial correlation between fractal dimension (FD), macroscopic safety surrogates such as sight distance conformity degree, anti-skid requirement conformity degree, harsh braking, jerk, and yaw rate), and crash data (i.e., absolute crash count and CR normalized by traffic volume), as well as the ability to identify high-risk road sections. The vehicle trajectories were derived from real-world telematics data collected between June and September 2023 from in-vehicle sensing devices mounted on 724 floating trucks in the Guilin City area, with a sampling frequency of 1HZ. The vehicles were equipped with onboard cameras and managed by Guilin Highway Company, from which vehicles affected by surrounding vehicles were filtered, making them representative for evaluating the spatial characteristics of the road. This article used the GPS vehicle trajectory error correction method developed by Arman et al. (Arman and Tampere, 2021) ensures that the GPS trajectory offset errors are controlled below the standard lane width. Additionally, geographic information matching technology (Ding et al., 2022) aligns vehicle trajectories with rich spatiotemporal data. To the best of the author's knowledge, there is currently little research on using large-scale high-frequency sampled vehicle trajectory data to analyze highway safety. This article aims to enrich the application scenarios of such trajectory data in highway safety analysis. In this work, the concept of vehicle trajectory fractal is introduced for the first time, and corresponding computational methods and parameter estimation techniques are developed. Through in-depth analysis of the geometric features of vehicle trajectories at different resolutions, we found that vehicle trajectories exhibit significant fractal characteristics. The results of incorporating fractal features into the negative binomial regression model are statistically significant, with locations having higher FD showing higher crash frequency. This finding provides an innovative, practical, and

theoretical tool for road safety assessment and prediction. In this paper, FD is specifically used as a surrogate for estimating large-scale spatiotemporal risks on highways, and its performance is evaluated by comparing Higuchi fractal dimension (HFD) with five commonly used surrogates. HFD reduces the false positive rate from 94 % (other surrogates) to 46 %, with a recall rate of 95 %, and FD demonstrates strong risk classification ability and better robustness under different road geometric designs. This finding offers a new perspective and methodological support for road safety analysis.

The remainder of this paper is organized as follows. In Section 2, we conduct a literature review. In Section 3, we introduce the concept of vehicle trajectory fractals and the calculation methods for FD. Then, in Section 4, we analyze the fractal characteristics of vehicle trajectories using trajectory data. In Section 5, we point out and discuss the theoretical and practical significance of FD, along with our new findings. Finally, we conclude the study in Section 6 and present its potential benefits for road traffic safety.

2. Literature review

2.1. The driving behavior safety surrogates from large-scale telematics data

The concept of surrogate safety measures (SSM) is mainly based on vehicle conflict datasets, such as TTC and PET (Sarkar et al., 2024). However, calculating these indicators is not an easy task as it requires complex sensing and tracking for object recognition and tracking to extract all relevant trajectories of potential interaction scenarios (Wang et al., 2021). Therefore, collecting TTC data in an uncontrolled, highway-scale environment requires advanced hardware and software. Given these challenges, Alrassy (Alrassy et al., 2023) expanded the scope of SSM by using single vehicle behaviors, such as kinematic triggers, as safety surrogates, allowing for the screening of safety–critical events. The data range and data localization accuracy required for these surrogates have been significantly improved in the context of advances in mobile sensors (Zhang et al., 2021). Compared to vehicle trajectory data collected by other devices, mobile sensors offer lower costs, richer information, and a broader understanding of crash causes. The vehicle kinematic triggers derived from these data include the frequency of dangerous driving events, longitudinal and lateral accelerations, deceleration rates, jerk, and yaw rates, which have been incorporated into the study of "crash surrogate events," having a significant impact (Kamla et al., 2019). For instance, Ziakopoulos et al. (Ziakopoulos, 2024) conducted an empirical study based on GPS-supported mobile devices, showing that locations with more instances of harsh braking and aggressive acceleration tend to have more crashes. Furthermore, the usage methods of these kinematic trigger surrogates have been further expanded. Specifically, Kim (Kim et al., 2024) demonstrated a stronger correlation between the further development of these kinematic surrogates collected from onboard recorders and crashes by applying new concepts such as variable thresholds, unstable driving, and safe/reliable driving. These kinematic surrogates reduce reliance on crash data.

Most of the relevant studies in the literature consistently suggest that the variation in vehicle kinematics is positively correlated with highway crashes. However, this conclusion varies significantly across geometric designs (Nikolaou et al., 2023). For example, longitudinal acceleration/ deceleration and longitudinal jerk are more suitable for identifying crashes on simple straight road segments, where these events are described as related to driving fluctuations associated with crashes (Ziakopoulos, 2024). Li (Li et al., 2021) based on GPS jerk data (the time derivative of acceleration), developed a predictive model to identify hazardous highway locations. However, they note that large-scale studies are needed to test the reliability of jerk-based models. Therefore, large-scale data is more crucial for the argument of safety surrogates. In curve sections, some studies have incorporated operational speed into geometric design, considering the interaction between

vehicle kinematic triggers and road geometry. Gavran et al. (Gargoum and El-Basyouny, 2020) investigated the difference between available sight distance (ASD) and theoretical sight distance, emphasizing the inclusion of operating speed in sight distance and assessing horizontal curve design risks based on non-compliant sight distance patterns. Alsaleh (Alsaleh et al., 2024) also used reliability theory to analyze sections of road that do not meet vehicle skid resistance requirements. The results showed that CR were higher in these non-compliant areas. In specific road sections such as interchanges and work zones, vehicles passing through these areas may be disturbed by mandatory lane change operations (Feknssa et al., 2023). It is more appropriate to consider the lateral acceleration/deceleration of disturbed vehicles as a safety surrogate. Yan (Yan et al., 2024) demonstrated the relationship between lateral acceleration of vehicles in these areas and crashes, showing that when lateral acceleration exceeds 0.7 g, the potential risk of crashes increases by 24 times. However, several studies have tested using smartphone GPS data, indicating that the use of yaw rate in these areas is more significant in predicting CR than measures based on lateral acceleration. This is because yaw rate describes the driver's steering operations when deviating from the road centerline. In avoiding crashes, drivers typically steer first and then brake, suggesting that steering operations may be more critical in crashes avoidance (Naude et al., 2019). Overall, the robustness of these kinematic trigger surrogates under different road geometries is relatively low, likely due to an insufficient understanding of the interaction between road geometry design and driver behavior. Additionally, in complex road geometries, there are multiple complex patterns such as lane changes, yawing, and vehicle oscillation (Ding et al., 2022). Current vehicle kinematic alternatives only focus on a single dimension of vehicle motion, and the feature extraction of trajectories themselves has yet to be fully explored.

2.2. Fractal theory

Fractal theory focuses on the complexity of trajectory geometric features and the high-dimensional nature of analysis methods (Yilmaz and Unal, 2020). In the field of road safety and traffic, Chand (Chand and Dixit, 2018) demonstrated that the Hurst exponent, measured using a rescaling method, is significantly correlated with CR. However, the Hurst exponent is only used to describe the overall trend and fluctuations of one-dimensional data, making it unsuitable for capturing the complex behaviors of two-dimensional vehicle trajectories. The FD, derived from rescaling methods, provides a means to quantify the fractal characteristics of two-dimensional vehicle trajectories (Karimui, 2021). Fractal characteristics can indicate the scaling characteristics of fractals in different signals of complex systems, providing a powerful tool for understanding the underlying hierarchical structures and multiscale behaviors in complex systems (Gvozdarev and Parovik, 2023). Compared to deep learning methods, FD has been proven to be an effective feature-based approach. Feature-based recognition does not result in high computational complexity, making it more attractive for applications with limited resources, such as the Internet of Things (Li et al., 2024a). Previous studies have shown that combining FD with other features can enhance the accuracy of signal recognition. For example, Li (Li et al., 2024b) demonstrated that combining FD with other radiative characteristics improves the identification accuracy of wireless devices. Therefore, the most accurate and comprehensive results often require integration with other features. As a nonlinear feature, FD can also represent physical properties. For example, Zhang et al. (Zhang et al., 2023a) demonstrated that FD, as a parameter for characterizing discontinuous roughness, can be used to predict the specific physical properties of heterogeneous materials. These studies highlight the commonality of fractal methods: fractal techniques refine the granularity of data analysis and enhance the dimensionality of feature analysis, providing valuable insights for the safety analysis of vehicle trajectory data. Furthermore, the fluctuation range of vehicle trajectory shapes remains within the road boundaries, which indirectly

reflects the road's geometric alignment. Therefore, fractal methods take into account the interaction between driving behavior and road geometric design, making them a potentially more effective indicator for quantifying the road safety.

There are several methods for calculating FD, but the most widely accepted ones are Katz (Li et al., 2024a), box-counting (Sui et al., 2022), and Higuchi (Yilmaz and Unal, 2020). Compared to Katz and box-counting methods, the Higuchi method has an advantage in capturing trajectory complexity, integrating multiscale features, and nonlinear behaviors (Yilmaz and Unal, 2020). This method does not rely on pre-determined scales, offering high adaptability, and is especially suitable for estimating the FD of two-dimensional trajectory sequences. Additionally, the sampling frequency of the data is a critical aspect in calculating FD. Karimui (Karimui, 2021) compared trajectories obtained from the Weierstrass cosine function at sampling frequencies of 1000 Hz and 3000 Hz and found that trajectories generated from low-frequency sampling were unsuitable for validating the estimator. Another challenge of the Higuchi method is that several parameters must be applied to it, and improper parameter selection can lead to incorrect computation of fractal features (Bahramizadeh-Sajadi et al., 2022). These fractal parameters include scale and reduction size. Multiscale parameters are a multi-channel representation of the complexity of time series, considering the local structure and overall distribution features of fractal objects. Yilmaz et al. (Yilmaz and Unal, 2020), inspired by multiscale sample entropy, proposed a multiscale method capable of accurately identifying specific categories of random and chaotic time series. However, as the reduction size increases, this traditional multiscale process shortens the length of coarse-grained sequences, thus reducing stability. Therefore, Li et al. (Li et al., 2024c) applied a fine composite multiscale process to FD and proposed a hierarchical fine composite multidimensional FD, which improved the performance of SRN in feature extraction. Moreover, the reduction size parameter represents the sampling length of subsequences at different scales. If the maximum value is not correctly selected, the method will be compromised from the start (Wanliss and Wanliss, 2022). A common approach to determining the maximum value involves finding a position relative to the range where the computed values approach a local maximum or asymptote (Macek et al., 2023). In summary, the selection process requires continuous testing.

The evaluation of safety surrogates in highway includes three points (Pinnow et al., 2021): First, the ability to identify crash locations as indicated by historical data or observational results, where the data or observations are not recognized through the surrogates. Second, the safety surrogates must be robust enough to reliably identify high-risk areas under varying external conditions (e.g., geometric design, time, driver heterogeneity). Third, when identifying high-risk areas in a specific context, the surrogates should demonstrate superior performance compared to other indicators. This study aims to argue for the promising application of fractal methods in road safety analysis from these perspectives.

3. Methodology

3.1. The concept of fractal vehicle trajectory

In the road plane, trajectories can be represented as tuples composed of x and y projection coordinate sequences. We define the Euclidean distance d_i between adjacent points on a trajectory as:

$$T = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\} \tag{1a}$$

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{1b}$$

The FD is related to the scale of the system, so we introduce a scale parameter ϵ to describe the geometric properties of trajectories at different resolutions. We investigate how the geometric features of the

trajectory change when reducing ϵ. Hausdorff method (Cheng et al., 2024) gives the general form of FD calculation, that is, a set of geometric structures (such as line segments, grids, spheres, etc.) can be used to represent the covered trajectory. If $N(\epsilon)$ is the minimum number required to cover the trajectory with a geometric structure of scale ϵ, then the core idea of FD calculation is to observe the rate of change of $N(\epsilon)$ as ϵ shrinks. Therefore, vehicle trajectories exhibit fractal characteristics, equivalent to the existence of the limit of the following equation:

$$N(\epsilon) \propto \epsilon^{-FD} \tag{2a}$$

$$FD = \lim_{\epsilon \to 0} \frac{logN(\epsilon)}{log(1/\epsilon)} \tag{2b}$$

The self-similarity of fractal structures can be verified through scaling laws (Li et al., 2022). If the FD calculated through different scales ϵ is stable, then the trajectory has fractal characteristics; If there is a drastic change with the change of ϵ, the trajectory does not have or has weak fractal characteristics. Therefore, use the following equation to determine:

$$P(d) \ d^{-\alpha} \tag{3}$$

Here α is a scaling index. By calculating the distance distribution between trajectory points, if the distribution exhibits power-law characteristics, then the trajectory has fractal characteristics.

The calculation methods of fractals mainly include the ratio method (Katz), the coverage method (box-counting), and the trajectory length scaling law (Higuchi). The driving trajectory of vehicles passing through different road sections is influenced by road alignment, structures, and traffic flow status, which intuitively reflects the different FD of the vehicle in different road sections. A higher FD reflects more complex vehicle behavior, which can indirectly reveal the influence of spatio-temporal factors on traffic flow characteristics.

3.2. Quantitative methods for fractal characteristics

3.2.1 Katz method

Katz fractal dimension(KFD) is a method of evaluating the complexity of a path by comparing its total length with its diameter. Its algorithm is relatively simple, easy to implement, and suitable for coarse-grained analysis. The calculation process of KFD is as follows:

For the time series T, calculate the length of the sequence, which is defined as L. Calculate the direction vectors of the trajectory as a whole and the road according to Equation (4), as shown in Fig. 1. Then, the maximum distance d between the first point and the i-th point is:

$$\overrightarrow{Trajectory_{dr}} = (x_N - x_1, y_N - y_1) \tag{4a}$$

$$\overrightarrow{Road_{dr}} = (X_{road}, Y_{road}) \tag{4b}$$

$$d = \frac{\overrightarrow{Trajectory_{dr}}}{\|\overrightarrow{Trajectory_{dr}}\|} \overrightarrow{Road_{dr}} \tag{4c}$$

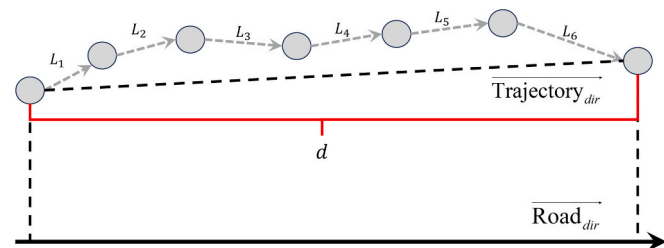**Fig. 1.** *KFD* calculation method.

KFD can be calculated according to the following equation:

$$KFD_{Trajectory} = \frac{log(N)}{log(N) + log(\frac{d}{L})} \tag{5}$$

3.2.2. Box-counting method

The box-counting method estimates the FD by covering fractal structures in boxes of different sizes and calculating the required number of boxes. Estimate the FD of road spatial trajectory through pre-defined box sizes using image methods. A grid with uniform intervals of δ covers the designated road space. As shown in Fig. 2, for different values of δ, determine the number of boxes ($N(\delta)$) related to fracture, and then calculate Box fractal dimension (BFD) using the Equation (6):

$$N(\delta) \ \delta^{-FD} \tag{6a}$$

$$BFD = -\lim_{\delta \to 0}(\frac{lnN(\delta)}{ln\delta}) \tag{6b}$$

The application of logarithms to the above equation reveals a linear correlation between the total number of boxes required to fully cover the road trajectory and the box size (unit of measurement).

The parameters of the Box algorithm must determine the minimum and maximum box sizes, as well as the method of increasing box sizes. We select the length of the analyzed road unit and the interval between the boundary lines of the selected road horizontally as the maximum box size, and then set the degree of box increase proportionally.

3.2.3. Higuchi method

The Higuchi method first calculates the average length (L_k) of the sub trajectories, which is obtained by reducing the main trajectory in a specific set of scales ($k = k_1, ..., k_2$), equivalent to the average of the actions that occur in the sub trajectories extracted for a specific scale (k).

Set the time series $X = \{x_i, i = 1, 2, ..., N\}$ and $Y = \{y_i, i = 1, 2, ..., N\}$ with a length of N, reconstruct the time series using the delay method, and obtain the matrix $T_{m,k}$ in the form of:

$$T_{m,k} = \begin{pmatrix} X(m) & Y(m) \\ X(m+k) & Y(m+k) \\ X(m+2k) & Y(m+2k) \\ ... & ... \\ X(m+int(\frac{N-m}{k})k) & Y(m+int(\frac{N-m}{k})k) \end{pmatrix}, \ m = 1, 2, ..., k \tag{7}$$

Where k is the delay time and m is the index of the starting point of each subsequence.

Then calculate the curve length $L_m(k)$ for each X_k^m, which can be obtained through Equation (8):

$$L_k(T_{m,k}) = \frac{1}{k} \sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} \|T(m+ik) - T(m+(i-1)k)\| \frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k} \tag{8}$$

Among them, $\|T(m+ik) - T(m+(i-1)k)\|$ represents the Euclidean distance between adjacent trajectory points, $\lfloor * \rfloor$ represents rounding downwards, and $\frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k}$ is a normalization factor used to adjust the curve length so that it corresponds to the overall length of the time series.

Generate the average value $L(k)$ of the curve length $L_k(T_{m,k})$ for all sequences $T_{m,k}$:

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_k(T_{m,k}) \tag{9}$$

Finally, obtain a set of data on $L(k)$ with different values of k; Draw the curve $lb(L(k)) \ lb(k)$. By fitting the line $lb(L(k)) = HFD \times lb(k) + C$,
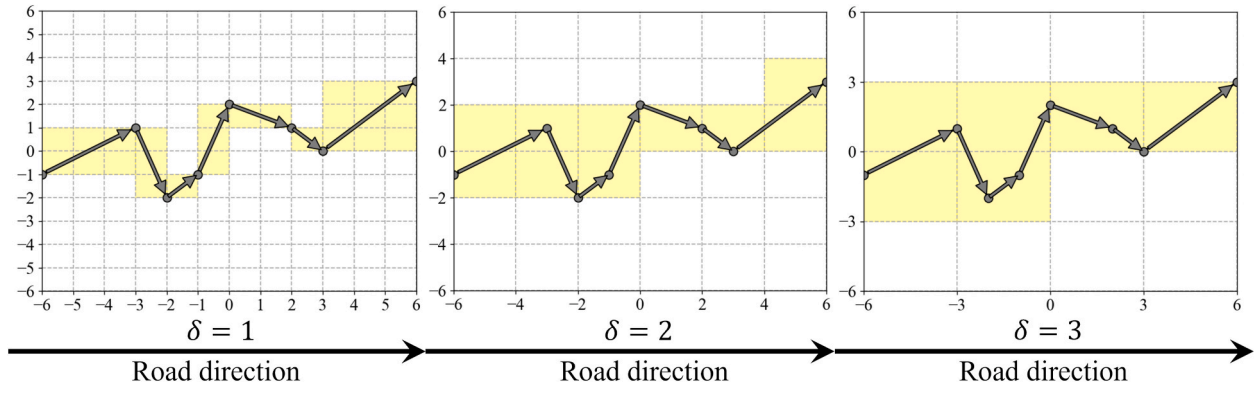
Fig. 2. *BFD* calculation method.

obtain the *HFD* values of the time series; The value of *HFD* is estimated by the following equation:

$$HFD = - \lim_{k \to \infty} \frac{log(L(k))}{log(k)} \tag{10}$$

Therefore, *HFD* is an effective tool for analyzing the impact of different driving maneuvers on the complexity of driving behavior. But we need to analyze its parameter changes and develop specific methods to control the parameters, which is a prerequisite for applying it to fractal analysis of road trajectories.

3.3. Parameter estimation

The maximum value for reducing the size k (k_2) is an important parameter, as selecting an inappropriate value can significantly increase errors. This empirical value comes from the relationship graph between *HFD* and the range of k_2 drawn, and then selects the appropriate k_2 at the position where the calculated *HFD* approaches the local maximum or asymptote. In addition, k_1 refers to the minimum value of the sample selected for fitting length changes at different scales. If k_1 exceeds half of the window length, the observation is insufficient, causing the length algorithm of the sequence to be interrupted prematurely, resulting in underfitting of the line. Therefore, k_1 is not greater than twice the value of k_2. Some studies have also shown that if the minimum value k_1 of the window length is too small, the linear variation of the fitted $log(L_k)$ value with the proportion of k weakens, which can lead to oscillation and instability of the FD value. The estimation of *HFD* is considered as a function of k_2 and k_1, as shown in the following equation:

$$HFD(k_1, k_2) = \lim_{k_1 \leq k \leq k_2} \frac{log(L(k))}{log(k)} \tag{11}$$

Scale (*s*) is an important parameter in fractal analysis. If *s* is too small, the amplitude of the vehicle trajectory swinging in adjacent time intervals will increase, resulting in unstable trajectory signals, which will increase the interference of noise on fractal calculations; In addition, if *s* is too large, it will also overwrite the changing properties of the trajectory itself, which means that objects created at a large scale may not be fractal objects, making it impossible to reflect the complexity of the trajectory. Therefore, we need to set different scales for the trajectory, and the algorithm is as follows:

Step 1: Divide the time series $T = (X, Y) = \{t_1, t_2, ..., t_N\}$ into new subsequences by setting different scales:

$$t_j^{(s)} = \frac{1}{s} \sum_{(j-1)s+1}^{js} t_i, 1 \leq j \leq \frac{N}{s} \tag{12}$$

It is worth noting that when $s = 1$, the subsequence is the same as the original sequence.

Step 2: Input each subsequence into Section 3.2.3 to calculate the *HFD* value at scale *s*:

$$MHFD(T, s) = \frac{1}{s} \sum_{d=1}^{s} HFD(T^{(s)}) \tag{13}$$

And $T^{(s)} = \{t_1^{(s)}, t_2^{(s)}, ..., t_{\frac{N}{s}}^{(s)}\}$ is the sequence obtained by executing Equation (12) at $T^{(s)}$. To better illustrate the multi-scale processing technique, we take Fig. 3 as an example. Since Scale 1 is the original time series, at Scale *s*, *s* points of continuous data are treated as one point to reconstruct the trajectory. Due to this downsampling, the data length at each scale has been shortened.

In order to determine the optimal parameters for road units of different lengths, we established a multi-objective evaluation function and considered the spatial autocorrelation, discriminability and reliability of the trajectory FD of different road units for automatic parameter selection.

According to the literature (Gedamu et al., 2024), considering the spatial autocorrelation of crash frequency, this paper optimizes fractal parameters by local Moran index to enhance the spatial autocorrelation of FD, so as to better simulate the distribution law of crashes. Therefore, the objective function J_1 is maximized.

$$J_1 = \frac{N}{W} \cdot \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(FD_j - \overline{FD})(FD_i - \overline{FD})}{S_{FD}} \tag{14a}$$

Where J_1 is the total autocorrelation measure. N is the total number of road spaces. w_{ij} is the spatial weight of the *i*-th and *j*-th spaces. W is the sum of all weights w_{ij}, and the weights use Gaussian kernel function. FD_i is the average value of FD in the *i*-th space. $\overline{FD}$ is FD of all spaces. S_{FD} is the standard deviation of FD mean.

In order to better reflect similar road geometry designs or complex weather conditions, we refer to the idea of unsupervised learning (Chen et al., 2022a); The goal is to have high similarity between different trajectory FD within the same road section. In addition, in order to effectively distinguish the trajectory characteristics of different road sections, we set the FD of different road units to have significant differences as much as possible. Therefore, we use J_2 and J_3 to represent these two objectives.

$$J_2 = |\frac{1}{N_{risk}} \sum_{i=1}^{N_{risk}} \overline{FD_i^{risk}} - \frac{1}{N_{normal}} \sum_{i=1}^{N_{normal}} \overline{FD_i^{nomarl}}| \tag{14b}$$

$$J_3 = \sum_{i=1}^{K} |\frac{1}{N_i^{risk}} \sum_{j=1}^{N_i^{risk}} fd_{ij}^{risk} - \frac{1}{N - N_i^{risk}} \sum_{j=1}^{N-N_i^{risk}} fd_{ij}^{normal}| \tag{14c}$$

Among them, *fd* is a function of k_1, k_2, s, and N_{risk} is through crash prone road sections; And N_{risk} is the risk trajectory of the *i* section passing through the velocity marker.

In addition, the selection of the *s* parameter must satisfy the stationarity condition of FD, which provides a convergence criterion for evaluating saturated FD in the sense of least squares fitting error (Gao et al., 2023):
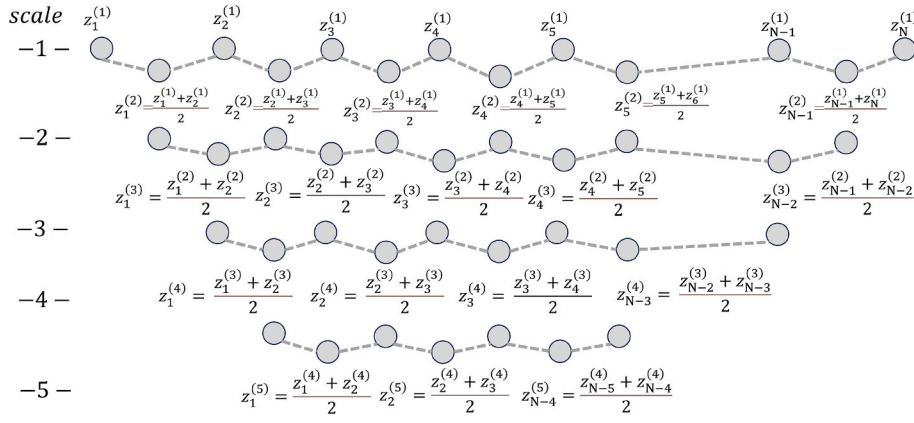
Fig. 3. Multi scale transformation process of trajectory.

$$|fd_{ij}(s,k_1,k_2) - fd_{ij}(s+1,k_1,k_2)| < e \tag{14d}$$

And the trajectory should have significant fractal characteristics, as shown in the following equation:

$$R(s,k_1,k_2) > 0.9 \tag{14e}$$

R is the degree of fit between $log(L_k)$ and k relationship under parameter selection. Therefore, we maximize the objective function J_4:

$$J_4 = \sum_{i=1}^{M}\sum_{j=1}^{n_i} R_{ij}(s,k_1,k_2) \tag{14f}$$

In addition, k_1, k_2, and s need to satisfy the following relationship in the vehicle trajectory data:

$$1 \le k_1 \le 0.5k_2 \tag{14g}$$

$$3 \le k_2 \le \frac{(11-s)L}{2\bar{v}}$$

$$1 \le s \le 12 \tag{14i}$$

Among them, k_2 cannot exceed half of the number of trajectory points, and the limitation of s is that the sampling frequency range of

trajectory data is 0.25HZ-10HZ. In summary, we establish the final multi-objective function as shown in Fig. 4.

The four items in the figure represent the spatial autocorrelation, within group variance, between group variance, and significance of fractal characteristics, where β_1, β_2, β_3, and β_4 are the weight co-efficients of each objective in the multi-objective function, with values ranging from 0 to 1 and a total of 1. Under different data sampling frequencies and analysis unit lengths, we can adjust the four weight coefficients to find the optimal FD.

3.4. Evaluation metrics

Based on the large-scale characteristics of the trajectory data of the floating vehicle, the CR (Alsaleh et al., 2024) is considered as the verification measure of the safety analysis, and is calculated according to Equation (15):

$$CR_i = \frac{\sum_{Year=1}^{5} CF_{Year,i}}{[\sum_{Year=1}^{5} AADT_{Year,i} \times L_i \times 365]/10^8} \tag{15}$$

Where $CF_{Year,i}$ is the number of crashes, L_i is the length of the curve (km), and $AADT_{Year,i}$ is the annual average daily traffic volume.
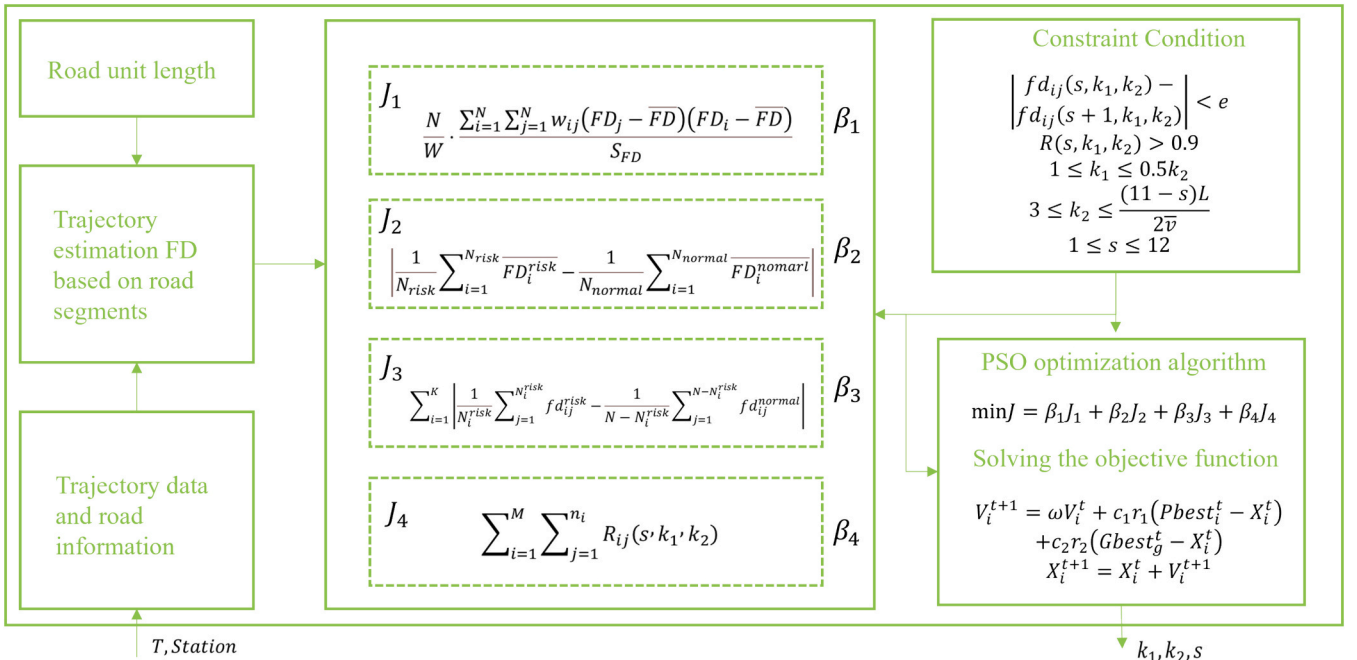

Fig. 4. Multi objective function optimization of fractal parameters.

In order to quantitatively analyze the prediction results, the road sections are divided into high-risk and low-risk groups according to the historical crash data, and the output of the model is analyzed by checking the recognition accuracy of high/low-risk road sections. Referring to another document (Liu et al., 2020), the average value of historical CR is positioned as the threshold of high risk and low risk. If the crash risk value of the road section is less than the threshold value, the road section is "the road section with relatively low crash risk" (Note: low risk); Otherwise, it is "road section with relatively high crash risk" (Note: high risk).

The confusion matrix is used to evaluate the performance of the current surrogates and the proposed metrics. The confusion matrix consists of four basic elements: true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). *TP* is the number of crash samples correctly predicted, *FN* is the number of crash samples incorrectly predicted as normal crash samples. *FP* is the number of normal samples that are incorrectly predicted as crash samples. *TN* is the normal number of samples correctly predicted. Based on the confusion matrix, three widely accepted indicators are calculated: recall rate (*Recall*), false alarm rate (*FAR*) and area under the curve (AUC). Recall refers to the proportion of correctly detected crashes in the total real crash samples, which is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{16a}$$

FAR measures the proportion of low-risk road sections that are incorrectly detected as high-risk road sections in all low-risk road sections, which is defined as:

$$FAR = \frac{FP}{TP + TN} \tag{16b}$$

The receiver operating characteristic (ROC) curve more intuitively evaluates the performance of the indicator (Wang et al., 2024c). The curve is constructed by plotting the relationship between Recall and false alarm rate (FAR) under different thresholds. The area under the ROC curve (AUC) is used as a quantitative index to compare the overall performance of different indicators. The maximum AUC value is 1, and the index with larger AUC value is considered to be a better method to predict the risk section.

4. Experimental result

4.1. Data collection and processing

The trajectory data we use comes from a management company of a highway in China. The company operates a fleet of trucks (724 vehicles), each equipped with GPS and forward-facing cameras. These trucks are deployed to the company's jurisdiction to carry out freight tasks. The recorded data is continuously uploaded to the company's data management platform, as shown in Fig. 5(a). Fig. 5(b) illustrates that the GPS vehicle trajectories spatially cover over 160 km of the highway mainline and 38 interchanges. These sections include crash event data from the past five years, totaling 4,645 incidents. This data has been manually input into the system by traffic police, road management departments, or company personnel. The crash event data for most sections has a location accuracy of 1 km. To investigate the accurate relationship between FD and road safety, we divided the entire section into 1-kilometer sub-sections, resulting in a total of 314 sub-sections. Table 1 presents basic information about the trajectory data and other related datasets.

The time span of the vehicle trajectory data covers from June 1, 2023, to September 1, 2023. The number of vehicle trajectories collected for each road segment ranges from 304 to 756. These trajectories are continuous and evenly distributed within the collection period, with a penetration rate ranging from 2 % to 5 %. Furthermore, based on the traffic volume data recorded hourly by traffic measurement stations or gantry systems, the road section approximates a free-flow traffic state, and the trajectories of merging and diverting vehicles are not included in the analysis. To further enhance the spatial representativeness of the trajectory data, we excluded vehicle trajectories influenced by surrounding vehicles, crash events, and work zones, using data from the vehicles' onboard cameras and event records.

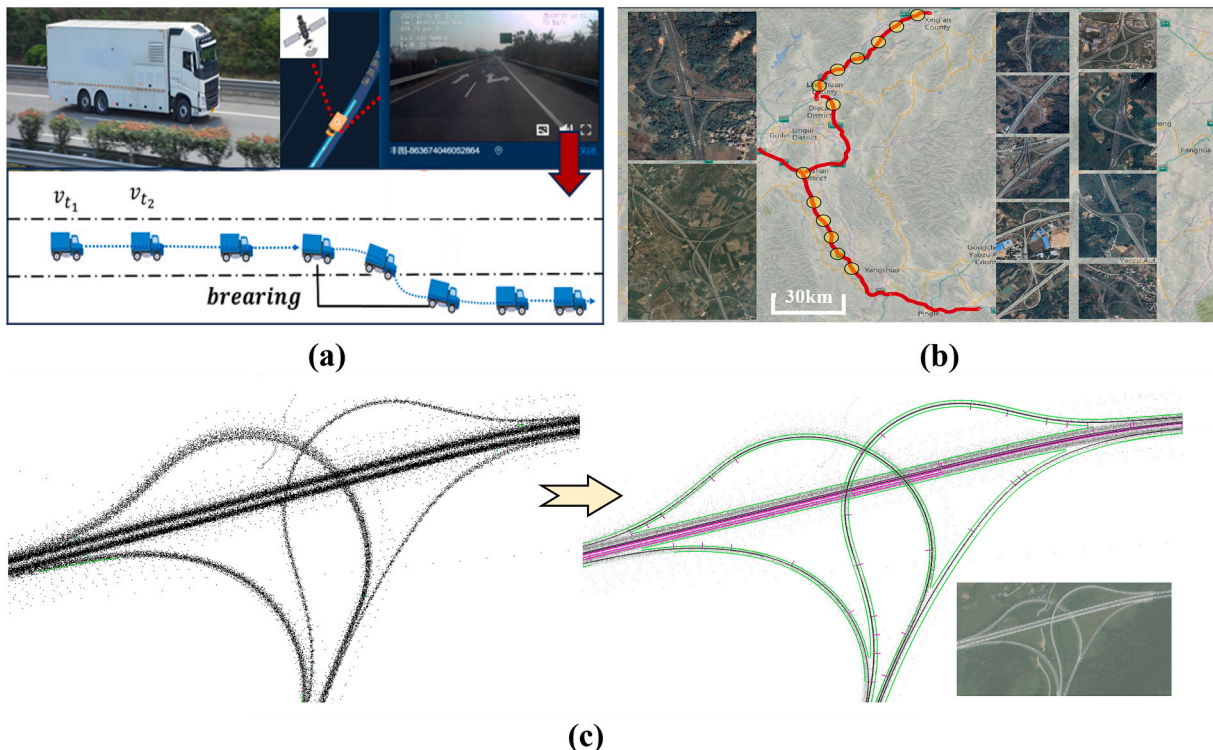Our data includes basic information such as latitude and longitude,

(a)

(b)

(c)

Fig. 5. Large-scale telematics trajectory data: (a) Trajectory data collection; (b) Large scale space–time; (c) Road line fitting and match map.

Table 1
Basic information statistics of data.

Elements	Max	Min	Mean	Sum
Frequency of crash occurring on road sections within 5 years	89	0	11	4645
Length of road segments (*km*)	1	1	1	160
The number of road segments	–	–	–	314
The trajectories time span of each road segment (day)	90	90	90	90
Sample size of trajectory for road sections	756	304	546	–
Speed limit value (*km/h*)	120	40	–	–
Number of lanes	2	2	2	–
The number of tunnels	1	0	0.03	3
The number of interchanges	1	0	0.19	38
The number of service area	1	0	0.08	15
1 (Contains reverse curve),0 (Excluding reverse curve)	1	0	0.12	24
Curvature radius (*m*)	8515	750	4695	–
Number of curves	1	0	0.58	158

vehicle ID, vehicle speed, and heading angle. As shown in Fig. 5(c), we use the "Pyautocad" library in Python to display trajectory points on a CAD map, and then perform coordinate projection transformation to align with the Ovi map. Our road sections contain rich geometric information, including diverse horizontal and vertical elements as well as highway nodes, which can be matched to vehicle trajectories through coordinates. Compared to the HighD (Zhang et al., 2023b) and NGSIM (Chauhan et al., 2022) datasets, our trajectory data stands out for its large-scale coverage, full-time domain, ease of access, and abundant information.

The trajectory data has a sampling frequency of 1 Hz, with positioning accuracy within a range of 5 m. Anomalies in the GPS data include large temporal delay errors, GPS positioning errors, and signal loss. We applied a local anomaly factor method to eliminate inherent GPS anomalies (Arman and Tampere, 2022). Additionally, for issues related to system lateral offset caused by GPS signal in certain road sections, we used the method proposed in the literature to address this problem (Arman and Tampere, 2021). Specifically, as shown in Fig. 5(c), we determined the road centerline based on a dissimilarity matrix (Fréchet distance between trajectories) and employed a two-dimensional clustering method based on Gaussian Mixture Models to identify road lanes based on actual trajectories. After applying these strategies, we were able to reduce the lateral GPS error to below the width of standard lanes in the highway network.

4.2. Trajectory fractal characteristic extraction

4.2.1. Verification of fractal characteristics

Due to the need to divide the trajectory into subsequences at different scales for fractal computation, if the length of the road unit is too small, the number of sampled points in the trajectory sequence with a low sampling frequency may not meet the minimum required for calculating the FD. To address this issue while preserving data authenticity, we use linear interpolation to fill data. The trajectory is then transformed into the Frenet coordinate system, as shown in Fig. 6(b), to eliminate the influence of fixed road curvature changes on the vehicle trajectory shape (Wang et al., 2024d).

Fig. 6 illustrates the process of calculating the FD of vehicle trajectories. We computed the FD of 50 vehicle trajectories over a 1 km road segment and selected one trajectory for analysis using both the box-counting and Higuchi methods. As shown in Fig. 6(a), the relationship between $Log(\delta)$ and $Log(N(\delta))$ was fitted to analyze the variation of the geometric characteristics of vehicle trajectories at different resolutions. The slope of the fitted line indicates that the p-value is less than 0.05, suggesting that the trajectory exhibits significant fractal characteristics. Similarly, we verified that all the trajectories from the road segments demonstrated significant fractal characteristics.

4.2.2 Fractal parameter estimation

Parameter estimation is an important component of FD calculation, and we adopt the fractal parameter estimation method in Section 3.3. We used a road unit length of 1 km as an example for parameter estimation. The main reason for choosing k_1 is to prevent a too small k_1 from weakening the linear variation of the fitted logarithm (L_k) value with the proportion of k. But in our parameter estimation, k_1 is not the main parameter, it varies between 1–3 with the objective function (The best k_1 of the data in this paper is set to 2). The parameter s usually represents the sampling interval (scale of the data), and the smaller the parameter s, the larger the sampling interval. Specifically, we selected a scale of 1 to 11 for analysis after data interpolation, which represents the frequency range of 10 Hz to 0.25 Hz. Therefore, Table 2 lists the effects of changes in the main parameters s and k_2 on the objective function, where we selected harsh braking and harsh acceleration behaviors as the risk trajectory criteria for J_2 and J_3 (Xian et al., 2023). In addition to the objective function, we also use Wasserstein distance as an evaluation metric to test the similarity between vehicle trajectories within the same road segment and the discrimination between different road segments in
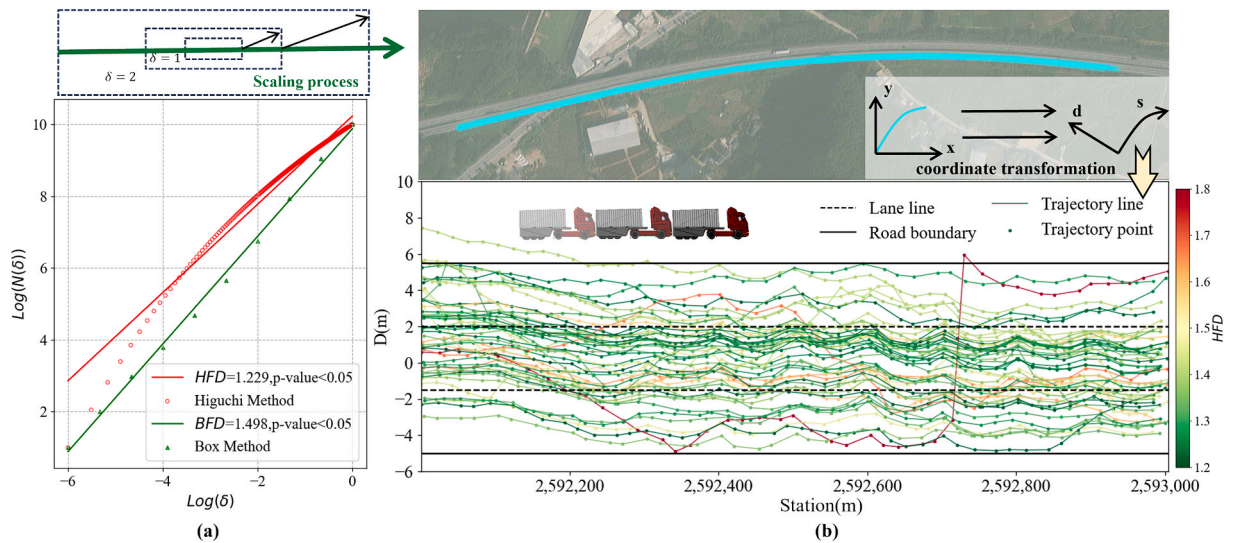

Fig. 6. Example of measuring FD of vehicle trajectory: (a) Fractal scaling process; (b) Vehicle trajectory FD of 1 km road length.

Table 2
Evaluation of the influence of fractal parameters on fractal estimation.

	FD distribution		Objective function			
s	within-subjects	between groups	J_1	J_2	J_3	J_4
4	0.01	0.86	0.03	0.37	0.18	0.08
6	0.02	0.46	0.04	0.40	0.13	0.04
8	**0.17**	**0.18**	**0.06**	**0.39**	**0.05**	**0.27**
10	0.49	0.06	0.11	0.22	0.01	0.71
12	0.25	0.01	0.08	0.18	0.03	1.27
k_2	within-subjects	between groups	J_1	J_2	J_3	J_4
3	0.10	2.46	0.11	0.43	0.48	1.13
6	0.14	2.15	0.12	0.41	0.28	1.26
9	**0.21**	**1.83**	**0.15**	**0.40**	**0.19**	**1.32**
12	0.31	1.35	0.17	0.37	0.06	1.36
15	0.41	1.57	0.20	0.33	0.09	1.55

FD. The smaller the value of the Wasserstein distance indicator, the greater the similarity between the two distributions. However, *HFD* is also difficult to choose between these two relationships. We tested several different values of s and k_2 in the Higuchi model. In order to minimize the overall objective function, s and k_2 are set to intermediate values of 8 and 9, respectively.

To validate our proposed parameter estimation approach, we used a negative binomial regression (NBR) model (Dzinyela et al., 2024) to predict crash frequency. In this example, the *HFD* mean and standard deviation in the road section are used as explanatory variables in the NBR model, with crash frequency as the dependent variable. The calculation of the *HFD* mean and standard deviation in the road section has undergone Monte Carlo error testing (Zhu et al., 2022), which means that as the sample size increases, their deviations will decrease, and the sampling volume will ultimately reduce the final deviation to below 3 %. Table 3 displays the output of the NBR model. The p-values indicate that the model's parameters are statistically significant. Moreover, the *HFD* mean has a stronger effect on crashes than the *HFD* standard deviation in the road section, with the *HFD* standard deviation being significant only at the 0.1 level. We use this model to examine the impact of fractal parameters on the model.

As shown in Fig. 7, smaller s will reach a minimum at smaller k_2, for example, the mean square error (MSE) value of curves such as s set to 4 decreases as k_2 increases from 0 to 6; However, the MSE of the model with s set to 12 continues to decrease when k_2 is set to 0–11. These phenomena are due to the fact that the smaller value of s increases the local details of the trajectory, and a smaller k_2 can improve the model's ability to capture fractal characteristics; The local morphology of the trajectory observed under larger s is coarse-grained, which requires a larger k_2 to improve the fitting degree. Overall, the *HFD* measurement model with k_2 of 9 and s of 8 has the best performance, and the parameter estimation approach has been well validated. However, further increase in k_2 after the minimum may reduce prediction accuracy and even cause fluctuations in MSE, indicating that larger k_2 is not a reliable choice for estimation. In most cases, k_2 is suboptimal between 6–10. Moreover, the choice of s is even more important. The model when s set to 12 seems unstable because the MSE value is very high, and even choosing the best k_2 will not perform well. Therefore, the s parameter is very important, which can be adjusted by adjusting the number of interpolation points or using a higher frequency data acquisition device

Table 3
The results of the negative binomial regression model.

variables	Coef.	St. Err.	t-value	p-value	Sig
The *HFD* mean in the road section	10.52	1.69	3.54	0.002	***
The *HFD* standard deviation in the road section	2.05	0.47	1.83	0.011	*
Constant	1.24	0.20	45.61	0.000	***

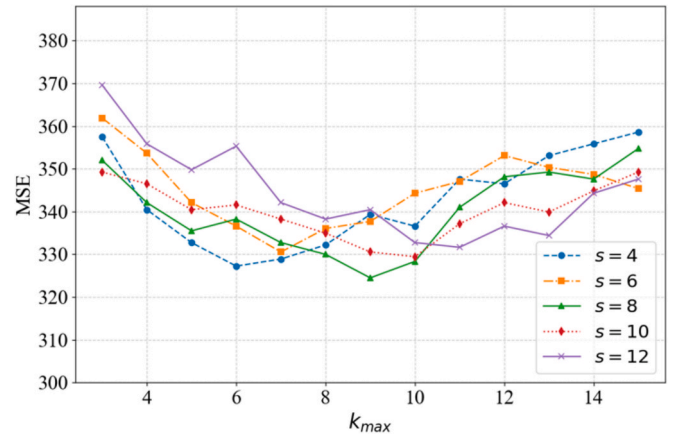***p < 0.01, **p < 0.05,*p < 0.1.

Fig. 7. Sensitivity analysis of fractal parameters for crash prediction.

to adjust s.

4.2.3. Comparison of fractal methods

We conducted an experiment to investigate the relationship between different types of FD and vehicle trajectory features. We selected representative features from the velocity category of the trajectory, including the standard deviation of speed (SD_V), the standard deviation of acceleration (SD_{ACC}), the 85th percentile speed (V_{85}) (Kim et al., 2024), as well as vehicle position transformation features, such as yaw rate (*YAW*) (Naude et al., 2019). These velocity indicators are used as longitudinal motion characteristics, while *YAW* is used as lateral motion characteristics.

As shown in Fig. 8, a comparison of the trajectory feature indicators mean in the road section and the correlation matrix of individual vehicle trajectory features reveals the relationship between fractal trajectory features and kinematic features. In Fig. 8(a), the *HFD* mean in the road section shows a positive correlation of 0.65 with SD_V, but in the results of Fig. 8(b), the correlation between individual vehicle trajectory *HFD* and SD_V is only 0.24. This highlights the distinguishing power of *HFD* from kinematic triggers. In the results shown in Fig. 8(b), *BFD* and *KFD* are most strongly correlated with V_{85} (−0.74) and SD_{ACC} (0.38), respectively. This sugg1ests that *KFD* reflects the amplitude of acceleration fluctuations in the trajectory and is consistent with the complexity of the trajectory path described by the method. *BFD* may be closely related to road congestion indicators, such as vehicle travel time, which is similar to the findings of Hurst Exponent (Chand and Dixit, 2018). Among the three types of FD, the *HFD* mean in the road section has the highest correlation with CR, reaching 0.26, while the correlation of other fractal features with CR is low. Therefore, this paper focuses on using *HFD* as a tool for safety analysis.

4.3. The results of fractal characteristics in identifying high-risk road sections

4.3.1 Visual analysis and variance test of fractal characteristics

This section conducts an empirical analysis of the relationship between the trajectory features mean in the road section and high-risk road segments. Road segment 2 (C2), identified as a high-risk segment, is compared visually with six normal low-risk segments (C1-C6). As shown in Fig. 9(a) and Fig. 9(b), low-risk segments exhibit similar driving characteristics to C2 in terms of SD_V and *YAW*. Fig. 9(c) reveals that the *HFD* of C2 significantly exceeds the values of the low-risk segments. This finding suggests that fractal features may serve as a potential tool for identifying high CR segments.

The *HFD* mean in high-risk and low-risk sections is shown in Fig. 10 (a). Apply analysis of variance to determine the correlation between the
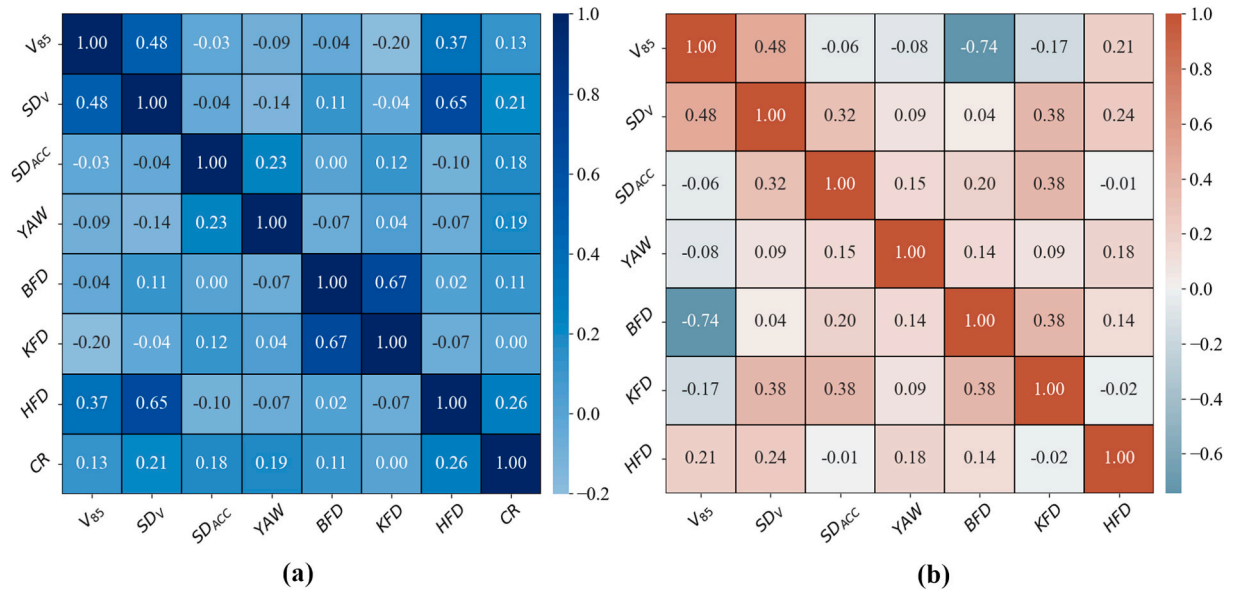
Fig. 8. Comparison of Correlation between Trajectory Features: (a) The trajectory features mean in the road section; (b) Single trajectory features.
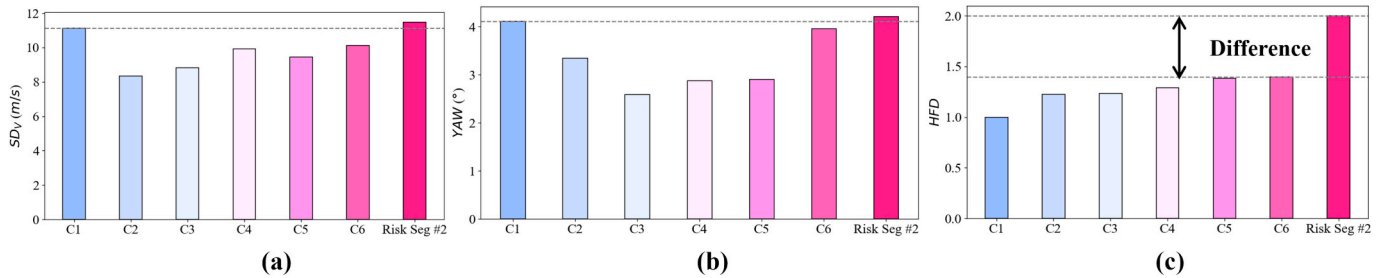

Fig. 9. Comparison results between section 2 (C2) and low-risk study section (C1-C26): (a) SD_V; (b) *YAW*; (c) *HFD*.
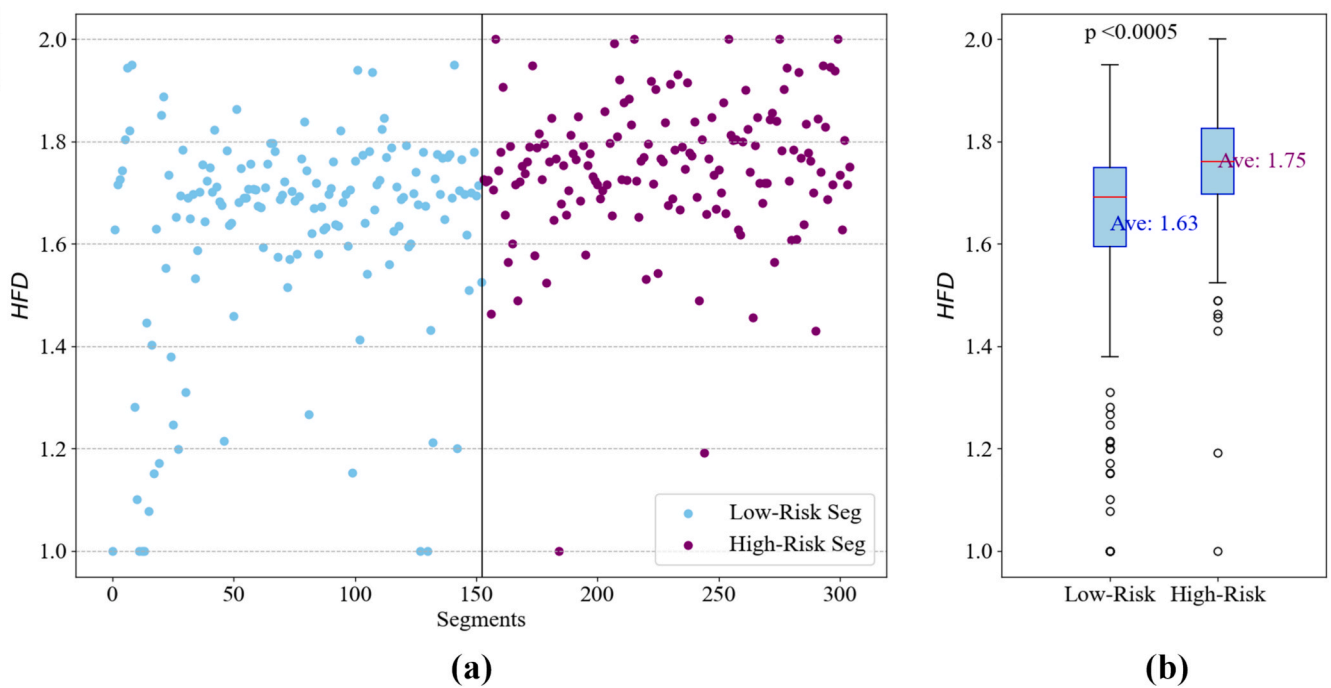

Fig. 10. (a) A graphical representation of the of *HFD* mean for all road segments; (b) the results of the analysis of variance test.

of *HFD* mean in the road section and the risk of highway crash. The results of Fig. 10(b) indicate that the *HFD* mean (1.75) in high-risk road segments is significantly higher than low-risk segments, indicating that the higher the *HFD* of vehicle trajectories, the higher frequency of crashes on the road sections.

In addition, six low-risk road sections (i.e. C7-C12) were selected, but the difference is that the *HFD* of these sections are higher than C2. Except for C12, its SD_V and *YAW* values are similar to those of C2. Fig. 11 (c) shows that in these low-risk road sections, the standard deviation of *HFD* is lower than that of C2. The standard deviation of *HFD* reflects the diversity of vehicle driving behavior in the same section, and higher values may indicate significant differences in driver behavior.

To confirm this statement, a variance analysis test was conducted, and the results are shown in Fig. 12. The research results indicate that the standard deviation of *HFD* in high-risk road sections (1.31) is significantly higher than that in low-risk road sections (1.23). This result means that when the vehicle trajectory *HFD* is more variable, the frequency of crashes on the road section.

4.3.2. Comparison of macroscopic safety surrogates

To explain the differences in identifying high-risk road sections among different surrogates, we selected two types of macroscopic surrogates as comparison baselines: surrogates related to road geometry design, vehicle kinematic trigger surrogates. The surrogates related to road geometry design has chosen sight distance conformity degree (CD_{SD}) and anti-skid requirement conformity degree (CD_{SK}) (Alsaleh et al., 2024). CD_{SD} is the percentage of the ASD of the vehicle that can meet the parking stop sight distance requirements (Chen et al., 2022b); The calculation of CD_{SD} takes into account the operating speed, road radius, and longitudinal slope. The calculation of CD_{SK} takes into account road radius, superelevation, and vehicle operating speed to reflect vehicle rollover, stability, and driving comfort. The relationship between these patterns and safety outcomes (i.e. crashes) is significant.

There are many kinematic trigger surrogates. Regarding them, Kim

(Kim et al., 2024) listed a table that includes velocity characteristics, position offset feature and dangerous driving events. We generate SHAP values based on the XGBoost model (Wu et al., 2024) to select the top ranked features. Specifically, negative jerk (*LPJ*) is selected as the speed feature, harsh braking event (*HB*) (Ziakopoulos, 2024) is selected as the frequency of dangerous driving events, and yaw rate (*YAW*) (Alruwaili and Xie, 2024) is selected as position offset feature. The jerk of trajectory speed reflects the variability of vehicle driving, and this information can be successfully applied to crash risk analysis without relying on threshold determination and road geometry features. *HB* is classified as a decrease in speed, with the threshold for *HB* varying from 0.2 g (1.96) to 0.86 g (8.43 m/s2), where 0.51 g and 0.66 g correspond to medium risk and high-risk road segments (Kamla et al., 2019); There will be more frequent dangerous driving behaviors on high-risk road sections. According to the correlation test of risk road sections, *HB* selected 0.55 g as the optimal thresholds for classifying hazardous events. *YAW* is related to the steering wheel angle and describes the driver's lateral positional deviation on the road to avoid crashes.

Fig. 13 shows that *HFD* exhibits significant stability in high-risk road sections, with limited fluctuations in its values. If 1.6 is chosen as the threshold, only 8 % of the scatter points below 1.6 are present. In contrast, other surrogates on high-risk road sections exhibit significant volatility. This finding suggests that the selected surrogates may be influenced by other factors that have not been considered, and fractal methods can effectively explain crash risk by enhancing the trajectory analysis dimension to comprehensively reflect the trajectory driving characteristics, thereby obtaining reliable results. The threshold may affect the performance of surrogates. Relaxing the threshold may lead to a higher Recall for high-risk road sections, although it may also increase the FAR for low-risk road sections. Due to the spatial limitations of previous trajectory data, there is currently no reference for evaluating thresholds. Therefore, we have set all surrogates thresholds within the range of minimum to maximum values.
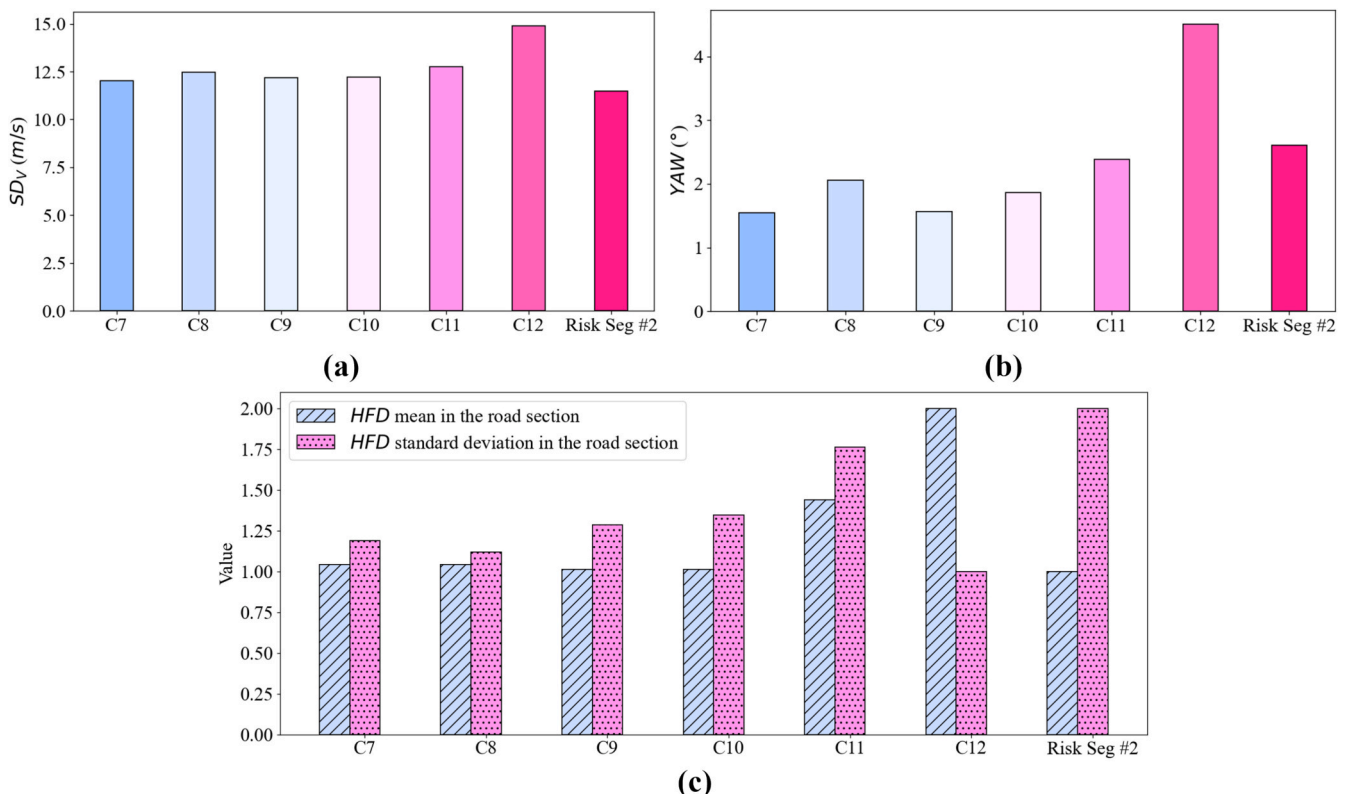

Fig. 11. Comparison results between C2 and low-risk study section (C7-C12): (a) SD_V; (b) *YAW*; (c) *HFD*.
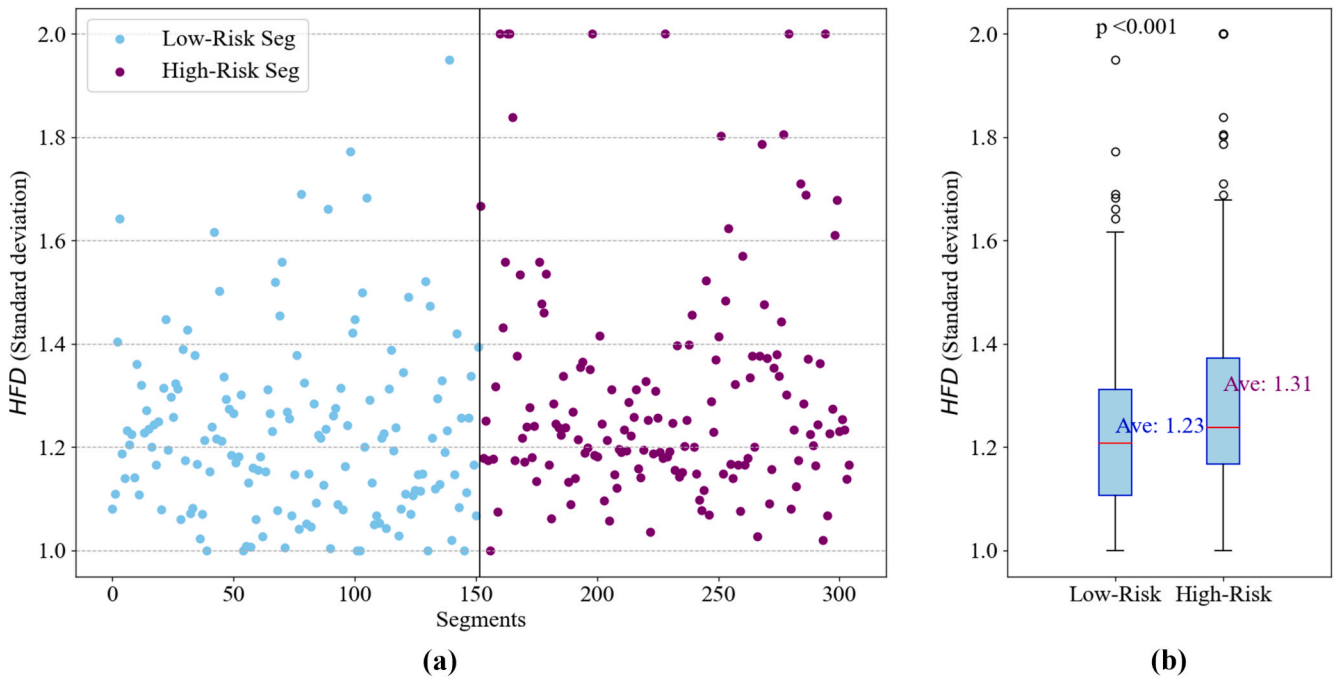
Fig. 12. (a) Diagram of *HFD* standard deviation for all road sections; (b) the results of the analysis of variance test.
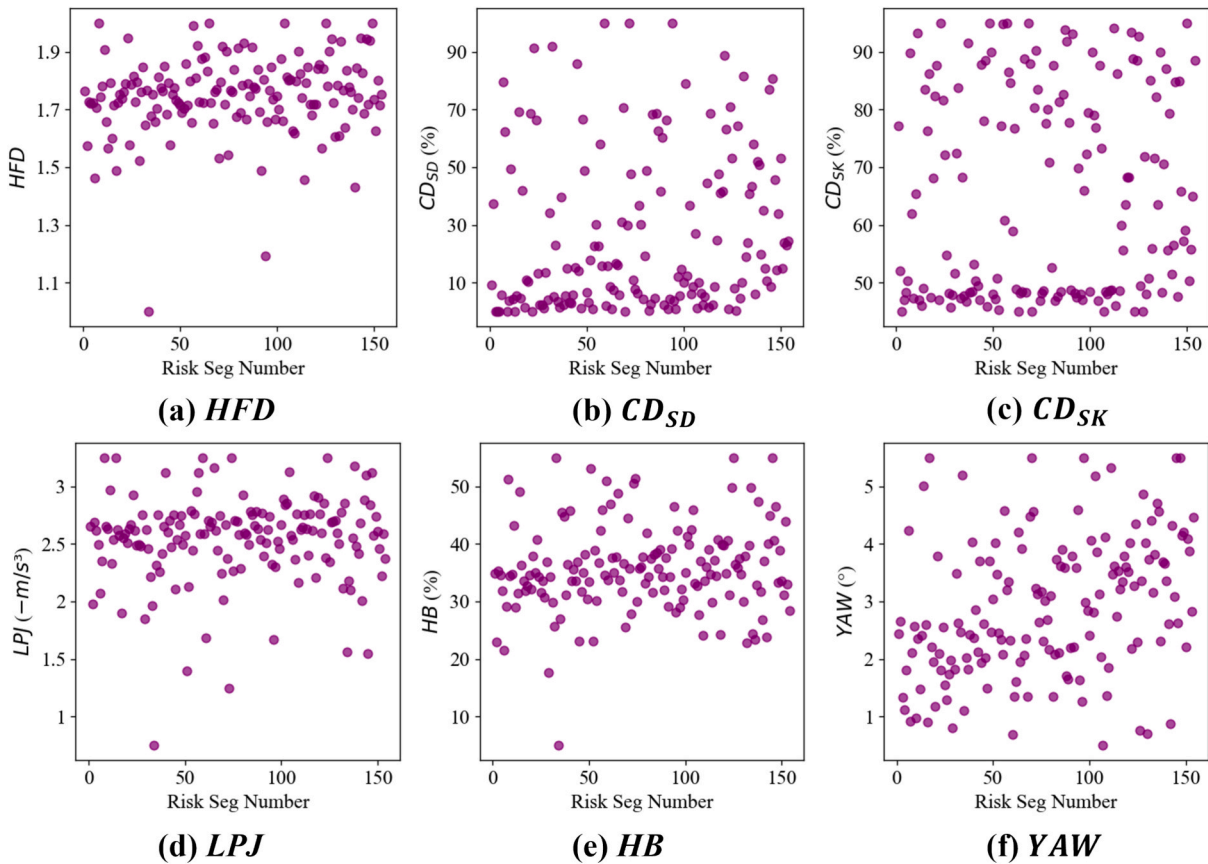

Fig. 13. The values of *HFD* and survey surrogates in high-risk road sections.

4.3.3 The tradeoff between Recall and FAR

Fig. 14 shows the Recall and FAR of each surrogate at different thresholds. The optimal threshold needs to ensure the desired Recall (set to 95 % in this study) while minimizing the FAR in low-risk road segment detection. As shown in Fig. 14(a), a threshold of 1.68 for *HFD* achieves a 95 % Recall and a minimal FAR of 46 %, making 1.68 the optimal threshold for *HFD*. When the Recall is set to 95 %, *HFD* has the
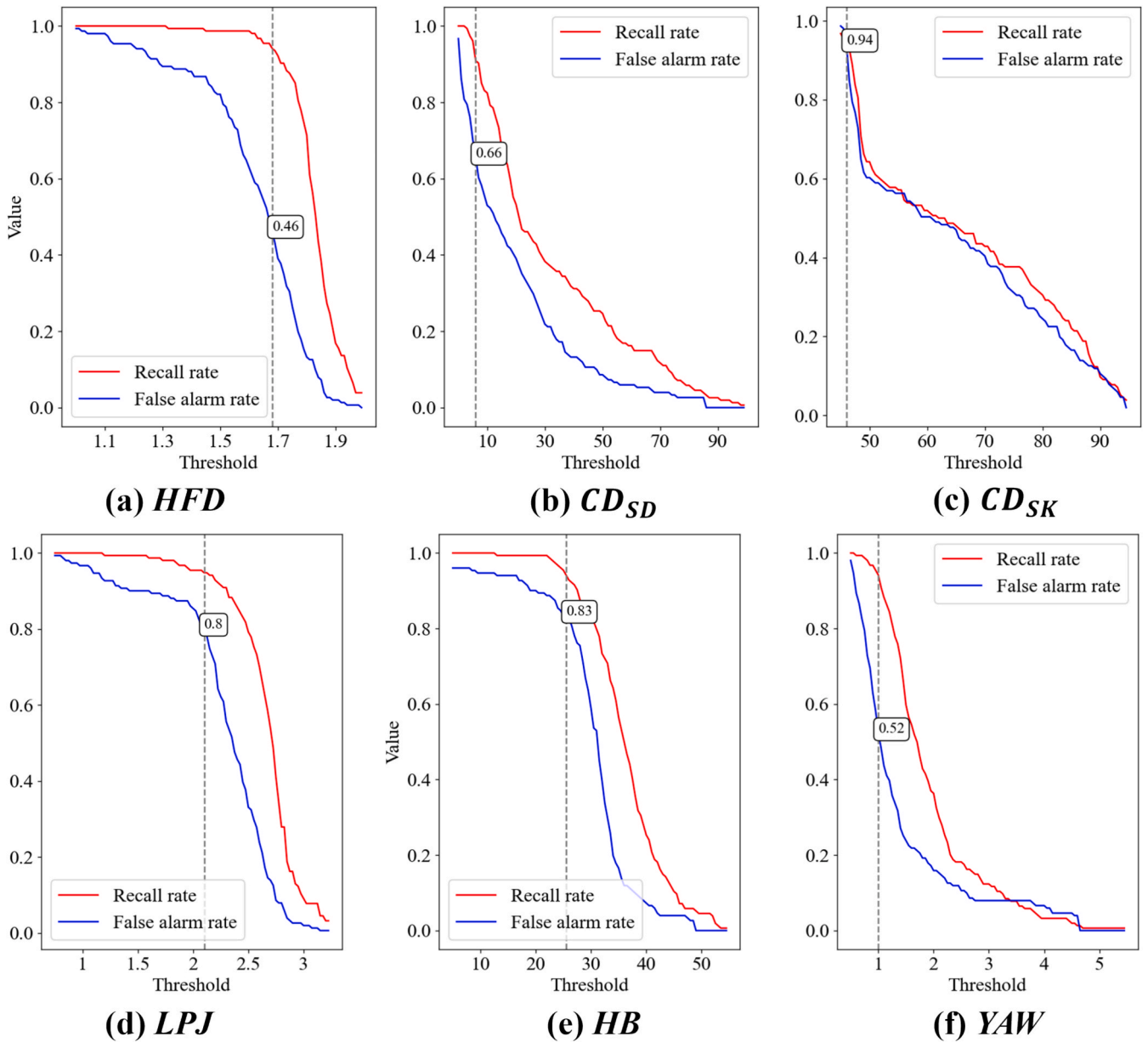
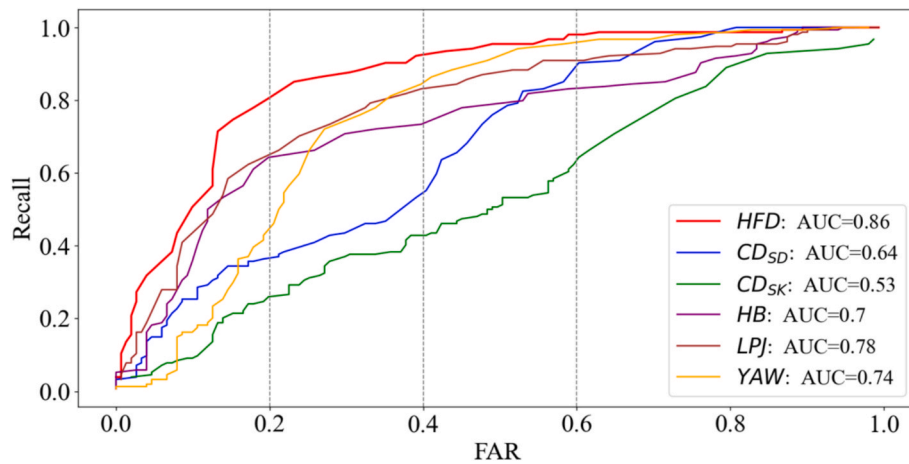**Fig. 14.** The relationship between the Recall of various surrogates and FAR.

Fig. 15. ROC curves of *HFD* and surrogates.

lowest FAR (46 %) compared to CD_{SD} (72 %), CD_{SK} (94 %), LPJ (80 %), HB (83 %), and YAW (52 %). Implementing a stricter threshold reduces FAR but comes at the cost of Recall. For example, when the desired Recall is relaxed to 90 %, the FAR values for HFD, CD_{SD}, CD_{SK}, HB, LPJ, and YAW decrease to 28 %, 58 %, 79 %, 54 %, 75 %, and 44 %, respectively.

Fig. 15 displays the ROC curve and the corresponding AUC values, which visually compare the performance of these metrics. The AUC value of HFD is the highest at 0.86, surpassing CD_{SD} (0.64), CD_{SK} (0.53), LPJ (0.78), HB (0.70), and YAW (0.74). Our results indicate that HFD performs the best in predicting highway crash risk. Furthermore, HFD outperforms the selected surrogates in maximizing Recall while maintaining a specific FAR. HFD significantly improves Recall, especially under conditions with lower FAR requirements. In contrast, when higher FAR requirements are imposed, the extent to which HFD improves Recall is reduced. This phenomenon can be attributed to a stringent threshold, which eliminates most low-risk segments. Comparative analysis shows that HFD effectively controls FAR while maintaining a high Recall, highlighting its practical utility and applicability.

4.3.4 Example analysis and application of fractal method

Based on Fig. 15, when the FAR is set to 0.5, CD_{SD}, CD_{SK}, LPJ, HB, and YAW identified 100, 83, 122, 110, and 116 high-risk segments, respectively. In comparison, HFD identified 135 high-risk segments. Compared to the selected surrogates, HFD demonstrated higher sensitivity in identifying high-risk segments. Two examples were chosen to explain and compare the identification performance of the surrogates. Fig. 16 includes 14 segments, along with their geometric features and trajectory fractal characteristics. According to CR, three high-risk areas were identified: Risk area 1 and 3 each contain one segment, while Risk area 2 contains three segments. The scatter points above and below the risk threshold line represent high-risk and low-risk segments identified by each normalized surrogate under the condition that FAR is set to 0.5, respectively. As shown in Fig. 17, an interchange area was also selected, where 6 high-risk regions were marked based on recorded crashes numbers, and a comparison was made between HFD and the selected surrogates. The red areas in the surrogates can be regarded as high-risk

segments identified by the surrogates. The interpretation of the results for both examples is as follows:

1) **Surrogates related to geometric design**: As shown in Fig. 16, the advantage of CD_{SD} and CD_{SK} lies in their ability to better identify high-risk road segments in highway curves, even in areas (such as before S-shaped curves and in the middle of interchanges) where kinematic triggers are not significantly noticeable. Our data shows that locations with non-compliant sight distance are far more prevalent than those with non-compliant skid resistance, which is why CD_{SD} identifies more high-risk locations than CD_{SK}. However, their performance is not ideal in diverging (Stations: 1057500–1058000) and merging (Stations: 1058500–1059000) zones, because they do not account for high-frequency weaving operations and interference from other lane-changing vehicles in these special areas (Rim et al., 2023). Fig. 17 shows that, possibly due to the continuity of road geometric design, the high-risk segments identified by CD_{SD} and CD_{SK} exhibit some spatial concentration. However, this does not consider the heterogeneity of driver behavior on restricted road segments. For example, although ramps are design-constrained, speed limit measures make drivers extremely cautious on these sections (Feknssa et al., 2023), leading to some false alarms in design-based surrogates.

2) **Kinematic trigger surrogates**: Fig. 16 shows that the kinematic surrogates identify the merging and diverging areas (risk area 2) as high-risk sections, as the motion characteristics of these regions exhibit significant fluctuations. Studies have shown that the crash location does not theoretically coincide with the traffic risk location; the "real" risk location may actually be before the crash point (Gu et al., 2023). For example, the true high-frequency crash area is located before the S-curve (risk area 1), possibly due to the large cognitive workload experienced by drivers when encountering the S-curve (Khan et al., 2023), but there is no significant Kinematic triggers in this region. Risk area 3 can be similarly explained. Furthermore, if frequent braking behavior occurs on a section, it only indicates that the driver is making rapid adjustments in response to the constantly changing road features.
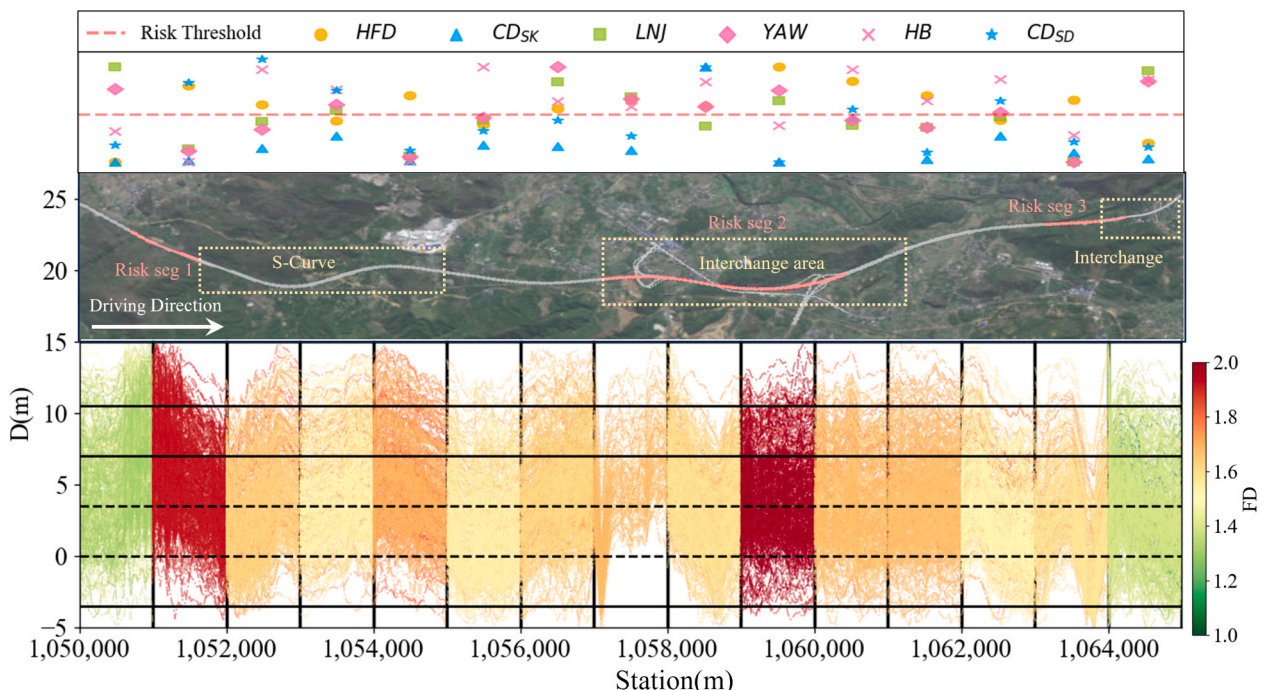

Fig. 16. Example 1: Comparison of risk identification ability of surrogates in Mainline Section.
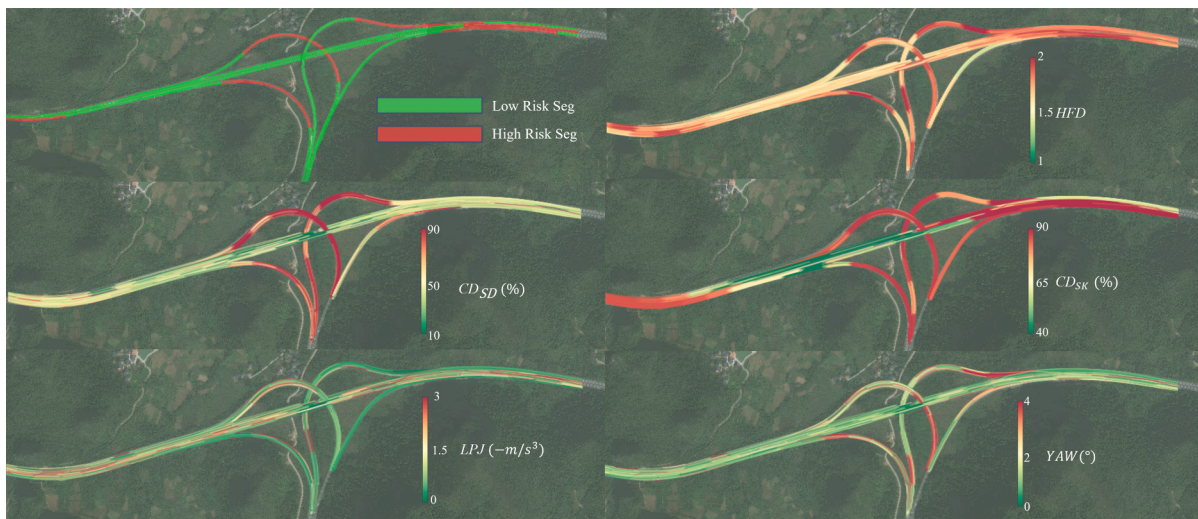
Fig. 17. Example 2: Comparison of surrogates' ability to identify risks in an interchange area.

The performance of *YAW* identification is superior to that of velocity characteristics in interchange sections. *YAW* shows stronger fluctuations in the interchange areas compared to velocity changes, likely because steering to avoid crashes is the first action, ahead of braking, which is consistent with previous studies (Naude et al., 2019). *HB* is a kinematic indicator with a threshold setting, but the threshold can only indicate the risk for drivers on specific sections (Pinnow et al., 2021), and in many cases, these are false alarms. Fig. 17 shows that due to the speed limit of the ramp, velocity fluctuations are inevitable when vehicles enter the ramp from the mainline, which is one of the sources of false alarms. *YAW* performs relatively well, as it identifies many extreme values, which often align more closely with the actual conditions of interchange areas. In conclusion, the kinematic patterns of a section are easily influenced by extreme values, making the difference in kinematic triggers between high-risk and low-risk sections less obvious (Fig. 9).

Compared to the baseline surrogates, the high-risk segments identified by *HFD* under the condition of a FAR set to 0.5 are more numerous, with a miss rate of only 12.9 %. More importantly, the differences in recognition performance across road sections with different geometric features are minimal (Fig. 13), indicating that *HFD* exhibits stronger robustness than other safety surrogates. One possible explanation is that the fractal method scales and observes the entire trajectory sequence, fundamentally improving the granularity of road safety analysis based on vehicle trajectories, thereby enhancing the accuracy of feature analysis. On the other hand, fractal features better reflect the interaction between road geometry and actual vehicle displacement, enabling fractal features to cover multiple dimensions of vehicle driving characteristics analysis, even in highly complex road segments. In addition, identifying low-risk road sections as high-risk road sections may result in certain resource waste. The data used in this article identifies many low-risk road sections as high-risk road sections. One possible explanation is that there are some factors that can cause surrogates triggers in these road sections, but do not increase the risk of crashes. This interesting finding deserves further investigation, as the evolution patterns of these independent crash surrogates may differ from crash related factors (Zhang et al., 2021). If we can distinguish abnormal surrogate triggers, it is likely to make the model's output more in line with the actual road crash risk. In summary, fractal features are significantly superior to other features in terms of misjudgment, highlighting their advantages in practical applications. Therefore, for newly built roads or roads lacking crash statistics data, relevant departments can deploy the experimental method developed in this study to analyze the patterns of trajectory data and make preliminary judgments on high-risk crash sections.

5. Discussion

5.1. Theoretical significance and main findings

This paper explores the application of vehicle trajectory fractal theory in road safety analysis. We propose three FD measurement methods for vehicle trajectory motion: the Ratio Method (Katz), the Covering Method (box-counting), and the Trajectory Length Scaling Method (Higuchi). By analyzing the variation of geometric features at different resolutions, it is found that vehicle trajectories exhibit distinct fractal characteristics. The fractal process of vehicle trajectories analyzes the non-linear and complex behaviors of vehicles at a finer granularity, capturing the multi-level structure of the trajectory. The analysis of variance tested significant differences in the *HFD* mean and standard deviation between high-risk and low-risk road sections. The *HFD* mean in the road section is more significant, and the *HFD* standard deviation in the road section can explain some high-risk road sections with low *HFD* mean. This result implies that the increased complexity and variability of vehicle trajectories will increase the risk of crashes. This suggests that *HFD* can serve as a safety surrogate, specifically used to assess crash risks associated with highway spatiotemporal factors based on large-scale telematics trajectory data.

We developed a method to estimate fractal parameters and discussed the impact of the key parameters k_2 and s. To minimize the overall objective function value, s and k_2 were set to intermediate values of 8 and 9, respectively. Based on this, a negative binomial regression model for crash frequency prediction was established using the *HFD* mean and standard deviation of the road section as independent variables to examine the sensitivity of parameter variations. The calculation of *HFD* mean and standard deviation in the road section was tested for Monte Carlo errors, and the model parameters were statistically significant in most cases. The range of k_2 between 6 and 10 ensures good estimation performance. Furthermore, the choice of s is more critical. In other words, the analysis length of the trajectory and the sampling frequency are essential, which means that to achieve the best fractal parameter estimation, higher frequency data acquisition devices or data interpolation might be necessary. It is noteworthy that the fractal characteristics of the same vehicle trajectory at different sampling frequencies may reflect similar road safety patterns, but their FD values are not directly comparable, as the richness of the motion details of vehicle trajectories varies with different sampling frequencies. This paper considers common FD such as *HFD*, *BFD*, and *KFD*, but correlation analysis shows that only *HFD* is associated with crash risk. *BFD* and *KFD*, on the other hand, are primarily related to some kinematic triggers. Therefore, further

research is needed on the application of *BFD* and *KFD* in traffic flow analysis.

This paper selects five widely used macroscopic surrogates as baselines. Compared to the selected surrogates, road segments with an *HFD* value below 1.6 account for only 8 %, and high-risk road segments show strong stability in their *HFD* values. This finding suggests that the selected surrogates may be influenced by more unconsidered factors. The paper analyzes the Recall and FAR balance for each surrogate. *HFD* increases the AUC value from 0.78 to 0.86 (for other surrogates). In summary, *HFD* outperforms the selected surrogates in maximizing Recall while maintaining a specific FAR, highlighting its practical utility and applicability.

Due to the interactive influence of road design elements on driving, frequent kinematic triggers do not necessarily indicate that the road segment is of high risk. In contrast, *HFD* has stronger recognition ability and robustness in different geometric design sections. One possible explanation is that the fractal method scales and observes the entire trajectory sequence, fundamentally improving the granularity of road safety analysis based on vehicle trajectories (Yilmaz and Unal, 2020), thereby enhancing the accuracy of feature analysis. On the other hand, fractal features better reflect the interaction between road geometry and actual driving displacement, enabling fractal features to cover multiple dimensions of vehicle driving characteristics analysis, even in highly complex road segments.

It is important to note that the method developed in this study can only estimate the overall risk tendency of road segments, and different types of drivers and micro-level behaviors may exhibit different *HFD* values in these high-risk areas. In other words, there is unobservable heterogeneity when calculating FD due to the differences in more detailed driving behavior and road scenarios. When we are able to capture these heterogeneities, we can identify high-risk behaviors in micro-driving actions. Since fractal methods enhance the feature capturing capability at a finer granularity compared to traditional trajectory sequence analysis techniques, and also have lower computational resource demands, fractal theory holds promise for applications in traffic conflict detection technologies, potentially offering greater assistance in crash prevention and management. Although this paper does not address traffic conflicts directly, fractal methods could serve as a surrogate estimate for conflicts under conditions where vehicle-equipped data collection devices are limited. In conclusion, the focus of this paper is to demonstrate the advantages of fractal methods in identifying high-risk highway segments within the context of large-scale telematics data. Importantly, fractal analysis enhances the diversity of data and expands new perspectives for analyzing vehicle trajectory data. With the improvement of fractal methods, it is expected to use fractal methods to identify more accurate risk points on shorter road sections in the future.

5.2. Realistic meaning

Against the backdrop of significant advancements in the positioning accuracy of mobile sensors and the convenience of data collection, we analyzed the relationship between the FD of vehicle trajectories and road safety using large-scale telematics trajectory data. To the best of our knowledge, few studies have conducted highway safety analysis based on such large-scale high-frequency trajectory data. Our research further enriches the application of large-scale trajectory data in highway safety studies. Our theoretical study contributes to the enrichment of analytical methods for trajectory big data, offering new insights for road safety research. The fractal theory of vehicle trajectories indicates that the scaling behavior of vehicle trajectories reflects the micro and macro levels of vehicle motion patterns, despite the lack of obvious periodicity in the motion patterns of individual vehicles. For newly built roads, relevant departments only need to establish a road model consistent with the design drawings in simulation software, and then establish vehicle models and driving control models to obtain trajectory fractal

data and make preliminary judgments on high-risk crash sections. For existing roads lacking crash data, based on the identification results, the department can develop targeted measures to improve road safety. This study has a limitation, although the spatial range of the data used in this article is large, detailed geometric feature data such as traffic signs, markings, roadside obstacles, etc. cannot be collected (Kim et al., 2024). Due to these issues, this study aims to eliminate the influence as much as possible through sufficient observation times. If more data can be obtained in the future and more crash data can be matched, it is expected to provide more reliable results in the validation process of the method. Human factors have a huge impact on the occurrence of crashes. If the driver information of vehicles can be matched in the future, it is expected that fractal methods for vehicle trajectories that reflect human factors can be developed. The vehicle trajectories used in this study do not include those involving lane changes, diverging, or merging, and these trajectories are sourced from highways. Therefore, it remains to be further investigated whether the research findings are applicable to urban roads or special road sections. Finally, it is recommended to evaluate the crash risk based on vehicle driving trajectories in the context of Connected and Automated Vehicle (CAV) environments (Dong et al., 2023).

6. Conclusion

This paper explores the application of vehicle trajectory fractal theory in road safety analysis using large-scale telematics trajectory data. A total of 90,000 vehicle trajectories were extracted from 314 road sections. These data were derived from GPS records of freight vehicles on highways in China. Using this data, we validated the significant relationship between trajectory FD and highway crash risks, and explored the advantages of fractal characteristics over the selected surrogates. In this comprehensive analysis, we made some new findings and further discussed the significance of vehicle trajectory fractals. We hope these findings will help enhance the dimension of trajectory feature analysis using trajectory data and refine the granularity of road safety analysis, enriching the analytical methods for vehicle trajectory data and providing methodological support for improving road geometric design and enhancing road safety.

CRediT authorship contribution statement

Yuhan Nie: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Min Zhang:** Writing – review & editing, Supervision. **Bo Wang:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Chi Zhang:** Writing – review & editing, Supervision. **Yijing Zhao:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Data availability

The authors do not have permission to share data.

References

Alrassy, P., Smyth, A.W., Jang, J., 2023. Driver behavior indices from large-scale fleet telematics data as surrogate safety measures. In Accident Analysis and Prevention (Vol. 179). Pergamon-Elsevier Science LTD. https://doi.org/10.1016/j.aap.2022.106879.

Alruwaili, A., Xie, K., 2024. Modeling the influence of connected vehicles on driving behaviors and safety outcomes in highway crash scenarios across varied weather conditions: a multigroup structural equation modeling analysis using a driving simulator experiment. Accident Analysis and Prevention 199. https://doi.org/10.1016/j.aap.2024.107514. Pergamon-Elsevier Science Ltd.

Alsaleh, R., Lanzaro, G., Sayed, T., 2024. Incorporating design consistency into risk-based geometric design of horizontal curves: a reliability-based optimization framework. In Transportmetrica A-Transport Science (Vol. 20, Issue 2). Taylor & Francis LTD. https://doi.org/10.1080/23249935.2023.2174356.

Arman, M.A., Tampere, C.M.J., 2021. Lane-level routable digital map reconstruction for motorway networks using low-precision GPS data. Transportation Research Part C-Emerging Technologies 129. https://doi.org/10.1016/j.trc.2021.103234. Pergamon-Elsevier Science Ltd.

Arman, M.A., Tampere, C.M.J., 2022. Lane-level trajectory reconstruction based on data-fusion. Transportation Research Part C-Emerging Technologies 145. https://doi.org/10.1016/j.trc.2022.103906. Pergamon-Elsevier Science Ltd.

Arman, M.A., Tampere, M.J., 2024. Choice-based macroscopic lane-change prediction model for weaving areas. Transportation Research Part C-Emerging Technologies 169. https://doi.org/10.1016/j.trc.2024.104871. Pergamon-Elsevier Science Ltd.

Bahramizadeh-Sajadi, S., Katoozian, H.R., Mehrabbeik, M., Baradaran-Rafii, A., Jadidi, K., Jafari, S., 2022. A fractal approach to nonlinear topographical features of healthy and keratoconus corneas pre- and post-operation of intracorneal implants. Fractal and Fractional 6 (11). https://doi.org/10.3390/fractalfract6110688.

Chand, S., Dixit, V.V., 2018. Application of Fractal theory for crash rate prediction: Insights from random parameters and latent class tobit models. Accident Analysis and Prevention 112, 30–38. https://doi.org/10.1016/j.aap.2017.12.023. Pergamon-Elsevier Science Ltd.

Chauhan, P., Kanagaraj, V., Asaithambi, G., 2022. Understanding the mechanism of lane changing process and dynamics using microscopic traffic data. Physica A-Statistical Mechanics and its Applications 593. https://doi.org/10.1016/j.physa.2022.126981. Elsevier.

Chen, Y., Li, G., Li, S., Wang, W., Li, S.E., Cheng, B., 2022. Exploring Behavioral Patterns of Lane Change Maneuvers for Human-Like Autonomous Driving. IEEE Transactions on Intelligent Transportation Systems (Vol. 23, Issue 9, pp. 14322–14335). IEEE-Inst Electrical Electronics Engineers Inc. https://doi.org/10.1109/TITS.2021.3127491.

Chen, Y., Shi, T., Yu, S., Shi, Q., He, J., Bian, Y., 2022b. Setting the speed limit for highway horizontal curves: a revision of inferred design speed based on vehicle system dynamics. Safety Science 151. https://doi.org/10.1016/j.ssci.2022.105729. Elsevier.

Cheng, D., Li, Z., Selmi, B., 2024. On the general fractal dimensions of hyperspace of compact sets. Fuzzy Sets and Systems (Vol. 488). Elsevier. https://doi.org/10.1016/j.fss.2024.108998.

Dai, Z., Pan, C., Xiong, W., Ding, R., Zhang, H., Xu, J., 2022. Research on vehicle trajectory deviation characteristics on freeways using natural driving trajectory data. International Journal of Environmental Research and Public Health 19 (22). https://doi.org/10.3390/ijerph192214695. MDPI.

Ding, R., Pan, C., Dai, Z., Xu, J., 2022. Lateral oscillation characteristics of vehicle trajectories on the straight sections of freeways. Applied Sciences-Basel 12 (22). https://doi.org/10.3390/app122211498. MDPI.

Dong, C., Xing, L., Wang, H., Yu, X., Liu, Y., Ni, D., 2023. Iterative learning control for lane-changing trajectories upstream off-ramp bottlenecks and safety evaluation. Accident Analysis and Prevention 183. https://doi.org/10.1016/j.aap.2023.106970. Pergamon-Elsevier Science Ltd.

Dzinyela, R., Shirazi, M., Das, S., Lord, D., 2024. The negative binomial-lindley model with time-dependent parameters: accounting for temporal variations and excess zero observations in crash data. In Accident Analysis and Prevention (Vol. 207). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.aap.2024.107711.

Feknssa, N., Venkataraman, N., Shankar, V., Ghebrab, T., 2023. Unobserved heterogeneity in ramp crashes due to alignment, interchange geometry and truck volume: Insights from a random parameter model. In Analytic Methods IN Accident Research (Vol. 37). Elsevier. https://doi.org/10.1016/j.amar.2022.100254.

Gao, S.-J., Mei, C.-L., Xu, Q.-X., Zhang, Z., 2023. Non-iterative multiscale estimation for spatial autoregressive geographically weighted regression models. Entropy 25 (2). https://doi.org/10.3390/e25020320. MDPI.

Gargoum, S.A., El-Basyouny, K., 2020. Analyzing the ability of crash-prone highways to handle stochastically modelled driver demand for stopping sight distance. Accident Analysis and Prevention 136. https://doi.org/10.1016/j.aap.2019.105395. Pergamon-Elsevier Science Ltd.

Gedamu, W.T., Plank-Wiedenbeck, U., Wodajo, B.T., 2024. A spatial autocorrelation analysis of road traffic crash by severity using Moran's I spatial statistics: a comparative study of Addis Ababa and Berlin cities. In Accident Analysis and Prevention (Vol. 200). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.aap.2024.107535.

Gu, Y., Liu, D., Arvin, R., Khattak, A.J., Han, L.D., 2023. Predicting intersection crash frequency using connected vehicle data: a framework for geographical random forest. Accident Analysis and Prevention 179. https://doi.org/10.1016/j.aap.2022.106880. Pergamon-Elsevier Science Ltd.

Gvozdarev, A., Parovik, R., 2023. On the relationship between the fractal dimension of geomagnetic variations at altay and the space weather characteristics. Mathematics 11 (16). https://doi.org/10.3390/math11163449. MDPI.

Hamedi, H., Shad, R., 2022. Context-aware similarity measurement of lane-changing trajectories. Expert Systems with Applications 209. https://doi.org/10.1016/j.eswa.2022.118289. Pergamon-Elsevier Science Ltd.

Kamla, J., Parry, T., Dawson, A., 2019. Analysing truck harsh braking incidents to study roundabout accident risk. Accident Analysis and Prevention 122, 365–377. https://doi.org/10.1016/j.aap.2018.04.031. Pergamon-Elsevier Science Ltd.

Karimui, R.Y., 2021. A new approach to measure the fractal dimension of a trajectory in the high-dimensional phase space. In Chaos Solitons & Fractals (Vol. 151). Pergamon-Elsevier Science LTD. https://doi.org/10.1016/j.chaos.2021.111239.

Khan, S.A., Yasmin, S., Haque, M.M., 2023. Effects of design consistency measures and roadside hazard types on run-off-road crash severity: application of random parameters hierarchical ordered probit model. Analytic Methods in Accident Research 40. https://doi.org/10.1016/j.amar.2023.100300. ELSEVIER.

Kim, Y., Kang, K., Park, J., Oh, C., 2024. A methodology for prioritizing safety indicators using individual vehicle trajectory data. Journal of Transportation Safety & Security (Vol. 16, Issue 1, pp. 18–42). Taylor & Francis Inc. https://doi.org/10.1080/19439962.2023.2178567.

Lee, S.-H., Park, C.-M., 2021. A new measure to characterize the self-similarity of binary time series and its application. IEEE Access 9, 73799–73807. https://doi.org/10.1109/ACCESS.2021.3081400.

Li, H., Liu, Y., Wang, Y., Liao, H., 2023. Estimation method of ideal fractal parameters for multi-scale measurement of polished surface topography. Fractal and Fractional 7 (1). https://doi.org/10.3390/fractalfract7010017. MDPI.

Li, X., Chen, Y., Zhu, J., Zeng, S., Jiang, X., Shen, Y., Zhang, D., 2024. Fractal dimension of DSSS frame preamble: radiometric feature for wireless device identification. In IEEE Transactions on Mobile Computing (Vol. 23, Issue 2, pp. 1416–1430). IEEE Computer Soc. https://doi.org/10.1109/TMC.2023.3235497.

Li, X., Mousavi, S. M., Dadashova, B., Lord, D., Wolshon, B., 2021. Toward a crowdsourcing solution to identify high-risk highway segments through mining driving jerks. In Accident Analysis and Prevention (Vol. 155). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.aap.2021.106101.

Li, Y., Zhang, S., Liang, L., Ding, Q., 2024b. Multivariate multiscale higuchi fractal dimension and its application to mechanical signals. Fractal and Fractional 8 (1). https://doi.org/10.3390/fractalfract8010056. MDPI.

Li, Y., Zhou, Y., Jiao, S., 2024c. Variable-step multiscale katz fractal dimension: a new nonlinear dynamic metric for ship-radiated noise analysis. Fractal and Fractional 8 (1). https://doi.org/10.3390/fractalfract8010009. MDPI.

Li, Z., Wang, J., Yuan, M., Wang, Z., Feng, P., Feng, F., 2022. An indicator to quantify the complexity of signals and surfaces based on scaling behaviors transcending fractal. In Chaos Solitons & Fractals (Vol. 163). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.chaos.2022.112556.

Liu, Z., He, J., Zhang, C., Xing, L., Zhou, B., 2020. The impact of road alignment characteristics on different types of traffic accidents. In Journal of Transportation Safety & Security (Vol. 12, Issue 5, pp. 697–726). Taylor & Francis inc. https://doi.org/10.1080/19439962.2018.1538173.

Macek, W., Branco, R., Podulka, P., Nejad, R.M., Costa, J. D., Ferreira, J.A.M., Capela, C., 2023. The correlation of fractal dimension to fracture surface slope for fatigue crack initiation analysis under bending-torsion loading in high-strength steels. In Measurement (Vol. 218). Elsevier Sci Ltd. https://doi.org/10.1016/j.measurement.2023.113169.

Moriguchi, K., 2023. Estimation of fractal dimension of trees using LiDAR point data with sequential data decimation. Remote Sensing of Environment 295. https://doi.org/10.1016/j.rse.2023.113722. Elsevier Science Inc.

Naude, C., Serre, T., Dubois-Lounis, M., Fournier, J.-Y., Lechner, D., Guilbot, M., Ledoux, V., 2019. Acquisition and analysis of road incidents based on vehicle dynamics. Accident Analysis and Prevention 130 (SI), 117–124. https://doi.org/10.1016/j.aap.2017.02.021. Pergamon-Elsevier Science Ltd.

Nikolaou, D., Dragomanovits, A., Ziakopoulos, A., Deliali, A., Handanos, I., Karadimas, C., Kostoulas, G., Frantzola, E.K., Yannis, G., 2023. Exploiting surrogate safety measures and road design characteristics towards crash investigations in motorway segments. Infrastructures 8 (3). https://doi.org/10.3390/infrastructures8030040. MDPI.

Pinnow, J., Masoud, M., Elhenawy, M., Glaser, S., 2021. A review of naturalistic driving study surrogates and surrogate indicator viability within the context of different road geometries. In Accident Analysis and Prevention (Vol. 157). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.aap.2021.106185.

Rim, H., Abdel-Aty, M., Mahmoud, N., 2023. Multi-vehicle safety functions for freeway weaving segments using lane-level traffic data. Accident Analysis and Prevention 188. https://doi.org/10.1016/j.aap.2023.107113. Pergamon-Elsevier Science Ltd.

Sarkar, D.R., Rao, K.R., Chatterjee, N., 2024. A review of surrogate safety measures on road safety at unsignalized intersections in developing countries. In Accident Analysis and Prevention (Vol. 195). Pergamon-Elsevier Science Ltd. https://doi.org/10.1016/j.aap.2023.107380.

Sui, L., Wang, H., Wu, J., Zhang, J., Yu, J., Ma, X., Sun, Q., 2022. Fractal description of rock fracture networks based on the space syntax metric. Fractal and Fractional 6 (7). https://doi.org/10.3390/fractalfract6070353. MDPI.

Wang, B., Chen, T., Zhang, C., Wong, Y.D., Zhang, H., Zhou, Y., 2024. Toward safer highway work zones: an empirical analysis of crash risks using improved safety potential field and machine learning techniques. In Accident Analysis and Prevention (Vol. 194). Pergamon-Elsevier Science ltd. https://doi.org/10.1016/j.aap.2023.107361.

Wang, B., Wong, Y.D., Zhang, C., Zhang, H., Gao, Y., 2024. Exploring the impact of rainfall on vehicle trajectory patterns and sideslip risk: an empirical investigation. In Journal of Advanced Transportation (Vol. 2024). Wiley-Hindawi. https://doi.org/10.1155/2024/3138719.

Wang, C., Xie, Y., Huang, H., Liu, P., 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modelling. Accident Analysis and Prevention 157. https://doi.org/10.1016/j.aap.2021.106157. Pergamon-Elsevier Science Ltd.

Wang, Y., Gloudemans, D., Ji, J., Teoh, Z.N., Liu, L., Zachar, G., Barbour, W., Work, D., 2024c. Automatic vehicle trajectory data reconstruction at scale. Transportation Research Part C-Emerging Technologies 160. https://doi.org/10.1016/j.trc.2024.104520. Pergamon-Elsevier Science Ltd.

Wang, Y., Li, Z., Liu, P., Xu, C., Chen, K., 2024d. Surrogate safety measures for traffic oscillations based on empirical vehicle trajectories prior to crashes. Transportation Research Part C-Emerging Technologies 161. https://doi.org/10.1016/j.trc.2024.104543. Pergamon-Elsevier Science Ltd.

Wanliss, J.A., Wanliss, G.E., 2022. Efficient calculation of fractal properties via the Higuchi method. In Nonlinear Dynamics (Vol. 109, Issue 4, pp. 2893–2904). Springer. https://doi.org/10.1007/s11071-022-07353-2.

Wu, D., Zhang, Y., Xiang, Q., 2024. Geographically weighted random forests for macro-level crash frequency prediction. Accident Analysis and Prevention 194. https://doi.org/10.1016/j.aap.2023.107370. Pergamon-Elsevier Science Ltd.

Xian, H., Hou, Y., Wang, Y., Dong, S., Kou, J., Li, Z., 2023. Influence of risky driving behavior and road section type on urban expressway driving safety. Sustainability 15 (1). https://doi.org/10.3390/su15010398. MDPI.

Yan, X., He, J., Wu, G., Sun, S., Wang, C., Fang, Z., Zhang, C., 2024. Driving risk identification of urban arterial and collector roads based on multi-scale data. Accident Analysis and Prevention 206. https://doi.org/10.1016/j.aap.2024.107712.

Ye, W., Xu, Y., Shi, X., Shiwakoti, N., Ye, Z., Zheng, Y., 2024. A macroscopic safety indicator for road segment: application of entropy theory. Physica A-Statistical Mechanics and its Applications 642. https://doi.org/10.1016/j.physa.2024.129787. Elsevier.

Yilmaz, A., Unal, G., 2020. Multiscale Higuchi's fractal dimension method. Nonlinear Dynamics 101 (2), 1441–1455. https://doi.org/10.1007/s11071-020-05826-w. Springer.

Zhang, C., He, J., King, M., Liu, Z., Chen, Y., Yan, X., Xing, L., Zhang, H., 2021. A crash risk identification method for freeway segments with horizontal curvature based on real-time vehicle kinetic response. Accident Analysis and Prevention 150. Pergamon-Elsevier Science Ltd.

Zhang, Q., Pei, Y., Shen, Y., Wang, X., Lai, J., Wang, M., 2023a. Fracture surface. Fractal and Fractional 7 (7). https://doi.org/10.3390/fractalfract7070496. MDPI.

Zhang, Y., Zou, Y., Selpi, Y., Zhang, Y., Wu, L., 2023. Spatiotemporal interaction pattern recognition and risk evolution analysis during lane changes. In IEEE Transactions On Intelligent Transportation Systems (Vol. 24, Issue 6, pp. 6663–6673). IEEE-Inst Electrical Electronics Engineers Inc. https://doi.org/10.1109/TITS.2022.3233809.

Zhao, Y., Wang, C., Ning, L., Zhao, H., Bi, J., 2022. Pore and fracture development in coal under stress conditions based on nuclear magnetic resonance and fractal theory. In Fuel (Vol. 309). Elsevier Sci Ltd. https://doi.org/10.1016/j.fuel.2021.122112.

Zhu, J., Ma, Y., Lou, Y., 2022. Multi-vehicle interaction safety of connected automated vehicles in merging area: a real-time risk assessment approach. Accident Analysis and Prevention 166. https://doi.org/10.1016/j.aap.2021.106546. Pergamon-Elsevier Science Ltd.

Ziakopoulos, A., 2024. Analysis of harsh braking and harsh acceleration occurrence via explainable imbalanced machine learning using high-resolution smartphone telematics and traffic data. Accident Analysis and Prevention 207. https://doi.org/10.1016/j.aap.2024.107743. Pergamon-Elsevier Science Ltd.