SCIENCE CHINA
Technological Sciences

Special Topic: Neurodynamics and Network Science
• Article •

July 2022 Vol. 65 No. 7: 1470–1481
https://doi.org/10.1007/s11431-021-1972-7

CPSO-XGBoost segmented regression model for asphalt pavement deflection basin area prediction

LI ZhuoXuan[1], SHI XinLi[2], CAO JinDe[1,3*], WANG XuDong[4] & HUANG Wei[5]

[1] *School of Mathematics, Southeast University, Nanjing 210096, China;*
[2] *School of Cyber Science & Engineering, Southeast University, Nanjing 210096, China;*
[3] *Yonsei Frontier Lab, Yonsei University, Seoul 120-749, Korea;*
[4] *Research Institute of Highway, Ministry of Transport, Beijing 100088, China;*
[5] *Intelligent Transportation System Research Center, Southeast University, Nanjing 210096, China*

Received August 18, 2021; accepted November 29, 2021; published online June 9, 2022

The use of non-destructive testing (NDT) equipment, such as the falling weight deflectometer (FWD), provides important estimates of road health and helps to optimize road management regimes. However, periodic road testing and post-processing of the collected data are cumbersome and require much expertise, a considerable amount of time, money, and other resources. This study attempts to develop a reliable prediction method for estimating the deflection basin area of different asphalt pavements using road temperature, load time, and load pressure as main characteristics. The data are obtained from 19 kinds of asphalt pavements on a 2.038-km-long full-scale field accelerated pavement testing track named RIOHTrack (Research Institute of Highway Track) in Tongzhou, Beijing. In addition, a chaotic particle swarm algorithm (CPSO) and a segmented regression strategy are proposed in this paper to optimize the XGBoost model. The experiment results of the proposed method are compared with those of classical machine learning algorithms and achieve an average of mean square error and mean absolute error respectively by 5.80 and 1.59. The experiments demonstrate the superiority of the XGBoost algorithm over classical machine learning methods in dealing with nonlinear problems in road engineering. Significantly, the method can reduce the frequency of deflection tests without affecting its estimation accuracy, which is a promising alternative way to facilitate the rapid assessment of pavement conditions.

FWD, pavement management systems, RIOHTrack, deflection basin area, CPSO, XGBoost, segmented regression model

Citation: Li Z X, Shi X L, Cao J D, et al. CPSO-XGBoost segmented regression model for asphalt pavement deflection basin area prediction. Sci China Tech Sci, 2022, 65: 1470–1481, https://doi.org/10.1007/s11431-021-1972-7

1 Introduction

Pavement structural damage has always affected road service life. With the increase of traffic volume and vehicle axle load, developing asphalt pavement with long service life has been a serious challenge faced by researchers. Long-life asphalt pavement becomes an important pavement development trend, due to fatigue cracking, permanent deformation, and other structural damage of the traditional asphalt pavement. Long-life asphalt pavement refers to the asphalt pavement with the designed service life of 40 to 50 years whose characteristic is that the damage of the pavement only occurs on the surface during the service period, and no structural damage will appear [1]. The pavement performance can be guaranteed through periodic maintenance. Under relatively heavy traffic conditions, long-life asphalt pavement helps to alleviate frequent pavement reconstruction, thus reducing the cost of pavement maintenance [2, 3]. Pavement evaluation can be divided into four aspects: pavement function evaluation, pavement structure ability evaluation, pavement damage evaluation, and safety evaluation. As an important index to evaluate the bearing capacity of asphalt pavement structures, the surface deflection value is vital to detect and evaluate the

*Corresponding author (email: jdcao@seu.edu.cn)

health condition of asphalt pavement [4].

The detection methods used to evaluate the pavement structure include destructive testing and non-destructive testing (NDT) [5]. Destructive testing not only causes man-made damage to the original pavement surface but also consumes time and energy. Meanwhile, the limitations of the field sampling test lead to the deviation between test results and the original material. The method of NDT is simple and fast as the main basis of pavement structure design and strength evaluation for a long time. Therefore, non-destructive testing methods based on deflection such as alling weight deflectometer (FWD) have become an important topic in evaluating pavement structural performance.

Data of the FWD test are used to back-calculate the modulus of pavement structure layer for further analysis of the pavement structure [6]. However, FWD test experiments require a lot of time and resources. In addition, the practice of frequent data collection is really time-consuming, disruptive to traffic, and expensive, making it difficult to repeat. These may be some common reasons for ignoring structural aspects of pavements when choosing maintenance or repair decisions. Therefore, to cope with these limitations, many efforts have been made by researchers to find relatively fast alternative analysis methods and obtain a comprehensive understanding of the asphalt pavement deflection basin area. Recently, advanced learning algorithms have been introduced to model engineering applications. Artificial intelligence (AI) techniques, such as expert systems, artificial neural networks (ANN), genetic algorithms (GA), and hybrid systems, provide reasonable approximations for analyses dealing with the complex interrelationships of traffic loads and climatic conditions.

As experiments at RIOHTrack are very expensive, a data-driven approach to high precision simulations is essential. The study in this paper aims at developing and designing a hybrid model called CPSO-XGBoost segmented regression model for predicting deflection basin area of 19 kinds of asphalt pavements with high precision. The model in this paper is first proposed and used for deflection basin area prediction. We propose a segmented regression strategy based on the experimental operation of RIOHTrack. For the task of deflection basin prediction, the CPSO algorithm is used to optimize the whole model and jump out the local optimum by chaotic mapping to obtain higher accuracy. Significantly, the method can reduce the frequency of deflection tests without affecting its estimation accuracy under the current experimental design of RIOHTrack, which is a promising alternative way to facilitate the rapid assessment of pavement conditions. We compare the proposed CPSO-XGBoost segmented regression model with several other methods including the k-nearest neighbor model (KNN) and XGBoost. The performance of these models is statistically evaluated and validated on a dataset from the Institute of Highway Science, Ministry of Transportation, Beijing [7]. The contributions are summarized as follows.

- This is the first time that the XGBoost algorithm is applied to the prediction problem of deflection basins. It outperforms mainstream machine learning algorithms such as random forest (RF), KNN, and support vector regression (SVR) in terms of mean square error (MSE) and mean absolute error (MAE), indicating that the XGBoost algorithm has good performance in the nonlinear fitting. The performance and computational speed are also better than those of the compared algorithms.
- For the characteristics of the measured data of the deflection basin, a segmentation regression strategy is proposed and the rate of load change is used as the segmentation point.
- Optimization of the XGBoost algorithm using the CPSO algorithm combined with the segmented regression strategy can improve the accuracy by more than 1200% compared with the unoptimized one, and more than 10% compared with the PSO algorithm combined with the segmented regression strategy.
- Combining the AI algorithm with the prediction problem of the deflection basin area can increase the prediction accuracy and reduce the number of deflection experiments in the future, and the deflection basin area of the road condition with high accuracy.

The rest of this paper is organized as follows. In Sect. 2, we review the related work. In Sect. 3, the fundamental algorithms are introduced for our work. In Sect. 4, the CPSO and a segmented regression model are proposed to improve the XGBoost for deflection basin area prediction and describe the complete prediction algorithm. In Sect. 5, experimental results are discussed and the prediction performance of our algorithm is analyzed. In Sect. 6, the conclusion of our work is drawn and the future work prospected.

2 Related work

Several NDT methods such as ultrasonic pulse-echo, ground-penetrating radar, infrared differential thermal, and FWD have been proposed to analyze the stress-strain response of asphalt pavement [8–11]. The authors use pavement deflection measurements under FWD loading for a variety of purposes. Gedafa et al. [12] attempt to propose a nonlinear model to predict cracking of flexible pavements by relating the deflection basin data under FWD loading with parameters such as condition survey data, effective structural number, layer thickness, and equivalent axle load. Similarly, Garbowski et al. [13] attempt to relate deflections under FWD load by different models with parameters such as thicknesses of the model layers, layer material properties, and cross-anisotropy in all layers. Furthermore, Habbouche et al. [14] design the mechanistic-empirical (ME) rehabilitation to evaluate the pavement condition in the field also requires an FWD test. Ma et al. [15] propose a new smart pavement model for conventional drop-weight deflectometer that can only characterize the overall load-carrying capacity of the pavement, in which built-in sensors are incorporated to monitor the pavement stress and strain response under aircraft loading. In 2015, the Institute of Highway Science, Ministry of Transportation constructed a 2.038 km long full-size field accelerated pavement test track named RIOHTrack to develop a long-life pavement suitable for China. Zhang et al. [16] perform statistical analysis of stress/strain, rutting depth, etc. for the experimental data obtained from RIOHTrack and summarized the advantages of different types of pavement structures. Wang et al. [17] use sensor detection technology to explore the actual usage environment and mechanical response law of asphalt pavement structures. The influence laws of external factors such as temperature and load on the mechanical response are analyzed to reveal the temperature and load dependence of the real mechanical response. The actual structural mechanical response of the pavement required to establish the calculation system and analysis theory of asphalt pavement is obtained.

Recently, advanced learning algorithms have been introduced to model pavement engineering applications [18–22]. Fakhri and Shahni Dezfoulian [23] provide a satisfactory correlation between international roughness index (IRI), pavement surface evaluation and rating index (PASER), and structural indices based on deflection measurements by ANN and regression models. Huang et al. [4] evaluate the statistical characteristics of random variables in asphalt pavements based on field data collected from asphalt pavement projects. Based on the numerical simulation results, unique formula with fairly good accuracy is found using multiple linear regression to facilitate the calculation of pavement deflection in reliability analysis. Yang and Deng [24] use KNN to model crack type and crack width information of chemically stabilized asphalt pavement from FWD test data. Hussain et al. [25] use the statistical techniques of multiple nonlinear regression and ANN to effectively model the asphalt pavement analyzer rut distress. New aggregate indices designated as "aggregate source index" and "aggregate gradation index" are developed. Karballaeezadeh et al. [26] consider that FWD and ground-penetrating radar are expensive tests. Besides, the back-calculation methods have some inherent disadvantages compared with the exact methods because they use a trial and error approach. Based on these two points, random forest algorithm cups are used to predict the number of structures in flexible pavements and tested with 759 flexible pavement sections in the Semnan and Khuzestan provinces of Iran. This study shows that using machine learning methods instead of backward computation can improve the quality and accuracy of the computational process. Machine learning algorithms provide a data-driven approach for pavement engineering research, which can effectively reduce the error of pavement back-calculation methods and improve the calculation efficiency and accuracy, but they need the support of a large amount of actual measurement data.

In addition, based on the research of machine learning in pavement engineering, some researchers have proposed fusion models combining evolutionary algorithms and machine learning models to achieve better results. Xu et al. [27] propose an optimization method combining orthogonal experimental design, ANN, and GA. The steel-epoxy asphalt pavement structure of the Sutong Yangtse River Bridge is optimized by this method. The method is confirmed to improve fatigue reliability. Li and Wang [28] combine an ANN program with the GA to predict the surface deflection response of asphalt pavements under FWD loading. The program optimized with the GA is trained and validated using a synthetic database. The soft computing model shows better prediction accuracy than traditional multivariate regression methods. Zhang and Ji [30] also use ANN-GA to back-calculate the flexible pavement layer modulus from the FWD test with certain advantages, such as elimination of seed modulus and consideration of complex material properties. More importantly, the back-calculated pavement layer parameters can be directly used in the mechanical-empirical design of pavement overlays. Li and Wang [29] propose a new gray model-based method (GM(1,1|sin)) to predict the roughness of pavements. The particle swarm optimization (PSO) algorithm is used to select the best parameters for the GM(1,1|sin) model. The method uses only historical IRI data for prediction, thus saving the cost of collecting a large amount of pavement condition data. Kaloop et al. [31] focus on the development of a particle swarm optimization-based extreme learning machine (PSO-ELM) to predict the performance of stabilized aggregate bases subjected to wet and dry cycles. PSO-ELM not only possesses higher accuracy but also lies in the fact that the predicted resilient modulus (M_r) usually produces the same distribution and trend as the observed M_r. Liang et al. [32] develop a new multi-objective PSO (MOPSO) algorithm that uses a Gaussian process regression (GPR) with a machine learning approach to solve asphalt mixture ratio design. In the optimization step, the metaheuristic algorithm based on adaptive weight MOPSO (AWMOPSO) is used to achieve the global optimal solution. After optimization by the proposed GPR-AWMOPSO algorithm, the comprehensive performance of the pavement is improved in terms of high-temperature permanent deformation resistance, low-temperature crack resistance, and moisture stability. Evolutionary algorithms have good applications in optimization problems. They are combined with machine learning algorithms to improve the prediction accuracy of machine learning algorithms for problems related to pavement engineering. An automated and intelligent method for asphalt pavement proportioning design is also provided. This method is more effective for the study of pavement engineering compared with the traditional method.

Different from the above work, this paper develops a high-precision model for predicting the basin area of deflection under FWD testing for asphalt pavements, which can simulate the basin area data of deflection in FWD testing of asphalt pavement roads conducted on RIOHTrack at the Institute of Highway Science, Ministry of Transportation, Beijing. The model provides a method to solve the problem of high cost and long cycle time associated with FWD testing, and provides highly accurate, credible experimental fit data for asphalt pavement research.

3 Fundamental algorithms: XGBoost and CPSO

3.1 XGBoost

The XGBoost algorithm [33] is an integrated learning algorithm based on the idea of Boosting. It is developed based on the gradient boosting decision tree (GBDT) algorithm [34]. It is not only improved in speed, but also in the improvement of accuracy. Specifically, the XGBoost algorithm performs a second-order Taylor expansion of the cost function by introducing regularization terms to avoid overfitting. The integrated tree model is

$$\hat{y}_i = \theta(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i),\ f_k \in F, \tag{1}$$

where K is the total number of sub-models; $F = \{f|f(\mathbf{x}) = \mathbf{w}_{q(\mathbf{x})}\}$ is the set of all regression trees, and $\mathbf{w}_{q(\mathbf{x})}$ is the weight vector composed of the weights of all leaf nodes of the regression tree; $\hat{y}_i$ is the sample prediction value; $\mathbf{x}_i$ is the sample input feature; f_k is the k-th regression tree, each regression tree has a separate leaf weight w and tree structure q. The objective function is introduced as

$$O = l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{2}$$

in which $l(y_i, \hat{y}_i)$ is the loss function, representing the difference between the predicted value $\hat{y}_i$ and the actual value y_i; Ω is the regularization term, used to smooth the final learning weight to avoid overfitting. By carrying out multiple iterations through the addition strategy, defining $\hat{y}_i^t$ as the predicted value of sample i in the t-th iteration, synthesizing the loss function and regularization term, and combining eqs. (1) and (2) , which can deduce that

$$\begin{aligned} &O^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)) + \Omega(f_t), \\ &g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y^{(t-1)}}),\ h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}), \end{aligned} \tag{3}$$

in which N is the number of trees. The main implementation process is organized as follows:

(1) establish a new decision tree;

(2) calculate g_i and h_i of each training sample according to the objective function shown in eq. (3) and start iteration;

(3) use the approximate greedy algorithm to find the best split point to obtain the decision tree structure $f_t(\mathbf{x})$, where t represents the t-th iteration;

(4) add $f_t(\mathbf{x})$ to the integrated tree model;

(5) follow steps (1)–(4) to perform multiple iterations to obtain the final classification model.

3.2 CPSO

The PSO algorithm is a robust metaheuristic technique proposed by Eberhart and Kennedy (1995) based on the behavior of the particles/social animals, like birds in a swarm [35]. To find the optimal solution, the operation process of the PSO algorithm is summarized [36].

In the initial period of the algorithm, n particles are first randomly initialized in the D-dimensional search space. After k iterations, the position $(x_{j1}^k, x_{j2}^k, ..., x_{jD}^k)$ and speed $(v_{j1}^k, v_{j2}^k, ..., v_{jD}^k)$ of the particle j move in the search space at a certain speed. Meanwhile, the individual extreme value $p_{l\text{best}}^k$ and group extreme value $p_{g\text{best}}^k$ are recorded, and then the states of the population are updated. In each iteration, the velocity and position of particle j can be updated as

$$v_{jd}^{k+1} = \omega v_{jd}^k + c_1 r(p_{l\text{best}}^k - x_{jd}) + c_2 R(p_{g\text{best}}^k - x_{jd}), \tag{4}$$

$$x_{jd}^{k+1} = x_{jd}^k + v_{jd}^k, \tag{5}$$

where v_{jd}^k is the velocity of the j-th particle in the d-dimensional of the search space at the k-th iteration, v_{jd}^{k+1} is the velocity at the $(k+1)$-th iteration, x_{jd}^k is the position of the j-th particle in the d-dimensional of the search space at the k-th iteration, x_{jd}^{k+1} is the position at the $(k+1)$-th iteration, ω is the inertial weight, c_1 and c_2 are acceleration factors, r and R are random numbers on [0,1], $p_{l\text{best}}^k$ is the historical optimal position of the j-th particle at the k-th iteration, $p_{g\text{best}}^k$ is the historical optimal position of the population at the k-th iteration.

In PSO, the setting of the inertial weight affects the particle's exploration and search capabilities, and thus plays a vital role in the performance of the algorithm. By adjusting ω, the global and local search capabilities of particles can be adjusted to achieve a balance. When ω is large, the particle has strong global searchability. Otherwise, the particle has strong local searchability. The setting of the inertia weight usually adopts a strategy of linearly decreasing within a certain interval as the number of iterations increases, and hence the calculation formula is given as follows:

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \frac{k}{\text{maxgen}}, \tag{6}$$

where $\omega_{\max}$ and $\omega_{\min}$ are the upper and lower limits of the inertial weight, respectively, k is the current iteration number and maxgen is the maximum iteration number.

In fact, the value of this linearly decreasing method can be relatively large at the initial stage. During this period, the global search ability of the particles is strong, while the local search ability is relatively weak, which is beneficial to quickly locate the approximate position of the optimal solution. However, as the number of iterations increases, the value of ω gradually decreases, and the population convergence rate becomes slower. At this stage, the local search ability of the particles is stronger, and the global search ability is weaker because the particles have only the best flight position and population to their own. The characteristics of the flight position learning of the best particles in the algorithm make the algorithm prone to the premature phenomenon. By simply relying on linear transformation, the weight of particles can not be accurately adjusted due to the diversity of the population [37].

Then, when the algorithm falls into the local optima, the population particles gather around the local optimal position and repeat the similar optimization trajectory, making it difficult to escape the local optima [38–40]. Chaos is a nonlinear natural phenomenon, which has the characteristics of randomness, ergodicity, etc., and can be searched for optimization [41, 42]. This paper draws on the characteristics of randomness and ergodicity of chaotic phenomena and uses them to escape the local optima.

The logistic equation of a commonly used chaos model is given as follows:

$$z_{n+1} = \mu z_n(1 - z_n), n = 0, 1, 2, ..., \tag{7}$$

where μ is the control parameter. When $0 \le z_0 \le 1$, $\mu = 4$, logistic equation is in a completely chaotic state. Eq. (7) is an evolutionary form of the following logistic equation [43]:

$$cx_i^{t+1} = 4cx_i^t(1 - cx_i^t), i = 0, 1, 2..., \tag{8}$$

$$cx_i = (x_i - a)/(b - a), \tag{9}$$

$$x_i^{'} = a + cx_i(b - a), \tag{10}$$

in which cx_i^t is the value of chaos variable cx_i after the t-th iteration. When $x_i \in [0, 1]$ and $cx_i \in \{0.25, 0.5, 0.75\}$, chaos will occur. The solution variable $x_i \in [a, b]$ can be mapped back and forth with cx_i by eqs. (9) and (10), where a and b are the maximum and minimum values of the solution space.

Based on the characteristics of randomness and ergodicity of chaotic phenomena, this paper proposes a method to jump out the local optima based on chaos optimization. By comparing the number of successive iterations SG that the best position of the population is not updated with the threshold $SG_{\max}$, it can determine whether the algorithm falls into the local extreme value. If $SG > SG_{\max}$, the algorithm is considered to have fallen into a local optimum. When the algorithm is judged to be trapped in a local optimal position $\mathbf{G}$, it will first use eq. (9) to map $\mathbf{G}$ to the domain of chaos variable as $[0, 1]$, and then use eq. (8) to perform iterative operations to obtain m chaos position $(G_1^c, G_2^c, ..., G_m^c)$, and finally use eq. (10) to perform inverse mapping to obtain the n' new extreme position $(G^{1'}, G^{2'}, ..., G^{n'})$. Since the particles complete self-updating by chasing the extreme positions of individuals and groups, when the algorithm falls into the local optima, the position of the group extreme value must be at the local extreme value position. The new extreme value position of the group combined with eq. (4) can change the trajectory of the particle optimization. When the i-th particle self-updates by chasing the new extreme value position of the new group, it can be performed to search for new neighbor areas and paths outside the neighborhood of the local optima. Therefore, a better solution can be found with a higher probability, thereby increasing the possibility of the algorithm escaping the local extremum.

In general, MSE can be used as the fitness function to evaluate the degree of evolution of particles which is listed

$$E = \frac{\sum_{i=1}^{M}(Y_i - y_i^*)^2}{M}, \tag{11}$$

where M represents the number of samples, Y_i is the actual value of the sample data, and y_i^* is the predicted value of the sample.

4 CPSO-XGBoost segmented regression model in asphalt pavement deflection basin area prediction

4.1 Segmented regression model

The deformation process of the whole process of the deflection basin area in the same cycle is inconsistent at different periods, and one or one type of digital model is not described. It is inevitable to use several different types of digital models to express this. The demand for segment regression is highlighted. In the specific modeling process of segmented regression, a specific task is the determination of the demarcation point.

The demarcation point is the turning point of the change of the defamation law and the inflection point of the deformation curve. Of course, it becomes the conversion point of the change of the type of the deformation model. The cutoff point is both the endpoint of the former deformation curve and the start point of the latter deformation curve; the deformation numerical model before the cutoff point is one type, and the deformation numerical model after the cutoff point is another one. The correctness of the cutoff point directly affects the quality of modeling and ultimately affects the accuracy of deflection basin area prediction.

According to the change law of the deflection basin area, it is found that the load level will change suddenly at a certain stage. This article uses it as a segmentation point, and two XGBoost models named XGBoost_1 and XGBoost_2 are proposed to predict and analyze the area of the deflection basin in T_1 and T_2 respectively, which can be formulated as

$$y = \begin{cases} \alpha \text{XGBoost}_1(\text{index}), \text{index} \in T_1, \\ \beta \text{XGBoost}_2(\text{index}), \text{index} \in T_2. \end{cases} \tag{12}$$

In the formula, y is the predicted values, α and β are the correction coefficients of two models, respectively, and index is the periodic position.

4.2 CPSO-XGBoost segmented regression model

The XGBoost segmented regression model has 8 main parameters, the learning rate of the two XGBoost models named as l_r^1 and l_r^2; the maximum depth of the tree named as $d_{\max}^1$ and $d_{\max}^2$; the minimum leaf weight named as $lw_{\min}^1$ and $lw_{\min}^2$. The correction parameters are α and β. Different parameters have different functions. The values of parameters have an important impact on the performance of the model. Tuning parameters usually depend on empirical judgment and traversal experiments. Traditional methods are not effective and lack a theoretical basis. Therefore, this paper optimizes parameters based on the improved CPSO algorithm, taking the MSE as the fitness function, and retains the group optimal solution and individual optimal solution for each iteration. Through the interaction of these two types of information, it will evolve towards the global optimal. Because traditional PSO may not jump out from the local optimum, CPSO can make particles out of the local optimum after a successful chaos map, and thus performs a better global optimization ability than the original PSO. Now the constitution of the training module and prediction module is illustrated in Figure 1 in detail. Specific steps are given as follows.

Input: Algorithm iteration number T, population size M, initial maximum position $x_{\max}$, initial maximum velocity $v_{\max}$, training dataset $\text{data}_{\text{train}}$ and test dataset $\text{data}_{\text{test}}$.

Output: optimal position $\mathbf{G}$, optimal fitness function value E.

(1) Perform segmentation processing on the training dataset $\text{data}_{\text{train}}$ and the test dataset $\text{data}_{\text{test}}$.

(2) Randomly initialize the speed and position of particles in the population, and initialize the number of iterations, counters and local extreme value judgment thresholds. The values are $t = 0$, $SG = 0$ and $SG_{\max} = 7$ [44].

(3) Define the relevant parameters of the XGBoost segmented regression model, pass the initial position of the particles to the relevant parameters as the initial parameters, and then train the model to find the initial fitness value e of the particles, find the initial $p_{l\text{best}}^t$ and initial $p_{g\text{best}}^t$ of the population.

(4) Perform the following operations on all particles in the population: ①. update the weight ω_{ti} of the particles according to eq. (6). ②. update the particle velocity and position according to eqs. (4) and (5). ③. calculate the particle fitness value e, and update the $p_{l\text{best}}^t$ and $p_{g\text{best}}^t$ of the particle. If the $p_{g\text{best}}^t$ has not been updated, set $SG = SG + 1$; otherwise, set $SG = 0$.

(5) If $SG \le SG_{\max}$, use eqs. (8)–(10) to generate chaos optimization for $(G^{1'}, G^{2'}, ..., G^{n'})$, and set $SG = 0$.

(6) Set $t = t + 1$. If $t < T$, go to step (4); otherwise, go to step (7).

(7) Output $\mathbf{G}$, E, the algorithm ends.

5 Experiment and analysis

The measured deflection basin data comes from the full-scale pavement test loop road project of the Ministry of Transport. It is located in the southwest corner of the Beijing Highway Traffic Test Field. It has a 2.038-km-long full-scale field accelerated pavement testing track named as RIOHTrack. The full-scale ring road includes 25 types of asphalt pavement structures. The layout of the pavement structure is shown in Figure 2.

Nineteen main test pavement structures were set up in the test ring road to study and compare the long-term performance and evolution of asphalt pavement structures and materials with different combinations of structural rigidity. The thickness of the asphalt concrete structure layer of these pavement structures is 12, 18, 24, 28, 36, 48 cm (or 52 cm), which basically covers the thickness of all asphalt concrete

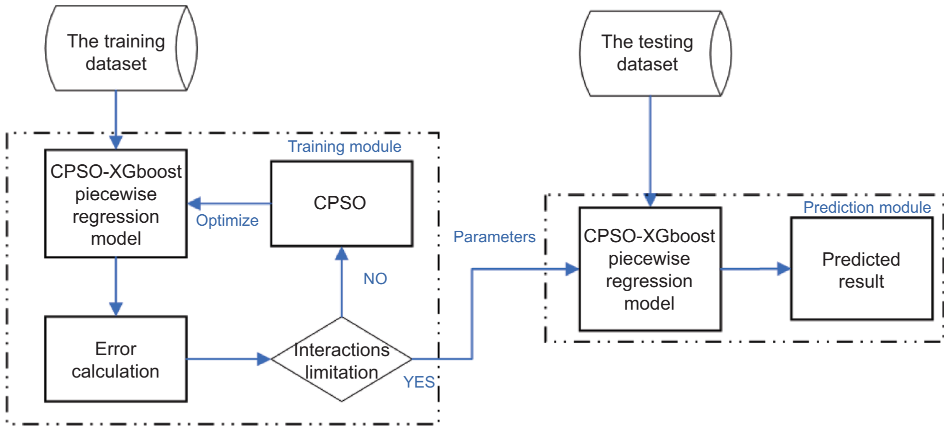

Figure 1 (Color online) The structure of CPSO-XGBoost segmented regression model.

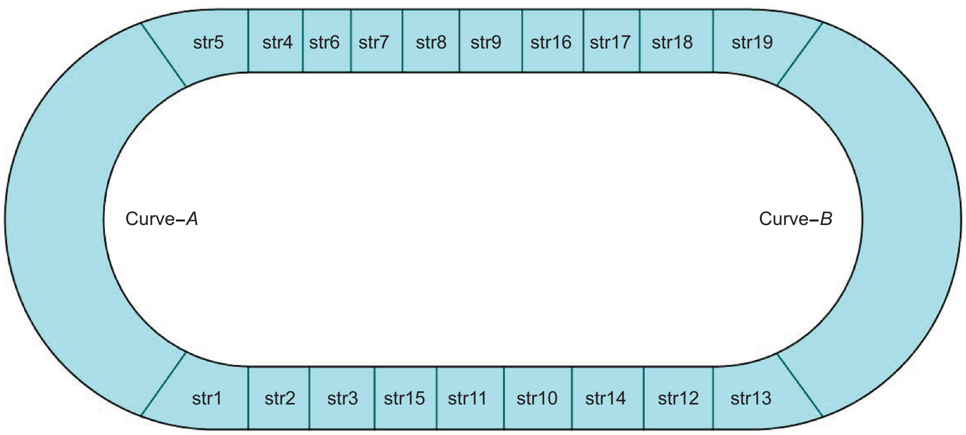

Figure 2 (Color online) RIOHTRACK testing sections.

structure layers of China's high-grade highways, as well as flexible base thickness of thick asphalt pavement. From the perspective of the type of base structure, it includes four typical structures: rigid base structure, semi-rigid base structure, flexible base structure, and full-thick asphalt pavement structure [7].

In this paper, the measured data of the full-scale road surface structure of 19 kinds of asphalt pavement are used as the data source. The goal is to deflect the basin area value from the two feature fields, a total of 5016 data samples, of which 3800 are used for training, and the remaining data are used for testing. A CPSO-XGBoost segmental regression model is modeled for each of the 19 kinds of asphalt pavements. The MSE and the MAE are used to evaluate the performance of algorithms. These metrics are used to reflect the difference between the real value and the predicted value. And, they are the most commonly used performance indices in regression tasks. The smaller the indices are the more accurate the prediction will be. The definitions of these accuracy indices are shown in Table 1.

Table 1 Brief description of the datasets used

Index	Formula
MAE	$\frac{1}{M}\sum_{m=1}^{M}\lvert y_m-\hat{y}_m\rvert$
MSE	$\frac{1}{M}\sum_{m=1}^{M}(y_m-\hat{y}_m)^2$

5.1 Model building

The framework of the deflection basin area prediction model is shown in Figure 3, which contains three parts: new feature extraction, data processing, and data fitting.

Feature engineering is a process that transforms the original data into training data. Its purpose is to obtain better training features and make a machine learning approach to achieve the upper limit of the model. The main feature engineering of this paper includes new feature extraction and data processing.

The initial data of 19 kinds of asphalt pavement mainly include cumulative load axis times, load level, pavement temperature, and deflection basin area. Some features may have missing values. The average value is used to fill in the numerical features with fewer missing values, and then the cycle is calculated. The related load axis time, load axis time change rate, load level change rate, temperature change rate, temperature change difference are shown in Table 2. In the experiments, they are compressed and handled with log-transformed.

Since the deflection basin area data are collected periodically, there is a certain amount of noise, which has a certain negative impact on the prediction of the deflection basin area. To reduce the effects of noise, the smoothing coefficient is taken to perform the first-order exponential smoothing on the deflection basin area. Figure 4 shows typical data and smoothed data. Each algorithm in this paper uses the same data set for training and testing.

5.2 Analysis of segmented regression results

Now, XGBoost, and XGBoost segmented regression are used to predict the area of deflection basin for asphalt pavement STR1. The associated load axis time, load axis time rate of change, load level rate of change, temperature rate of change, and temperature change difference were used to predict the deflection basin area.

Three load experiment cycles of data are used as the training set and one load experiment cycle of data is used as the test set. The correction coefficients of multiple sets of XGBoost segmented regression are set for comparison, and the correction coefficients are shown in Table 3. Also, the prediction results are compared with those of XGBoost with default parameters to obtain the prediction results, as shown in Figure 5. The performance comparison of the prediction results is shown in Table 4.

Table 4 shows that the accuracy of the XGBoost segmented regression model is improved significantly by adjusting the correction coefficients. Note that the best-performing coefficients are difficult to be found by artificial settings. To solve this problem, the CPSO algorithm is proposed in the next to perform an intelligent search for better performance.

5.3 Performance of CPSO-XGBoost Segmented Regression Model

We will predict the deflection basin area of STR1 asphalt pavement in one cycle. Firstly, the performance of CPSO is analyzed. According to the characteristics of the improved CPSO algorithm, combined with the parameter range of XGBoost and the characteristics of the asphalt pavement structure data, the corresponding settings are made. For CPSO optimization, the initial parameters are set as follows: maxi-

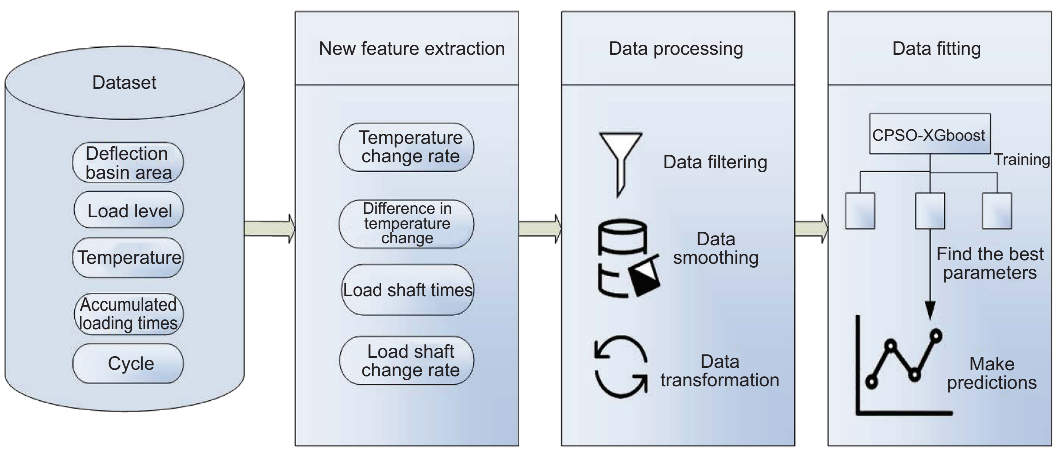

Figure 3 (Color online) Flow chart of deflection basin area prediction.

Table 2 The initial data (part)

Loading period	Accumulated loading shaft times	Load level	. . .	Temperature change rate	Difference in temperature change
1	28396	56.47	. . .	0	0
2	297079	56.47	. . .	–0.586462274	–4.72
3	483685	56.47	. . .	0.30940138	3.61
4	670604	56.47	. . .	0.566987355	15.26
5	854651	56.47	. . .	0.108741099	3.28
⋮	⋮	⋮	⋮	⋮	⋮
N72	21838286	48.09	. . .	–0.845282119	–6.81

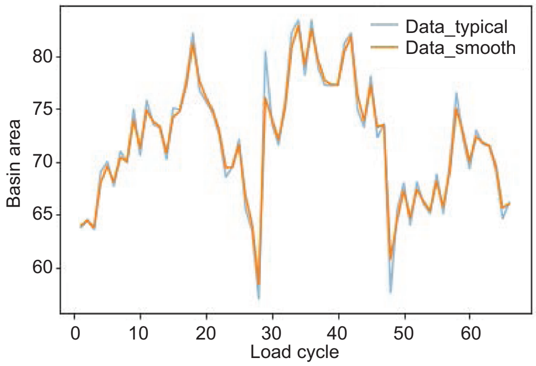

Figure 4 Typical data and smoothed data.

mum iteration $T = 100$, number of population $M = 40$, $x_{\max} = 5$, $v_{\max} = 0.05$ and $SG_{\max} = 7$. Because the proposed CPSO algorithm randomly generates the initial population with certain randomness, it is possible to find the suboptimal or even optimal parameters with high probability through multiple experiments. In this paper, three experiments are performed for different random initializations named experiment I, II, III. The results of three CPSO experiments are shown in Figure 6.

Figure 6 shows that the CPSO algorithm converges quickly in about 20 iterations and escapes the local optimal point by chaotic mapping when it falls into a local optimum, which improves the optimization search accuracy. However, the final accuracy of the CPSO algorithm is influenced by the initial distribution.

Table 3 The correction coefficients of XGBoost segmented regression

Group number	α	β
No. 1	1.3	0.8
No. 2	1.2	0.8
No. 3	1.3	0.7

Table 4 Performance of prediction results in MSE and MAE

Algorithm	MSE	MAE
XGBoost	46.2770	6.4225
No. 3	52.1759	4.7416
No. 2	40.3601	5.7187
No. 1	**6.2744**	**1.7624**

Moreover, the CPSO is compared with some existing PSO algorithms with fixed weight [35] and linear decreasing weight [45] with the same initial particle distribution. The corresponding fitness graph of different PSO algorithms is shown in Figure 7. In the iterative process of these algorithms, we make some of the particles in the swarm mutate and escape from the initial range, to avoid falling into the local optimum. The performance is shown in Table 5.

The results show that the CPSO algorithm proposed in this

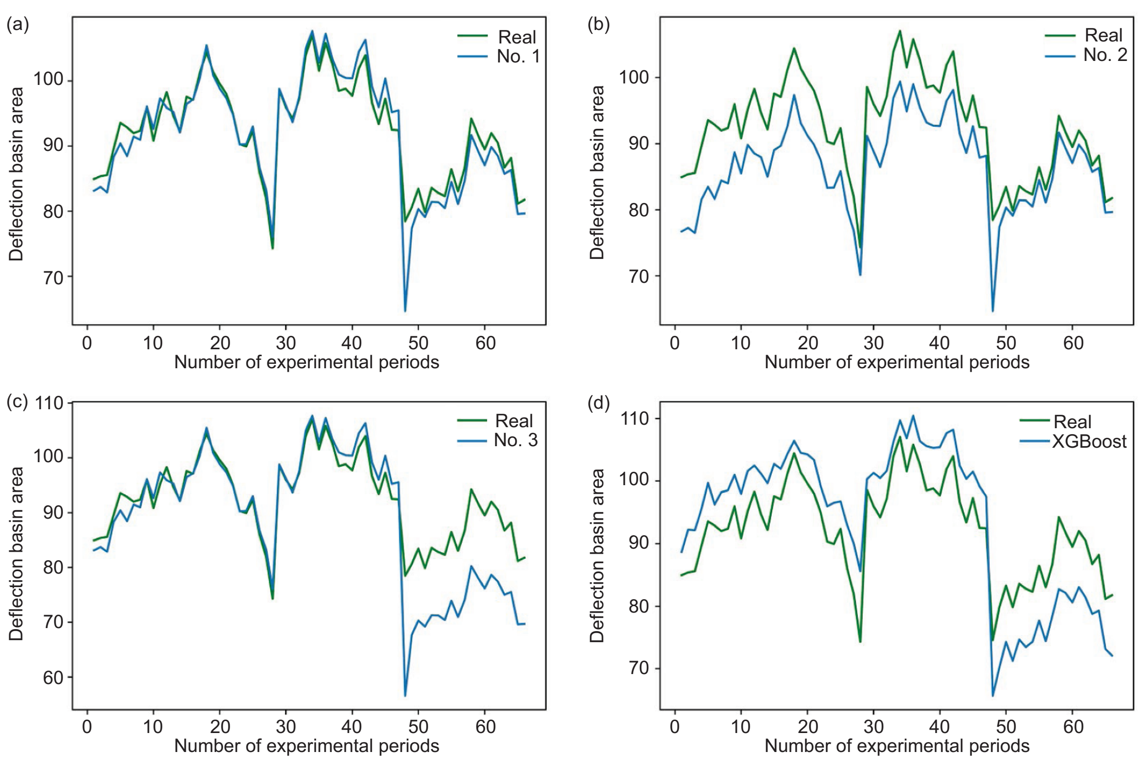

Figure 5 Prediction results of deflection basin area for STR1 with different correction factors and XGBoost algorithm. (a) No. 1; (b) No. 2; (c) No. 3; (d) XGBoost.

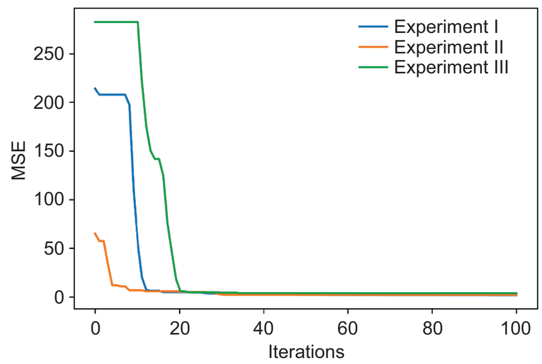

Figure 6 CPSO search process (three experiments).

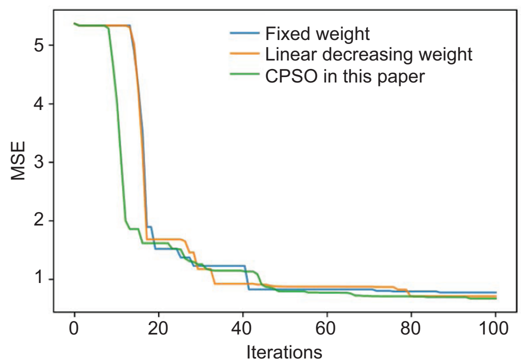

Figure 7 Fitness graph of different PSO.

Table 5 Optimization comparison among different PSO

Algorithm	MSE
Fixed weight	2.1770
Linear decreasing weight	2.0417
CPSO in this paper	**1.9659**

paper has the minimum fitness function value of all particles in the last iteration compared with other PSO algorithms. It jumps out of the local optimum at the later stage of the search by the method of chaotic mapping, and obtains a more schooled fitness function value.

5.4 Deflection basin area prediction for different types of asphalt pavements

To illustrate the generality of our algorithm, the prediction of the deflection basin area under load tests on 19 types of asphalt pavements is considered. To verify the effectiveness of the CPSO-XGBoost segmented regression model, it is compared with the commonly used models in the prediction of the deflection basin area. XGBoost, SVR, KNN, RF, long short-term memory (LSTM) [46] and gated recurrent units neural network methods (GRU) [47] are used as the baseline models for comparison. The baseline models which are machine learning algorithms are implemented using the Scikit learn [48] library for python, and the parameters use the default recommended parameters. As for the LSTM and GRU, we set 7 input nodes, 128 hidden nodes, and 1 output node. Then train the models by the same train dataset with a learning rate of 0.01. We leverage a batch size of 128 and the max epoch is set to 500. stochastic gradient descent (SGD) algorithm solves for the parameters of models.

The hardware environment of machine learning algorithms used in the test uses Intel(R) Core(TM)i5-6500 (3.19 GHz, 4 cores) microprocessor, 8 GB DDR3 memory, and the operating system uses 64-bit Windows 10 Chinese Professional Edition. The hardware environment of LSTM and GRU used in the test uses Intel(R) Xeon(R) E5-2620-v4 (2.10 GHz, 8 cores) microprocessor, Tesla P100-PCIE GPU, and the operating system uses Linux. We use Python 3.6 for experimental simulations.

In the paper, the performances of CPSO-XGBoost segmented regression model, PSO-XGBoost segmented regression model, XGBoost, SVR, KNN, RF, LSTM, and GRU on MSE and MAE are compared and shown in Figure 8.

The comparison results of each algorithm are shown in Table 6. Table 6 shows that the algorithm of this paper has the best performance in MSE and MAE on 19 kinds of asphalt pavement datasets. XGBoost is better than KNN, SVR, and RF algorithms in terms of prediction accuracy. The algorithm in this paper improves 1267.97% in MSE compared with the XGBoost algorithm and is better than deep learning algorithms such as LSTM and GRU. Compared with the algorithm optimized with PSO, it improves 10.58% in MSE. The algorithm of this paper improves 3508.62% in MAE compared with the comparison algorithm. Therefore, the algorithm in this paper has good generalization and can be adapted to extremely complex pavement structures with satisfactory prediction results.

6 Conclusion

Given the weak generalization ability and low prediction

Table 6 The performance of algorithms in MSE and MAE

Algorithm	Mean MSE	Mean MAE
XGBoost	79.4201	7.1966
SVR	82.2908	5.7743
KNN	112.2139	8.1189
RF	196.8466	11.3883
LSTM	61.3932	5.6529
GRU	65.4412	5.2346
PSO-XGBoost segmented regression model	6.4199	1.6066
CPSO-XGBoost segmented regression model	**5.8057**	**1.5962**

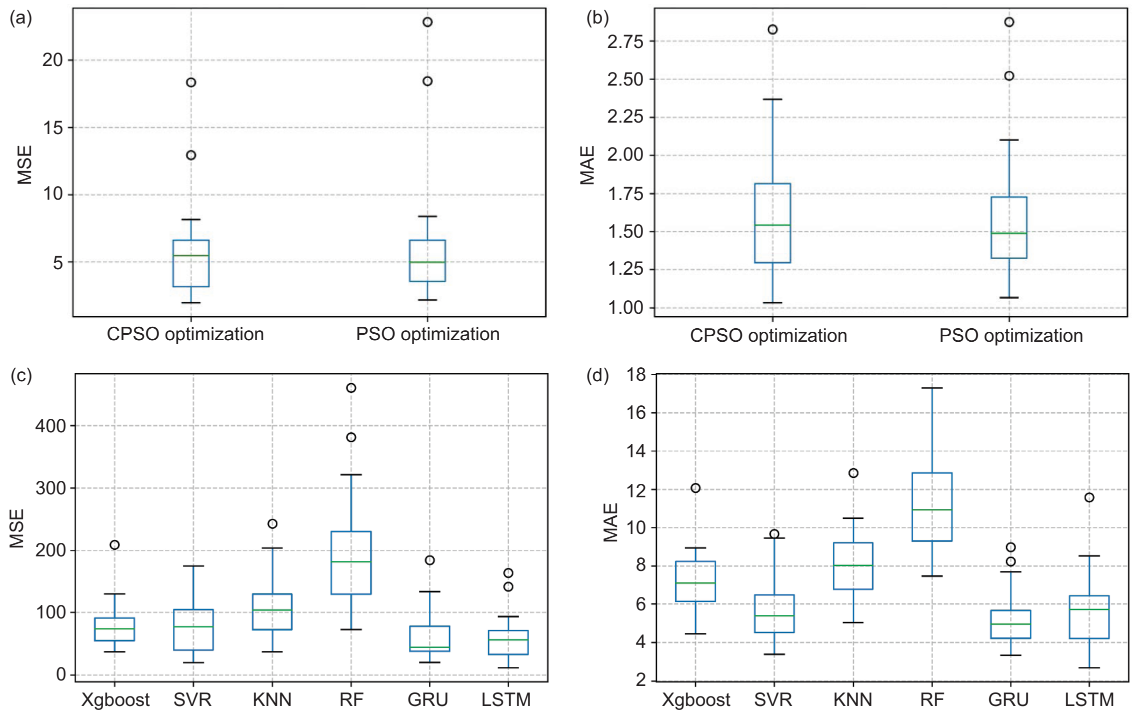

Figure 8 (Color online) Boxplots of MSE and MAE with different algorithms. (a) MSE of different PSO; (b) MAE of different PSO; (c) MAE of comparison algorithm; (d) MAE of comparison algorithm.

accuracy of the XGBoost-based prediction model in asphalt pavement deflection basin area prediction, the CPSO ensemble segmental regression method was introduced into XGBoost named by CPSO-XGBoost segmental regression model, the parameters of XGBoost, and the correction coefficients in the segmental regression method were optimized and improved with the help of CPSO.

In this study, we used 19 datasets to test the predictive performance of the CPSO-XGBoost segmented regression model. The performance of this algorithm outperformed the other compared algorithms on all datasets in terms of MSE and MAE. The MSE performance of this algorithm also outperforms the other compared algorithms. Because the training sample size has a great impact on the prediction performance of the computational intelligence method, its prediction performance can be improved by increasing the training data size. The test results show that the CPSO-XGBoost segmental regression model is effective in predicting the deflection basin area of asphalt pavements. From the results of the comparative analysis, it can be found that although the algorithm in this paper shows the best optimization performance among all the algorithms, its prediction error still has a small fluctuation. For this reason, how to better adapt the evolutionary algorithm and optimize the XGBoost model will be the next problem to be studied and solved. Meanwhile, improving the prediction accuracy and adaptive tuning of the parameters are also important issues to be investigated.

This work was supported by the National Key Research and Development Program of China (Grant No. 2020YFA0714300), the National Natural Science Foundation of China (Grant Nos. 61833005 and 62003084), and the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20200355).

1 Zheng J, Lü S, Liu C. Technical system, key scientific problems and technical frontier of long-life pavement. Chin Sci Bull, 2020, 65: 3219–3229
2 Wang Y, Chong D. Long-life flexible pavement: myth, reality, and the way forward. In: New Frontiers in Road and Airport Engineering. ASCE, Shanghai, 2015. 268–283
3 Liu A, Li H, Zhang P. Long-term performance study of long life pavement pilot section in Jiangsu, China. In: Proceedings of the Transportation Research Congress 2016: Innovations in Transportation Research Infrastructure. ASCE, Beijing, 2018. 353–363
4 Huang W, Liang S M, Wei Y. Surface deflection-based reliability analysis of asphalt pavement design. Sci China Tech Sci, 2020, 63: 1824–1836
5 Hong F, Chen D. Evaluation of asphalt overlay permanent deformation based on ground-penetrating radar technology. J Test Eval, 2016, 44: 20130241
6 Fu G, Zhao Y, Zhou C. Determination of effective frequency range excited by falling weight deflectometer loading history for asphalt pavement. Const Building Mater, 2020, 235: 117792
7 Wang X D. Design of pavement structure and material for full-scale test track. J Highw & Transp Res & Dev, 2017, 34-6: 30–37
8 Simonin J M, Villain G. Detection and survey of interface defects within a pavement structure with ultrasonic pulse echo. In: Proceedings of the 8th RILEM International Conference on Mechanisms of Cracking and Debonding in Pavements. Nantes: Springer, 2016. 673–678
9 Zhao S, Al-Qadi I L, Wang S. Prediction of thin asphalt concrete overlay thickness and density using nonlinear optimization of GPR data.

NDT E Int, 2018, 100: 20–30

10 Wang D, Shi J. Study on infrared differential thermal non-destructive testing technology of the permeability of hot mix asphalt pavements. IOP Conf Ser-Earth Environ Sci, 2017, 69: 012109

11 Loganathan K, et al. Mechanistic empirical estimation of remaining service life of flexible pavements based on simple deflection parameters: A case study for the state of Texas. In: Airfield and Highway Pavements 2019: Design, Construction, Condition Evaluation, and Management of Pavements. ASCE, Chicago, 2019. 294–305

12 Gedafa D S, et al. Relationship between flexible pavement cracking and surface deflections. In: Proceedings of the Transportation Research Board 91st Annual Meeting. Transportation Research Board, 2012. 12-1350

13 Garbowski T, PoŻarycki A. Multi-level backcalculation algorithm for robust determination of pavement layers parameters. Inverse Problems Sci Eng, 2017, 25: 674–693

14 Habbouche J, Hajj E Y, Sebaaly P E. Damage assessment for ME rehabilitation design of modified asphalt pavements: Challenges and findings. Transp Res Record, 2018, 2672: 228–241

15 Ma X, Dong Z, Yu X. Monitoring the structural capacity of airfield pavement with built-in sensors and modulus back-calculation algorithm. Const Building Mater, 2018, 175: 552–561

16 Zhang L, Zhou X, Wang X. Research progress of long-life asphalt pavement behavior based on the RIOHTrack full-scale accelerated loading test. Chin Sci Bull, 2020, 65: 3247–3258

17 Wang X, Zhou X, Guan W. Characteristics and analysis of the mechanical response inside the structure of asphalt pavement. Chin Sci Bull, 2020, 65: 3298–3307

18 Wang Y B, Zheng P, Peng T. Smart additive manufacturing: Current artificial intelligence-enabled methods and future perspectives. Sci China Tech Sci, 2020, 63: 1600–1611

19 Di Q F, Wu Z H, Chen T. Artificial intelligence method for predicting the maximum stress of an off-center casing under non-uniform ground stress with support vector machine. Sci China Tech Sci, 2020, 63: 2553–2561

20 Wang R X, Gao X, Gao J M. An artificial immune and incremental learning inspired novel framework for performance pattern identification of complex electromechanical systems. Sci China Tech Sci, 2020, 63: 1–13

21 Zhang W J, Qin J, Mei F. Short-term power load forecasting using integrated methods based on long short-term memory. Sci China Tech Sci, 2020, 63: 614–624

22 Xiong J, Zhang T Y, Shi S Q. Machine learning of mechanical properties of steels. Sci China Tech Sci, 2020, 63: 1247–1255

23 Fakhri M, Shahni Dezfoulian R. Pavement structural evaluation based on roughness and surface distress survey using neural network model. Const Building Mater., 2019, 204: 768–780

24 Yang Q, Deng Y. Evaluation of cracking in asphalt pavement with stabilized base course based on statistical pattern recognition. Int J Pavement Eng, 2019, 20: 417–424

25 Hussan S, Kamal M A, Hafeez I. Modelling asphalt pavement analyzer rut depth using different statistical techniques. Road Mater Pavement Des, 2020, 21: 117–142

26 Karballaeezadeh N, Ghasemzadeh Tehrani H, Mohammadzadeh Shadmehri D. Estimation of flexible pavement structural capacity using machine learning techniques. Front Struct Civ Eng, 2020, 14: 1083–1096

27 Xu X, Gu Y, Huang W. Structural optimization of steel-epoxy asphalt pavement based on orthogonal design and GA-BP algorithm. Crystals, 2021, 11: 417

28 Li M, Wang H. Prediction of asphalt pavement responses from FWD surface deflections using soft computing methods. J Transp Eng Part B-Pave, 2018, 144: 04018014

29 Li M, Wang H. Development of ANN-GA program for backcalculation of pavement moduli under FWD testing with viscoelastic and nonlinear parameters. Int J Pavement Eng, 2019, 20: 490–498

30 Zhang X, Ji C. Asphalt pavement roughness prediction based on gray GM(1,1—sin) model. Int J Comput Intell Syst, 2019, 12: 897

31 Kaloop M R, Kumar D, Samui P. Particle swarm optimization algorithm-extreme learning machine (PSO-ELM) model for predicting resilient modulus of stabilized aggregate bases. Appl Sci, 2019, 9: 3221

32 Liang C, Xu X, Chen H. Machine learning approach to develop a novel multi-objective optimization method for pavement material proportion. Appl Sci, 2021, 11: 835

33 Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, 2016. 785–794

34 Friedman J H. Greedy function approximation: A gradient boosting machine. Ann Stat, 2001, 29: 1189–1232

35 Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of the ICNN'95-International Conference on Neural Networks. Perth, 1995. 1942–1948

36 Shi Y. Particle swarm optimization: Developments, applications and resources. In: Proceedings of the 2001 Congress on Evolutionary Computation. Seoul, 2001. 81–86

37 Shi Y, Eberhart R. A modified particle swarm optimizer. In: Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings. Anchorage, 1998. 69–73

38 Zhu H, Wu Y. A PSO algorithm with high speed convergence. Control Decis, 2010, 25: 20–24

39 Yang C H, Gu L S, Gui W H. Particle swarm optimization algorithm with adaptive mutation. Comput Eng, 2008, 16: 188–190

40 Tang X, Zhou W, Zhang H, et al. Robot soccer defensive strategy based on multi-objective chaotic PSO. J Syst Simul, 2014, 26: 51–55

41 Song Y, Chen Z, Yuan Z. New chaotic PSO-based neural network predictive control for nonlinear process. IEEE Trans Neural Netw, 2007, 18: 595–601

42 Zhou K, Qin J. PID controller parameters tuning of main steam temperature based on chaotic particle swarm optimization. In: Proceedings of the 2011 IEEE International Conference on Computer Science and Automation Engineering. Zhangjiajie, 2011. 647–650

43 Cheng M Y, Huang K Y, Chen H M. K-means particle swarm optimization with embedded chaotic search for solving multidimensional problems. Appl Math Comput, 2012, 219: 3091–3099

44 Xiang T, Liao X, Wong K. An improved particle swarm optimization algorithm combined with piecewise linear chaotic map. Appl Math Comput, 2007, 190: 1637–1645

45 Shi Y, Eberhart R C. Empirical study of particle swarm optimization. In: Proceedings of the 1999 Congress on Evolutionary Computation. Washington D.C., 1999. 1945–1950

46 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput, 1997, 9: 1735–1780

47 Li C, Xiao F, Fan Y. An approach to state of charge estimation of lithium-ion batteries based on recurrent neural networks with gated recurrent unit. Energies, 2019, 12: 1592

48 Pedregosa F, Varoquaux G, Gramfortet A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res, 2011, 12: 2825–2830