Journal of Highway and Transportation Research and Development
2024, 18 (2): 14-25
ISSN 2095-6215
DOI: 10.26599/HTRD.2024.9480010

# Cumulative Distribution-Based Method for Pavement Performance Modeling

Chenhao Tu*[1], Wenya Ye[1], Rui Zhang[2], Xudong Qian[3], Qun Yang**[2]

(1. School of Civil and Transportation Engineering, Ningbo University of Technology, Ningbo, Zhejiang 315021, China;
2. The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China;
3. Ningbo Road and Bridge Engineering Construction Co., Ltd., Ningbo, Zhejiang 315100, China)

**Abstract**: Pavement performance prediction is the basis for maintenance decisions. Predicting future pavement conditions accurately and efficiently helps determine the optimal maintenance time, select appropriate measures, and allocate rehabilitation funds effectively. However, limited to the instability and variability in pavement condition data collection, deterministic models are not always reliable for all pavement situations. On the other hand, probabilistic-based models are influenced by environmental factors that are challenging to quantify. Recognizing the limitations of the above two methods, this paper proposes a cumulative distribution-based technique for developing pavement performance prediction models. First, after comparing performance metrics such as pile-by-pile single-point, probability density, and cumulative distribution, it is evident that the cumulative distribution is the most reliable method for describing pavement conditions. A continuous distribution function is created from a limited set of discrete observed field pavement condition data using the sampling theorem. With cumulative distribution-based deterioration curves changing over time, it is possible to predict future pavement deterioration rates. A case study is presented at last. Analyses of the predicted curve and observed pavement performance indicate that the cumulative distribution-based technique is effective in modeling pavement performance and can provide reliable predictive results.

**Key words**: road engineering; pavement performance prediction; cumulative distribution; Montecarlo simulation; sampling theorem; pavement management system

## 1 Introduction

During the service process of the pavement, its performance decrease over time due to combined effects of repeated loadings and environment cycles, leading to damage[1-2]. To enhance performance, decrease the deterioration rate, and prolong the service life of pavement, predicting future pavement conditions accurately and efficiently is of primary importance[3]. It allows the Pavement Management System (PMS) to make maintenance and rehabilitation plans ahead, including determining the optimal maintenance time, selecting appropriate measures, and allocating rehabilitation funds, so as to reduce future expenses and minimize costs[4-5].

Since the 1950s, the issue of pavement performance deterioration and prediction has gradually attracted the attention of scholars, and great progress has been made. There are basically two types of pavement prediction models, deterministic and probabilistic models[6-7]. The Pavement Management Guide[8] classifies them into deterministic, probabilistic, Bayesian, and subjective (or expert-based) models. Similarly, UDDIN[9] classified the models in to regression analysis techniques, Artificial Neural Networks (ANN), and Probabilistic performance models, with Markov and Bayesian approaches been the main methodologies in this category. JUSTO-SILVA, et al.[10] classified them into deterministic, probabilistic, and hybrid (including fuzzy logic, ANN, and neuro-fuzzy). Nevertheless, the deterministic and the probabilistic models are the most widely used and they are recognized as the basic groups[11-15].

Regression models fit variables and parameters using traditional regression analysis approaches based on amounts of observed pavement condition data[16-17]. However, the reliability of regression analysis is greatly affected by the amount of observed data. Grey theories-

Received 13 November 2023; Revised 26 January 2024; Accepted 2 February 2024

* *E-mail address*: 956250283@qq.com

** Corresponding author, yang667tj@163.com

based prediction[18] is suitable for use when data is scarce and information is poor. Probabilistic models are often used to predict network pavement condition states. By generating a set of Markovian transition probability matrices, realizing the prediction of pavement deterioration rate[19-20]. However, the above methods are subject to environmental factors, which are difficult to quantify. The neural network has a good nonlinear quality and is suitable for handling road performance prediction problems with complex environmental information[21], but it has low model generalization capability for pavement engineering applications[22].

If the prediction is based on the average value of highway pavement performance indicators, it means that 50% of the actual conditions may be better or worse than the prediction[23]. Even if a standard road section of 1km is used for prediction, the workload is too large.

In pavement evaluation and the prediction variability should be paid attention. Uncertainty in pavement conditions, measurement errors, nonlinearities in pavement deterioration trends, and heterogeneity due to unobserved influencing factors can all negatively impact model accuracy[24-25]. Inaccurate models can lead to a poor prediction of pavement conditions and inefficient pavement management.

Probability distributions provide an alternative to the above existing problems. Since the beginning of the 21st century, probability density evolution theory is established and developed for the analysis of civil engineering structural systems under the excitation of non-stationary stochastic processes to deal with extremely strong nonlinearity and randomness[26-27]. For pavement conditions, the mean and the variance of the predicted performance indicators is increasing over time, and the probability distribution function can describe pavement conditions at any reliability level[28]. In addition, probabilistic pavement performance curves allow the effects of preventive maintenance to be incorporated into probabilistic pavement performance models without historical preventive maintenance data, thereby reducing bias in predicted pavement conditions[29]. However, probabilistic prediction methods, whether using distributions or Markov chains, require prior knowledge and design of the input distribution[4]. The inability to use a uniform distribution form to represent different performance indicators and road surface distress makes some work inconvenient to a certain extent.

Therefore, this study considers the inherent uncertainty of pavement condition data and uses the MC (Monte Carlo) method to determine the most stable representation of road segment performance that handles data variability. On this basis, a set of pavement performance evaluation and prediction methods that do not need to determine the exact distribution form are developed.

This paper is structured into three sections. The first section provides detailed stability comparisons among three performance expressions: pile-by-pile single-point, probability density, and cumulative distribution. The second section presents the cumulative distribution-based pavement performance deterioration functions. The final section outlines the prediction technique for cumulative distribution-based models and analyzes the deterioration curve.

## 2 Methodology

Researchers and transportation agencies around the country have developed a host of indexes to measure the functional, structural and material integrity of pavements. These indexes are an aggregation of several distress types (e.g., cracking, rutting, bleeding, etc. in asphalt pavement; spalling, cracking, faulting, etc. in concrete pavement) and other physical measurements (such as surface roughness and friction). According to the *Highway Performance Assessment Standards* (JTG 5210—2018)[30], *PCI* (Pavement Condition Index) and *RQI* (Pavement Riding Quality Index) are two of the most important indexes to evaluate the pavement condition. The *PCI* and *RQI* scale ranges from 0 to 100, with 100 representing the perfect score (i.e., a pavement in good condition). Therefore, this paper used *PCI* and *RQI* as pavement performance indicators for subsequent studies. According to the (JTG 5210—2018), the continuous *PCI* data from the east to west direction to evaluated the pavement performance were observed on a section of an expressway located in Shanghai (China) in the four years from 2017 to 2020. The observed road is 83.51 km in length. The observed road was divided into one unit evaluation section per km, with a total of 85-

unit evaluation sections due to bridges and non-whole pile numbers.

### 2.1 Stability analysis

The stability of the three pavement expressions is studied in three steps, graphically represented in Figure 1.

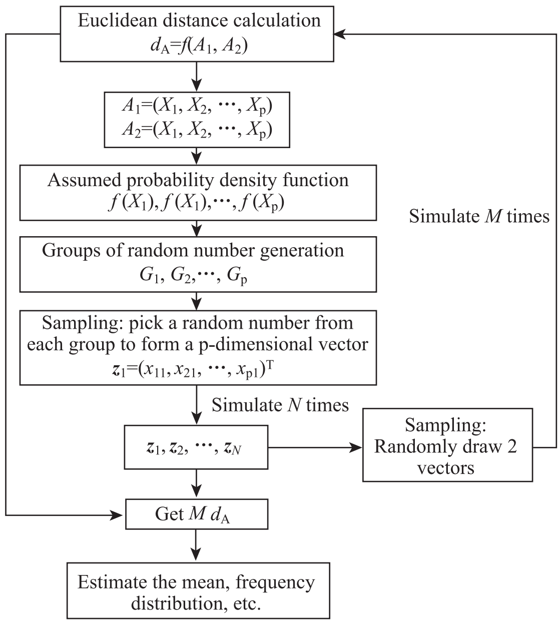



**Fig. 1 Flowchart of the simulation model**

where $p$ represents the number of unit section; $x_i$ represents the field measured value of the pavement performance indicators pile-by-pile single-point, $i=1,2,\cdots p$; $A_1$ and $A_2$ represent two p-dimensional vectors; $X_i$ represents random variable of the pavement performance indicators pile-by-pile single-point, following a normal distribution $N(\mu, \sigma^2)$, $\mu$ represents the mean, $\sigma$ represents the standard deviation; $G_i$ represents generated p-point groups with 10 000 data per group; $x_{ij}$ represents the generated value sampled from $G_i$, $j=1,2,\cdots,N$, $N$ represents simulation times; $Z_j$ represents $p$-dimensional vector consists of $x_{ij}$; $d_A$ represents Euclidean distance.

Firstly, a large amount of pavement condition simulation data based on their mean value and assumed standard deviation are generated using MATLAB. Secondly, sampling from the simulated data to form multiple groups of pile-by-pile single-point data vectors, probability density vectors, and cumulative distribution vectors. It shows how to get pile by pile data vectors in Figure 1. Finally, the Euclidean distance between groups for each performance expression was calculated, and the results were statistically analyzed to compare stability.

#### 2.1.1 Data generation

Considering the influence of noise such as environment and measurement, this study assumes that the pile-by-pile data for unit section pavement performance index are random variables following a normal distribution, with the mean $\mu$ being the field measured value, and the standard deviation $\sigma$ can be set to 0.5, 1, 2 according to the demand or actual situation. This study assumes that the random variable, *RQI* pile-by-pile single point test results follow a normal distribution, with the mean being the measured value for that year and a standard deviation of 1.

The mileage of a highway is $p$ km. The field observed pavement condition data is gained as single point data per kilometer, denoted as $x_1, x_2, \cdots, x_p$. The measured *RQI* values for each unit section are denoted as $\mu_1$, $\mu_2$, $\cdots$, $\mu_p$. On each single-point data per kilometer, normal distribution $N(x_i, \sigma)$ is used to generate 10 000 random numbers. Therefore $p$-point groups with 10 000 data per group are generated, denoted as $G_1, G_2, \cdots, G_p$. The meaning of $G_i$ is a point group containing 10,000 data simulated by the observed performance value of $i$-th kilometer.

#### 2.1.2 Random Sampling

Sequentially and randomly pick a point from each point group. It means a point is extracted from $G_1$, a point is extracted from $G_2$, $\cdots$, a point is extracted from $G_p$, and $p$ new points are obtained, which are recorded as group $g$. Repeat the above step 10 000 times to obtain 10 000 new groups, denoted as $g_1, g_2, \cdots, g_{10\,000}$. The meaning of $g_i$ is the simulation group consisting of the $p$ single points simulation points randomly sampled from $G_1, G_2, \cdots, G_p$ for the $i$-th time. For $g_i$, the frequency data is calculated at an interval of 0.01, and the probability density simulation group is obtained, denoted as $p_i$. Also, for $g_i$, the cumulative frequency data is calculated, and the cumulative distribution simulation group is obtained, denoted as $c_i$.

#### 2.1.3 Quantitative comparison

From 10 000 single-point simulation groups, $g_1, g_2, \cdots, g_{10\,000}$, randomly pick two groups. Treat each group as a high-dimensional vector, and calculate the Euclidean distance between the two vectors, so as to measure the proximity and similarity between the two vectors. Repeat the random sampling 10 000 times, and get 10 000 values of Euclidean distance, denoted as $d_{a-1}, d_{a-2}, \cdots, d_{a-i}, \cdots, d_{a-10\,000}$. The meaning of $d_{a-i}$ is the Euclidean distance calculated by the two groups randomly selected from the single point simulation group $g_1, g_2, \cdots, g_{10\,000}$ for the $i$-th time. Similarly, calculate the Euclidean distance between the probability density groups and between the cumulative distribution groups, denoted as $d_{b-i}$ and $d_{c-i}$, respectively. The obtained Euclidean distances were then subjected to statistical analysis. Through the calculation and comparison of statistics such as mean, standard deviation, maximum, minimum, and median, the stability of different pavement performance expressions is analyzed.

The data were normalized before calculating the Euclidean distance. The purpose is to prevent the dimension and order of magnitude of different pavement performance expressions from affecting the comparison of distances. All data were placed in the 0-1 interval using the Max-Min normalization method. The conversion method is as follows. For a sequence containing $n$ data, $x_1, x_2, \cdots, x_i, \cdots, x_n$,

$$y_i = \frac{x_i - \min\limits_{1 \leqslant j \leqslant n}\{x_j\}}{\max\limits_{1 \leqslant j \leqslant n}\{x_j\} - \min\limits_{1 \leqslant j \leqslant n}\{x_j\}}, \tag{1}$$

where $y_i$ is the transformed dimensionless standard value.

After normalizing the data, calculate the Euclidean distance. Let two $z$-dimensional vectors $\boldsymbol{x}_k = (x_1, x_k, \cdots, x_n)^{\mathrm{T}}$ and $\boldsymbol{y}_k = (y_1, y_k, \cdots, y_n)^{\mathrm{T}}$ represent two objects respectively, and their Euclidean distance is:

$$d(X,Y) = \| X - Y \| = (\sum_{k=1}^{n} |\boldsymbol{x}_k - \boldsymbol{y}_k|^2)^{\frac{1}{2}}. \tag{2}$$

## 2.2 Cumulative distribution-based performance function

Generally, for pavement performance indicators such as *PCI* and *RQI*, they are averaged from the amount of on-site observation pavement condition data. Because they are averaged, it is easy to compress the information and hide the local road conditions. The cumulative distribution is not only the most stable pavement method, but also effectively utilizes all detection data. This section presents a method for converting finite observational data into a continuous distribution function. First, the interpolation function is introduced. Then, the observation point data is inversely sampled to obtain the degradation function based on the continuous cumulative distribution. Finally, discuss the abnormal situations that occur in the process.

### 2.2.1 Interpolation unction

Using the interpolation function, the discrete pavement condition observed data can be interpolated to achieve the continuity of the cumulative distribution for pavement performance indicators so that the distribution proportion of any level of performance can be obtained, providing a large amount of data for the subsequent distribution prediction.

Interpolation is not the same as linear. A suitable interpolation formula or interpolation function needs to be selected, in order to calculate the specific value of the middle point of the two sampling points. Interpolated uncertainty can be viewed as an oscillation, that is, the rate of change between two points of a function. The Fourier transform reflects the frequency composition of the signal in the frequency domain. High frequencies are associated with fast oscillations, and low frequencies are associated with slow oscillations. To account for the uncertainty of the interpolation, specify the maximum frequency at which the function is allowed to oscillate. Then, removing the high frequency after Fourier transform is equivalent to removing the fast oscillation, that is, removing this uncertainty.

For a function $f(x)$ with limited bandwidth, if its Fourier transform is always zero outside a certain frequency band, as follows:

$$\mathcal{F} f(x) \equiv 0, \text{ for } |s| \geqslant \frac{p}{2}, \tag{3}$$

where, $p$ is called the bandwidth; $s$ represents any value outside the certain frequency band with $p$.

Then the sampling theorem and interpolation function are derived:

$$f(t) = \sum_{k=-\infty}^{\infty} f(\frac{k}{p}) \sin c[p(t - \frac{k}{p})],$$

$$\mathcal{F} f(s) \equiv 0, \text{ for } |s| \geqslant \frac{p}{2}, \tag{4}$$

$$\sin c(x)=\frac{\sin(\pi x)}{\pi x}, \tag{5}$$

where $\mathrm{sinc}(x)$ is sinc function.

For a function with a bandwidth of $p$, if the sampling interval is $1/p$, and the value of the sampling points $f(k/p)$, $k=0, \pm1, \pm2, \cdots$, are known. Then all the values of the original function $f(t)$ can be obtained by interpolation through this equation.

### 2.2.2 Pavement performance modeling

According to the sampling theorem, the range of the independent variable $x$ of the cumulative distribution function of pavement performance is set to $[S, 100]$, the bandwidth $P=100-S$, and the sampling interval is 0.1, that is, $1/p=0.1$, $p=10$. The number of sampling points is $\frac{100-S}{0.1}$.

Substitute the measured value of the cumulative distribution of pavement performance into Equation (4):

$$F(t)=F(0.1)\sin c[10\times(t-0.1)]+\cdots+F(100)\sin c[10\times(t-P)], \tag{6}$$

Let $t=x-S$, then

$$F(x)=F(S+0.1)\sin c[10\times(x-S-0.1)]+\cdots+F(100)\sin c[10\times(x-S-P)]. \tag{7}$$

Finally, the cumulative distribution of road performance in this section has been continuous. Any road performance index level is substituted into the function to solve numerically, and the proportion of road sections below the performance index level can be obtained.

### 2.2.3 Tail oscillation discussion

The observed pavement condition data and continuity function of the cumulative distribution for pavement performance are both plotted in Figure 2. As can be seen from the figure, except for the tail, the continuity function is almost coinciding with the observed value. It shows that the accuracy of modeling is high.

Discuss the tail concussion and whether it affects the serialization results:

(1) Whether the range of the independent variable is required to be infinite: The equation of the sampling theorem depends on the infinite sampling values of the

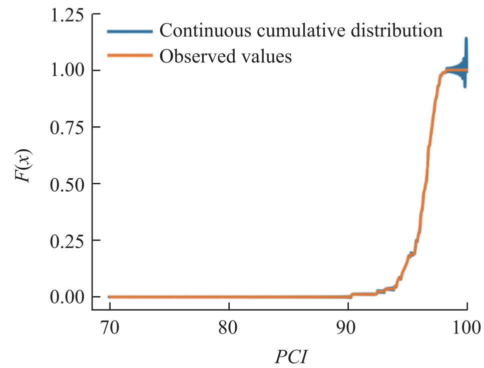



**Fig. 2 Comparison of the observed value and the continuous function**

function. According to these sampling values, the original function $f(t)$ can be obtained by Equation (4). However, in practical applications, only a limited number of sampled values can be obtained, which will lead to errors in calculation. To fully recover the signal, the sampling theorem states that the sampling rate must be at least twice the maximum bandwidth of the analog signal[31]. After the correction according to this theorem, the tail effect is still not eliminated, so this reason is ruled out.

(2) Applicable conditions are inconsistent: During the derivation of this equation, the sampled raw data must be a function of limited bandwidth. The frequency is zero outside a limited area. The sampled data, that is, the measured value of the cumulative distribution of road performance, is subjected to the fast Fourier transform to obtain a spectrogram (Figure 3): It can be seen that it does not meet the conditions.

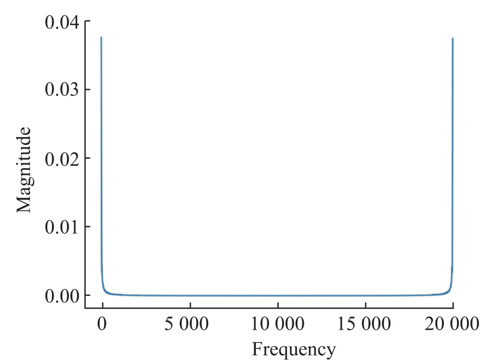



**Fig. 3 Spectrogram of pavement condition observed values after normalization**

According to the boundedness of the cumulative distribution, Equation (8), modify the continuous graph of the cumulative distribution of pavement performance:

$$\lim_{x \to 0} F_X(x) = 0 \text{ and } \lim_{x \to 100} F_X(x) = 1. \quad (8)$$

The continuous cumulative distribution of pavement performance after modification can be obtained as shown in Figure 4. It can be seen from the figure that the measured data and the continuous curve are almost completely fitted.

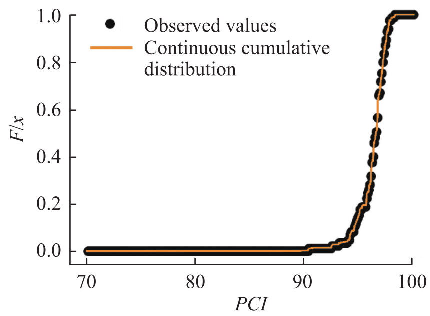



**Fig. 4 Continuous cumulative distribution function after correction**

## 2.3 Model evolution

In this part, the method to construct a cumulative distribution-based pavement performance prediction model is proposed. First, the rules of the cumulative distribution itself are introduced. And the rules presented when the cumulative distribution is applied to model the pavement performance are also pointed out. These rules will be used in subsequent predicting work. Second, using two means of sampling and grey theory, the prediction of the future cumulative distribution-based performance curve is presented. Finally, the discussion of the change of the cumulative distribution-based performance curve and the analysis of the situation of pavement performance deterioration are presented.

### 2.3.1 Cumulative distribution rules

For the continuous cumulative distribution function, it has two important rules as shown below:

(1) Boundedness. The value range of the pavement performance index is [0, 100]. As shown in Equation (6), when $x \to 100$, $X$ can take any value, then the probability of $X \leqslant x$ should be close to 1; when $x \to 0$, $X$ can take any value, then the probability of $X \leqslant x$ should be close to 0.

(2) Non-decreasing. $F_X$ is a monotone invariant function: if $y > x$, then $F_X(y) \geqslant F_X(x)$.

When the continuous cumulative distribution function is applied in describing pavement performance, it has an essential law in the case of pavement natural decay deterioration. As shown in Figure 5, the cumulative distribution curves of pavement performance at different times will not intersect. Because no matter how fast the degradation rate is, in the natural deterioration process, pavement performance of all road sections is deteriorating at the same time. It means the cumulative distribution value of one single point is non-decreasing.

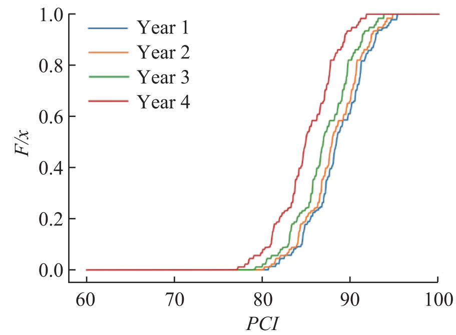



**Fig. 5 Pavement performance distribution curve under natural deterioration**

For example, on the $x$-axis $PCI = 85$, the ordinate represents the proportion of road sections with pavement performance less than 85. In the process of deterioration decay, the ordinate is always increasing, and no crossover occurs. Only when maintenance intervention is carried out, the proportion of road sections with pavement performance less than 85 will decrease, and the ordinate value becomes smaller, will they intersect.

### 2.3.2 Prediction technique

Actually, it is difficult to directly predict a new cumulative distribution-based function for the pavement performance over time. Because the distribution function does not follow simple laws to evolve. So high-density sampling of the function is still required. Specifically, it can be divided into four steps:

(1) Based on the observed data of pavement performance at different times, the continuous cumulative distribution-based deterioration functions at different times are established according to the method presented before.

(2) Sampling on the continuous cumulative distribution functions of each time in a very small interval, obtain multiple sets of time series data in different intervals.

(3) For each set of time series data, the short-term prediction of pavement performance is carried out based on grey theory, and the results are improved according to the rules of cumulative distribution described before.

(4) Reconvert the discrete data obtained after prediction into a continuous function by interpolation function. So far, a future cumulative distribution-based deterioration function is achieved.

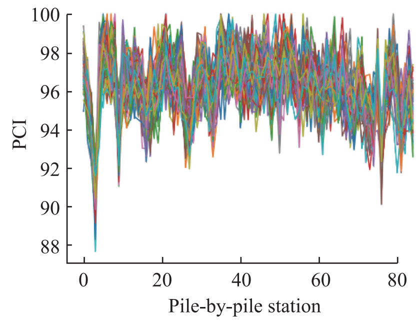
## 3 Results and discussion

### 3.1 Stability analysis

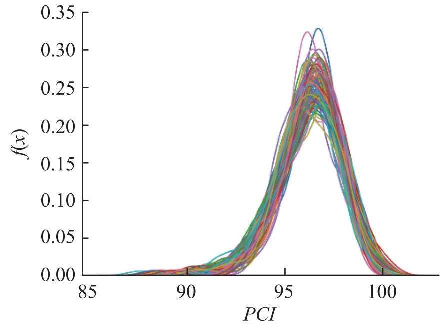
According to the the *PCI* and *RQI* data from the east to west direction observed on a section of an expressway located in Shanghai (China), set the standard deviation is 0.5, 1, and 2 respectively. Based on the field observed pavement condition data, obtain simulation random number groups, $G_1, G_2, \cdots, G_{85}$.

After random sampling, get 10 000 single point groups, probability density group, and cumulative distribution group respectively, and the supposed mean for *PCI* is 95.8. Plot $g_i$, $p_i$ and $c_i$ ($i$ from 1 to 10 000) respectively in Figure 6.

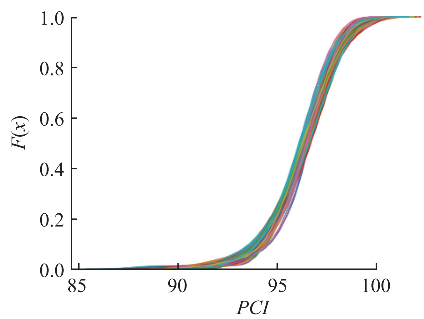
The Euclidean distance is calculated by randomly sampling from the simulation group. Obtain 10 000 $d_{a-i}$, $d_{b-i}$ and $d_{c-i}$, respectively. Statistical analysis was performed on the Euclidean distance results of three different pavement performance expressions. The results are summarized in Table 1, and the frequency plots are drawn in Figure 7.



(a) Single point group



(b) Probability density group



(c) Cumulative distribution group

**Fig. 6 Simulation Groups as three performance expressions**

**Table 1 Statistical analysis results of Euclidean distance**

| Index | Assumed $\sigma$ | Expressions | Mean | Standard deviation | Minimum | Median | Maximum |
|---|---|---|---|---|---|---|---|
| *RQI* | 0.5 | *a* | 0.876 8 | 0.133 4 | 0.566 0 | 0.853 6 | 1.732 1 |
| | | *b* | 0.311 7 | 0.132 7 | 0 | 0.290 3 | 0.921 7 |
| | | *c* | 0.077 0 | 0.027 4 | 0.011 8 | 0.073 5 | 0.217 6 |
| *RQI* | 1 | *a* | 1.532 6 | 0.214 3 | 0.992 7 | 1.506 6 | 2.795 4 |
| | | *b* | 0.459 7 | 0.151 5 | 0 | 0.450 5 | 1.130 8 |
| | | c | 0.107 6 | 0.037 1 | 0 | 0.102 6 | 0.284 1 |
| *RQI* | 2 | *a* | 2.475 6 | 0.277 3 | 0 | 2.472 0 | 3.484 5 |
| | | *b* | 0.716 2 | 0.177 6 | 0 | 0.713 5 | 1.431 2 |
| | | *c* | 0.147 8 | 0.054 6 | 0.028 8 | 0.139 2 | 0.446 3 |
| *PCI* | 1 | *a* | 0.751 7 | 0.072 1 | 0 | 0.746 8 | 1.127 3 |
| | | *b* | 0.750 9 | 0.174 5 | 0 | 0.742 3 | 1.451 7 |
| | | *c* | 0.108 5 | 0.027 2 | 0 | 0.105 9 | 0.248 2 |

Note: *a*-pile by pile single point; *b*-probability density; *c*-cumulative distribution

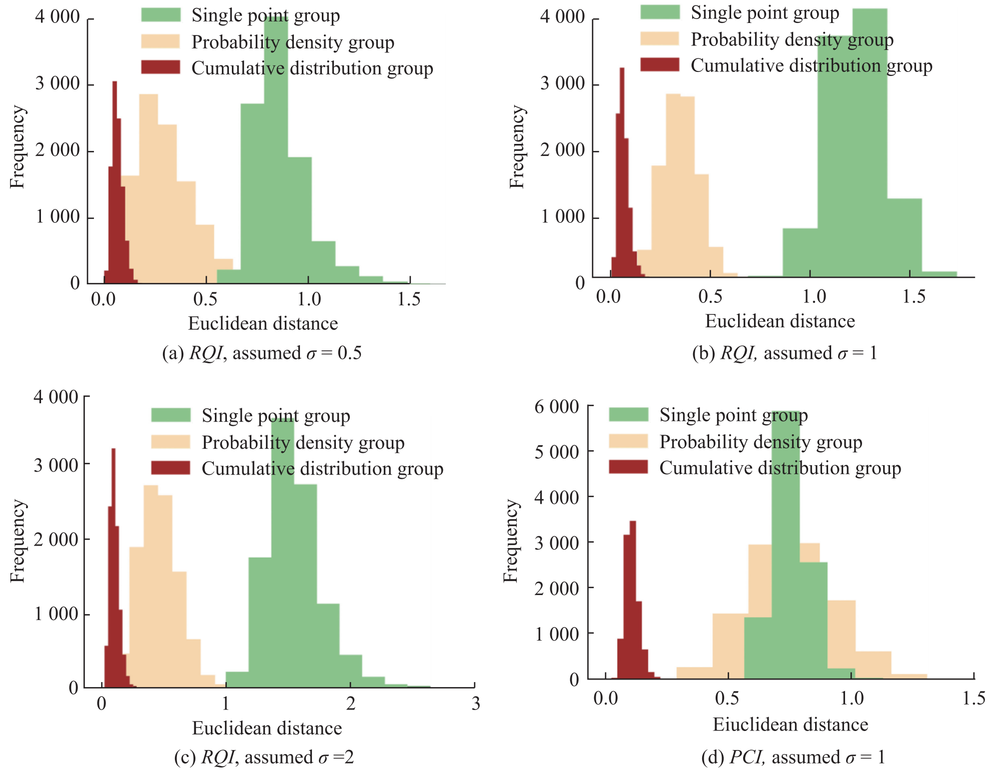



**Fig. 7 Euclidean distance for each pavement performance expressions**

It can be seen from Table 1 and Figure 7 that, regardless of the standard deviation value set during the simulation, for *RQI*, the Euclidean distance between the pile-by-pile vectors of the pavement performance index is larger than the other two expressions. This shows that the degree of similarity between the pile-by-post data is low, the variability is large, and the most unstable. The Euclidean distance between the cumulative distribution vectors is the smallest, the similarity between the data is high, and the variability is the lowest.

For *PCI*, the variability between the pile-by-pile expression and the distribution probability expression of pavement performance is not much different. But the Euclidean distance of the cumulative distribution expression is still much smaller than the other two. Therefore, the variability of the cumulative distribution is the lowest, its ability to resist the uncertainty of the detection data is the best, and it is the most stable way of expressing the pavement performance.

### 3.2 Model evolution

The cumulative distribution-based function of pavement performance is established respectively. And according to Equation (6), the tail oscillation is corrected. Plot four-year continuous cumulative distribution curve of pavement in Figure 8.

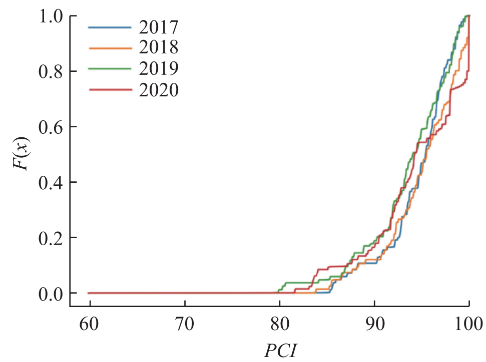



**Fig. 8 The four-year continuous cumulative distribution curve**

Secondly, with a minimum interval of 0.01, sampling is carried out on the continuous cumulative distribution function of the pavement performance each year. The sampling interval is (60,100], and the number of sampling points is 4 000. Thus, 4 000 sets of time series data are obtained, denoted as $I_1, I_2, \cdots, I_{4\,000}$. The meaning of $I_i$ is: when the pavement performance value is $(60+0.001i)$, the $y$ value of the period from 2017 to 2020 cumulative distribution function at this point are taken, shown as Equation(9).

$$I_i = \{I_{i-2017}, I_{i-2018}, I_{i-2019}, I_{i-2020}\} \tag{9}$$

Thirdly, short-term prediction of pavement performance based on grey theory is carried out for each set of time series data, as shown in Equation (10). Multiple groups of time series after prediction are obtained, noted as $\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_i, \cdots, \hat{Y}_{4\,000}$. The composition of $\hat{Y}_i$ is shown in Equation (11).

$$\frac{\mathrm{d}X_{(k)}^{(0)}}{\mathrm{d}t} + aX_{(k)}^{(1)} = u, \tag{10}$$

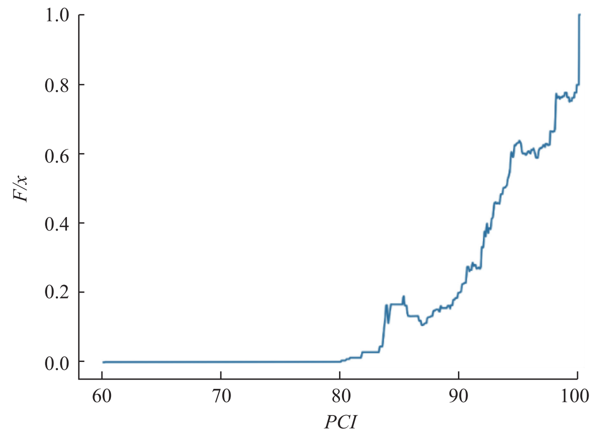
where, $a$ and $u$ are endogenous variables, namely undetermined coefficients; $X_{(k)}^{(0)}$ is raw data column; $X_{(k)}^{(1)}$ is the new data column obtained by accumulating the original data column.

$$\hat{Y}_i = \{\hat{Y}_{i-2017}, \hat{Y}_{i-2018}, \hat{Y}_{i-2019}, \hat{Y}_{i-2020}, \hat{Y}_{2021}\}. \tag{11}$$

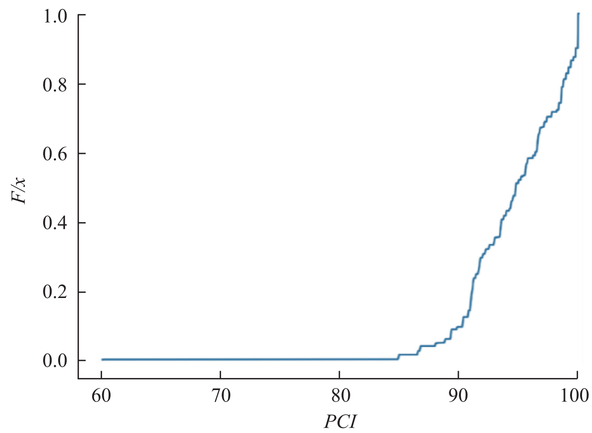
Calculate the mean square error ratio $c$ and small error probability $p$ of each group to test the precision. The mean of the mean square error ratio is 0.445 1 and the mean of the probability of small error is 0.888 3, both of which belong to the second level of precision, indicating that the prediction precision meets the requirements.

The predicted cumulative distribution of *PCI* in 2021 is plotted in Figure 9(a). It can be seen from the figure that the predicted cumulative distribution curve does not conform to the characteristics of the cumulative distribution itself. To revise the cumulative distribution curve in Figure 9(a), the value of the rightmost point (P1) of the cumulative distribution curve is first determined. It is then determined whether the value of its neighbouring point (left, P2) satisfies the non-decreasing characteristics of the cumulative distribution curve (i. e. the value of P2 is less than or equal to the value of P1). If so (i. e. the value of P2 is less than or equal to P1), the value of P2 is obtained, and if not (i. e. the value of P2 is greater than P1), the value of P1 is assigned to P2. And so on, gradually advancing the measured value to the left along the curve until the leftmost value was determined. The mean square error ($E_{\mathrm{MS}}$) and goodness of fit ($R^2$) were then carried out based on the revised *PCI* predicted data and the measured value to determine the validity of the fit results. The revised results are shown in Figure 9(b).

Based on the revised *PCI* cumulative distribution



(a) Before the correction



(b) After the correction

**Fig. 9 Prediction curve of pavement performance cumulative distribution in 2021**

data in 2021 and the measured cumulative distribution data, the mean square error and goodness of fit between the predicted value and the true value were performed, following the equations below.

$$E_{\mathrm{MS}} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, \tag{12}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}, \tag{13}$$

The results of $E_{\mathrm{MS}}$ and are 0.003 1 and 0.945 9 respectively. It can be seen from the results that the mean square error is close to 0, and the goodness of fit is close to 1, indicating that the predicted value fits well with the measured value.

Finally, Convert the corrected *PCI* discrete data to a continuous function through an interpolation function according to Equation (14).

$$F(x) = \hat{Y}(60.1)\sin c[10 \times (x - 60 - 0.1)] + \cdots + \hat{Y}(100)\sin c[10 \times (x - 60 - 40)]. \quad (14)$$

The cumulative distribution of pavement performance from 2017 to 2021 is plotted in Figure 10. It can be found from the figure that the cumulative distribution curve of the measured pavement performance is at a righter position than the predicted cumulative distribution curve. At the same time, compared with the road performance distribution curve in 2020, the measured value in 2021 is shifted to the right. This shows that the measured road performance level in 2021 is not only better than the predicted level but also better than the 2020 road performance level. Combined with the analysis of maintenance data, maintenance measures will be implemented on some road sections in 2021, especially the road sections with a pavement performance level lower than 85, resulting in an improvement in the pavement performance level.

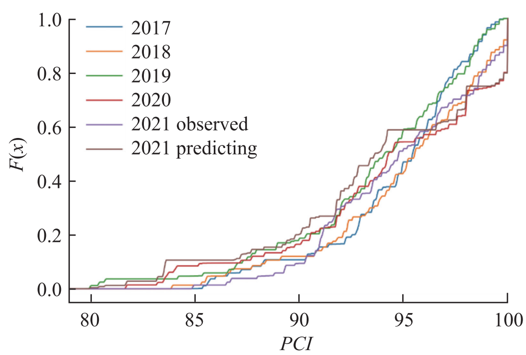



**Fig. 10 Comparison of observed data and prediction curve**

The cumulative distribution curve of the pavement obtained through the prediction, because the input time series of the pavement performance change over the years shows more of a decaying trend, the prediction is also more focused on the result of the pavement performance decay without maintenance intervention. Therefore, there will be some differences between the predicted results and the obvious improvement of the pavement performance level in 2021 under the maintenance intervention, and it also shows the effectiveness of the maintenance activities implemented in 2021.

The cumulative distribution curves of pavement performance at different points do not intersect in the case of natural decay of the pavement. It is because all sections are decaying during the natural decay process, regardless of the speed of decline. It means that the cumulative distribution of values at a single point is non-decreasing. However, due to maintenance activities, the rate of decay of pavement performance is altered, increasing the proportion of good sections, so that a number of cumulative distribution curves will be closer to each other around a certain value, which is defined as the congestion area. As shown in Figure 11, the congestion area occurs at 93, with the left side of the congestion area representing the percentage of *PCI* less than 93 and the right side of the congestion area indicating the percentage of *PCI* greater than 93. The further to the right of the congestion area, the more effective the conservation activity is.

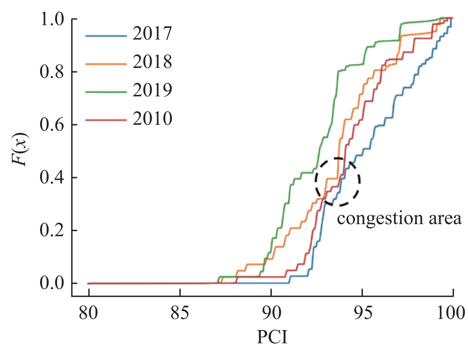



**Fig. 11 Example of a congestion area**

For natural deteration:

(1) The cumulative distribution curve of pavement performance at different times has a congestion area (an example is shown in Figure 11) in the horizontal direction, indicating that the decay rates of different road sections are different, and the pavement performance is still controllable.

(2) The cumulative distribution curves of pavement performance at different times move in the horizontal direction as a whole, and there is no congestion area, indicating that the decay of all road sections has entered a stage of rapid overall development and needs to be maintained. It is the judgment-point of maintenance time.

### 3.3 After maintenance

(1) The cumulative distribution function of the pavement performance after maintenance will move to the right, which can be used to qualitatively judge the effectiveness of the maintenance measures.

(2) One year after maintenance, if there is a congestion area, the effectiveness of the maintenance measures can be qualitatively judged. For quantitative evaluation, the area between the curves can be calculated.

## 4 Conclusions

This study introduces a novel approach for predicting pavement performance utilizing cumulative distribution analysis. The key findings of this investigation are outlined as follows:

(1) The study compares various representations of pavement performance, including pile-by-pile single-point data, probability density data, and cumulative distribution data. It employs the Euclidean distance metric to quantitatively assess the stability of these different expressions. The results reveal that the cumulative distribution offers superior stability in evaluating pavement performance, preserving the granularity of road section information, thereby facilitating more targeted maintenance strategies.

(2) Addressing the challenge of describing the distribution with a unified equation, this research proposes a discrete facts modeling approach for pavement performance indicators, leveraging the sampling theorem. This approach enables the conversion of limited, discrete measured data into a continuous representation, utilizing sampling functions. This not only allows for the determination of the proportion of road sections at various levels of pavement performance but also provides the necessary foundation for extensive data requirements in distribution prediction, serving as the cornerstone for evaluation and prediction methodologies.

(3) An analysis of the cumulative distribution curve of pavement performance, considering both natural decay and maintenance interventions at different time points, demonstrates that the evolution of the distribution offers valuable insights into the necessity and effectiveness of maintenance measures, as well as the timing of such interventions. This analysis underscores the practical utility of the proposed method in guiding pavement management decisions.

## Acknowledgement

This work was supported by 2022 Ningbo Transportation Science and Technology Plan Project (Grant No. 202216).

## Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

### References

[1] HONG F, PROZZI J A. Estimation of Pavement Performance Deterioration Using Bayesian Approach [J]. Journal of Infrastructure Systems, 2006, 12 (2): 77-86.

[2] DONEV V, HOFFMANN M. Condition Prediction and Estimation of Service Life in the Presence of Data Censoring and Dependent Competing Risks [J]. International Journal of Pavement Engineering, 2019, 20 (3): 313-331.

[3] REMEK L, MIKOLAJ J, RUTTKAY L. Pavement Performance Prediction as a Key to Sustainable Road Infrastructure Construction [C] //15th International Multidisciplinary Scientific Geoconference (SGEM). Albena, Bulgaria: Curran Associates, Inc. , 2015.

[4] ABAZA K A, ASHUR, S A. Optimum Microscopic Pavement Management Model Using Constrained Integer Linear Programming [J]. International Journal of Pavement Engineering, 2009, 10 (3): 149-160.

[5] OUYANG Y, MADANAT S. Optimal Scheduling of Rehabilitation Activities for Multiple Pavement Facilities: Exact and Approximate Solutions [J]. Transportation Research Part a-Policy and Practice, 2004, 38 (5): 347-365.

[6] LYTTON R L. Concepts of Pavement Performance Prediction and Modelling [M]. 2nd ed. Toronto, Canada: North American Conference on Managing Pavements, 1987.

[7] HAAS R, HUDSON W R, ZANIEWSKI J P. Modern Pavement Management. Malabar [M]. India: Krieger Publishing Company, 1994.

[8] AASHTO. Pavement Management Guide [S]. 2nd ed. Washington, DC: AASHTO, 2012.

[9] UDDIN. Pavement Management Systems [M] //FWA T F. The Handbook of Highway Engineering. Boca Raton: Taylor & Francis, 2005.

[10] JUSTO-SILVA R, FERREIRA A, FLINTSCH G. Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models [J]. Sustainability, 2021, 13 (9): 5248. .

[11] ABAZA K A. Back-calculation of Transition Probabilities for Markovian-based Pavement Performance Prediction Models [J]. International Journal of Pavement Engineering, 2016, 17 (3): 253-264.

[12] ABAZA K A. Empirical Markovian-based Models For Rehabilitated Pavement Performance Used in a Life Cycle Analysis Approach [J]. Structure and Infrastructure Engineering, 2017, 13 (5): 625-636.

[13] PÉREZ-ACEBO H, MINDRA N, RAILEAN A, et al. Rigid Pavement Performance Models by Means of Markov Chains with Half-year Step Time [J]. International Journal of Pavement Engineering, 2019, 20 (7): 830-843.

[14] PÉREZ-ACEBO H, GONZALO-ORDEN H, FINDLEY D J, et al. Modeling the International Roughness Index Performance on Semi-rigid Pavements in Single Carriageway Roads [J]. Construction and Building Materials, 2021, 272: 121665.

[15] PÉREZ-ACEBO H, MONTES-REDONDO M, APPELT A, et al. A Simplified Skid Resistance Predicting Model for a Freeway Network to be Used in a Pavement Management System [J]. International Journal of Pavement Engineering, 2023, 24 (2): 2020266.

[16] AASHTO Guide. Guide for Design of Pavement Structures [M]. Washington, DC: American Association of State Highway and Transportation Officials, 1993.

[17] SUN Li-jun, LIU Xue-ping. General Deterioration Equation for Pavement Performance [J]. Journal of Tongji University, 1995, 23 (5): 512-518. (in Chinese)

[18] WANG K C P, LI Q. Pavement Smoothness Prediction Based on Fuzzy and Gray Theories [J]. Computer-Aided Civil and Infrastructure Engineering, 2011, 26 (1): 69-76.

[19] LI N, HAAS R, XIE W C. Development of a New Asphalt Pavement Performance Prediction Model [J]. Canadian Journal of Civil Engineering, 1997, 24 (4): 547-559.

[20] OSORIO-LIRD A, CHAMORRO A, VIDELA C, et al. Application of Markov Chains and Monte Carlo Simulations for Developing Pavement Performance Models for Urban Network Management [J]. Structure and Infrastructure Engineering, 2018, 14 (9): 1169-1181.

[21] ATTOH-OKINE N O. Grouping Pavement Condition Variables for Performance Modeling Using Self - Organizing Maps [J]. Computer-Aided Civil and Infrastructure Engineering, 2001, 16 (2): 112-125.

[22] MARCELINO P, de LURDES ANTUNES M, FORTUNATO E, et al. Machine Learning Approach for Pavement Performance Prediction [J]. International Journal of Pavement Engineering, 2021, 22 (3): 341-354.

[23] ROSE S, MATHEW B S, ISAAC K P, et al. Risk Based Probabilistic Pavement Deterioration Prediction Models for Low Volume Roads [J]. International Journal of Pavement Engineering, 2018, 19 (1): 88-97.

[24] HONG F, PROZZI J A. Roughness Model Accounting for Heterogeneity Based on In-service Pavement Performance Data [J]. Journal of Transportation Engineering, 2010, 136 (3): 205-213.

[25] YAMANY M S, SAEED T U, VOLOVSKI M, et al. Characterizing the Performance of Interstate Flexible Pavements Using Artificial Neural Networks and Random Parameters Regression [J]. Journal of Infrastructure Systems, 2020, 26 (2): 04020010.

[26] LI Jie, CHEN Jian-bing. The Probability Density Evolution Method for Analysis of Dynamic Nonlinear Response of Stochastic Structures [J]. Acta Mechanica Sinica, 2003, 35 (6): 722-728. (in Chinese)

[27] CHEN J B, LI J. The Extreme Value Distribution and Dynamic Reliability Analysis of Nonlinear Structures with Uncertain Parameters [J]. Structural Safety, 2007, 29 (2): 77-93.

[28] ABED A, THOM N, NEVES L. Probabilistic Prediction of Asphalt Pavement Performance [J]. Road Materials and Pavement Design, 2019, 20 (sup1): S247-S264.

[29] YAMANY M S, ABRAHAM D M. Hybrid Approach to Incorporate Preventive Maintenance Effectiveness into Probabilistic Pavement Performance Models [J]. Journal of Transportation Engineering, Part B: Pavements, 2021, 147 (1): 04020077.

[30] JTG 5210—2018, Highway Performance Assessment Standards [S]. (in Chinese)

[31] ZHA Z, WEN B, YUAN X, et al. Learning Nonlocal Sparse and Low-Rank Models for Image Compressive Sensing: Nonlocal Sparse and Low-rank Modeling [J]. IEEE Signal Processing Magazine, 2023, 40 (1): 32-44.